Proportional Mean Regression Models for Censored Data

Sujit K. Ghosh and Subhashis Ghosal

Department of Statistics
North Carolina State University
Raleigh, NC 27695-8203

Institute of Statistics Mimeo Series# 2563

Abstract

A novel semiparametric regression model for censored data is proposed as an alternative to the widely used proportional hazards survival model. The proposed regression model for censored data turns out to be flexible and practically meaningful. Features include physical interpretation of the regression coefficients through the mean response time instead of the hazard functions, and a rigorous proof of consistency of the posterior distribution. It is shown that the regression model obtained by a mixture of parametric families, has a *proportional mean* structure (as in an accelerated failure time models). The statistical inference is based on a nonparametric Bayesian approach that uses a Dirichlet process prior for the mixing distribution. Consistency of the posterior distribution of the regression parameters in the Euclidean metric is established. Finite sample parameter estimates along with associated measure of uncertainties can be computed by a MCMC method. Simulation studies are presented to provide empirical validation of the new method. Some real data examples are provided to show the easy applicability of the proposed method.

AMS 2000 *subject classification.* Primary 62N01, 62G99; secondary 62G20.

Key Words and phrases: Censored Data; Dirichlet Process Prior; Gibbs Sampling; Markov Chain Monte Carlo; Mixture Models; Posterior consistency, Regression; Semiparametric Models; Survival Analysis.

1 Introduction

The framework of generalized linear models (GLM) popularized by McCullagh and Nelder (1989) is arguably one of the most flexible statistical tool and is used extensively in almost all fields of applications. However, such a flexible framework is not widely available in the area of survival regression models for censored data. A key motivation of this article is to develop a flexible class of regression models that are similar in spirit to the GLM but retaining the semiparametric structure of the widely used proportional hazard (PH) models. It is shown that such an objective may be achieved by a class of mixture models where the mixing distribution is left unspecified.

An important feature of most of the semi-parametric regression methods with censored data is to keep the error distribution unspecified (and hence estimate it nonparametrically). A parametric form is usually specified for the regression function using some link function. However, for regression models with censored data, the link is specified through the hazard function of the response (usually the survival time) instead of the familiar mean response. The Cox proportional hazard regression model (Cox, 1972) and the associated partial likelihood theory of estimation was a breakthrough in developing a flexible method of regression for censored data. The huge success of PH models testify the many needs for this type of semiparametric regression models.

The structure of PH is quite different from the usual GLM for regression, in that the link function is not specified via the mean but rather through the hazard function. The so-called PH structure is interesting but it may be hard to interpret the regression coefficients. Sir D. Cox himself once remarked (see Reid, 1994) that

> *"Of course, another issue is the physical or substantive basis for the proportional hazards model. I think that's one of its weakness, that accelerated life models are in many ways more appealing because of their quite direct physical interpretation, particularly in an engineering context."*

An *Accelerated Failure Time (AFT)* model is characterized by specifying the conditional survival function, $S(t|\boldsymbol{Z} = \boldsymbol{z})$ of survival time, T given the covariates, $\boldsymbol{Z} = \boldsymbol{z}$, as $S(t|\boldsymbol{z}) = S_0(tg(\boldsymbol{z}^T\boldsymbol{\beta}))$, where $S_0(\cdot)$ is an unspecified baseline survival function. If $S_0(\cdot)$ is specified parametrically (with possibly some unknown parameters), then we get a parametric AFT model. Under such parametric models

estimates of the regression coefficient $\boldsymbol{\beta}$ can be obtained by maximum likelihood or Bayesian method, for a given completely known link function $g(\cdot)$. The parametric AFT models have such direct physical interpretation but the parametric models lack the semi-parametric appeal of the PH models. Therefore, other types of regression models for censored data that retain simple physical interpretation and a semi-parametric appeal, are desired.

Among the various extensions of the traditional linear model, the AFT models and the least squares method to accommodate the censored data seems most appealing, simply because the model is well known, widely used, well understood and well tested (see Wei, 1992). Some early attempts of such extensions are due to Miller (1976) and Buckley and James (1979). However, no rigorous proof of consistency was presented and the iterations in the computation of the estimate may not converge. Koul *et al.* (1981) developed a simple least square estimation method based on "synthetic data", as termed by Leurgans (1987). The main idea of this approach is to replace the censored observations by a set of estimated "responses" and then obtain the usual least square estimate for regression parameters based on the such estimated responses. Using a U-statistic representation, Koul *et al.* (1981) showed that their estimates are consistent and asymptotically normal under some regularity conditions. Following on this simple idea of using synthetic data several extensions of the method have appeared in the literature that uses a more efficient way to obtain estimated responses (see Zeng, 1984, Lai et al., 1995, Zhou, 1992 and references therein). These developments have been very exciting but generally lacks stability of the estimators and hence are not as widely used as the PH model.

A flexible class of semiparametric regression models have been developed in this article. The proposed method of parameter estimation is quite different as compared to the "synthetic data" approach. But the proposed method is similar in spirit to the previous approaches in terms of having a physical interpretation of the regression estimates. It is shown that a class of mixture models can be used to obtain *proportional mean* models. This is also satisfied by the familiar accelerated failure time models if the baseline survival function, $S_(\cdot)$ is left unspecified and is estimated nonparametrically (see Jin et al., 2003). However it is generally not straightforward to obtain an estimate of the asymptotic variance-covariance matrix of the regression parameter, $\boldsymbol{\beta}$ and Jin *et al.* (2003) and Park and Wei (2003) used a resampling method to obtain such an estimate. For our proposed mod-

els, such estimates can be computed easily from the posterior variance-covariance matrix of $\boldsymbol{\beta}$. Another class of models related to our approach would be the transformation models (see Chen, Jin and Ying, 2002 and Cheng, Wei and Ying, 1995), where the survival time is transformed using an unspecified monotone function but keeping the error distribution completely known. However under this model the estimated transformation may not be a smooth function and hence lacks straightforward interpretation enjoyed by AFT models. Because of the straightforward implementation of the Markov Chain Monte Carlo (MCMC) method in a mixture model, we find that a nonparametric Bayesian perspective with a mixture model is particularly attractive. Moreover, we show that such computationally intensive methods can be easily implemented using freely available softwares like `WinBUGS` (Spiegelhalter et al., 1999). In addition to the computational stability and simplicity of the proposed method, using the results similar to that of Ghosal *et al.* (1999), we show that the posterior distribution is consistent.

The rest of this article has been organized as follows. In Section 2, the mixture model is described along with proportional mean interpretation for the regression parameters. In Section 3, we discuss how Markov Chain Monte Carlo (MCMC) method can be easily implemented to obtain the estimates based on posterior distribution of the parameters. We discuss posterior consistency of the proposed model is presented in Section 4. Simulation studies are presented in Section 5. Finally, in Section 6 we present couple of real data examples to illustrate our method. Proofs are presented in the Appendix.

2 The Proportional Mean Regression Model

Mixture models have been used for a variety of inference problems including density estimation, clustering analysis and robust estimation; see Lindsay (1995), McLachlan and Basford (1988), Banfield and Raftery (1993), Robert (1996) and Roeder and Wasserman (1997). In the Bayesian context, mixture models for density estimation were introduced by Ferguson (1983) and Lo (1984) who used a Dirichlet process prior on the mixing distribution and obtained expressions for Bayes estimates. Escobar and West (1995) developed this idea further and provided MCMC algorithms for the computation of the posterior distribution of parameters of a normal mixture model. The choice of the kernel depends on the range space of the variable of interest. For the entire real line, the normal kernel is usually

used. A beta kernel has been considered for densities on the unit interval (see Petrone, 1999), while a gamma, Weibull or lognormal kernel seems to be appropriate on the positive half line (see Kottas and Gelfand, 2001). More generally, Feller approximation techniques can be used as shown by Petrone and Veronese (2002). Consistency and rates of convergence issues for Bayesian density estimation in mixture models have been studied by Ghosal *et al.* (1999), Ghosal and van der Vaart (2001), Ghosal (2001) and Petrone and Wasserman (2002). For survival models without covariates, consistency has been studied by Ghosh and Ramamoorthi (1994), Ghosh *et al.* (1999) and Kim and Lee (2001). In this article we propose using an infinite mixture of standard parametric distributions (such as the Weibull distributions) to model the conditional distribution of a survival time given a set of covariates.

For each subject $i = 1, \ldots, n$, let T_i denote the failure time and C_i denote the censoring time. The observed survival data are $X_i = \min(T_i, C_i)$ and $\Delta_i = I(T_i \leq C_i)$, where $I(\cdot)$ is the indicator function; these and all other variables are independent across i. Let $\boldsymbol{Z}_i$ denote a p-dimensional vector of covariates associated with subject i. Assume that for each subject i the conditional survival function of T given $\boldsymbol{Z} = \boldsymbol{z}$ (suppressing the subscript i) is given by,

$$S(t|\boldsymbol{z}, \boldsymbol{\beta}) = \int_0^\infty S_b(t/\mu g(\boldsymbol{z}^T\boldsymbol{\beta}))dH(\mu), \tag{1}$$

where $S_b(\cdot)$ is a survival function with a specified functional form, but may involve additional unknown parameters. The link function $g : \mathbb{R} \to [0, \infty)$ is completely specified. The mixing distribution function $H(\cdot)$ is left unspecified, which will be estimated nonparametrically with the constraint $H(0) = 0$. It follows that the density of T given $\boldsymbol{Z} = \boldsymbol{z}$ is then given by

$$p(t|\boldsymbol{z}, \boldsymbol{\beta}) = \int_0^\infty p_b(t/\mu g(\boldsymbol{z}^T\boldsymbol{\beta}))(\mu g(\boldsymbol{z}^T\boldsymbol{\beta}))^{-1}dH(\mu), \tag{2}$$

where p_b is the density function corresponding to the survival function S_b. Notice that (**??**) leads to an AFT model with $S_0(t) = \int_0^\infty S_b(\frac{t}{\mu})\, dH(\mu)$. In other words, the baseline survival function is modeled as a mixture of parametric survival function with an unknown mixing distribution.

Let $m = \int_0^\infty t p_b(t)dt = \int_0^\infty S_b(t)dt$ stand for the mean of the distribution with survival function S_b. Then

$$\mathrm{E}(T|\boldsymbol{z}, \alpha, \boldsymbol{\beta}) = m g(\boldsymbol{z}^T\boldsymbol{\beta}) \int_0^\infty \mu dH(\mu), \tag{3}$$

and hence we obtain the *proportional mean* model,

$$\frac{\mathrm{E}(T|\boldsymbol{z}_1,\boldsymbol{\beta})}{\mathrm{E}(T|\boldsymbol{z}_2,\boldsymbol{\beta})} = \frac{g(\boldsymbol{z}_1^T\boldsymbol{\beta})}{g(\boldsymbol{z}_2^T\boldsymbol{\beta})}. \tag{4}$$

As a consequence of the above property in equation (**??**), we will call the proposed model as *Proportional Mean Regression* (PMR) model. In most applications a logarithmic link is chosen for g^{-1}, that is, $g(w) = e^w$. Putting in the equation (**??**), we obtain,

$$\log\left(\frac{\mathrm{E}(T|\boldsymbol{z}_1,\boldsymbol{\beta})}{\mathrm{E}(T|\boldsymbol{z}_2,\boldsymbol{\beta})}\right) = (\boldsymbol{z}_1-\boldsymbol{z}_2)^T\boldsymbol{\beta}$$

which provides an easy interpretation of the parameter $\boldsymbol{\beta}$ as the unit change in the logarithm of the mean response with respect to a unit change in the covariate. Henceforth, we shall assume the logarithmic link function. Note that the mixing distribution $H(\cdot)$ has been left completely unspecified.

We shall use a Weibull survival function with an unknown shape parameter as a choice for S_b, i.e., we assume that $S_b(t,\alpha) = e^{-t^\alpha}$. The distribution with survival function $e^{-t^\alpha/\lambda^\alpha}$ will be referred to as Weibull (α,λ) distribution, the Weibull distribution with scale parameter λ and shape parameter α. One may use other parametric families such as log normal or gamma. More generally, any parametric family supported on the positive half line may be considered. All these families, particularly the Weibull and the gamma, form very flexible classes of life distributions. However, unlike the gamma or the lognormal family, the Weibull distribution enjoys the advantage that the survival function does not involve transcendental functions, and hence it will be easier to explicitly write down the likelihood for censored data. It is interesting to notice that mixture of Weibull distribution (as in (**??**)) can be seen as a location mixture model on a log-scale. More specifically, for Weibull mixtures, when we choose a log-link for g^{-1}, the conditional density of $Y = \log(T)$ given $\boldsymbol{Z} = \boldsymbol{z}$ is given by

$$\begin{aligned} p(y|\boldsymbol{z},\boldsymbol{\beta}) &= \int_0^\infty e^y p_b(e^y/\mu g(\boldsymbol{z}^T\boldsymbol{\beta}))(\mu g(\boldsymbol{z}^T\boldsymbol{\beta}))^{-1} dH(\mu), \\ &= \int_0^\infty \alpha f_b(\alpha(y-\log(\mu)-\boldsymbol{z}^T\boldsymbol{\beta}))(\mu g(\boldsymbol{z}^T\boldsymbol{\beta}))^{-1} dH(\mu) \end{aligned}$$

where $f_b(t) = \exp\{t - e^t\}$ is the Gumbel's Extreme value density. Thus $\log(\mu)$ takes the role of the location parameter and α the scale parameter in the above location mixture. Therefore, letting $\alpha \to \infty$, mixtures over μ can approximate any arbitrary density in the total variation distance. As the total variation distance is invariant

under monotone transformations, the same conclusion holds in the original scale. Thus, Weibull mixture can adequately represent an arbitrary density in (**??**).

A likelihood based method to estimate the parameter $\boldsymbol{\beta}$ would require maximization of over the space of all distributions $H(\cdot)$ with the property $H(0) = 0$ and the shape parameter α. Such an approach may yield efficient estimates for $\boldsymbol{\beta}$ and under regularity conditions a large sample estimate of the standard error of the estimate can be obtained. Optimizations over such large space of parameters may be computationally challenging. Alternatively, one may use a set of estimating equations (see Jin et al., 2003) to obtain parameter estimates. However it turns out that for semiparametric AFT models such estimating equations can be non-monotone and leads to multiple solutions. In our nonparametric Bayesian approach, we use MCMC method to obtain the marginal posterior distribution of $\boldsymbol{\beta}$ by integrating out the nuisance parameters $H(\cdot)$ and α.

Prior distributions play a crucial role in Bayesian inference. Informative priors can be used if deemed plausible. However, in the present set up, it is desirable to choose the prior distribution by some default mechanism. A priori, it is assumed that $(\alpha, \boldsymbol{\beta})$ is independent of $H(\cdot)$, and that $H(\cdot)$ has a Dirichlet Process (DP) Prior, denoted by $DP(M, H_b(\cdot))$ with base measure $H_b(\cdot)$ and precision parameter M. The proposed model can be equivalently written as a Bayes hierarchical model as follows:

$$\begin{aligned} T_i|\boldsymbol{Z}_i, \alpha, \mu_i, \boldsymbol{\beta} &\sim \text{Weibull } (\alpha, \mu_i e^{\boldsymbol{Z}_i^T \boldsymbol{\beta}}) \\ \mu_i|H(\cdot) &\overset{i.i.d}{\sim} H(\cdot) \\ H(\cdot) &\sim DP(M, H_b(\cdot)) \\ (\alpha, \boldsymbol{\beta}) &\sim \pi(\alpha, \boldsymbol{\beta}) \end{aligned} \tag{5}$$

where the joint density $\pi(\alpha, \boldsymbol{\beta})$ is usually chosen to be of the product type and often taken to be the diffuse uniform prior on $\mathbb{R}^p \times (0, \infty)$. A moderate value of M is usually chosen. Let Π denote the joint prior distribution of $(\alpha, \boldsymbol{\beta}, H)$. Note that from the properties of a DP it follows that,

$$\mathrm{E}(H(\cdot)|\mu_1, \ldots, \mu_n) = \frac{M}{M+n} H_b(\cdot) + \frac{n}{M+n} H_n(\cdot), \tag{6}$$

where $H_n(\cdot)$ is the empirical distribution function of $\mu_1, \ldots, \mu_n$. However, these latent variables $\mu_1, \ldots, \mu_n$ are not directly observable, and hence the posterior distribution may only be computed by MCMC methods. In the next section we

discuss how to obtain the posterior distribution of the parameters given the observed data $\{(X_i, \Delta_i, \boldsymbol{Z}_i), i = 1, \ldots, n.\}$

3 Model Fitting

For the proposed PMR model, the posterior distribution of the parameters can not be obtained in a closed form. Gibbs sampling with a data augmentation step will be used to obtain MCMC samples.

First assume that T_i's are indeed observed. In this case, a simple strategy to obtain samples from the posterior distribution of $(\boldsymbol{\beta}, \alpha, H(\cdot))$ can be described as follows (see Dey, Muller and Sinha, 1998):

Gibbs Sampling: Initialize the parameters $\boldsymbol{\beta}^{(0)}$ and $\alpha^{(0)}$. At the k-th iteration of the Gibbs sampler the full conditional density of μ_i given $(\boldsymbol{\beta}^{(k-1)}, \alpha^{(k-1)}, \mu_l^{(k-1)}, l \neq i, T_i, \boldsymbol{Z}_i)$ is given by,

$$\begin{aligned} & p_b(T_i|\boldsymbol{Z}_i, \mu, \alpha^{(k-1)}, \boldsymbol{\beta}^{(k-1)}) H_b'(\mu), \qquad \text{with probability } q_{0i}^{(k)} \\ & \mu_l^{(k-1)} I(l > i) + \mu_l^{(k)} I(l < i), \quad \text{with probability } q_{li}^{(k)},\ l \neq i, \end{aligned}$$

where

$$q_{li}^{(k)} \propto \begin{cases} \mu_l^{(k)} p_b(T_i|\boldsymbol{Z}_i, \mu_l^{(k)}, \alpha^{(k-1)}, \boldsymbol{\beta}^{(k-1)}) & \text{if } l < i \\ \mu_l^{(k-1)} p_b(T_i|\boldsymbol{Z}_i, \mu_l^{(k-1)}, \alpha^{(k-1)}, \boldsymbol{\beta}^{(k-1)}) & \text{if } l > i \end{cases}$$

and

$$q_{0i}^{(k)} \propto M \int_0^\infty p_b(T_i|\boldsymbol{Z}_i, \mu, \alpha^{(k-1)}, \boldsymbol{\beta}^{(k-1)}) dH_b(\mu),$$

such that $q_{0i}^{(k)} + \sum_{l \neq i} q_{li}^{(k)} = 1$. Note that if we choose a conjugate base measure $H_b(\cdot)$, the above integral can be evaluated analytically and hence samples can be obtained by an inversion method. Otherwise a numerical one-dimensional integral can be used to compute q_{0i}^k's. For our model we may achieve conjugacy by choosing an appropriate gamma distribution for μ_i^α. The full conditional density of $(\boldsymbol{\beta}, \alpha)$ given $\{(\mu_i^{(k)}, T_i, \boldsymbol{Z}_i),\ i = 1, \ldots, n\}$ is given by,

$$\prod_{i=1}^n p_b(T_i|\boldsymbol{Z}_i, \mu_i^{(k)}, \alpha, \boldsymbol{\beta}) H_b'(\mu_i^{(k)})$$

which is a log-concave density and hence a adaptive rejection sampling (see Gilks, 1992) can be used to sample from the above density. Thus a Gibbs sampler as described above can be used to obtain approximate samples from the joint posterior

distribution of the parameters given the observed data. If desired the posterior mean of $H(\cdot)$ can be obtained from the MCMC samples using (**??**).

When censoring is present, an imputation method can be used to "replace" the censored data by the imputed values. This is done as follows. If $\Delta_i = 1$, set $T_i = X_i$ otherwise set $T_i = (U_i + X_i^{\frac{1}{\alpha}})^{\alpha}$, where U_i's are independently exponentially distributed with means $\mu_i e^{\boldsymbol{Z}_i^T \boldsymbol{\beta}}, i = 1, \ldots, n$. Alternatively a finite approximation for DP (see Ishawaran and Zarepour, 2002) can be used within the software `BUGS` to implement the Gibbs sampling. In fact, for our simulation study we use a finite approximation technique to implement the required Gibbs sampling using `BUGS`. This is achieved by introducing latent variables $\boldsymbol{L} = (L_1, \ldots, L_n)$ which indicate the group membership for the hidden variables μ_i's. More precisely, (**??**) can be written as,

$$\begin{aligned}
T_i | L_i, \alpha, \boldsymbol{\beta} &\sim \text{Weibull}(\alpha, \mu_{L_i} e^{\boldsymbol{Z}_i^T \boldsymbol{\beta}}) \\
L_i | \boldsymbol{w} &\overset{i.i.d}{\sim} \text{Multinomial}(\{1, \ldots, N\}, \boldsymbol{w}) \\
\mu_l &\overset{i.i.d}{\sim} H_b(\cdot),\ l = 1, \ldots, N \\
\boldsymbol{w} &\sim \text{Dirichlet}(\frac{M}{N}, \ldots, \frac{M}{N}) \\
(\alpha, \boldsymbol{\beta}) &\sim \pi(\alpha, \boldsymbol{\beta})
\end{aligned}$$

where N is chosen appropriately to approximate the DP. Notice that, $L_i \in \{1, \ldots, N\}$ takes random multinomial values with $\Pr[L_i = l] = w_l$ for $l = 1, \ldots, N$. The above hierarchical framework which uses a finite dimensional Dirichlet distribution (with N possibly depending on n) can be used for programming in `BUGS`. For instance, we may use $N = \sqrt{n}$ for large n and $N = n$ for small n (see Section 5 of Ishwaran and Zarepour, 2002 for more details).

4 Consistency of Posterior Distribution

Consistency is an important desirable large sample property of the posterior distribution which provides a useful validation of a particular Bayesian method in use. For discussion and examples of a of inconsistency, the readers are referred to Diaconis and Freedman (1986) and Ghosh and Ramamoorthi (2003). A celebrated theorem of Schwartz (1965) gives sufficient conditions for posterior consistency in terms of conditions involving the existence of appropriate tests and the prior

positivity of a neighborhood defined by the Kullback-Leibler divergence. Useful extensions of Schwartz's theorem are given by Barron *et al.* (1999) and Ghosal *et al.* (1999). The goal of the present section is to justify our Bayesian analysis of the PMR model by posterior consistency.

We shall assume that the domains of $\boldsymbol{Z}$, $\boldsymbol{\beta}$ and α, and the support of H are compact. The compactness assumption is agreeably somewhat restrictive. However, unbounded scale mixtures are often known to misbehave in terms of asymptotic properties. Also, we assume, without loss of generality that 0 is a possible value of the covariate $\boldsymbol{Z}$. If not, we may shift the covariates to satisfy this condition. We further assume that the true density p_0 of T given $\boldsymbol{Z} = \boldsymbol{z}$ is actually a scale mixture of Weibull:

$$p_0(t|\boldsymbol{z}) = \int \alpha_0(\mu e^{\boldsymbol{\beta}_0^T \boldsymbol{z}})^{-\alpha_0} t^{\alpha_0 - 1} \exp(-(t/\mu e^{\boldsymbol{\beta}_0^T \boldsymbol{z}})^{\alpha_0}) dH_0(\mu),$$

so that α_0, $\boldsymbol{\beta}_0$ and H_0 are respectively the true values of the parameters α, $\boldsymbol{\beta}$ and H. The covariates are assumed to be i.i.d. with an absolutely continuous distribution supporting the vector 0 on $\mathbb{R}^p$. The density of $\boldsymbol{Z}$ at $\boldsymbol{z}$ will be denoted by $q(\boldsymbol{z})$. Let $f_{\alpha,\boldsymbol{\beta},H}(x, \delta, \boldsymbol{z})$ stand for the joint density of $(X, \Delta, \boldsymbol{Z})$, that is,

$$f_{\alpha,\boldsymbol{\beta},H}(x, \delta, \boldsymbol{z}) = \begin{cases} p(x|\alpha, \boldsymbol{\beta}, \boldsymbol{z}) q(\boldsymbol{z}), & \text{if } \Delta = 1, \\ S(x|\alpha, \boldsymbol{\beta}, \boldsymbol{z}) q(\boldsymbol{z}), & \text{if } \Delta = 0, \end{cases}$$

where $p(x|\alpha, \boldsymbol{\beta}, \boldsymbol{z})$ and $S(x|\alpha, \boldsymbol{\beta}, \boldsymbol{z})$ are as defined in equations (2) and (1), respectively. Note that the class of distributions that are supported in a given compact domain is also compact with respect to the weak topology on the space of probability measures. Hence the parameter space of $(\alpha, \boldsymbol{\beta}, H)$ with respect to the product of Euclidean, Euclidean and weak topology, is also compact.

The following is the main theorem of this paper.

THEOREM 1. *Suppose that the prior density $\pi(\alpha, \boldsymbol{\beta})$ for $(\alpha, \boldsymbol{\beta})$ has compact support containing $(\alpha_0, \boldsymbol{\beta}_0)$ and the base measure H_b of the Dirichlet process has compact support that contains the support of H_0. Then the posterior distribution $\Pi((\alpha, \boldsymbol{\beta}, H) \in \cdot | (X_1, \Delta_1), \ldots, (X_n, \Delta_n))$ of $(\alpha, \boldsymbol{\beta}, H)$ given $(X_1, \Delta_1), \ldots, (X_n, \Delta_n)$ is consistent with respect to the Euclidean distances on $\boldsymbol{\beta}$ and α and the weak topology on H, that is, given any $\varepsilon > 0$ and a weak neighborhood $\mathcal{N}$ of H_0,*

$$\Pi\{(\alpha, \boldsymbol{\beta}, H) : |\alpha - \alpha_0| < \varepsilon,\ |\boldsymbol{\beta} - \boldsymbol{\beta}_0| < \varepsilon,\ H \in \mathcal{N} | (X_1, \Delta_1), \ldots, (X_n, \Delta_n)\} \to 1$$

almost surely in $P^{\infty}_{(\alpha_0, \boldsymbol{\beta}_0, H_0)}$-probability.

This gives a large sample justification of our procedure. The proof of Theorem **??** is given in the Appendix.

If the covariates are not random, but arise deterministically from a design, then by largely a similar analysis and using posterior consistency results for independent, non-identically distributed observations as in Amewou-Atisso *et al.* (2003), consistency will follow if the values of the covariate gradually fill up its range space as the sample size increases. The details are omitted here.

5 A Simulation Study

We performed extensive simulations to explore the sampling properties of the Bayes estimates obtained by the proposed method. We present only some of the significant findings from our simulation experiments. In order to perform the simulation experiments, we generate the data from the model (**??**). In particular we use the following data generation scheme:

Data Generation Process (DGP): Fix true values $\boldsymbol{\beta}_0 = (-0.5, 0.5), \alpha_0 = 1.0$ and the sample size $n = 50, 100$ and 200.

1. Generate $U_i \overset{iid}{\sim} Beta(3,3)$ and set $\boldsymbol{Z}_i = 6U_i - 3$. This ensures that support of $\boldsymbol{Z}_i$'s is compact.

2. Generate $\mu_i \overset{iid}{\sim} Weibull(\alpha^* = 1, \mu^* = 2)$. This means that the mixing distribution is also a Weibull distribution.

3. Generate $T_i \sim Weibull(\alpha_0 = 1, \mu_i e^{\boldsymbol{Z}_i^T \boldsymbol{\beta}_0})$

4. Generate $C_i \overset{iid}{\sim} Weibull(\alpha^*, 2\mu^*)$. We also perform simulation where C_i's are allowed to depend on $\boldsymbol{Z}_i$'s. In particular we generate $C_i \overset{indep}{\sim} Weibull(\alpha^*, 4.5e^{\boldsymbol{Z}_i^T \boldsymbol{\beta}})$. These choices of censoring variable ensure that the censoring rate is about 30% on average.

5. Set $X_i = \min(T_i, C_i)$ and $\Delta_i = I(T_i \leq C_i)$

6. Observed data: $\{(X_i, \Delta_i, \boldsymbol{Z}_i);\ i = 1, \ldots, n\}$

We fit the semiparametric Bayesian model to the data generated from above DGP and repeat the method 500 times with a fixed burn-in time of 500 iterations followed by a sample of size 1000 from the posterior distribution of $\boldsymbol{\beta}$ and α. We

have tried several combinations of true values, mixing distributions and censoring variables for our simulations. In this article we present results based on only one set of true values for four different samples sizes and two different censoring mechanisms.

The results are presented in Tables 1 and 2 by increasing sample size. In each table we present the average five-number posterior summary values, viz. posterior mean, posterior sd, posterior 2.5% and 97.5% percentiles and the posterior median with associated standard error (s.e.) from 500 Monte Carlo repetitions. In addition, we also present the coverage probability (cp) based on a 95% equal-tail posterior interval obtained by posterior 2.5% and 97.5% percentiles. For example, in Table 1, for $n = 100$, the entry -0.491 for β_1 is the average of 500 posterior means with s.e. 0.172 (which is the standard deviation of 500 posterior means). Similarly the entry 0.166 is the average of 500 posterior sd's with s.e. 0.027. Finally, a cp of 0.935 means that 93.5% of 500 posterior 95% intervals (based on posterior 2.5% and 97.5% percentiles) contained the true value of $\beta_1 = -0.5$. We observe that at an average censoring rate of about 30% the Bayes estimates (i.e. posterior mean and median) are nearly unbiased as sample size increases and also the coverage probability of a 95% posterior interval approaches the nominal value of 95%. In addition, we see that the average length of a 95% posterior interval decreases with increasing sample size, thus making the intervals tighter. For instance, for β_2 the average length of 95% interval drops from 0.977 (for $n = 50$) to an average length of 0.467 (for $n = 200$). Similar observations can be made based on Table 2. These results numerically assert that consistency of posterior distribution of $\boldsymbol{\beta}$ and α. We repeated similar studies for other mixing distributions and for lower and higher censoring rates, the results were very similar to what we see in Tables 1 and 2.

6 Applications to real data sets

We now apply our proposed models to some popular real data sets that have been analyzed by other authors. In particular we fit our semi-parametric models to ovarian cancer data set (with $n = 26$, small sample size) and multiple myeloma data set (with $n = 65$, moderate sample size).

	$n = 50$					
	mean	sd	2.5%L	97.5%U	median	cp
β_1	-0.505	0.244	-1.000	-0.036	-0.501	0.927
$s.e.$	0.273	0.053	0.313	0.288	0.271	
β_2	0.516	0.248	0.055	1.032	0.534	0.928
$s.e.$	0.253	0.055	0.269	0.289	0.253	
α	1.116	0.241	0.721	1.641	1.134	0.948
$s.e.$	0.224	0.051	0.148	0.289	0.233	
	$n = 100$					
	mean	sd	2.5%L	97.5%U	median	cp
β_1	-0.491	0.166	-0.821	-0.165	-0.483	0.935
$s.e.$	0.172	0.027	0.192	0.170	0.173	
β_2	0.494	0.167	0.172	0.828	0.491	0.938
$s.e.$	0.177	0.027	0.178	0.197	0.176	
α	1.032	0.188	0.716	1.430	1.021	0.954
$s.e.$	0.176	0.042	0.115	0.235	0.183	
	$n = 200$					
	mean	sd	2.5%L	97.5%U	median	cp
β_1	-0.502	0.121	-0.753	-0.279	-0.504	0.949
$s.e.$	0.123	0.009	0.113	0.098	0.102	
β_2	0.497	0.118	0.261	0.728	0.496	0.948
$s.e.$	0.120	0.013	0.131	0.139	0.129	
α	0.996	0.152	0.727	1.305	0.967	0.951
$s.e.$	0.145	0.034	0.072	0.168	0.128	

Average observed censoring 29.25% (range 10% - 40%)

Table 1: Sampling properties of Bayes estimates when the censoring mechanism does not depend on covariates. True values: $\beta_1 = 0.5, \beta_2 = 0.5, \alpha = 1.0$. The $s.e.$ denotes the standard deviation of the corresponding posterior summaries (e.g., mean, sd etc.) over 500 replications.

	$n = 50$					
	mean	sd	2.5%L	97.5%U	median	cp
β_1	-0.487	0.234	-0.953	-0.028	-0.487	0.925
$s.e.$	0.258	0.053	0.283	0.282	0.258	
β_2	0.502	0.233	0.037	0.956	0.503	0.928
$s.e.$	0.248	0.052	0.266	0.278	0.248	
α	1.121	0.244	0.723	1.654	1.137	0.945
$s.e.$	0.239	0.053	0.153	0.309	0.247	
	$n = 100$					
	mean	sd	2.5%L	97.5%U	median	cp
β_1	-0.489	0.159	-0.819	-0.175	-0.489	0.939
$s.e.$	0.165	0.024	0.188	0.168	0.169	
β_2	0.495	0.163	0.179	0.817	0.493	0.940
$s.e.$	0.167	0.021	0.172	0.190	0.171	
α	1.029	0.179	0.710	1.422	1.018	0.953
$s.e.$	0.168	0.038	0.108	0.230	0.177	
	$n = 200$					
	mean	sd	2.5%L	97.5%U	median	cp
β_1	-0.502	0.115	-0.746	-0.280	-0.503	0.950
$s.e.$	0.123	0.009	0.113	0.098	0.102	
β_2	0.501	0.105	0.287	0.709	0.500	0.949
$s.e.$	0.110	0.010	0.126	0.133	0.124	
α	1.004	0.144	0.719	1.294	0.990	0.949
$s.e.$	0.140	0.028	0.067	0.160	0.121	

Average observed censoring 32.34% (range 15% - 42%)

Table 2: Sampling properties of Bayes estimates when the censoring mechanism depends on covariates. True values: $\beta_1 = 0.5, \beta_2 = 0.5, \alpha = 1.0$. The $s.e.$ denotes the standard deviation of the corresponding posterior summaries (e.g., mean, sd etc.) over 500 replications.

	mean	sd	2.5%L	97.5%U	median
β	-0.922	0.243	-1.460	-0.487	-0.911
α	2.470	0.640	1.330	3.760	2.440

Table 3: Posterior summaries of the regression parameter (β) and the shape parameter (α)

6.1 Analysis of ovarian cancer data

Consider a study on ovarian cancer as reported in Edmunson *et al.* (1979). In this study $n = 26$ patients were monitored and age for each patient were also recorded. Let X_i denotes the number of days patient i was on the study. For each patient, let A_i denote the age (recorded as number of days per 365.25). It was of interest to find a relation between the X_i's and the A_i's using statistical regression methods. However in this study 53.84% of the observations were censored. We obtained the posterior distribution of β, under the model $\log[E(T|Z)] = A\beta$.

Using the proposed semiparametric method based on Weibull mixture models, we obtain the summary of the posterior distribution of (β, α) using the Gibbs sampling algorithm described in Section 3. In order to maintain numerical stability, we transformed the X_i to $X_i/500$ and also standardized the covariate age to $Z_i = (A_i - \bar{A})/sd(A)$, where $\bar{A}$ and $sd(A) = 10.1$ denote the sample mean and standard deviation of A_i's. The results are summarized in Table 3. In addition, we can obtain the entire posterior density of α and β based on MCMC samples. In Figure 1, we plot the trace and kernel density estimates of (α, β) based 2000 burn-ins and 5000 samples from three independent chains. From the plots it appears that the chains mixed well and there were no apparent problems with convergence to the stationary distributions. The MCMC diagnostic software `CODA` available in `Splus` and `R` was also used to check convergence. The Gelman-Rubin 50% and 97.5% shrink factors (based on three dispersed starting values) was found to be 1.00 and 1.01, respectively for β, which indicates good mixing (as also evident from the trace plots in Figure 1).

From Table 3, it is apparent that age has a significant negative effect on the mean survival time. More precisely, for unit increase in age (in days/365.25), the expected survival time (on the log scale) decreases by $44(= 0.9 * 500/10.1)$ days for patients with ovarian cancer. This type of straight forward interpretation is

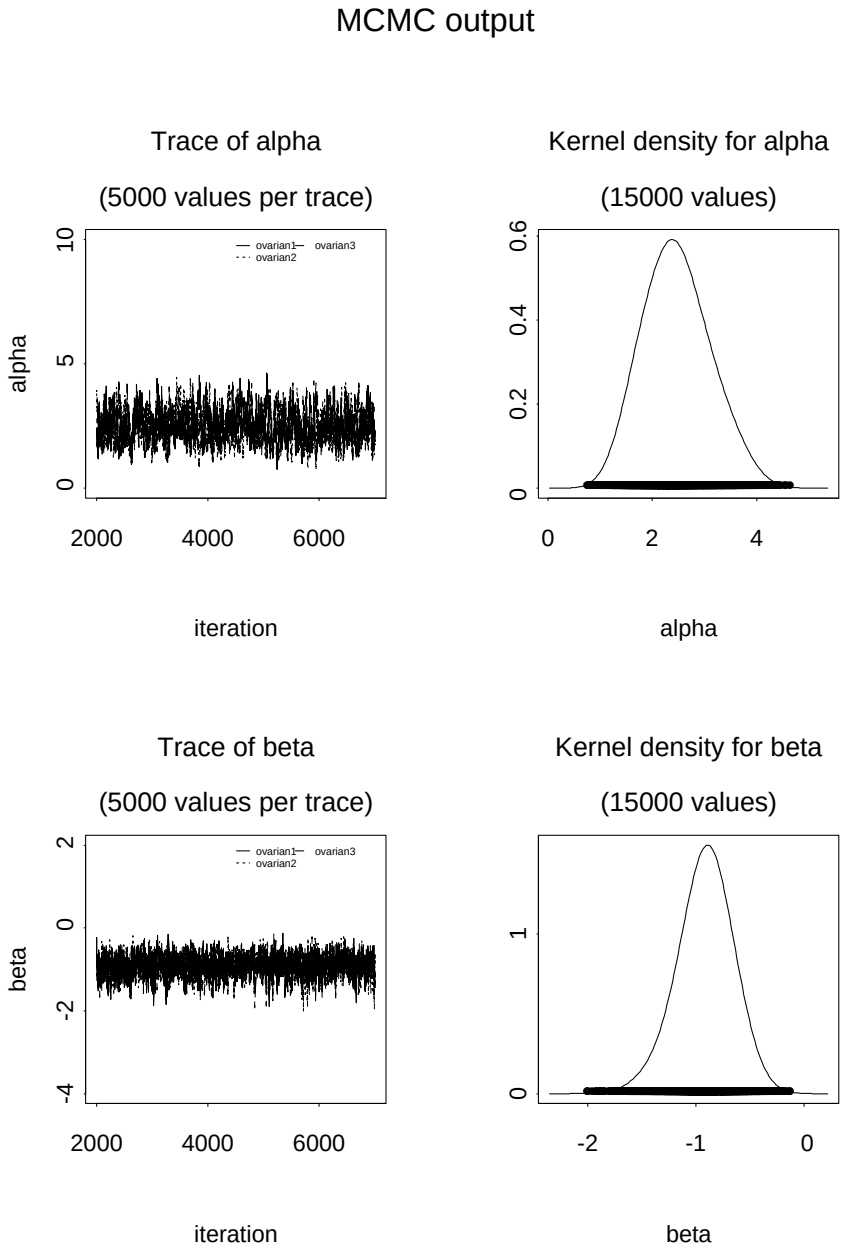

Figure 1: Trace and density plots based on MCMC samples from the posterior distribution of (α, β) for the ovarian data

possible only for these kind of AFT models.

6.2 Analysis of multiple myeloma data

We analyze a data set on multiple myeloma reported by Krall *et al.* (1975). This data has recently been analyzed by Jin *et al.* (2003) using a semiparametric AFT model. We compare our findings to that of Jin *et al.* (2003). In this study $n = 65$ patients were treated with alkylating agents and survival time (in months) were monitored along with several prognostic variables.

In order to compare our results to that of Jin *et al.* (2003) we select log of the blood urea nitrogen measurement and the hemoglobin measurement, both measured at diagnosis as the covariates and standardize them to bring numerical stability. In other words, we fit a model such that $\log(\mathrm{E}(T|\boldsymbol{Z})) = \boldsymbol{Z}^T\boldsymbol{\beta}$, where $\boldsymbol{Z} = (Z_1, Z_2)^T$ denotes the vector of standardized covariates and $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$ denotes the vector of regression coefficients to be estimated. Similar to the previous application we ran 3 independent parallel chains to draw MCMC samples from the posterior distribution of $(\alpha, \boldsymbol{\beta})$. In Figure 2, we plot the trace and density estimates of $(\alpha, \boldsymbol{\beta})$ based 5000 samples from three parallel chains after a burn-in of 2000 samples. The Gelman-Rubin 50% and 97.5% shrink factors (based on three dispersed starting values) was found to be 1.00 and 1.01, respectively for $\boldsymbol{\beta}$, which indicates good mixing (as also evident from the trace plots in Figure 2). It appears that there is no problem of convergence of the samples to stationary (posterior) distribution.

Based on the final 15000 samples we compute the posterior summaries of $(\alpha, \boldsymbol{\beta})$ and present these values in Table 4. First notice that the posterior means and median of $\boldsymbol{\beta}$ are quite similar to those obtained by Jin *et al.* (2003), which are presented in $(\cdot)$. Also from the posterior MCMC samples of $\boldsymbol{\beta}$ we can easily obtain the posterior correlation coefficient of $\boldsymbol{\beta}$ to be $Corr(\beta_1, \beta_2|data) = -0.187$. This is clearly an advantage of a Bayesian method over the frequentist method, as almost any posterior summary can be computed from the posterior samples generated by the MCMC method. Note that, Jin *et al.* (2003) had to use a resampling method to compute the asymptotic variance covariance matrix of $\hat{\boldsymbol{\beta}}$. In contrast, our method produces the entire posterior density of $\boldsymbol{\beta}$ using a relatively simple code in `WinBUGS` (that can be obtained from the first author). Further, our interval estimates of $\boldsymbol{\beta}$ as presented in Table 4, are based on the exact (finite sample) posterior percentiles as compared to large sample normal percentiles of Jin *et al.* (2003).

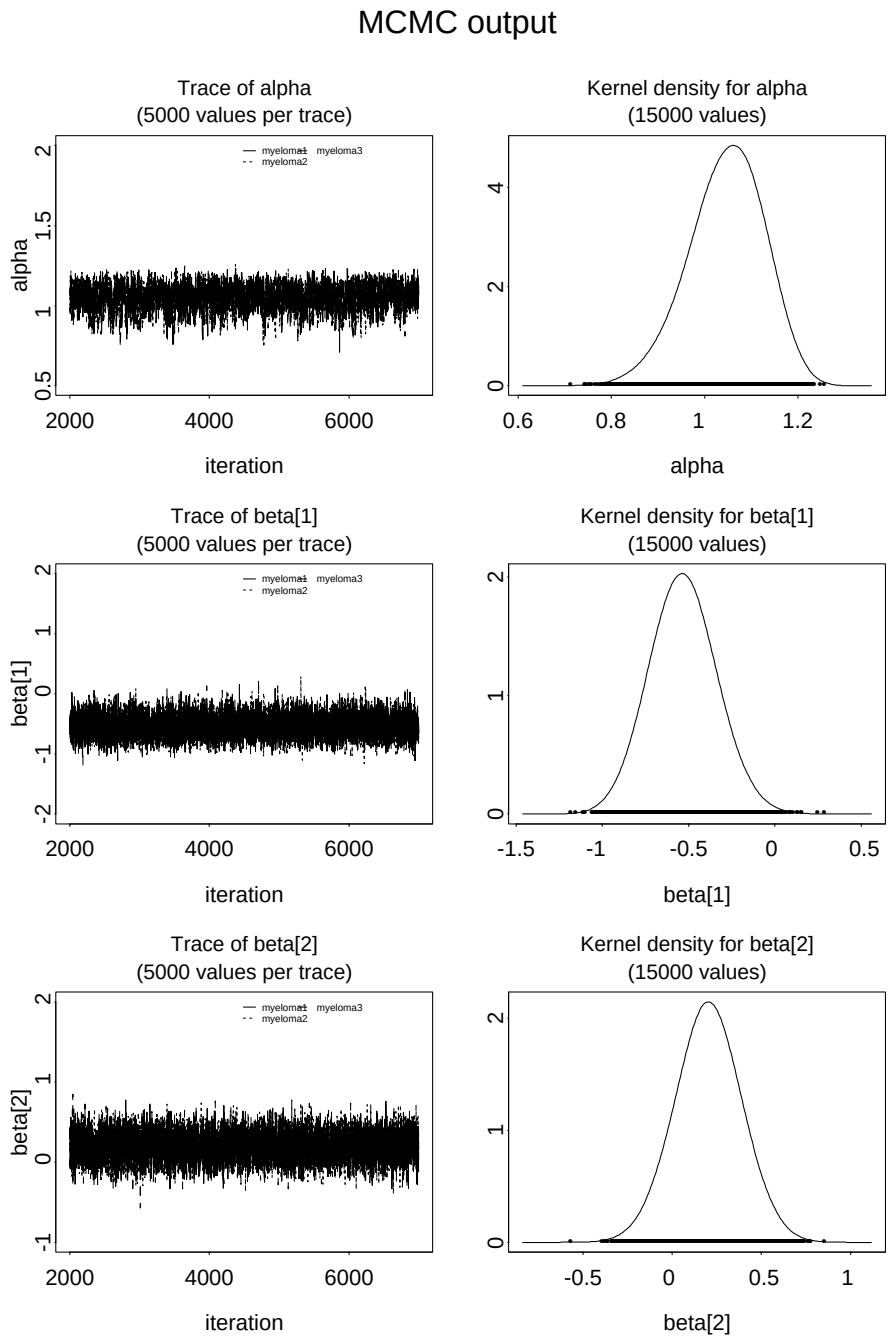

Figure 2: Trace and density plots based on MCMC samples from the posterior distribution of $(\alpha, \boldsymbol{\beta})$ for the myeloma data

	mean	sd	2.5%L	97.5%U	median
β_1	-0.534	0.175	-0.870	-0.190	-0.531
freq est.	(-0.532)	(0.146)	(-0.818)	(-0.246)	
β_2	0.202	0.168	-0.128	0.528	0.199
freq est.	(0.292)	(0.169)	(-0.039)	(0.623)	
α	1.040	0.074	0.885	1.170	1.050

Table 4: Posterior summaries of the regression parameter ($\boldsymbol{\beta}$) and the shape parameter (α). Within $(\cdot)$ corresponding frequentist estimates obtained by Jin *et al.* (2003) are presented.

For this study, it appears that log of the blood urea nitrogen measurement has a significant negative effect on the log of the mean survival time, whereas hemoglobin does not seem to have a significant effect. Based on our study we may conclude that, for unit increase in log blood urea nitrogen, the expected survival time decreases by $5.53(= \exp(-0.534/0.312))$ months for patients with multiple myeloma. This again illustrates the straight forward interpretation of the regression coefficients based on semiparametric PMR/AFT models.

Appendix

In this section, we prove the main theorem by verifying the conditions of the following general result .

THEOREM 2. *Let $Y_1, \ldots, Y_n$ be i.i.d. with density p lying in a class $\mathcal{P}$ of probability densities which is compact under the total variating metric. Let Π be a prior on $\mathcal{P}$ and p_0 be the true value of the density and let P_0 be the probability measure corresponding to p_0. If for all $\varepsilon > 0$,*

$$\Pi\left\{p \in \mathcal{P} : \int p_0 \log \frac{p_0}{p} < \varepsilon\right\} > 0,$$

then the posterior distribution is consistent at p_0 under the total variation (or the L_1) distance a.s. $[P_0^\infty]$, that is, for all $\varepsilon > 0$,

$$\Pi\left\{p \in \mathcal{P} : \int |p - p_0| < \varepsilon \middle| Y_1, \ldots, Y_n\right\} \to 1 \quad a.s.\ P_0^\infty.$$

PROOF. Observe that a compact set has finite metric entropy. Taking $\mathcal{P}$ itself as the sieve at each stage, (ii) of Theorem 2 of Ghosal *et al.* (1999) is satisfied. Alternatively, as $\mathcal{P}$ is compact under the variation distance, which gives a topology stronger than the weak one, the two topologies actually coincide on $\mathcal{P}$. Thus, total variation neighborhoods are also weak neighborhoods. As Schwartz's (1965) testing condition is satisfied for the weak topology (see Theorem 4.4.2 of Ghosh and Ramamoorthi, 2003), the result follows. □

To verify the conditions of Theorem **??**, we shall need the following three lemmas.

LEMMA 1. *The model is identifiable.*

PROOF. Let $f_{\alpha_1,\boldsymbol{\beta}_1,H_1}(x,\delta,\boldsymbol{z}) = f_{\alpha_2,\boldsymbol{\beta}_2,H_2}(x,\delta,\boldsymbol{z})$ for all $(x,\delta,\boldsymbol{z})$. Putting $\delta = 1$, we can work only with the first term in the definition of $f_{\alpha,\boldsymbol{\beta},H}$. Hence for all $(x,\boldsymbol{z})$,

$$\begin{aligned}&\int_0^\infty \alpha_1(\mu e^{\boldsymbol{z}^T\boldsymbol{\beta}_1})^{-\alpha_1}x^{\alpha_1-1}\exp[-(x/\mu e^{\boldsymbol{z}^T\boldsymbol{\beta}_1})^{\alpha_1}]dH_1(\mu)\\ &= \int_0^\infty \alpha_2(\mu e^{\boldsymbol{z}^T\boldsymbol{\beta}_2})^{-\alpha_2}x^{\alpha_2-1}\exp[-(x/\mu e^{\boldsymbol{z}^T\boldsymbol{\beta}_2})^{\alpha_2}]dH_2(\mu). \qquad (7)\end{aligned}$$

Putting $\boldsymbol{z} = 0$, (**??**) reduces to

$$\begin{aligned}&\int_0^\infty \alpha_1\mu^{-\alpha_1}x^{\alpha_1-1}\exp[-(x/\mu)^{\alpha_1}]dH_1(\mu)\\ &= \int_0^\infty \alpha_2\mu^{-\alpha_2}x^{\alpha_2-1}\exp[-(x/\mu)^{\alpha_2}]dH_2(\mu) \qquad (8)\end{aligned}$$

for all $x > 0$. Taking ratio, letting $x \to 0$ and using the dominated convergence theorem (DCT), we obtain

$$\lim_{x\to 0} x^{\alpha_2-\alpha_1} = \frac{\int_0^\infty \alpha_1\mu^{-\alpha_1}dH_1(\mu)}{\int_0^\infty \alpha_2\mu^{-\alpha_2}dH_2(\mu)}. \qquad (9)$$

Because of compact supports, the right hand side (RHS) of (**??**) is a finite positive number, while the left hand side (LHS) is so only when $\alpha_1 = \alpha_2$. Thus it follows that $\alpha_1 = \alpha_2 = \alpha$, say. Using the transformation $\mu \mapsto \mu^{-\alpha}$ and letting $\tilde{H}_1$ and $\tilde{H}_1$ respectively as the probability measures induced from H_1 and H_2 by this transformation, we obtain, for all $x > 0$,

$$\int_0^\infty \lambda\exp[-\lambda x^\alpha]d\tilde{H}_1(\lambda) = \int_0^\infty \lambda\exp[-\lambda x^\alpha]d\tilde{H}_2(\lambda). \qquad (10)$$

Thus the Laplace transforms of the two measures $\lambda d\tilde{H}_1(\lambda)$ and $\lambda d\tilde{H}_1(\lambda)$ agree everywhere, implying that the measures are themselves equal. Hence $\tilde{H}_1$ and $\tilde{H}_2$ are equal and so are H_1 and H_2.

Now going back to (**??**), letting $x \to 0$ and using the DCT, we obtain, for all $\boldsymbol{z}$, $e^{-\boldsymbol{z}^T\boldsymbol{\beta}_1} = e^{-\boldsymbol{z}^T\boldsymbol{\beta}_2}$, and so $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$. This completes the proof. □

LEMMA 2. *Consider the product topology on $(\alpha, \boldsymbol{\beta}, H)$, where α and $\boldsymbol{\beta}$ are given the usual Euclidean topology and H the weak topology. On the densities $f_{\alpha,\boldsymbol{\beta},H}(x, \delta, \boldsymbol{z})$, put the total variation (or the L_1) distance defined as*

$$\|f_{\alpha_1,\boldsymbol{\beta}_1,H_1} - f_{\alpha_2,\boldsymbol{\beta}_2,H_2}\| = \int\int |f_{\alpha_1,\boldsymbol{\beta}_1,H_1}(x, \delta, \boldsymbol{z}) - f_{\alpha_2,\boldsymbol{\beta}_2,H_2}(x, \delta, \boldsymbol{z})|\, dx d\boldsymbol{z}. \quad (11)$$

Then

$$\|f_{\alpha_\nu,\boldsymbol{\beta}_\nu,H_\nu} - f_{\alpha,\boldsymbol{\beta},H}\| \to 0$$

if and only if $(\alpha_\nu, \boldsymbol{\beta}_\nu, H_\nu) \to (\alpha, \boldsymbol{\beta}, H)$. In other words, the variation topology on the densities is equivalent to the product topology on the indexing parameters.

PROOF. For the "if" part, it suffices to show that the densities converge pointwise and then apply Scheffe's theorem. Fix $(x, \delta, \boldsymbol{z})$. We shall show the proof only for $\delta = 1$.

As μ has a fixed compact range, the integrand $\alpha x^{\alpha-1}(\mu e^{\boldsymbol{z}^T\boldsymbol{\beta}})^{-\alpha} \exp[-(x/\mu e^{\boldsymbol{z}^T\boldsymbol{\beta}})^\alpha]$ in the definition of $f_{\alpha,\boldsymbol{\beta},H}$, as a family of functions of $(\alpha, \boldsymbol{\beta})$ indexed by μ is equicontinuous. For a given $\varepsilon > 0$, find ν large enough so that the integrands are ε-close for all μ. For such a ν, in the decomposition,

$$|f_{\alpha_\nu,\boldsymbol{\beta}_\nu,H_\nu} - f_{\alpha,\boldsymbol{\beta},H}| \le |f_{\alpha_\nu,\boldsymbol{\beta}_\nu,H_\nu} - f_{\alpha,\boldsymbol{\beta},H_\nu}| + |f_{\alpha,\boldsymbol{\beta},H_\nu} - f_{\alpha,\boldsymbol{\beta},H}|, \quad (12)$$

the first term on the RHS is clearly less than ε. The second term on the RHS converges to 0 as $\nu \to \infty$ since H_ν converges weakly to H and the integrand is a bounded continuous function of μ.

To prove the "only if" part, consider any subsequence of the original sequence. By compactness of the domains of α, $\boldsymbol{\beta}$ and μ, we can extract a further subsequence along which α_ν, $\boldsymbol{\beta}_\nu$ and H_ν converges in the respective topologies to α^*, $\boldsymbol{\beta}^*$ and H^*, say. Thus by the "if" part, along that subsequence, $\|f_{\alpha_\nu,\boldsymbol{\beta}_\nu,H_\nu} - f_{\alpha^*,\boldsymbol{\beta}^*,H^*}\| \to 0$. Hence $f_{\alpha,\boldsymbol{\beta},H} = f_{\alpha^*,\boldsymbol{\beta}^*,H^*}$, and so by Proposition **??**, $(\alpha, \boldsymbol{\beta}, H) = (\alpha^*, \boldsymbol{\beta}^*, H^*)$. Therefore the original sequence must converge to $(\alpha, \boldsymbol{\beta}, H)$. □

LEMMA 3. *For all* $\varepsilon > 0$,

$$\Pi\left\{(\alpha, \boldsymbol{\beta}, H) : \int\int f_{\alpha_0,\boldsymbol{\beta}_0,H_0} \log \frac{f_{\alpha_0,\boldsymbol{\beta}_0,H_0}}{f_{\alpha,\boldsymbol{\beta},H}}\, dxd\boldsymbol{z} < \varepsilon\right\} > 0.$$

PROOF. The log likelihood ratio is given by

$$\Lambda(\alpha, \boldsymbol{\beta}, H) = \begin{cases} \log \frac{p_{\alpha,\boldsymbol{\beta},H}(x,\boldsymbol{z},\delta)}{p_{\alpha_0,\boldsymbol{\beta}_0,H_0}(x,\boldsymbol{z},\delta)}, & \text{if } \delta = 1, \\ \log \frac{S_{\alpha,\boldsymbol{\beta},H}(x,\boldsymbol{z},\delta)}{S_{\alpha_0,\boldsymbol{\beta}_0,H_0}(x,\boldsymbol{z},\delta)}, & \text{if } \delta = 0. \end{cases}$$

We shall show for $\delta = 1$; the other one is similar (in fact, simpler). Note that $\boldsymbol{\beta}$, α, $\boldsymbol{z}$ are all bounded. Thus the integrand defining $f_{\alpha,\boldsymbol{\beta},H}$ is bounded above and below by functions of the form $Cx^{\kappa_1}e^{cx^{\kappa_2}}$. Taking ratio and then logarithm, Λ in the tails (in x) is bounded by a multiple of a power of x. This is integrable with respect to the true density, and hence we can find a central region, outside which, the contribution to the integral $\int\int f_{\alpha_0,\boldsymbol{\beta}_0,H_0} \log \frac{f_{\alpha_0,\boldsymbol{\beta}_0,H_0}}{f_{\alpha,\boldsymbol{\beta},H}}\, dxd\boldsymbol{z}$ is small. In the central region, the family indexed by $(x, \boldsymbol{z})$ is equicontinuous in the parameters. Rest of the proof follows as in that of Theorem 3 of Ghosal *et al.* (1999). □

PROOF OF THEOREM **??**. Lemma **??** shows that it suffices to consider neighborhoods with respect to the distance (**??**). Also, the space is compact. The condition of prior positivity has been verified in Lemma **??**. The proof therefore follows. □

References

AMEWOU-ATISSO, M., GHOSAL, S., GHOSH, J. K. and RAMAMOORTHI, R. V. (2003). Posterior consistency for semiparametric regression problems. *Bernoulli* **9** 291–312.

BANFIELD, J. and RAFTERY, A. (1993). Model based Gaussian and non-Gaussian clustering. *Biometrics* **49** 803–821.

BARRON, A., SCHERVISH, M. and WASSERMAN, L. (1999). The consistency of posterior distributions in nonparametric problems. *Ann. Statist.* **27** 536–561.

Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika*, **66**, 429–436.

Chen, K., Jin, Z. and Ying, Z. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika*, **89**, 659-668.

Cheng, S. C., Wei, L. J. and Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika*, **82**, 835-845.

Cox, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc., Ser. B*, **34**, 187–202.

Dey, D., Müller, P. and Sinha (1998). *Practical Nonparametric and Semiparametric Bayesian Statistics*, vol. 133 Lecture Notes in Statistics, Springer-Verlag, New York.

Diaconis, P. and Freedman, D. (1986). On the consistency of Bayes estimates (with discussion). *Ann. Statist.* **14** 1–67.

Edmunson, J. H., Fleming, T. R., Decker, G. D., Malkasam, G. D., Kvols, L. K. (1979). Different chemotherapeutic sensitivities and host factors affecting prognosis in advanced ovarian carcinoma versus minimal disease residual. *Cancer Treatment Reports*, **63**, 241-247.

Escobar, M. and West, M. (1995). Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.* **90** 577–588.

Ferguson, T. S. (1983). Bayesian density estimation by mixtures of normal distributions. In *Recent Advances in Statistics* (Rizvi M., Rustagi, J. and Siegmund, D., Eds.) 287–302.

Ghosal, S. (2001). Convergence rates for density estimation with Bernstein polynomials. *Ann. Statist.* **29** 1264–1280.

Ghosal, S., Ghosh, J. K. and Ramamoorthi, R. V. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *Ann. Statist.* **27** 143–158.

Ghosal, S. and van der Vaart, A. W. (2001). Entropies and rates of convergence of maximum likelihood and Bayes estimation for mixtures of normal densities. *Ann. Statist.* **29** 1233–1263.

Ghosh, J. K. and Ramamoorthi, R. V. (1994). Consistency of Bayesian inference for survival analysis with or without censoring. In *Analysis of Censored Data* (Pune, 1994/1995), 95–103, IMS Lecture Notes Monogr. Ser., **27**, Inst. Math. Statist., Hayward, CA.

Ghosh, J. K. and Ramamoorthi, R. V. (2003). *Bayesian Nonparametrics*, Springer-Verlag, New York.

Ghosh, J. K., Ramamoorthi, R. V. and Srikanth, K. R. (1999). Bayesian analysis of censored data. Special issue in memory of V. Susarla. *Statist. Probab. Lett.* **41** 255–265.

Gilks, W. R. (1992). Derivative-free adaptive rejection sampling for Gibbs sampling. In *Bayesian Statistics 4* (eds Bernardo, Berger, Dawid and Smith), 641–649, Oxford University Press, Oxford.

Ishwaran, H. and Zarepour, M. (2002). Dirichlet prior sieves in finite normal mixtures. *Statistica Sinica*, **12**, 941–963.

Jin, Z. Lin, D. Y., Wei, L. J. and Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika*, **90**, 341–353.

Kim, Y. and Lee, J. (2001). On posterior consistency of survival models. *Ann. Statist.* **29** 666–686.

Kottas A. and Gelfand, A. E. (2002). Bayesian Semiparametric Median Regression Modeling. *J. Amer. Statist. Assoc.*, **96**, 1458–1468

Koul, H., Sursula, V. and Van Ryzin, J. (1981). Regression analysis with randomly right censored data. *Ann. Statist.*, **9**, 1276–1288.

Krall, J.N., Uthoff, V. A. and Harley, J. B.. (1975). A set-up procedure for selecting variables associated with survival. *Biometrics*, **31**, 49–57.

Lai, T. L., Ying, Z. and Zheng, Z. (1992). Asymptotic normality of a class of adaptive statistics with applications to synthetic data methods for censored regression. *J. Multivariate Anal.*, **52**, 259–279.

Leurgans, S. (1976). Linear models, random censoring and synthetic data. *Biometrika*, **74**, 301–309.

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

Lo, A. Y. (1994). On a class of Bayesian nonparametric estimates I: Density estimates. *Ann. Statist.* **12** 351–357.

McCullagh P. and Nelder, J. A. (1989). *Generalized Linear Models* (2nd Edition), Chapman & Hall, London.

McLachlan, G. and Basford, K. (1988). *Mixture Models: Inference and Applications to Clustering*. Marcel-Dekker, New York.

Miller, R. G. (1976). Least squares regression with censored data. *Biometrika*, **63**, 449-464.

Miller, R. and Halpern, J. (1982). Regression with censored data. *Biometrika*, **69**, 521-531.

Park Y. and sc Wei, L. J. (2003). Estimating subject-specific survival functions under the accelerated failure time model. *Biometrika*, **90**, 717–723.

Petrone, S. (1999). Bayesian density estimation using Bernstein polynomials. *Canad. J. Statist.* **26** 373–393.

Petrone, S. and Veronese, P. (2002). Nonparametric mixture priors based on an exponential random scheme. *Statistical Methods and Applications*, **11**, 1–20.

Petrone, S. and Wasserman, L. (2002). Consistency of Bernstein polynomial posteriors. *J. Roy. Statist. Soc., Ser. B* **64** 79–100.

Reid, N. (1994). A conversation with Sir David Cox. *Statist. Sci.*, **9**, 439-455.

Robert, C. (1996). Mixtures of distributions: inference and estimation. In *Markov Chain Monte Carlo in Practice* (W. Gilks, S. Richardson and D. Spiegelhalter, eds.), Chapman and Hall, London, 441–464.

Roeder, K. and Wasserman, L. (1997). Practical Bayesian density estimation using mixtures of normals. *J. Amer. Statist. Assoc.* **92** 894–902.

Schwartz, L. (1965). On Bayes procedures. *Z. Wahr. Verw. Gebiete* **4** 10–26.

Spiegelhalter, D. J., Thomas, A. and Best, N. (1999). *WinBUGS Version 1.2 User Manual*, MRC Biostatistics Unit.
`http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml`

Wei, L. J. (1992). The accelerated failure time model: A useful alternative to the Cox regression model in survival analysis. *Statistics in Medicine*, **11**, 1871-1879.

Zheng, Z. (1984). *Regression with Randomly Censored Data.* Ph.D Thesis, Columbia University, NY.

Zhou, M. (1992). Asymptotic normality of the 'synthetic data' regression estimator for censored survival data. *Ann. Statist.*, **20**, 1002-1021.