Bayesian Nonparametric Biostatistics

Bayesian Nonparametric Biostatistics

Wesley O. Johnson and Miguel de Carvalho

Abstract We discuss some typical applications of Bayesian nonparametrics in biostatistics. The chosen applications highlight how Bayesian nonparametrics can contribute to addressing some fundamental questions that arise in biomedical research. In particular, we review some modern Bayesian semi- and nonparametric approaches for modeling longitudinal, survival, and medical diagnostic outcome data. Our discussion includes methods for longitudinal data analysis, non-proportional hazards survival analysis, joint modeling of longitudinal and survival data, longitudinal diagnostic test outcome data, and receiver operating characteristic curves. Throughout, we make comparisons among competing BNP models for the various data types considered.

2.1 Introduction

"Why Bayesian nonparametrics?" Motivation for Bayesian nonparametrics encompasses model flexibility and robustness, as parametric models are often inadequate due to their constraints. Bayesian nonparametric models that embed parametric families of distributions in broader families seem eminently sensible since they allow for flexibility and robustness beyond the constrained parametric family. The models we consider here are in fact richly parametric (formally, using an infinite-dimensional parameter space) rather than nonparametric, which is an unfortunate misnomer

W.O. Johnson (✉)
University of California, Irvine, CA, USA
e-mail: wjohnson@uci.edu

M. de Carvalho
Departamento de Estadística, Pontifica Universidad Católica de Chile, Santiago, Chile
e-mail: mdecarvalho@mat.puc.cl

that we will not attempt to rectify. Bayesian nonparametric models involve placing prior distributions on broad families of probability distributions; examples considered here include Mixtures of Polya trees (MPT) and Dirichlet Processes mixtures (DPM).

The MPT will be seen to be a clear extension of a selected parametric family for data. The DPM is more ambiguous but in some instances could be viewed in the same way. A popular theme in much of the Bayesian nonparametrics literature is to regard a parametric approach as a reference, while allowing data that are modeled nonparametrically to inform a subsequent analysis about the adequacy of the parametric model.

Other Bayesian nonparametric approaches involve the use of Gaussian process priors and consist of probability models over spaces of functions. For these the natural probabilistic concept is that of a random function; conceptually, random functions can be regarded as stochastic processes, and are the subject of Part IV of this volume.

2.1.1 Organization of this Chapter

Section 2.2 Comments on the DPM and MPT. In this section we discuss some features of Dirichlet and Polya tree processes; a technical introduction to these and other prior processes can be found in Chap. 1 of this volume (Mitra and Müller 2015).

Section 2.3 Longitudinal Data: Semiparametric Autoregressive Modeling. Here we discuss a model that generalizes standard mixed models for longitudinal data, and which includes a functional mean function, and allows for compound symmetry (CS) and autoregressive (AR) covariance structures. The AR structure is specified through a Gaussian process (GP) with an exponential covariance function, which allows observations to be more correlated if they are observed closer in time than if they are observed farther apart. Quintana et al. (2015) generalize this model by considering a DPM of Gaussian processes. In Sect. 2.3.2 we discuss their analysis of data from the Study of Women's Health across the Nation (SWAN) that involves longitudinal outcomes of hormone data for women experiencing the menopausal transition.

Section. 2.4 Survival Data: Nonparametric and Semiparametric Modeling. We discuss Bayesian non and semi-parametric modeling for survival regression data; Sect. 2.4 provides some preparation for Part III of this volume, which is entirely dedicated to survival analysis. We first give a selective historical perspective of the development of nonparametric Bayesian survival regression methods (Sect. 2.4.1). We discuss an analysis of time to abortion in dairy cattle with fixed covariates, and then discuss models for time dependent regression survival data, followed by analyses of the Stanford Heart Transplant data and a data set involving the timing of

cerebral edema in children diagnosed with ketoacidosis. We end the section with a presentation of a Bayesian nonparametric survival model that allows survival curves to cross, and a subsequent analysis of breast cancer data where survival curves are expected to cross.

Section 2.5 Joint Modeling of Longitudinal and Survival Data. We consider the joint modeling of survival data and a longitudinal process. In Sect. 2.4, we discussed a number of survival regression models with time dependent covariates where we fixed the time dependent covariates (TDC) in the same sense that we fix covariates in regression. However, Prentice (1982) pointed out that fixing the TDCs rather than modeling them could bias final estimates. The general rule has been to use the last observation carried forward (LOCF) in the TDC process, despite the fact that the last observation might have occurred some time ago, suggesting that it may not well represent the current value of the process. In Sect. 2.5 we discuss a data analysis performed by Hanson et al. (2011b), which uses the models and methods in Hanson et al. (2009) in conjunction with longitudinal modeling to develop joint models for longitudinal-survival data.

Section 2.6 Medical Diagnostic Data. In Sect. 2.6.1 we discuss the subject of Receiver Operating Characteristic (ROC) curve regression, and in Sect. 2.6.2 we consider the issue of Bayesian semi-parametric estimation in ROC regression settings that lack availability of a gold standard test, i.e., when there is no available test that could perfectly classify subjects as diseased and non-diseased. Related literature is reviewed in detail in Chap. 16 (Inácio de Carvalho et al. 2015). We illustrate methods by assessing the potential of a soluble isoform of the epidermal growth factor receptor (sEGFR) for use as a diagnostic biomarker for lung cancer in men, and we assess the effect of age on the discriminatory ability of sEGFR to classify diseased and non-diseased individuals. In Sect. 2.6.3 we discuss joint longitudinal diagnostic outcome modeling and analysis, and we illustrate with longitudinal cow serology and fecal culture data.

In Sect. 2.7 we briefly comment on other types of data that are of interest in biomedical research, and on some current Bayesian nonparametric approaches for modeling.

2.2 Comments on the DPM and MPT

We briefly comment on two mainstream prior processes for data analysis: The Dirichlet and Polya tree processes. By themselves, they are perhaps not practical models for data analysis, but it is their mixture forms that are. The Dirichlet Process Mixture (DPM) and the Mixture of Polya Trees (MPT) have been established to be practical tools for data analysis. Models that employ the DPM in various forms are by far the most popular for a variety of reasons including the fact that the DP has been in the literature since at least Ferguson (1973), and DPMs have been developed

extensively for use in analyzing data since at least Escobar (1994). Polya Trees have been around since at least Ferguson (1974), but did not seem to be particularly noticed until the 1990s (Mauldin et al. 1992; Lavine 1992, 1994), and were not given a lot of attention until Berger and Guglielmi (2001), and Hanson and Johnson (2002) and Hanson (2006), who developed MPTs for survival analysis and beyond.

A key property of the DPM is that any inferential object that is modeled as a DPM of continuous parametric densities is smooth. Moreover, under some conditions, the DPM of location-scale normal densities has been shown to have strong posterior consistency for the true density (Tokdar 2006). There are many other theoretical works of this type, including Amewou-Atisso et al. (2003), who established large sample consistency properties for semiparametric linear regression models with error distributions that are modeled with median zero processes based on both PTs and DPMs. The original and continuing appeal to DPMs was and is at least partly based on the ease of marginalizing over the DP when performing numerical calculations. The marginalization led to computationally straightforward schemes involving the Polya Urn scheme that researchers often describe as a Chinese restaurant process. Neal (2000) improved upon previous computational schemes pioneered by Escobar (1994); Escobar and West (1995), and MacEachern and Müller (1998), among others. In addition, there are many extensions of the DPM, including the Dependent Dirichlet Process (DDP) (MacEachern 2000), the Nested DP (NDP) (Rodriguez et al. 2008), and the Hierarchical DP (HDP) (Tomlinson and Escobar 1999; Teh et al. 2006), among others, many discussed in Chap. 1 of this volume (Mitra and Müller 2015).[1] The Sethuraman (1994) representation of the DP facilitated the development of all of these, and it provided an easy understanding of the precise meaning of the DP and the DPM. In addition, it facilitated the extension to more general stick-breaking processes, for example the Dunson and Park (2008) application to density regression, among others. The point here is that there is now a wealth of papers that have developed, extended, and used various forms of and which stem from the DP, and which have used these tools to analyze data of all complexities. The DPM is clearly here to stay.

The MPT has many positives as well. It can be selected to be absolutely continuous with probability one, so it is possible to use it directly as a model for data. When used as a model for the error distribution in a linear regression, it is easy to specify that the MPT has median zero with probability one, resulting in a semiparametric median regression model. In Sect. 2.4 we discuss such models for survival data. In addition, it is a flexible model, allowing for multimodality, skewness, etc. It is straightforward to perform MCMC computations for many complex models (Hanson 2006), and there is no need to marginalize the process to make computations simpler. From our point of view, a major positive feature of the Mixture of Finite Polya Tree (MFPT) prior is that it not only allows for a broad/flexible class of distributions but that it has a parametric family of distributions for the data embedded in it, and that the embedding is natural. Thus, if a scientist has previous experience or information that suggests that a log normal family of distributions

[1] See also Müller and Mitra (2013) for a recent survey.

might be appropriate for their survival data, they could hedge their bets by embedding that family as the centering family of an MFPT. Moreover, if they also had scientific information about the log normal family parameters, they could construct an informative prior for those parameters. See Bedrick et al. (1996) and Bedrick et al. (2000) for illustrations of informative prior specification for generalized linear models and for survival models. Thus far, we are not aware of any such nice properties for specifying prior distributions on the parameters of the base distribution in the DPM. Berger and Guglielmi (2001) also took advantage of the fact that a parametric family can be embedded in the PT family in developing a method to test the adequacy of the parametric family to fit data.

A possible advantage of the DPM over the MPT is the ease of extending the DPM to multivariate data, which is straightforward for the DPM. Hanson (2006) has developed MPT methods for multivariate data, and Jara et al. (2009) and Hanson et al. (2011a) improved them. While no comparison between the methods has been performed to date, Hanson reports that the MPT-based method would perform well for joint density estimation, and clearly better for "irregular densities" (personal communication). Another advantage is the smoothness of the DPM. When the weight associated with the MPT is small, density estimates can be quite jagged, despite the fact that Hanson and Johnson (2002, Thm. 2) proved that predictive densities in the context of the semiparametric model that they develop are differentiable under some conditions. For applications, an important issue is prior elicitation for the DPM; cf. Hanson et al. (2005).

In the illustrations below, we take examples that use the DPM, DDP, and MPT. For the MPT based models, we always use a truncated version, which is termed an MFPT. The truncation is at some level, usually termed M, of the basic tree structure. In addition, MPTs have weights, c, just like the DPM, whereas small weight corresponds to the model being 'more nonparametric.' Some models discussed below, e.g. Hanson and Johnson (2002, 2004), and De Iorio et al. (2009), can be fit using the `R` package `DPpackage` (Jara et al. 2011).

2.3 Longitudinal Data: Semiparametric Autoregressive Modeling

2.3.1 The Semiparametric Model

Assume that observations are made on individual i at times $\{t_{i1},\ldots,t_{in_i}\}$, namely $Y_i = \{Y_{ij} : j = 1,\ldots,n_i\}$. At time t_{ij} we allow for a vector of possibly time-dependent covariates $x_{ij}^{\mathrm{T}} = (1, x_{i1}(t_{ij}),\ldots,x_{ip}(t_{ij}))$, and assume that $\mathrm{E}(Y_{ij}) = x_{ij}^{\mathrm{T}}\beta$. Define the $n_i \times (p+1)$ design matrix $X_i = (x_{i1},\ldots,x_{in_i})^{\mathrm{T}}$, leading to an assumed mean vector $\mathrm{E}(Y_i) = X_i\beta$. Then, allow for a corresponding $n_i \times q$ design matrix Z_i, with $q \leqslant p$ and with the column space of Z_i restricted to be contained in the column space of X_i.

The starting point for the model to be discussed is a well-known linear mixed model (Diggle 1988) that also allows for AR structure, namely

$$Y_i = X_i\beta + f_i(t) + Z_i b_i + w_i + \varepsilon_i, \quad b_i \mid \xi \sim \mathrm{N}_r(0, D(\xi)), \quad w_i \mid \phi \sim \mathrm{N}_{n_i}(0, H_i(\phi)), \tag{2.1}$$

Here, $H_i(\phi)$ is $n_i \times n_i$ and has a structural form, $\varepsilon_i \sim \mathrm{N}(0, \sigma^2 I_{n_i})$, and $f_i(t)$ is a function evaluated at subject-specific times t_{ij} for individual i; in addition, ξ and ϕ contain variance–covariance parameters for b_i and w_i, respectively.

The w_i are generated by zero-mean Gaussian processes, $\{w_i(t) : t > 0\}$. If $\mathrm{Cov}(w_i(t+s), w_i(t)) = \sigma_w^2 \rho(s)$, with $\rho(s) = \rho^s$, the resulting stationary process is an Ornstein–Uhlenbeck process (Rasmussen and Williams 2006), which yields an exponential covariance function and induces AR structure.[2] The combination of choosing which terms to include in (2.1)—and making particular choices for $H(\phi)$ and $D(\xi)$—when the corresponding effects are included in the model, determines the covariance structure for the data.

The semiparametric autoregressive model extends (2.1) by introducing flexibility beyond the exponential covariance structure. Consider first the GP, w_i, for the ith subject, with covariance matrix of the form $H_i(\phi) = \sigma_w^2 \tilde{H}_i(\rho)$, where $\phi = (\sigma_w^2, \rho)$ and $\{\tilde{H}_i(\rho)\}_{k,\ell} = \rho^{|t_{i\ell} - t_{ik}|}$. Let $\phi \mid G \sim G$ with $G \sim \mathrm{DP}(\alpha, G_0)$ so that

$$f(w_i \mid G) = \int \mathrm{N}(w_i \mid 0, \sigma_w^2 \tilde{H}_i(\rho))\, \mathrm{d}G(\phi) = \sum_{k=1}^{\infty} \pi_k \mathrm{N}_{n_i}(w_i \mid 0, \tilde{\sigma}_{wk}^2 \tilde{H}_i(\tilde{\rho}_k)), \tag{2.2}$$

is an infinite mixture of multivariate normal densities, where $(\tilde{\sigma}_{wk}^2, \tilde{\rho}_k) \overset{\text{iid}}{\sim} G_0$, and the $\pi_k = V_k \prod_{l<k}(1 - V_l)$, where $V_k \overset{\text{iid}}{\sim} \mathrm{Be}(1, \alpha)$; here, G_0 is the centering distribution and $\alpha > 0$ is the so-called precision parameter. A related spatial DP with exponential covariance function in the base distribution was developed by Gelfand et al. (2005).

Model (2.2) implies clustering on autocorrelation structure across subjects, and using the Sethuraman representation, it can be noticed that

$$\mathrm{Cov}(w_i(t+s), w_i(t) \mid G) = \sum_{k=1}^{\infty} \pi_k \tilde{\sigma}_{wk}^2 \tilde{\rho}_k^s.$$

Hence, if the ith subject has equally spaced times between observations, the corresponding covariance matrix has equal diagonals with decreasing correlations as s increases, but not necessarily at a geometric rate.

[2] Zeger and Diggle (1994) used $\rho(s) = \alpha + (1-\alpha)\rho^s$. There are additional choices, including the possibility that σ_w^2 could depend on t, resulting in a nonhomogeneous Ornstein–Uhlenbeck process (Zhang et al. 1998). Taylor et al. (1994) used an integrated Ornstein–Uhlenbeck process (integrating over an Ornstein–Uhlenbeck with exponential covariance function) that results in a covariance function that depends on both t and s. With structured covariance functions, the marginal covariance matrix for Y_i is $\mathrm{Cov}(Y_i) = \Sigma_i(\xi, \phi, \sigma^2) = Z_i D(\xi) Z_i^{\mathrm{T}} + H_i(\phi) + \sigma^2 I_{n_i}$.

It is useful to re-write the semiparametric autoregressive model (2.2) hierarchically based on latent parameters $\phi_1, \ldots, \phi_n$, i.e.

$$\begin{aligned} Y_i \mid \beta, b_i, w_i, \sigma^2 &\overset{\text{ind}}{\sim} \mathrm{N}_{n_i}(X_i\beta + f_i(t) + Z_i b_i + w_i, \sigma^2 I), \\ w_i \mid \phi_i = (\sigma^2_{wi}, \rho_i) &\overset{\text{ind}}{\sim} \mathrm{N}_{n_i}(0, \sigma^2_{wi}\tilde{H}_i(\rho_i)), \\ \phi_1, \ldots, \phi_n \mid G &\overset{\text{iid}}{\sim} G, \\ G &\sim \mathrm{DP}(\alpha, G_0), \\ b_i &\overset{\text{iid}}{\sim} \mathrm{N}(0, D(\xi)), \\ \sigma, \beta, \xi &\sim \mathrm{U}(0, A) \times \mathrm{N}(\beta_0, B) \times p(\xi), \end{aligned} \tag{2.3}$$

where w_i and b_i are assumed independent for $i = 1, \ldots, n$.

What about posterior sampling? It can be shown that $f(w_i \mid \phi_i)$ is easily obtained, by noting that $w_i \sim \mathrm{N}_{n_i}(0, \sigma^2_{wi}\tilde{H}(\rho_i))$. Then, with

$$r_{ik} = \rho_i^{|t_{i,k+1} - t_{ik}|}, \quad k = 1, \ldots, n_i - 1,$$

Quintana et al. show that

$$\begin{cases} w_{i1} \sim \mathrm{N}_1(0, \sigma_i^2), \\ w_{ik} \mid w_{i1} = \tilde{w}_1, \ldots, w_{ik-1} = \tilde{w}_{k-1} \sim \mathrm{N}_1\left(\tilde{w}_{k-1} r_{ik-1}, \sigma_i^2(1 - r^2_{ik-1})\right). \end{cases}$$

Thus, $f(w_i \mid \phi_i)$ is obtained as the product of n_i univariate normal probability densities, making it simple to obtain the full conditional distribution of w_i in a Gibbs sampling algorithm.

2.3.2 Model Specification for Hormone Data

Quintana et al. (2015) considered a small subset of data that were obtained from SWAN (Study of Woman Across the Nation, *www.swanstudy.org*). The data included 9 observations for each of 162 women, and contained no missing observations. The data were grouped according to age at the beginning of the study (under 46 and over 46 years), and according to four racial/ethnic groups (African American, Caucasian, Chinese, and Japanese).

The main interest was to model the annual follicle stimulating hormone (FSH) concentrations through the menopausal transition. Concentrations of FSH and other hormones had been modeled to increase according to a (four parameter) sigmoidal shape (Dennerstein et al. 2007). FSH concentrations were measured annually from serum samples in days two through five of the menstrual cycle for women who were still menstruating or on any day that women came in for their annual visit if they were postmenopausal. Times of observation were centered on the year of final menstrual period (FMP), namely $t_i = 0$ corresponds to the year in which the final

menses occurred, which is defined to be the actual time of last menses before a 12-month period in which there were none. Thus, year -3 is 3 years prior to the FMP, and year $+3$ is 3 years after. The data included women who started at year -8 continuing through year 0, and women starting at year -2 and continuing through year 6 (after FMP).

The functional part of the model involves a generalized sigmoid function allowing for greater flexibility than the Dennerstein et al. model. Each of the eight age by race-ethnicity groups was modeled with its own generalized sigmoid function. Let $c(i) \in \{1,\ldots,8\}$ be an indicator variable describing the particular combination of four races and two ages corresponding to subject i. Here, we set $\beta = (\beta_1,\ldots,\beta_8)$, where β_l is the vector of fixed parameters associated with combination l.

Quintana et al. used the five parameter generalized sigmoid curve that was discussed in Ricketts and Head (1999):

$$S(t \mid \beta) = \beta_1 + \frac{\beta_2}{1 + f_t \exp\{\beta_3(\beta_4 - t)\} + (1 - f_t)\exp\{\beta_5(\beta_4 - t)\}}, \tag{2.4}$$

where

$$f_t = \frac{1}{1 + \exp\{-C(\beta_4 - t)\}}, \quad C = \frac{2\beta_3\beta_5}{|\beta_3 + \beta_5|},$$

in which case the fixed effects become $f_i(t_{ij}) = S(t_{ij} \mid \beta_{c(i)})$. The parameters now five-dimensional and the curves defined by (2.4) are not restricted to be monotone, as would be the case of a pure sigmoidal curve. If β_3 and β_5 are however both positive, then (2.4) is monotone and increasing, and if both are negative, then it is decreasing. Using a model with fixed effects specified through (2.4), estimated mean profiles can be compared for the eight groups.

The data analysis just below is based on the specification:

$$Y_i = S(t_i \mid \beta_{c(i)}) + b_i\mathbf{1} + w_i + \varepsilon_i, \tag{2.5}$$

where $t_i^{\mathrm{T}} = (t_{i1},\ldots,t_{i9})$, $b_i \overset{\text{ind}}{\sim} \mathrm{N}(0,\sigma_b^2)$ are individual-specific random effects, $\mathbf{1}$ is a vector of ones, w_i is distributed as a DPM of Ornstein–Uhlenbeck (OU) processes, as specified in (2.3), and where $S(t_i \mid \beta_{c(i)})$ is a vector with entries $S(t_{ij} \mid \beta_{c(i)})$, for $j = 1,\ldots,9$.

Hormone Data Analysis

Quintana et al. (2015) fitted a total of six models to the data, including (2.5) above. The models considered included a parametric version of (2.5) without the OU process, model (2.5) with mixed and fixed linear terms replacing the sigmoid function, a model just like this one, except setting $\rho = 0$, model (2.5) again, but with $\rho = 0$, and finally model (2.5) without OU structure and with a general nonparametric Bayes mixture for the random effects.

They calculated log pseudo-marginal likelihood (LPML) statistics for each model; see Christensen et al. (2010, Sect. 4.9.2), or Gelfand and Dey (1994). This criterion for model selection was first introduced by Geisser and Eddy (1979) and has been used extensively for model selection in recent years; see, for example, Hanson, Branscum, and Johnson (2011b). The pseudo-marginal likelihood used was defined as $\prod_{i=1}^{n}\prod_{j=1}^{n_i} f(y_{ij} \mid y_{(ij)}, X_i, \mathscr{M})$, where $f(y_{ij} \mid y_{(ij)}, X_i, \mathscr{M})$ is the predictive density, under model $\mathscr{M}$, corresponding to individual i at time j based on the data minus y_{ij}. LPML value for model (2.5) was -5966, and the range for the other five models was -6673 to -6986; thus the sigmoid function with NP autoregressive structure was the clear winner. Leaving out the AR part of the model was simply not an option.

Plots of fitted values and corresponding probability bands (not shown) were virtually identical for (2.5) and its linear counterpart was virtually identical. The model with linear structure would have however been useless for prediction or for characterizing mean curves as can be seen in Fig. 2.1.

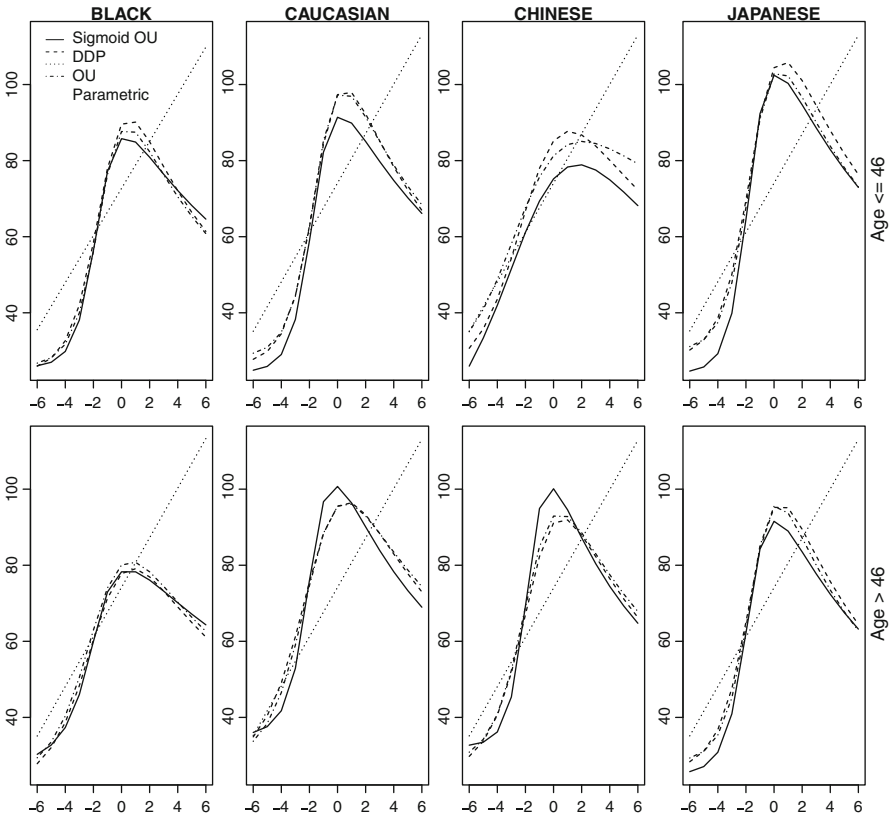

Fig. 2.1 Predictions of future hormone concentrations (y axis) for eight types of women, using (2.3) (*solid curve*), linear version of (2.3) (*dotted*), parametric sigmoid (*dot dash*), nonparametric random effects with sigmoid (*dashed*). Times of observation (x axis) are centered on the year of final menstrual period (FMP) ($t_i = 0$), so that year -3 is 3 years prior to the FMP, and year $+3$ is 3 years after

Figure 2.1 shows model-based future predictions (posterior mean curves) for the eight different types of patient, all on the same time scale. It thus makes sense to compare shapes and levels across race/ethnicity for the same age group, and between age groups for the same race/ethnicity. Generally speaking, all models that include sigmoid mean functions predict that women's FSH hormones will go up sigmoidally, and then curve downwards toward the end of the time frame, regardless of age-race/ethnicity category. On the other hand, the linear effects model, labeled as OU on the graph, predicts a simple linear increase in FSH hormone values in contrast to the others.

Quintana et al. also made inferences comparing the maximum level achieved, the timing of the maximum level achieved and the overall slope of increase in the 4 years before FMP. The most dramatic inference is that Chinese women who are 46 years old and under at baseline achieve their maximum approximately between 1 and 3 years after FMP with 95 % posterior probability, while corresponding intervals for younger women in the other race/ethnic groups are below this interval. Among older women at baseline, there is a 0.95 posterior probability that timing for African Americans is greater than for Caucasians. The posterior probability that the difference in timing comparing younger to older Chinese women is positive is one to four decimal places. There is a clear statistical difference in timing comparing age groups for Chinese women but not for the other groups.

Finally, they estimated correlations among repeated responses on a new patient with *equally spaced times* of observation based on the joint predictive distribution under (2.5). The estimated correlations for these times that were 1–8 years apart were respectively $\{0.43, 0.27, 0.21, 0.17, 0.15, 0.14, 0.14, 0.13\}$, which is quite distinct from an AR structure. Quintana et al. observe that, after about 4 years, the correlations flatten out around 0.14. With a typical AR structure, the estimated correlations would continue to decrease across time.

2.4 Survival Data: Nonparametric and Semiparametric Modeling

2.4.1 Nonparametric and Semiparametric Survival Regression: A Selective Historical Perspective

Survival modeling has a long and enduring history that continues. The field took its initial directions from the landmark papers by Kaplan and Meier (1958) (KM) and by Cox (1972).[3] The former paper developed the most famous nonparametric estimator of a survival function for time to event data with censoring called the product limit estimator. The second paper extended the field of survival analysis to semiparametric regression modeling of survival data; the model introduced there

[3] According to Ryan and Woodall (2005); Cox (1972) and Kaplan and Meier (1958) are the two most-cited statistical papers.

is termed the Cox proportional hazards (PH) model and is ubiquitous in medical research. There have literally been hundreds if not thousands of papers addressing various models and methods for performing survival analysis.

The main goal of a large proportion of these papers is to examine the relationship between the time to event, say T, and covariate information, say x, through the survivor function $S(t \mid x) \equiv \Pr(T > t \mid x)$. This is often done by starting with a model for T, like $\log(T) = x\beta + W$ where β is a vector of regression coefficients and W is modeled to have a mean zero error distribution.[4] Parametric models have W distributed as normal, or extreme value or logistic, resulting in parametric log normal, Weibull and log logistic survival models. These models are termed parametric accelerated failure time (AFT) models (Kalbfleisch and Prentice 2002, Sect. 2.3.3). If the distribution of W is parameterized to have median zero, which is automatic for the normal and the logistic and involves a slight modification for the extreme value distribution, then the median time to event is $\text{med}(T \mid x) = \mathrm{e}^{x\beta}$.

Models that allow for flexible distributions for W are termed semiparametric. Specifically, the AFT model with fixed covariates x discussed in Hanson and Johnson (2002) asserts $\log(T) = -x\beta + W$ with $\mathrm{e}^W \sim \text{MFPT}(M, c, F_\theta)$ and $\theta \sim p(\theta)$, where M is the truncation level for the tree structure and c is the weight that is associated with how much flexibility there will be about the parametric centering model, F_θ. The nonparametric model embeds the family of distributions $\{F_\theta : \theta \in \Theta\}$ in it, in the sense that $\mathrm{E}\{F_W(t) \mid \theta\} = F_\theta(t)$ for all θ and t. Here, for example F_θ could be a log normal distribution. The survivor function for this model is $S(t \mid x, \beta, S_0) = S_0(t\mathrm{e}^{x\beta})$ and the hazard is $h(t \mid x, \beta, h_0) = \mathrm{e}^{x\beta} h_0(t\mathrm{e}^{x\beta})$.

Alternatively, models can be constructed by considering hazard functions, which can be regarded as instantaneous failure rates, formally defined as $h(t \mid x) = \lim_{\Delta s \to 0} \Pr(T \in (t, t+\Delta s] \mid T > t, x)/\Delta s = f(t \mid x)/S(t \mid x)$, where $f(t \mid x)$ is the density for T. The Cox (1972) PH model is $h(t \mid x) = h_0(t)\mathrm{e}^{x\beta}$ where h_0 is an arbitrary baseline hazard function. For two distinct individuals, it follows that the ratio of their hazards involves the cancellation of the common baseline hazard and what remains is a constant (in t) that only depends on their covariate vectors and the regression coefficients, hence the PH model. The survival function can be written as $S(t \mid x, \beta, H_0) = \exp\{-\mathrm{e}^{x\beta} H_0(t)\}$, where $H_0(t) = \int_0^t h_0(s)\,\mathrm{d}s$, which is termed the baseline cumulative hazard. Defining $S_0(t) = \exp\{-H_0(t)\}$, the survival function can be expressed as $S(t \mid x, \beta, S_0) = S_0(t)^{\mathrm{e}^{x\beta}}$, where S_0 is termed the baseline survival function. Under the PH model, survival curves for individuals with distinct covariate values cannot cross. We see that there is a parametric part to the PH model involving β, and a nonparametric part involving the unknown baseline hazard function, or equivalently the corresponding cumulative hazard, or baseline survival distribution. Bayesian approaches place parametric priors on the former, and nonparametric priors on the latter.

Bayesian methods for survival analysis were somewhat constrained until the advent of modern MCMC methods. Susarla and Van Ryzin (1976) placed a DP prior on S, and derived the posterior mean with censored survival data resulting in

[4] For ease of notation, we often write $x\beta$ to denote of $x^{\boldsymbol{T}}\beta$.

the Bayesian analogue to the KM estimator in the no covariate case. There were many ensuing papers, including a paper by Johnson and Christensen (1986) who again placed a DP prior on S and provided analogous results for interval censored data. Kalbfleisch (1978) placed a gamma process prior distribution on H_0 in the PH model, and derived empirical Bayes (EB) results for that model by marginalizing over the gamma process and using the marginal likelihood to obtain estimates of β. Christensen and Johnson (1988) considered the AFT model, placed a DP prior on e^W, marginalized over this distribution and maximized the marginal likelihood to obtain EB estimates of regression parameters. Finally, Johnson and Christensen (1989) established the analytical intractability of a fully Bayesian approach to that model.

Subsequently, Kuo and Mallick (1997) developed a Bayesian semiparametric model for AFT data by modeling W with a DP mixture of normal distributions. They performed numerical approximations to posterior inferences using the basic ideas presented in Escobar (1994). Kottas and Gelfand (2001) then developed an AFT model with error distribution modeled as a DPM of split normals that was designed to have median zero and thus resulted in a regression model with $\text{med}(T \mid x, \beta) = e^{x\beta}$, a semiparametric median regression model. Then, Hanson and Johnson (2004) developed a fully Bayesian AFT model for interval censored regression data by placing a mixture of DP priors on e^W. While this model is analytically intractable, Hanson and Johnson were able to develop an MCMC algorithm for numerically approximating posterior distributions for all parameters of interest, including survival functions and regression coefficients. Hanson and Johnson (2002) modeled e^W with a mixture of finite Polya trees (MFPT).

Time-to-Abortion in Dairy Cattle Data Analysis

We illustrate the semiparametric AFT regression model with MFPT model for the error distribution. The model and analysis of these data were presented in Hanson and Johnson (2002). The data included $n = 1344$ dairy cattle that were observed to naturally abort their fetus prematurely. Nine herds from the central valley of California had been monitored and it was of interest to assess the relationship between two characteristics of the dam: Days open (DO), the number of days between the most recent previous birth and conception, and gravidity (GR), the number of previous pregnancies that the dam has had, and the timing to abortion. The herds were followed for 260 days; 16 dams aborted after the 260 days, and hence were right-censored. Hanson et al. (2003) also analyzed these data and determined that it was likely that the baseline densities and hazard functions were bimodal thus ruling out a standard parametric model.

The model used was:

$$\log T_{ij} = -\beta_0 - \beta_1 \text{DO}_{ij} - \beta_2 \text{GR}_{ij} - \gamma_i + W_{ij}, \quad W_{ij} \mid G \overset{\text{iid}}{\sim} G,$$

where T_{ij} is the fetal lifetime of the 1344 fetuses that aborted in each of the $i = 1, \ldots, 9$ herds, with $j = 1, \ldots, h_i$ dams observed to have aborted in herd i.

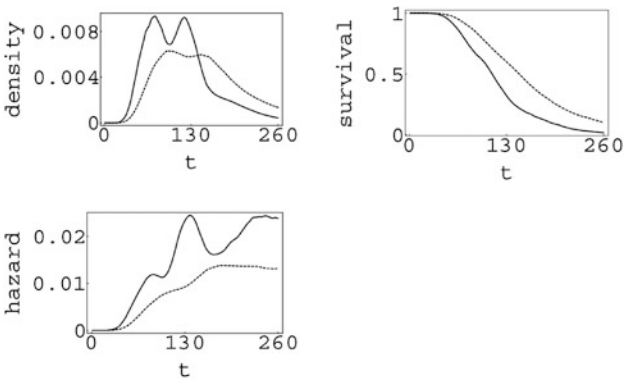

Fig. 2.2 Predictive densities, survival curves, and hazard curves for herds 4 (*solid*) and 9 (*dashed*); here t denotes time in days

The baseline G was modeled as a mixture of finite $(M = 10 \stackrel{\bullet}{=} \log_2 1344)$ Polya trees. The fixed effect for herd 1, γ_1, was fixed at zero and hence herd 1 has the baseline survival distribution. The mixture of Polya trees was centered about the family $G_\theta = \mathrm{N}(0, \theta^2)$ and $p(\beta) \propto 1$ and the prior for θ was taken to be $\propto \theta^{-2}$. The parameter w was fixed at 10, signifying relative comfort in the parametric log normal family, but small enough to allow for deviations from it. Table 2.1 displays the posterior regression effects. All probability intervals include zero, however there are herd differences. For example, fixing DO and GR, $\exp(\gamma_i - \gamma_j)$, with $j \neq i$, is the ratio of median survival times for herds j and i. The median and 95 % probability interval for $\exp(\gamma_4 - \gamma_9)$ is 1.3 (0.9, 2.0), that is, the median time-to-abortion of herd 9 is estimated to be 1.3 times that of herd 4, with a plausible range of 0.9 to 2.0.

Table 2.1 Posterior inference (posterior medians and 95 % probability intervals) for cow abortion data

Parameter	Posterior median	95 % Probability intervals
Intercept	-4.79	$(-4.89, -4.70)$
DO	-1.1×10^{-4}	$(-6.4 \times 10^{-4}, 3.3 \times 10^{-4})$
GR	0.01	$(-0.01, 0.03)$
γ_2	-0.01	$(-0.08, 0.05)$
γ_3	0.00	$(-0.12, 0.10)$
γ_4	0.09	$(-0.02, 0.21)$
γ_5	-0.03	$(-0.14, 0.07)$
γ_6	0.02	$(-0.16, 0.15)$
γ_7	0.05	$(-0.02, 0.14)$
γ_8	-0.01	$(-0.08, 0.06)$
γ_9	-0.20	$(-0.56, 0.16)$

Figure 2.2 compares the predictive densities, survival, and hazard functions for herds 4 and 9 evaluated at the population mean values of DO and GR. The predictive survival densities are both clearly bimodal as suggested by Hanson et al. (2003). The herd 4 hazard curve peaks at 86 days and 138 days. Hanson et al. (2003) described these peaks as possibly being related to difficulty in previous calving (the first peak) and the effect of leptospirosis infection (the second peak).

2.4.2 *Semiparametric Models for Survival Data with Time-Dependent Covariates*

A number of semiparametric regression models associating survival time with time-dependent covariates (TDC), have been proposed in the literature, including models due to Cox (1972), Prentice and Kalbfleisch (1979), Aalen (1980), Cox and Oakes (1984), and Sundaram (2006), among many others. In this section, we discuss the extension of the Hanson and Johnson (2002) model, the Sundaram (2006) proportional odds model, and the Cox PH model, to include TDCs, and we discuss the Cox and Oakes (1984, Chap. 8) model—to which we refer as the COTD model—which was designed to incorporate TDCs. This work is discussed in detail in Hanson et al. (2009).

Consider the time-dependent covariate process $\{x(t) : t \in (t_1, \ldots, t_k)\}$ where t_is are times of observation, and $x(t)$ is the possibly vector valued observation on the TDC process. Also define h_0 to be an arbitrary baseline hazard, and in particular, let it correspond to an individual with constant covariate process values of zero for all times. Let $S_0(t) = \exp\{-\int_0^t h_0(s)\,\mathrm{d}s\}$ be the corresponding baseline survivor function. Prentice and Kalbfleisch (1979) extended the AFT model to TDCs as

$$h(t \mid x(t), \beta, h_0) = \mathrm{e}^{x(t)\beta} h_0(t\mathrm{e}^{x(t)\beta}), \tag{2.6}$$

and Hanson et al. (2009) termed it as the PKTD model. The TD Cox model has hazard function

$$h(t \mid h_0(t), x(t), \beta) = \mathrm{e}^{x(t)\beta} h_0(t), \tag{2.7}$$

and we will call it the CTD model. The TD covariate version of the Sundaram (2006) proportional odds model is

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\{\frac{1 - S(t \mid X_t)}{S(t \mid X_t)}\right. = \mathrm{e}^{x(t)\beta}\frac{\mathrm{d}}{\mathrm{d}t}\left\{\frac{1 - S_0(t)}{S_0(t)}\right. \qquad X_t = \{x(s) : s \leqslant t\}, \tag{2.8}$$

and we will call it the POTD model. A generalization of the AFT model due to Cox and Oakes (1984) is

$$S(t \mid x_t, \beta, S_0) = S_0\left(\int_0^t \mathrm{e}^{x(s)\beta}\,\mathrm{d}s\right).$$

Hanson et al. show that $S(t \mid x(t), \beta, S_0)$ for all of these models can be written as easily computable functions of S_0 and β.

Hanson et al. (2009, 2011b) place the same MFPT prior on S_0 for all of these models and their model assumes independence of β and S_0; they use an improper uniform prior for β. It is however straightforward to incorporate the informative priors for β that are discussed in Bedrick et al. (2000) for fixed covariates. This is another nice feature of this semiparametric model.

Hanson et al. (2009) analyzed the classic Stanford Heart Transplant data (Crowley and Hu 1977), and data involving cerebral edema in children with diabetic ketoacidosis. We present parts of their analyses below.

Stanford Heart Transplant Data Analysis

These data involve the time to death from after entry into the study, which was designed to assess the effect of heart transplant on survival. Individuals entered the study and either received a donor heart at some point according to availability of an appropriate heart and a prioritization scheme, or they left the study and possibly died before a suitable heart was found. The main TDC considered was an indicator of having received a heart, yes or no, at each time t. The second and third TDCs were a mismatch score that indicated the quality of the match between donor and recipient hearts, which was centered at 0.5, and age at transplant (AT), which was centered at 35 years. These TDCs switched on when a heart was transplanted.

Crowley and Hu (1977) and Lin and Ying (1995) analyzed these data using the CTD and COTD models, respectively. Hanson et al. (2009) fit these models and the PKTD model using the same MFPT prior on the baseline survivor function with a log logistic base-measure. They truncated the trees at $M = 5$ levels, fixed the PT weight at one, and placed an improper constant prior on β.

Patients not receiving a new heart have TDC process for the heart transplant, age and mismatch score (MS) that are all zero for all t. Let z_i denote the time of transplant for individual i if they did receive a transplant, and define the TDCs

$$x_{i1}(t) = \begin{cases} 0, & \text{if } t < z_i, \\ 1, & \text{if } t \geqslant z_i, \end{cases}$$

and

$$x_{i2}(t) = \begin{cases} 0, & \text{if } t < z_i, \\ \text{AT} - 35, & \text{if } t \geqslant z_i, \end{cases} \quad x_{i3}(t) = \begin{cases} 0, & \text{if } t < z_i, \\ \text{MS} - 0.5, & \text{if } t \geqslant z_i. \end{cases}$$

Let $x_i(t) = (x_{i1}(t), x_{i2}(t), x_{i3}(t))^{\mathrm{T}}$. Results from the three posterior distributions are displayed in Table 2.2.

The models are decisively ranked in the order CTD, COTD, and PKTD, using the LPML criterion. The integrated Cox–Snell residual plots (not shown) were consistent with this ranking and showed nothing that could be construed as extreme lack of fit for any of the models. The CTD model shows statistical importance for status and age but not for mismatch, while the other models do not indicate the statistical importance of status.

Table 2.2 Posterior inference (posterior medians and 95 % probability intervals) for Stanford Heart Transplant data; the PKTD and CTD models are, respectively, based on (2.6) and (2.7)

	Model		
Parameter	PKTD	COTD	CTD
Status	$-1.76\ (-3.86, 1.57)$	$-1.10\ (-2.70, 0.50)$	$-1.04\ (-1.99, -0.17)$
AT−35	$0.10\ (-0.02, 0.26)$	$0.05\ (-.004, 0.13)$	$0.06\ (\ .015,\ 0.11)$
MS−0.5	$1.63\ (-0.38, 3.89)$	$0.64\ (-0.30, 1.52)$	$0.49\ (-0.09,\ 1.03)$
LPML	-468.0	-467.0	-464.1

AT denotes age at transplant while MS denotes mismatch score

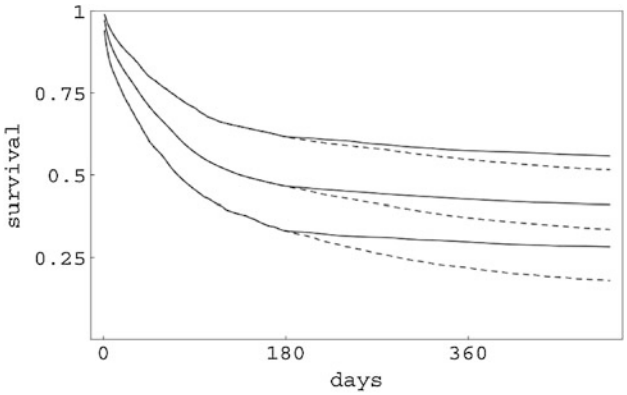

Fig. 2.3 Estimated survival curves and 95 % probability intervals for individuals with mismatch score 0.5 and age 35. *Solid line* is for individual with a heart transplant at 6 months and *dashed line* is for an individual with no heart transplant

Under the CTD model, Hanson et al. (2009) considered two individuals aged 35 years with mismatch scores of 0.5. The first individual did not receive an HTP while the second did after 6 months. The relative hazard comparing the individual with the no heart transplant to the one with the heart transplant is of course one from time zero to 6 months, and is $e^{-\beta_1}$ from that time on. A 95 % posterior probability interval for the relative hazard after 6 months is (1.19, 7.31), and the posterior median is 2.83. Figure 2.3 displays estimated survivor curves for these two individuals, and their 95 % limits. They also fitted the MFPT with a parametric exponential base that resulted in quite different estimates of regression coefficients. The LPML for this model was −486.3, much smaller than any value in Table 2.2. Chen et al. (2014) later found an AFT model that fit the Stanford data better.

Cerebral Edema Data Analysis

The data analyzed here were collected by Glaser et al. (2001), who assessed risk factors associated with the onset of cerebral edema (CE) in children with diabetic ketoacidosis. The description that follows is taken from Hanson et al. (2009):

> Cerebral edema is a dangerous complication associated with emergency department and in-patient hospital care of children with diabetic ketoacidosis. Children with symptoms of diabetic ketoacidosis are initially treated in the emergency department, then moved to the hospital, typically the pediatric intensive care unit, over the course of 24 h. The main purpose of treatment is to normalize blood serum chemistry and acid-base abnormalities. A major, but infrequent complication of children associated with diabetic ketoacidosis and its treatment is CE, or swelling in the brain, which may result in death or permanent neurological damage.

Hanson et al. consider only the children in that study who developed CE ($n = 58$). Their goal was to ascertain the effect of treatment procedures in time and fixed covariates on the timing of CE.

Upon admission, various treatments were recorded hourly for up to 24 h, and several initial measurements taken. The only fixed variable considered was age. Two types of TDCs are considered, the first involving the monitoring of biochemical variables over time; Hanson et al. considered serum bicarbonate (BIC) (concentration in the blood measured in mmol per liter) and blood urea nitrogen (BUN) (mg/deciliter). The second type involved actions by physicians; Hanson et al. used fluids administered (FL) (volume of fluids in ml/Kg/hour) and sodium administered (NA) (mEq/Kg/hour). None of the event times are censored. They again used the log logistic family to center the three MFPT survival models, and they set the number of levels for the finite tree to be $M = 4$ and the weight to be one. Table 2.3 gives posterior summaries of the analysis of all three models.

Table 2.3 Posterior inference (posterior medians and 95 % probability intervals) for cerebral edema data

	Model		
Parameter	PKTD	COTD	CTD
Age (Fixed)	0.028 $(-0.01, 0.08)$	0.021 $(-0.02, 0.07)$	0.044$(-0.02, 0.11)$
Serum-BUN (TD)	-0.005 $(-0.02, 0.01)$	-0.01 $(-0.022, 0.005)$	0.00 $(-0.03, 0.03)$
Serum-BIC (TD)	0.04^* $(-0.01, 0.13)$	0.05^* $(-0.02, 0.12)$	0.06^* $(-0.05, 0.17)$
Serum-BIC^2 (TD)	-0.005 $(-0.01, 0.006)$	-0.006 $(-0.02, 0.003)$	-0.007 $(-0.02, 0.005)$
Adm-FL (TD)	-0.03 $(-0.09, 0.03)$	-0.05 $(-0.10, 0.02)$	-0.05 $(-0.15, 0.04)$
Adm-NA (TD)	0.60^* $(0.16, 0.93)$	0.74^* $(0.18, 1.2)$	0.90^* $(0.19, 1.57)$
FL$\times$NA (TD)	-0.011^* $(-0.03, -0.00)$	-0.013^* $(-0.03, 0.001)$	-0.014^* $(-0.04, 0.003)$
LPML	-176	-176	-175

BUN denotes blood urea nitrogen, BIC denotes bicarbonate, while NA denotes sodium administered; the PKTD and CTD models are, respectively, based on (2.6) and (2.7)

Integrated Cox–Snell residual plots did not show radical departures from the assumption of a correct model for any of the three models. Table 2.3 gives LPML values for each model, and there is no obvious distinction among the models according to this criterion. Estimates of regression coefficients for all variables in the models have the same sign and general magnitude across models. Under all models, there is a 99 % posterior probability that the coefficient for Admin-NA is positive and at least a 96 % posterior probability that the coefficient for the interaction is negative. The Serum-BIC variable has at least a 94 % probability of being positive across models; thus the effect of sodium administration appears to be modified by fluids administration. However, the estimated relative hazard under the CTD model, comparing two patients identical in all respects, including the administration of k units of fluids and with the numerator patient having an increase of one unit in NA administration over the patient in the denominator, would be $\exp(0.9 - 0.014k)$. The effect modification of fluids is thus demonstrated. For small values of k, there would be little practical import.

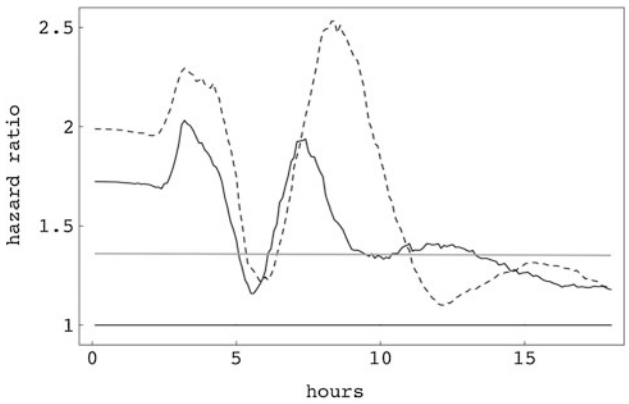

Fig. 2.4 Cerebral edema hazard ratio for subject with NA = 0.7 versus NA = 0.35; the *black dashed*, and *solid lines* correspond, respectively, to COTD and PKTD, whereas the *solid gray line* corresponds to CTD; the PKTD and CTD models are, respectively, based on (2.6) and (2.7)

Hence, according to all models, larger values of Serum-BIC are associated with earlier diagnosis of CE. For example, under the CTD model, comparing two children that are otherwise being treated the same over a period of time and who are of the same age, the hazard of cerebral edema for a child with a larger value for BIC will be greater than for one with a lower value.

The posterior density estimates and hazard functions for time to CE corresponding to patients with specified TDC profiles are simple to obtain. Consider hypothetical patients 1 and 2 of age 10, BUN = 35, fluids constant at 3.6, and BIC increasing from 5 to 22, as was the case for patient 5 in the data. Figure 2.4 presents an estimated relative hazard comparing hypothetical subject 1, who has NA constant at 0.7, to hypothetical subject 2, who has NA constant at 0.35. Observe that the CTD model gives a constant relative hazard since the only difference in the two subjects is a TDC that is remaining constant over time for both subjects. According to this

model, subject 1 is estimated to be about 1.35 times as much at risk of CE as subject 2 for all times. Under the PKTD and COTD models, subject 1 is usually at higher risk of CE, but the estimated relative risk varies considerably over the first 18 h. Observe the similarity of shapes of these two relative hazards, with both peaking twice.

2.4.3 A Nonparametric Survival Regression Model

We now discuss the approach by De Iorio et al. (2009) who model censored survival data using a DPM of linear regression models, and which can be shown to be a DDP model. We discuss their analysis of breast cancer data from a cancer clinical trial after describing the model. The model was developed because it was anticipated that survival curves for different treatments would cross each other, which would contraindicate the use of PH, AFT, and PO models.

If we were to posit a parametric survival regression model for the data, we could use the log normal, log logistic, or log extreme value families, among others. These models can be expressed as

$$\log(T) = x^{\mathrm{T}}\beta + \sigma W,$$

where x is a vector of covariates with a one in the first slot for the intercept. We could let W have an $\mathrm{N}(0,1)$, or $\mathrm{Logistic}(0,1)$, or an Extreme-Value(0,1) (re-parameterized to have median zero) distribution. Let $f(t \mid x, \beta, \sigma)$ be the density for an individual with covariate x from one of these models, and let

$$f(t \mid x, G) = \int f(t \mid x^{\mathrm{T}}\beta, \sigma)\, \mathrm{d}G(\beta, \sigma),$$

with $G \sim \mathrm{DP}(\alpha, G_\theta)$ and $\theta \sim p(\theta)$. This is a DPM of regression models where the base of the DP can possibly have unknown parameters and where a further distribution is placed on them.

For simplicity, consider the case with a simple binary covariate, v, and a single continuous covariate, z. Then $x^{\mathrm{T}} = (1, v, z)$ takes on the values $(1,0,z)$ or $(1,1,z)$. So the parametric version of this model would be an analysis of covariance model in the log of the response. Let x_i denote the covariate for individual i, for $i = 1, \ldots, n$. Then $x_i^{\mathrm{T}}\beta = \beta_0 + z_i\beta_2$ or $\beta_0 + \beta_1 + z_i\beta_2$. Let $\mathscr{X} = \{x_i : i = 1, \ldots, n\}$ and let G_{x_i} be the induced distribution on $x_i^{\mathrm{T}}\beta$ that is derived from the DP distribution on G. The collection $\{G_{x_i} : i = 1, \ldots, n\}$ is a DDP for which the DPM distributions corresponding to the n observations in the data are dependent. The model is termed a linear DDP by De Iorio et al. (2009), and interested readers can find details about the choice of G_θ and $p(\theta)$ there. Another nice feature of this model is that it can be fit in `DPpackage` (Jara et al. 2011).

Breast Cancer Data Analysis

De Iorio et al. (2009) illustrate the proposed approach using data on 761 women from a breast cancer clinical trial. Survival times in months are the times until death, relapse, or treatment-related cancer, or censoring. Fifty three percent of the 761 observations are censored. Interest lies in determining whether a high dose of the treatment is more effective overall for treating cancer compared with lower doses. High doses of the treatment are known to be more toxic. It was hoped that the initial risk associated with toxicity would be offset by a subsequent improvement in survival prospects. The main goal of the clinical trial was to compare high versus low dose survival rates.

Two categorical covariates were considered; treatment dose ($-1 =$ low, $1 =$ high) and estrogen receptor (ER) status ($-1 =$ negative, $1 =$ positive or unknown); standardized tumor size was also considered as a continuous covariate, and an interaction between treatment and ER was also included in the model. The centering distribution was log normal.

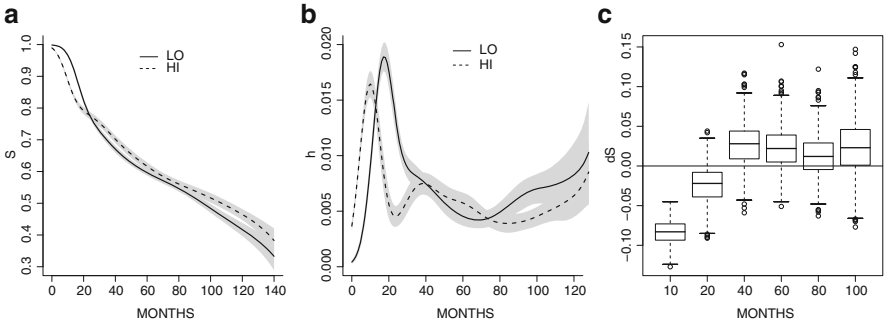

Fig. 2.5 Inference for high versus low dose. (**a**) Estimated survivor functions (*solid lines*) along with pointwise 50 % probability intervals (*grey bands*). (**b**) Estimated hazard functions (*solid lines*) along with pointwise 50 % probability intervals (*grey bands*). (**c**) Box-plots for posterior distribution of the difference in survival rates at 10, 20, 40, 60, 80, and 100 months between a patient who receives high treatment dose versus a patient who receives the low dose. Remark: (**a**), (**b**), and (**c**) correspond to positive ER status and tumour size equal 2.0

Figure 2.5a,b show the posterior survival and hazard function estimates with their corresponding posterior uncertainty for ER positive patients with tumor size 2.0 cm (equal to the first quartile). As expected, the survivor functions corresponding to the two treatment groups cross, showing a higher level of risk associated with high treatment dose in the first 20 months. Figure 2.5c shows box plots corresponding to posteriors for the difference in survival rates between the two treatment groups for positive ER status and tumour size equal 2.0 cm, across a range of times. There is a statistically important negative effect of high dose due to toxicity early in the study, and a non-statistically important positive effect later in the study. Ultimately, the high dose treatment was abandoned as a result of the study.

2.5 Joint Modeling of Longitudinal and Survival Data

Many studies entail an event/survival time of interest and measurements on longitudinal processes that might be associated with patient prognosis. Examples include:

- Blood pressure measurements in dialysis patients (event: Death).
- Daily fertility counts in Mediterranean fruit flies (event: Death).

In the former case, maintaining blood pressure to be sufficiently high plays a key role in long-term prognosis for dialysis patients. In the latter case, it has been argued in the literature that life span of fruit flies might be related to fertility (see Hanson et al. 2011b, for references).

Hanson et al. (2011b) developed a general Bayesian semiparametric methodology for joint analysis. They illustrated and compared Bayesian joint models in which the survival component was taken to be the POTD, CTD, or COTD models that were discussed in Sect. 2.4.2. Comparisons were made using the LPML criterion for model selection. In each instance, baseline survival functions were modeled, as in Sect. 2.4.2, with an MFPT prior.

Two-stage procedures involve modeling the observed longitudinal processes, as, for example, was done in Sect. 2.3.1. That model is then used to predict the 'true' underlying processes, namely the process without measurement error. The predicted processes are then used as if they were the observed TDCs in fitting the time to event data with the TDC survival models discussed in Sect. 2.4.2. Subsequently, we term analyses that condition on the observed processes using LOCF (last observation carried forward) as 'raw' analyses.

Drawbacks of raw and two-stage methods motivated a considerable flourish of research on joint models for longitudinal and survival data (see Tsiatis and Davidian 2004, for a review up to that time). Bayesian approaches to joint analysis include Faucett and Thomas (1996), Wang and Taylor (2001), and Brown and Ibrahim (2003), among others. Joint modeling would appear to be a good idea since one would expect potential benefits from modeling all of the stochastic data, especially when there is the possibility of considerable measurement error, which would be the case when measuring blood pressure, and also beneficial when observations on the process are spaced out in time.

A joint analysis, on the other hand, involves simultaneously modeling longitudinal and survival data and making inferences about the effect of the true process on survival in a single stage of analysis. Let $y(t)$ be the observed vector process. This can be regarded as the vector TDC process discussed in Sect. 2.4.2, only now we consider modeling it rather than simply conditioning on it. Since we expect most processes to be observed with error, let $x(t)$ denote the 'true' (vector) process. In the absence of measurement error $y(t) = x(t)$.

A joint model involving a single process proceeds as follows. All of the models considered involve a baseline survivor curve, S_0, and a regression coefficient vector, β. In each instance, we specify

$$S_0 \mid \theta \sim \text{MFPT}(M, c, G_\theta), \quad \theta \sim p(\theta), \quad p(\beta) \propto \text{const},$$

namely the baseline survivor function has an MFPT prior and the regression coefficients have an improper constant prior distribution. The data consist of $\{(T_i, y_i, t_i) : i = 1, \ldots, n\}$ where T_i is the minimum of the event time and the censoring time, t_i is the vector of observed times, and y_i is the corresponding vector of observations on the process $y_i(t)$, for individual i, in a sample of size n. We assume that $x_i(t)$ is the 'true' process and that

$$y_i(t) = x_i(t) + \varepsilon_i, \quad \varepsilon_i \sim F_\lambda.$$

If we let F_λ be the distribution function of an $\mathrm{N}(0, \sigma^2)$ distribution, one can glean the particular $x_i(t)$ for model (2.1) in Sect. 2.3.1.

Survival modeling is conditional on the longitudinal process. We model the survivor function for individual i, $S_i(t \mid x(t_i), S_0, \beta)$, using the POTD, CTD, and COTD models discussed in Sect. 2.4.2, and where $x(t_i) = (x(t_{i1}), \ldots, x(t_{ih_i}))^{\mathrm{T}}$. From this, we know the form of the hazard function and the density. Assuming a parametric model of the form $f(x_i \mid \Delta)$, then the full joint model for a non-censored observation is expressed as

$$f(T_i, y_i \mid x_i, \lambda, S_0, \beta) = f(T_i \mid x_i, S_0, \beta) f(y_i \mid x_i, \lambda) f(x_i \mid \Delta).$$

If an observation is censored, replace $f(T_i \mid x_i, S_0, \beta)$ with $S(T_i \mid x_i, S_0, \beta)$, making the usual assumption that event times and censoring times are independent. We have made the assumption that T_i is conditionally independent of the observed process given the true process and the parameters. Details on inference can be found in Hanson et al. (2011b).

2.5.1 Medfly Data Analysis

The data used for illustration came from a study reported in Carey et al. (1998) and further analyzed by Chiou et al. (2003) and Tseng et al. (2005). Tseng et al. (2005) analyzed a sample size of 251 Mediterranean fruit flies with lifetimes ranging from 22 to 99 days. The number of eggs produced per day was recorded throughout their lifespan. We removed the first 2 days from each trajectory since all flies have zero counts on those days.

We present some of the analysis presented in Hanson et al. (2011b). Our case study makes the point that joint or two-stage modeling may not predict as well as simply conditioning on the 'raw' process, for these data. For comparison with the analysis by Tseng et al. (2005), Hanson et al. used the same longitudinal model as they did, as well as some additional more flexible alternatives. Tseng et al. let $y_i(t) = \log\{N_i(t) + 1\}$, the natural log of one plus the number of eggs laid on day t, and modelled trajectories as

$$y_i(t) | (b_{i1}, b_{i2}), \tau \sim \mathrm{N}(b_{i1} \log(t) + b_{i2}(t-1), \tau^{-1}), \quad (b_{i1}, b_{i2}) \mid \mu, \Sigma \overset{\text{iid}}{\sim} \mathrm{N}_2(\mu, \Sigma).$$

where the mean is a log gamma function. Since there are no additional covariates for survival, a single regression coefficient β connects the survival model to the longitudinal process $x_i(t) = b_{i1}\log(t) + b_{i2}(t-1)$. The MFPT models used here set $M = 4$ and $c = 1$, with flat priors otherwise. About 16 observations fall into each of the 16 sets at level $M = 4$ if the log logistic family is approximately correct. They also considered the prior $c \sim \text{Gamma}(5,1)$ for a subset of models, obtaining LPML values slightly smaller than with fixed $c = 1$.

All models were fitted with both the MFPT with weight $c = 1$, and parametric log logistic model, corresponding to a weight that grows without bound. According to the LPML statistics presented in Table 2.4, the COTD model performs the worst in this data analysis, regardless of the method used to incorporate the longitudinal predictor (e.g., raw versus modeled) or whether parametric versus MFPT for S_0 was assumed. For the two types of raw analysis, the flexibility obtained from an MFPT generalization of the log logistic model improves predictive performance, though not dramatically so. Moreover, it is also clear that two-stage and joint methods predict almost identically but are inferior to simple raw analysis in this setting. Observe from Table 2.5 that point estimates of β under the POTD model are similar across types of analysis and that they are different for the COTD model.

From Table 2.4, the general conclusions about predictive model comparison are that a raw LOCF analysis is preferred to two-stage or joint methods, the POTD model is preferred over the COTD and CTD models, and that the COTD model might be excluded from further consideration. On the other hand, Tseng et al. (2005) rejected the CTD model based on a test involving Schoenfeld residuals and proposed the COTD model as a plausible alternative. Hanson et al. (2011b) discuss why these data might not be ideal for joint or two-stage modeling beyond the analysis performed here.

Table 2.4 LPML across models (larger is better) for medfly data; the POTD and CTD are, respectively, based on (2.8) and (2.7)

		Model		
Inference	Method	POTD	CTD	COTD
Parametric	Raw	−867	−870	−937
MFPT	Raw	−865	−866	−938
MFPT	Two-stage	−947	−959	−973
Parametric	Joint	−947	−959	−973
MFPT	Joint	−945	−956	−973

Hanson et al. (2011b) also pointed out that not all of the egg count trajectories fit the log gamma structure that is posited for these data. Consequently, they considered a more flexible longitudinal model that represents a compromise between the Tseng et al. approach and using the empirical egg counts (LOCF). They considered a B-spline longitudinal model in conjunction with the POTD model, which resulted in the largest LPML among all models considered, namely LPML $= -879$ for the parametric joint model, worse than parametric raw but much better than using the

basis $\{\log(t), t-1\}$. Hanson et al. also argue that preference for the POTD model over the CTD model in our analysis is tantamount to an acceptance that a change in egg laying behavior at a particular time is eventually forgotten.

Table 2.5 Posterior inference (posterior medians and 95 % probability intervals) across models for medfly data; the POTD and CTD are, respectively, based on (2.8) and (2.7)

	Model		
Method	POTD	CTD	COTD
Par/Raw	−0.75 (−1.02,−0.53)	−0.65 (−0.74,−0.56)	−0.36 (−0.44,−0.27)
MFPT/Raw	−0.74 (−0.85,−0.64)	−0.64 (−0.73,−0.55)	−0.37 (−0.45,−0.29)
MFPT/Two-Stage	−0.74 (−0.97,−0.52)	−0.37 (−0.52,−0.24)	0.16 (−0.01, 0.30)
Par/Joint	−0.78 (−1.02,−0.53)	−0.39 (−0.54,−0.25)	0.19 (0.01, 0.33)
MFPT/Joint	−0.79 (−1.00,−0.52)	−0.40 (−0.54,−0.24)	0.19 (0.01, 0.32)

2.6 Medical Diagnostic Data

2.6.1 ROC Regression

We consider the quality of a medical diagnostic test for its ability to discriminate between alternative states of health, generally referred to diseased/infected $(D+)$ and non-diseased/infected $(D-)$ states. In many settings of clinical interest, covariates can be used to supplement the information provided by a biomarker, and thus can help to discriminate between $D+$ and $D-$. For example, consider diabetes testing, where blood glucose levels are used to diagnose individuals with diabetes. The covariate, age, plays a key role as older subjects tend to have higher levels of glucose, without that necessarily meaning that there is a higher incidence of diabetes at greater ages. However, since the aging process is believed to be associated with relative insulin deficiency or resistance among the $D-$ individuals, it is relevant to adjust for age in the analysis; see Inácio de Carvalho et al. (2013) and the references therein. The general area we now discuss is called ROC regression.

But first briefly consider the no covariate case using a diagnostic marker T. It might be continuous, or dichotomous. If it is dichotomous, the marker outcomes are $T+$, or yes, the individual tested has the infection/disease, or $T-$, or no, they don't. In the case of a continuous marker, a cutoff, c, is selected and, without loss of generality, if the marker value exceeds the cutoff, the outcome is $T+$, and is $T-$ otherwise. In either case, observing the yes/no outcome is called a diagnostic test. The quality of the test is determined by considering two types of test accuracy. The sensitivity of the test is defined to be $\text{Se} = \Pr(T+ \mid D+)$, the proportion of the time that the test says yes when it should, and the specificity, $\text{Sp} = \Pr(T- \mid D-)$, the proportion of time the test says no when it should. In the continuous case, we write

$\mathrm{Se}(c)$, and $\mathrm{Sp}(c)$, and in this case, it is common to plot the false positive rate versus the true positive rate across all possible cutoffs. The ROC curve for a continuous biomarker is thus the plot $\{(1-\mathrm{Sp}(c), \mathrm{Se}(c)) : \text{for all } c\}$. It is possible to re-write this plot as $\{\mathrm{ROC}(t) : t \in [0,1]\}$, where $\mathrm{ROC}(t) = 1 - F_{D+}\{F_{D-}^{-1}(1-t)\}$, (Pepe 2003, Chap. 4), where $F_{D+}(\cdot)$ and $F_{D-}(\cdot)$ are the distribution functions for $D+$ and $D-$ individuals. We now extend this to include adjustment for covariates, x.

The key object of interest for modeling in Sect. 2.6.2 is the covariate-adjusted ROC curve, which can be defined just as in the no covariate case, only now $\mathrm{Se}(c)$ and $\mathrm{Sp}(c)$ are allowed to depend on covariates, x. So for every x, we have an ROC curve. Here, we define the three-dimensional ROC surface:

$$\{(t, x, \mathrm{ROC}(t \mid x)) : t \in [0,1], x \in \mathbb{R}^p\},$$

where

$$\mathrm{ROC}(t \mid x) = 1 - F_{D+}\{F_{D-}^{-1}(1-t \mid x) \mid x\}. \tag{2.9}$$

We now have two conditional distributions that are allowed to depend on covariates. They may depend on distinct covariates, or one may depend on covariates and the other not. The covariate-adjusted AUC is defined as

$$\mathrm{AUC}(x) = \int_0^1 \mathrm{ROC}(u \mid x)\, \mathrm{d}u,$$

and will be used as our preferred summary measure of covariate-adjusted discriminative power.

In some cases a 'perfect' or gold-standard (GS) test exists, i.e., a test that correctly classifies the subjects as $D+$ and $D-$. In this case, data consist of two samples, one known to be $D+$ and the other known to be $D-$. Observed outcomes for each unit consist of the pair

$$\{\text{Test Covariates}, \text{Test Scores}\};$$

we denote test covariates as x. A test score is a continuous diagnostic marker outcome, and a test covariate is simply a covariate that is, at least believed to be, related to a test score. With GS data, the model is identifiable regardless of the amount of separation between F_{D+} and F_{D-}; the case where a gold standard test exists is considered in detail in Chap. 16 (Inácio de Carvalho et al. 2015).

Section 2.6.2 focuses on ROC regression for the no gold-standard (NGS) case, thus there is no direct information on whether individual subjects in a study are $D+$ or $D-$. The data consist of a single mixed sample with disease status unknown. The NGS setting typically involves identification issues. However, if there are covariates that allow us to learn about the probability of disease, the model is identifiable under mild assumptions (see Branscum et al. 2015, Appendix 1). We refer to these as disease covariates, and denote them as x^*. Hence in this setting we assume that data consist of the triple,

$$\{\text{Disease Covariates}, \text{Test Covariates}, \text{Test Scores}\}.$$

The model discussed in Sect. 2.6.2 was proposed by Branscum et al. (2015), and it was built on the principle that Disease Covariates can be used to mitigate identification issues in the NGS setting. See Branscum et al. (2013) for another GS approach to this problem, and also see Branscum et al. (2008) for an approach that develops much of the machinery used here.

2.6.2 A Semiparametric ROC Regression Model in the Absence of a Gold Standard Test

Here we assume there are no test covariates available for $D-$ subjects. For $D+$ subjects we specify the model $Y_{D+} = x^{\mathrm{T}}\beta + \varepsilon_{D+}$, where x is a test covariate, β is a coefficient vector, and $\varepsilon_{D+} \sim F_{\varepsilon_{D+}}(\cdot)$. With this specification, (2.9) can be rewritten as

$$\mathrm{ROC}(t \mid x) = 1 - F_{\varepsilon_{D+}}\{F_{D-}^{-1}(1-t) - x^{\mathrm{T}}\beta \mid x\},$$

by noting that $F_{\varepsilon_{D+}}(y - x^{\mathrm{T}}\beta) = F_{D+}(y \mid x)$.

Suppose continuous marker scores (y_i) are obtained on n randomly sampled individuals from a population. Then let x_i^* denote the disease covariate outcome, and let z_i denote latent disease status for subject i, with $z_i = 1$ if they are $D+$, and $z_i = 0$ otherwise. Define π_i as the probability that subject i is $D+$, for $i = 1, \ldots, n$. The latent z_is are independent and Bern(π_i), with $\pi_i = G_0(x_i^{*\mathrm{T}}\alpha)$, with $\alpha = (\alpha_0, \ldots, \alpha_s)^{\mathrm{T}}$ and where G_0 is a standard distribution function, like normal, or logistic. These choices result in probit and logistic regression models for the z_is. Test scores are modeled according to a mixture distribution with conditional density,

$$f(y_i \mid z_i, x_i) = z_i f_{\varepsilon_{D+}}(y_i - x_i^{\mathrm{T}}\beta) + (1 - z_i) f_{D-}(y_i),$$

where $\beta = (\beta_0, \ldots, \beta_p)^{\mathrm{T}}$, $f_{\varepsilon_{D+}}$ is the density associated with $F_{\varepsilon_{D+}}$, and f_{D-} is the density associated with F_{D-}. The model for $D-$ subjects can also depend on covariates; test and disease covariates may overlap.

The nonparametric part of the model involves placing independent MFPT priors on $F_{\varepsilon_{D+}}$ and F_{D-}; here, $F_{\varepsilon_{D+}}$ is constrained to have median zero to alleviate confounding between β_0 and the location of $F_{\varepsilon_{D+}}$ (Hanson and Johnson 2002). Since the marker was log transformed, the MFPTs were centered on normal families, the former family having mean zero and the latter having an arbitrary mean. Weights for the PTs were either specified to be one, or given a diffuse gamma distribution. Parametric priors were placed on all hyperparameters. See Branscum et al. (2015) for further details.

Lung Cancer Data Analysis

Branscum et al. (2015) investigated the potential of a soluble isoform of the epidermal growth factor receptor (sEGFR) to be considered as a diagnostic biomarker for lung cancer in men. The data were gathered a case-control study that was conducted at the Mayo Clinic. The data included 88 controls and 139 lung cancer cases; see Baron et al. (1999, 2003) for further details. Branscum et al. (2015) analyzed the data as if disease status was unknown and used these data to assess the impact of age on the discriminatory ability of sEGFR to distinguish cases and controls. Age was used as a test covariate for controls, and as a disease covariate. They also analyzed the data using known disease status in a GS analysis of the same data for comparative purposes.

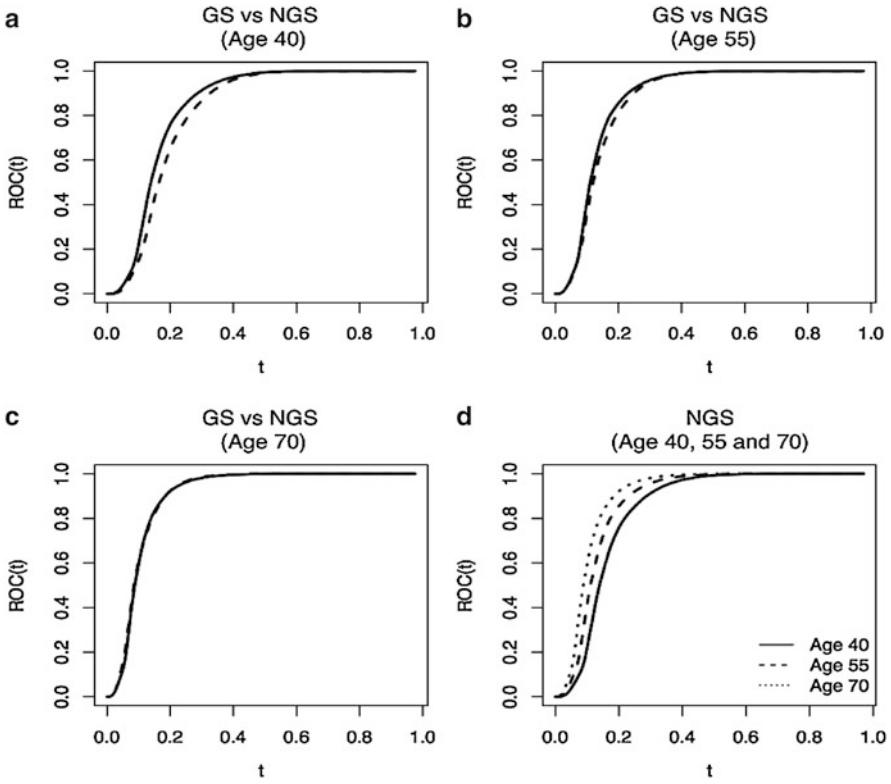

Fig. 2.6 GS and NGS semiparametric estimates of covariate adjusted ROC curves for ages 40, 55, and 70

The sampling model for the natural log transformed test scores and latent diseased status was:

$$z_i \sim \text{Bern}(\pi_i), \quad \log\left(\frac{\pi_i}{1-\pi_i}\right) = \alpha_0 + \alpha_1 x_i^*,$$
$$f(y_i \mid z_i, x_i) = z_i f_{\varepsilon_{D+}}(y_i - \beta_0 - \beta_1 x_i) + (1 - z_i) f_{D-}(y_i).$$

In Fig. 2.6 we plot semiparametric estimates of the covariate-adjusted ROC curves corresponding to ages 40, 55, and 70. Posterior inferences for covariate-adjusted AUCs for the same ages are displayed in Table 2.6. It is clear that it is easier to diagnose lung cancer in older men than in younger men, and that the NGS analysis provides a reasonable approximation to the GS analysis for these data. As expected, interval inferences are less certain in the NGS case than in the GS case.

LPML and corresponding pseudo Bayes factors were used to compare parametric and semi-parametric models. In the NGS setting, the LPML for the parametric normal model was -439, which was larger than the values for all semi-parametric models considered. The largest LPML statistic for all models considered was -422,

Table 2.6 Posterior inference (posterior medians and 95 % probability intervals) for the covariate-adjusted AUCs corresponding to ages 40, 55, and 70 based on GS and NGS analyses of the lung cancer data

	Analysis	
Parameter	GS	NGS
AUC_{40}	0.78 (0.72, 0.84)	0.79 (0.71, 0.86)
AUC_{55}	0.83 (0.77, 0.88)	0.83 (0.75, 0.89)
AUC_{70}	0.87 (0.81, 0.92)	0.86 (0.77, 0.92)

for a model with the two MFPTs truncated at four levels and with both weights equal to one. Compared to the parametric model, the pseudo Bayes factor of e^{17} provides strong evidence in favor of the selected semi-parametric model.

2.6.3 Joint Longitudinal Diagnostic Outcome Modeling and Analysis

Most diagnostic outcome data are cross-sectional, as was the case in the previous section. A main goal in those studies was to estimate sensitivity and specificity of one or more biomarker outcomes over a range of cutoffs, resulting in an estimate of the ROC curve. With cross-sectional data, by definition, sampled individuals include a cross-section of the population. Individuals in this population are either diseased/infected, $D+$, or not, and if they are $D+$, there will be a range of times at which the disease/infection was acquired. For many such maladies, the ability to detect will very much depend on the time of acquisition. For example, it is practically impossible to detect HIV infection in the near term after infection. However,

after some time has passed, ELISA and Western Blot tests are able to detect it. If the cross-sectional sample happened to include only newly infected individuals, the estimated sensitivity of the test would be quite low. The purpose of developing the model discussed below was to consider longitudinal or prospective diagnostic outcome data so that it would be possible to estimate the sensitivity of a dichotomous outcome test as a function of time from infection. A major difficulty faced in this endeavor is that it would rarely be known precisely when individuals in a population or sample become infected, or even in many instances if they had become infected. If a perfect/gold standard test is applied, the actual disease status could be known, but not the exact timing. The model developed below does not assume a gold standard and as a result, the latent status and timing of infection/disease are modeled.

Norris et al. (2009, 2014) developed a model for repeated observations in time on a yes/no diagnostic test outcome and a continuous biomarker for a disease. They analyzed longitudinal fecal culture and continuous serum ELISA outcomes for mycobacterium avium paratuberculosis (MAP), the causal agent for Johne's disease in dairy cattle. We discuss their model and analysis in the context of the cow data, but the model would apply to many other data sets as suggested by Norris et al. (2009, 2014).

Once an animal is infected, it is expected that, after some delay, serum antibody outcomes will increase. If animals are being monitored in time, as they are in the cow data set, antibodies should increase to a point that the ELISA outcome exceeds a cutoff, and thus becomes positive for MAP. If an animal is not infected during the study, their ELISA outcomes should remain steady but variable around some baseline value that depends on the cow. The model includes a latent disease status indicator for all cows, and a change point corresponding to time of infection, t^*, for animals with a positive disease indicator. The probability of a positive fecal test changes at the time of infection, but the rise in serology score occurs some time later. Norris et al. noted that there was literature that pointed to a 1 year lag after infection. Nonetheless, they modeled lag as an unknown parameter. After the lag, increase in antibodies was modeled to be linear. They also assumed that fecal and serology results are independent for several reasons discussed in their paper. The model takes account of the fact that the fecal test is viable soon after infection whereas the production of detectable serum antibodies involves a lag.

The model incorporates three latent states: (1) no infection during the entire screening period, (2) infection, but insufficient time to mount an antibody reaction during screening period (since "lag" has not elapsed when screening ends), and (3) infection with antibody reaction within screening period (since "lag" elapses before the end of screening period). They define the variable, $k_i \in \{1,2,3\}$, to denote the latent disease state of cow i, and they define t_{ij} to be the time of the jth screening for the ith subject; (S_{ij}, F_{ij}) are the serology and fecal culture outcomes of the ith subject at time t_{ij}; Se_F is the sensitivity of fecal culture; Sp_F is the specificity of fecal culture;lag is the time interval between infection and serology reaction, Θ denotes vector of all model parameters, and U is the vector of all model latents. Figure 2.7 describes the model, discussed below, for a cow with $k_i = 3$.

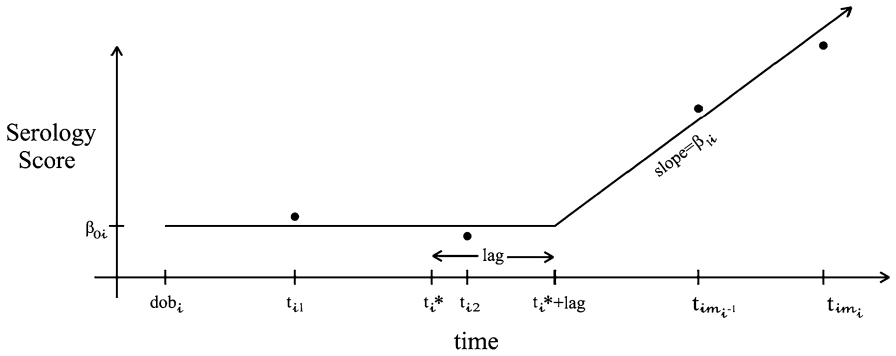

Fig. 2.7 Serology trajectory with data for cow with $k_i = 3$

The models for cows in latent states 1 and 2 are:

$$S_{ij} \mid \Theta, U, k_i = 1 \sim \beta_{0i} + \varepsilon_{ij}, \quad \perp \quad F_{ij} \mid \Theta, U, k_i = 1 \sim \text{Bern}(1 - \text{Sp}_F),$$
$$S_{ij} \mid \Theta, U, k_i = 2 \sim \beta_{0i} + \varepsilon_{ij} \quad \perp \quad F_{ij} \mid \Theta, U, k_i = 2 \sim \text{Bern}(\pi_{ij}),$$

where $\beta_{0i} \overset{\perp}{\sim} \text{N}(\beta_0, \tau_{\beta_0})$, $\varepsilon_{ij} \overset{\perp}{\sim} \text{N}(0, \tau_e)$, $\beta_{0i} \perp \varepsilon_{ij}$, and $\pi_{ij} = I(t_{ij} \geqslant t_i^*)\text{Se}_F + I(t_{ij} < t_i^*)(1 - \text{Sp}_F)$ for all i, j. The model for cows in latent state 3 incorporates a random cow-specific slope for the post-lag serology trajectory, allowing for differing rates of antibody production among infected cows. The function z^+ equals z if $z > 0$ and 0 otherwise. The model is:

$$S_{ij} \mid \Theta, U, k_i = 3 \sim \beta_{0i} + \beta_{1i}(t_{ij} - t_i^* - \text{lag})^+ + \varepsilon_{ij}, \ \perp \ F_{ij} \mid \Theta, U, k_i = 3 \sim \text{Bern}(\pi_{ij}),$$

with β_{1i},β_{0i}, and ε_{ij} pairwise independent; β_{1i} is zero until $t_{ij} = t_i^* +$ lag. Hence, the mean serology trajectory is a flat line until t_i^*+lag, then it increases linearly with slope β_{1i} as shown in Fig. 2.7. We refer the interested reader to Norris et al. for details about the change points, which were modeled with uniform distributions over appropriate ranges, and the disease status variable, which is a simple multinomial for each cow but requires reversible jump methodology to handle the fact that, from one iteration to the next of the Gibbs sampler, the dimension of the parameter space changes according to the (latent status) multinomial outcomes for all n cows.

Norris et al. (2009) analyzed the cow data using the above parametric model, and Norris et al. (2014) extended this model to allow for a DPM of slopes for $k_i = 3$ type cows. The scientific motivation for this was because it was believed that some infected cows may have a more gradual slope, while others a steeper slope after the infection time plus lag. Thus a DPM of slopes will allow for groups of cows with different slopes. Since biology also dictates that antibody slopes must be *non-decreasing* after infection slopes were constrained to be positive by modeling the log-slope as a DPM of normals as follows:

$$\log \beta_{1i} = \gamma_i \mid \mu_i, \tau_i \overset{\perp}{\sim} \mathrm{N}(\mu_i, \tau_i), \quad \text{for } i : k_i = 3,$$
$$(\mu_i, \tau_i) \mid G \overset{\perp}{\sim} G,$$
$$G \mid \alpha, G_0 \sim \mathrm{DP}(\alpha, G_0),$$

which can also be expressed as

$$\gamma_i \mid G \overset{\perp}{\sim} \int \mathrm{N}(\cdot \mid \mu_i, \tau_i) G(\mathrm{d}\mu_i, \mathrm{d}\tau_i), \quad G \mid \alpha, G_0 \sim \mathrm{DP}(\alpha, G_0).$$

Let (n_1, n_2, n_3) be the latent numbers of cows in each of the three latent states. Since G is discrete with probability one, at any given iteration of the Gibbs sampler, there will be, say r, clusters of distinct values among the n_3 realizations of $\theta_i = (\mu_i, \tau_i)$. Cows associated with each of these clusters will have different slopes. At the end of an MCMC run, cows will be belonged to different clusters and corresponding slopes will have changed from iteration to iteration. It is possible to monitor the number of clusters, and the number of modes, at each iteration of the Gibbs sampler and Norris et al. report those results, some of them reproduced below. However, it is impossible to define particular clusters precisely over the entire MCMC sample, due to lack of identifiability of the individual components in the DPM. Nonetheless, through post processing of output, it is possible to allocate cows to clusters that are associated with particular modes in the slope distribution for infected cows using ad hoc methods. The data analysis discussed below uses such a method to make inferences about the sensitivity of the ELISA test, with a particular cutoff, as a function of time since infection for groups of cows deemed to have distinct slopes.

Analysis of Longitudinal Cow Serology and Fecal Culture Data

The estimated sensitivity and specificity of the FC test were 0.57 (0.52, 0.63) and 0.976 (0.955, 0.990), respectively. The FC test is known to be highly specific. The estimated proportions of animals falling into the three latent status groups is (0.048, 0.25, 0.26), thus the estimated prevalence of MAP in the population sampled at the end of the study is 0.52. The estimated lag is 1.60 (1.32, 1.85), in years.

Figure 2.8 shows some iterates from the posterior log slope distribution; some are bimodal with global maximum near zero and a smaller mode less than zero. The posterior distribution of the number of modes showed a 0.62 probability of one and 0.30 of two modes.

ROC curves at selected times past infection for estimated high and low serology reaction groups are displayed in Fig. 2.9a. By analyzing the posterior iterates of the log-slope distribution shown in Fig. 2.8, Norris et al. obtained rough estimates of the mean and standard deviation of the high and low clusters. Many of the iterates suggest the low cluster is centered around -1.6 with a standard deviation of about 0.4 and the higher cluster is centered at about 0.6 with standard deviation of 0.9. The curves depicted in Fig. 2.9 show that discriminatory ability is very poor in the

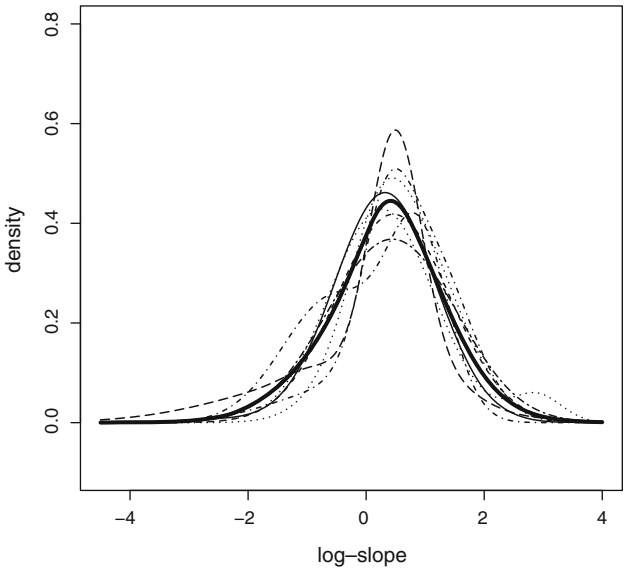

Fig. 2.8 Posterior iterates of log-slope distribution, with posterior mean in *bold*, for cow serology and fecal culture data

hypothetical low group, and can be very good in the hypothetical high group, and is especially so the longer it has been since infection.

The corresponding graph for low and high groups depicting estimated sensitivity of the dichotomized ELISA as a function of time past infection is shown in Fig. 2.9b. There is a large difference in performance of the ELISA between these two groups.

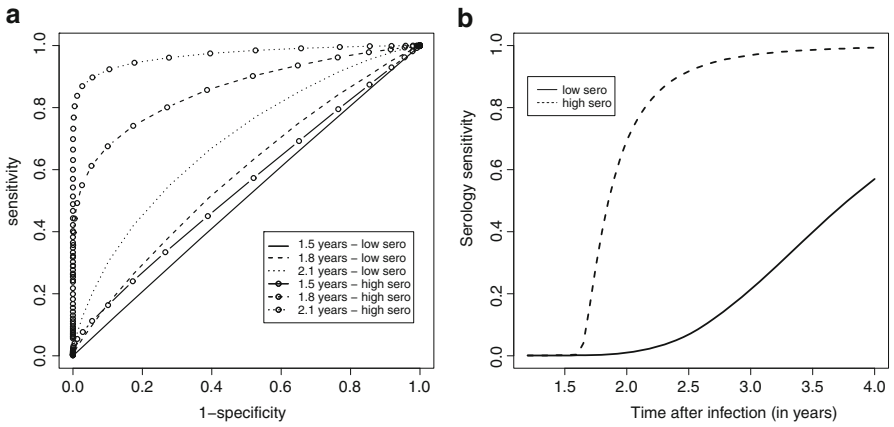

Fig. 2.9 (**a**) Estimated ROC curves for Johne's disease data for hypothetical groups at selected values of time past infection. (**b**) Estimated sensitivity as a function of time for hypothetical high and low serology groups, with a cutoff level of −1.29

At 3 years past infection, the ELISA applied to the 'low' group has estimated sensitivity less than 0.20, while it is one in the 'high' group.

More sophisticated methods of post processing allocation to clusters have been developed by Dahl (2006) and Bigelow and Dunson (2009).

2.7 Final Remarks

What is statistics all about? As put simply by A. Wald:

> The purpose of statistics,..., is to describe certain real phenomena.
> Wald (1952)

Different real phenomena lead us to different types of data, and beyond the ones we have seen above (survival, longitudinal, and medical diagnostic data) there is a wealth of other options arising naturally in biostatistics. These include, for instance:

- **Binary Diagnostic Outcome Data**: Binary diagnostic outcome data are ubiquitous in human and veterinary medicine. While many Bayesian parametric models have been developed, there appears to be a paucity of Bayesian nonparametric approaches in this setting.
- **Compositional Data**: Nonnegative-valued variables constrained to satisfy a unit-sum constraint also find their application in biostatistics. This type of data is known as compositional data; for an application in biostatistics, see Faes et al. (2011), who analyze the composition of outpatient antibiotic use through statistical methods for unit simplex data. Bernstein polynomial-based approaches are tailored for this setting; see, for instance, Petrone (1999) and Barrientos et al. (2015), and the references therein.
- **Functional Data**: Recent advances in technology have led to the development of more sophisticated medical diagnostic data, and, nowadays, applications where measurements are curves or images are becoming commonplace. Dunson (2010, Sect. 3) overviews some recent Bayesian nonparametric approaches for modeling functional data.
- **Missing Data**: In a recent paper at the *The New England Journal of Medicine*, Little et al. (2012) discuss how missing data can compromise inferences from clinical trials. In Chap. 21 (Daniels and Linero 2015) this important subject is considered in detail. An important question that remains after our chapter is: Can we conduct reliable inferences based on the prior processes discussed above, if we have missing data? In terms of Polya trees, Paddock (2002) provides an approach for multiple imputation of partially observed data. Imputation via the Bayesian bootstrap—which can be regarded as a non-informative version of the DP (Gasparini 1995, Theorem 2)—has also been widely applied; more details on the Bayesian bootstrap can be found in Chap. 16 (Inácio de Carvalho et al. 2015).
- **Spatial Data**: This is the subject of Part V of this volume.

- **Time Series Data**: Connected with the topic of longitudinal data is also that of time series data. In this direction some recently proposed models include Nieto-Barajas et al. (2012), Jara et al. (2013), and Nieto-Barajas et al. (2014).

This list continues with multivariate data, shape data, and many more topics, including combinations of the different types of data; see, for example, Chap. 11 (Zhou and Hanson 2015), where models for spatial-survival data are discussed.

We close this introductory part of *Nonparametric Bayesian Methods in Biostatistics and Bioinformatics* with the hope that the next chapters stimulate interaction between experts in Bayesian nonparametric biostatistics and bioinformatics, and that they are useful for those entering this important field of research.

Acknowledgements We thank the Editors for the invitation, and we are indebted to our 'partners in crime,' including Adam Branscum, Ron Christensen, Ian Gardner, Maria De Iorio, Alejandro Jara, Prakash Laud, Michelle Norris, Fernando Quintana, Gary Rosner, and Mark Thurmond. Special thanks go to Vanda Inácio de Carvalho, Tim Hanson, and Peter Müller, who made substantive contributions to the penultimate draft of this paper, in addition to their contributions to the work presented. M. de. C was supported by Fondecyt grant 11121186.

References

Aalen, O. (1980). A model for nonparametric regression analysis of counting processes. In: *Mathematical Statistics and Probability Theory, Lecture Notes in Statistics*, vol. 2, pp. 1–25. New York: Springer.

Amewou-Atisso, M., Ghosal, S., Ghosh, J. K., and Ramamoorthi, R. (2003). Posterior consistency for semi-parametric regression problems. *Bernoulli*, **9**, 291–312.

Baron, A. T., Lafky, J. M., Boardman, C. H., Balasubramaniam, S., Suman, V. J., Podratz, K. C., and Maihle, N. J. (1999). Serum sErbB1 and epidermal growth factor levels as tumor biomarkers in women with stage III or IV epithelial ovarian cancer. *Cancer Epidemiology Biomarkers and Prevention*, **8**, 129–137.

Baron, A. T., Cora, E. M., Lafky, J. M., Boardman, C. H., Buenafe, M. C., Rademaker, A., Liu, D., Fishman, D. A., Podratz, K. C., and Maihle, N. J. (2003). Soluble epidermal growth factor receptor (sEGFR/sErbB1) as a potential risk, screening, and diagnostic serum biomarker of epithelial ovarian cancer. *Cancer Epidemiology Biomarkers and Prevention*, **12**, 103–113.

Barrientos, A. F., Jara, A., and Quintana, F. A. (2015). Bayesian density estimation for compositional data using random Bernstein polynomials. *Journal of Statistical Planning and Inference (DOI: 10.1016/j.jspi.2015.01.006)*.

Bedrick, E. J., Christensen, R., and Johnson, W. (1996). A new perspective on priors for generalized linear models. *Journal of the American Statistical Association*, **91**, 1450–1460.

Bedrick, E. J., Christensen, R., and Johnson, W. O. (2000). Bayesian accelerated failure time analysis with application to veterinary epidemiology. *Statistics in Medicine*, **19**, 221–237.

Berger, J. O. and Guglielmi, A. (2001). Bayesian and conditional frequentist testing of a parametric model versus nonparametric alternatives. *Journal of the American Statistical Association*, **96**, 174–184.

Bigelow, J. L. and Dunson, D. B. (2009). Bayesian semiparametric joint models for functional predictors. *Journal of the American Statistical Association*, **104**, 26–36.

Branscum, A. J., Johnson, W. O., Hanson, T. E., and Gardner, I. A. (2008). Bayesian semiparametric ROC curve estimation and disease diagnosis. *Statistics in Medicine*, **27**, 2474–2496.

Branscum, A. J., Johnson, W. O., and Baron, A. T. (2013). Robust medical test evaluation using flexible Bayesian semiparametric regression models. *Epidemiology Research International*, **ID 131232**, 1–8.

Branscum, A. J., Johnson, W. O., Hanson, T. E., and Baron, A. T. (2015). Flexible regression models for ROC and risk analysis with or without a gold standard. *Submitted*.

Brown, E. R. and Ibrahim, J. G. (2003). A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics*, **59**, 221–228.

Carey, J. R., Liedo, P., Müller, H.-G., Wang, J.-L., and Chiou, J.-M. (1998). Relationship of age patterns of fecundity to mortality, longevity, and lifetime reproduction in a large cohort of mediterranean fruit fly females. *The Journals of Gerontology, Ser. A: Biological Sciences and Medical Sciences*, **53**, 245–251.

Chen, Y., Hanson, T., and Zhang, J. (2014). Accelerated hazards model based on parametric families generalized with Bernstein polynomials. *Biometrics*, **70**, 192–201.

Chiou, J.-M., Müller, H.-G., Wang, J.-L., and Carey, J. R. (2003). A functional multiplicative effects model for longitudinal data, with application to reproductive histories of female medflies. *Statistica Sinica*, **13**, 1119–1133.

Christensen, R. and Johnson, W. (1988). Modelling accelerated failure time with a Dirichlet process. *Biometrika*, **75**, 693–704.

Christensen, R., Johnson, W., Branscum, A., and Hanson, T. E. (2010). *Bayesian Ideas and Data Analysis*. Boca Raton, FL: Chapman & Hall/CRC.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Ser. B*, **34**, 187–220.

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*. Boca Raton, FL: Chapman & Hall/CRC.

Crowley, J. and Hu, M. (1977). Covariance analysis of heart transplant survival data. *Journal of the American Statistical Association*, **72**, 27–36.

Dahl, D. B. (2006). Model-based clustering for expression data via a Dirichlet process mixture model. In: Bayesian Inference for Gene Expression and Proteomics, Eds: Kim-Anh Do, Peter Müller & Marina Vannucci, New York: Springer, pp. 201–218.

Daniels, M. J. and Linero, A. R. (2015). Bayesian nonparametrics for missing data in longitudinal clinical trials. In: *Nonparametric Bayesian Methods in Biostatistics and Bioinformatics,* Eds: R. Mitra & P. Müller, New York: Springer.

De Iorio, M., Johnson, W. O., Müller, P., and Rosner, G. L. (2009). Bayesian nonparametric non-proportional hazards survival modelling. *Biometrics*, **65**, 762–771.

Dennerstein, L., Lehert, P., Burger, H., and Guthrie1, J. (2007). New findings from non-linear longitudinal modelling of menopausal hormone changes. *Human Reproduction Update*, **13**, 551–557.

Diggle, P. J. (1988). An approach to the analysis of repeated measurements. *Biometrics*, **44**, 959–971.

Dunson, D. B. (2010). Nonparametric Bayes applications to biostatistics. In: *Bayesian Nonparametrics*, Eds: N. L. Hjort et al., Cambridge UK: Cambridge University Press, pp. 223–273.

Dunson, D. B. and Park, J.-H. (2008). Kernel stick-breaking processes. *Biometrika*, **95**, 307–323.

Escobar, M. D. (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.

Faes, C., Molenberghs, G., Hens, N., Muller, A., Goossens, H., and Coenen, S. (2011). Analysing the composition of outpatient antibiotic use: A tutorial on compositional data analysis. *Journal of Antimicrobial Chemotherapy*, **66**, 89–94.

Faucett, C. L. and Thomas, D. C. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: A Gibbs sampling approach. *Statistics in Medicine*, **15**, 1663–1685.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, **1**, 209–230.

Ferguson, T. S. (1974). Prior distribution on the spaces of probability measures. *The Annals of Statistics*, **2**, 615–629.

Gasparini, M. (1995). Exact multivariate Bayesian bootstrap distributions of moments. *The Annals of Statistics*, **23**, 762–768.

Geisser, S. and Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, **74**, 153–160.

Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society, Ser. B*, **56**, 501–514.

Gelfand, A. E., Kottas, A., and MacEachern, S. N. (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association*, **100**, 1021–1035.

Glaser, N., Barnett, P., McCaslin, I., Nelson, D., Trainor, J., Louie, J., Kaufman, F., Quayle, K., Roback, M., Malley, R., *et al.* (2001). Risk factors for cerebral edema in children with diabetic ketoacidosis. *The New England Journal of Medicine*, **344**, 264–269.

Hanson, T. and Johnson, W. O. (2002). Modeling regression error with a mixture of Polya trees. *Journal of the American Statistical Association*, **97**, 1020–1033.

Hanson, T. and Johnson, W. O. (2004). A Bayesian semiparametric AFT model for interval-censored data. *Journal of Computational and Graphical Statistics*, **13**, 341–361.

Hanson, T., Bedrick, E. J., Johnson, W. O., and Thurmond, M. C. (2003). A mixture model for bovine abortion and foetal survival. *Statistics in Medicine*, **22**, 1725–1739.

Hanson, T., Sethuraman, J., and Xu, L. (2005). On choosing the centering distribution in Dirichlet process mixture models. *Statistics & Probability Letters*, **72**, 153–162.

Hanson, T., Johnson, W., and Laud, P. (2009). Semiparametric inference for survival models with step process covariates. *Canadian Journal of Statistics*, **37**, 60–79.

Hanson, T. E. (2006). Inference for mixtures of finite Polya tree models. *Journal of the American Statistical Association*, **101**, 1548–1565.

Hanson, T. E., Monteiro, J. V., and Jara, A. (2011a). The Polya tree sampler: Toward efficient and automatic independent Metropolis–Hastings proposals. *Journal of Computational and Graphical Statistics*, **20**, 41–62.

Hanson, T. E., Branscum, A. J., and Johnson, W. O. (2011b). Predictive comparison of joint longitudinal-survival modeling: A case study illustrating competing approaches (with discussion). *Lifetime Data Analysis*, **17**, 3–28.

Inácio de Carvalho, V., Jara, A., Hanson, T. E., and de Carvalho, M. (2013). Bayesian nonparametric ROC regression modeling. *Bayesian Analysis*, **8**, 623–646.

Inácio de Carvalho, V., Jara, A., and de Carvalho, M. (2015). Bayesian nonparametric approaches for ROC curve inference. In: *Nonparametric Bayesian Methods in Biostatistics and Bioinformatics*, Eds: R. Mitra & P. Müller, New York: Springer.

Jara, A., Hanson, T. E., and Lesaffre, E. (2009). Robustifying generalized linear mixed models using a new class of mixtures of multivariate Polya trees. *Journal of Computational and Graphical Statistics*, **18**, 838–860.

Jara, A., Hanson, T., Quintana, F., Müller, P., and Rosner, G. L. (2011). DPpackage: Bayesian semi- and nonparametric modeling in R. *Journal of Statistical Software*, **40**, 1–30.

Jara, A., Nieto-Barajas, L., and Quintana, F. (2013). A time series model for responses on the unit interval. *Bayesian Analysis*, **8**, 723–740.

Johnson, W. and Christensen, R. (1986). Bayesian nonparametric survival analysis for grouped data. *Canadian Journal of Statistics*, **14**, 307–314.

Johnson, W. and Christensen, R. (1989). Nonparametric Bayesian analysis of the accelerated failure time model. *Statistics & Probability Letters*, **8**, 179–184.

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 Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. New York: Wiley.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**, 457–481.

Kottas, A. and Gelfand, A. E. (2001). Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association*, **96**, 1458–1468.

Kuo, L. and Mallick, B. (1997). Bayesian semiparametric inference for the accelerated failure-time model. *Canadian Journal of Statistics*, **25**, 457–472.

Lavine, M. (1992). Some aspects of Polya tree distributions for statistical modelling. *The Annals of Statistics*, **20**, 1222–1235.

Lavine, M. (1994). More aspects of Polya tree distributions for statistical modelling. *The Annals of Statistics*, **22**, 1161–1176.

Lin, D. and Ying, Z. (1995). Semiparametric inference for the accelerated life model with time-dependent covariates. *Journal of Statistical Planning and Inference*, **44**, 47–63.

Little, R. J., D'Agostino, R., Cohen, M. L., Dickersin, K., Emerson, S. S., Farrar, J. T., Frangakis, C., Hogan, J. W., Molenberghs, G., Murphy, S. A., et al. (2012). The prevention and treatment of missing data in clinical trials. *The New England Journal of Medicine*, **367**, 1355–1360.

MacEachern, S. N. (2000). Dependent Dirichlet processes. Technical report, Department of Statistics, The Ohio State University.

MacEachern, S. N. and Müller, P. (1998). Estimating mixture of Dirichlet process models. *Journal of Computational and Graphical Statistics*, **7**, 223–238.

Mauldin, R. D., Sudderth, W. D., and Williams, S. (1992). Polya trees and random distributions. *The Annals of Statistics*, **20**, 1203–1221.

Mitra, R. and Müller, P. (2015). Bayesian nonparametric models. In: *Nonparametric Bayesian Methods in Biostatistics and Bioinformatics*, Eds: R. Mitra & P. Müller, New York: Springer.

Müller, P. and Mitra, R. (2013). Bayesian nonparametric inference—Why and how (with discussion). *Bayesian Analysis*, **8**, 269–302.

Neal, R. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, **9**, 249–265.

Nieto-Barajas, L. E., Müller, P., Ji, Y., Lu, Y., and Mills, G. B. (2012). A time-series DDP for functional proteomics profiles. *Biometrics*, **68**, 859–868.

Nieto-Barajas, L. E., Contreras-Cristán, A., *et al.* (2014). A Bayesian nonparametric approach for time series clustering. *Bayesian Analysis*, **9**, 147–170.

Norris, M., Johnson, W. O., and Gardner, I. A. (2009). Modeling bivariate longitudinal diagnostic outcome data in the absence of a gold standard. *Statistics and its Interface*, **2**, 171–185.

Norris, M., Johnson, W. O., and Gardner, I. A. (2014). Bayesian semi-parametric joint modeling of biomarker data with a latent changepoint: Assessing the temporal performance of Enzyme-Linked Immunosorbent Assay (ELISA) testing for paratuberculosis. *Statistics and its Interface*, **7**, 417–438.

Paddock, S. M. (2002). Bayesian nonparametric multiple imputation of partially observed data with ignorable nonresponse. *Biometrika*, **89**, 529–538.

Pepe, M. S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford, UK: Oxford University Press.

Petrone, S. (1999). Random Bernstein polynomials. *Scandinavian Journal of Statistics*, **26**, 373–393.

Prentice, R. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, **69**, 331–342.

Prentice, R. L. and Kalbfleisch, J. D. (1979). Hazard rate models with covariates. *Biometrics*, pages 25–39.

Quintana, F., Johnson, W. O., Waetjen, E., and Gold, E. (2015). Bayesian nonparametric longitudinal data analysis. *Submitted*.

Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press.

Ricketts, J. and Head, G. (1999). A five-parameter logistic equation for investigating asymmetry of curvature in baroreflex studies. *American Journal of Physiology—Regulatory, Integrative and Comparative Physiology*, **277**, R441–R454.

Rodriguez, A., Dunson, D. B., and Gelfand, A. E. (2008). The nested Dirichlet process (with discussion). *Journal of the American Statistical Association*, **103**, 1131–1154.

Ryan, T. P. and Woodall, W. H. (2005). The most-cited statistical papers. *Journal of Applied Statistics*, **32**, 461–474.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, **2**, 639–650.

Sundaram, R. (2006). Semiparametric inference for the proportional odds model with time-dependent covariates. *Journal of Statistical Planning and Inference*, **136**, 320–334.

Susarla, V. and Van Ryzin, J. (1976). Nonparametric Bayesian estimation of survival curves from incomplete observations. *Journal of the American Statistical Association*, **71**, 897–902.

Taylor, J., Cumberland, W., and Sy, J. (1994). A stochastic model for analysis of longitudinal AIDS data. *Journal of the American Statistical Association*, **89**, 727–736.

Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, **101**, 1566–1581.

Tokdar, S. T. (2006). Posterior consistency of Dirichlet location-scale mixture of normals in density estimation and regression. *Sankhyā: The Indian Journal of Statistics*, **68**, 90–110.

Tomlinson, G. and Escobar, M. (1999). Analysis of densities. Technical report, University of Toronto.

Tseng, Y.-K., Hsieh, F., and Wang, J.-L. (2005). Joint modelling of accelerated failure time and longitudinal data. *Biometrika*, **92**, 587–603.

Tsiatis, A. A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica*, **14**, 809–834.

Wald, A. (1952). On the principles of statistical inference. *Notre Dame Mathematical Lectures*, No. 1, Notre Dame, Ind.

Wang, Y. and Taylor, J. M. G. (2001). Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association*, **96**, 895–905.

Zeger, S. L. and Diggle, P. J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics*, **50**, 689–699.

Zhang, D., Lin, X., Raz, J., and Sowers, M. (1998). Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association*, **93**, 710–719.

Zhou, H. and Hanson, T. (2015). Bayesian spatial survival models. In: *Nonparametric Bayesian Methods in Biostatistics and Bioinformatics*, Eds: R. Mitra & P. Müller, New York: Springer.