Sankhyā : The Indian Journal of Statistics
2008, Volume 70, Part 3, pp. 0-46

Posterior Consistency for some Semi-parametric Problems

Yuefeng Wu
North Carolina State University, USA
Subhashis Ghosal
North Carolina State University, USA

Abstract

The Bayesian approach to analyzing semi-parametric models are gaining popularity in practice. For the Cox proportional hazard model, it has been shown recently that the posterior is consistent and leads to asymptotically accurate confidence intervals under a Lévy process prior on the cumulative hazard rate. The explicit expression of the posterior distribution together with independent increment structure of Lévy process play a key role in the development. However, except for one-dimensional linear regression with an unknown error distribution and binary response regression with unknown link function, even consistency of Bayesian procedures has not been studied for a general prior distribution. We consider consistency of Bayesian inference for several semi-parametric models including multiple linear regression with an unknown error distribution, exponential frailty model, generalized linear model with unknown link function, Cox proportional hazard model where the baseline hazard function is unknown, accelerated failure time models and partial linear regression model. We give sufficient conditions under which the posterior distribution of the parametric part is consistent in the Euclidean distance while the non-parametric part is consistent with respect to some topology such as the weak topology. Our results are obtained by verifying the conditions of an appropriate modification of a celebrated result of Schwartz. Our general consistency result applies also to the case of independent, non-identically distributed observations. Application of our theorem requires showing the existence of exponentially consistent tests for the complements of the neighborhoods of the "true" value of the parameter and the prior positivity of a Kullback-Leibler type of neighborhood of the true distribution of the observations. We construct the required tests and give sufficient conditions for positivity of prior probabilities of Kullback-Leibler neighborhoods in all the examples we consider in this paper.

AMS (2000) *subject classification.* Primary 62C10,62F15,60G09; secondary 62F03, 62H15.
Keywords and phrases. Consistency, exponentially consistent test, Kullback-Leibler property, multiple linear regression, exponentially frailty model, generalized linear model.

1 Introduction

Semiparametric models such as the linear regression model with unknown error distribution or Cox proportional hazard model, where the distribution of the data is governed by a finite dimensional parameter as well as an infinite dimensional parameter, are widely used in practice because of their flexibility and interpretability of parameters. In such models, typically the finite dimensional part is of interest, while the infinite dimensional part is treated as nuisance. Consistent and efficient methods of estimation of the parameters, especially estimation of the finite dimensional part at the $\sqrt{n}$ rate, have been developed in the literature; see Bickel *et al.* (1988) and van der Vaart and Wellner (1996) for details. The Bayesian approach has also gained recent popularity, although its asymptotic properties have not been understood well, partly because even the posterior of the parametric part is hard to analyze without analyzing the whole posterior, unless one can obtain closed form expressions for the posterior distribution. For the Cox proportional hazard model, recently Kim (2006) showed that the posterior is consistent and the posterior distribution of the parametric part β satisfies a Bernstein-von Mises theorem in the sense that the Bayesian and the frequentist distribution of $\sqrt{n}(\hat{\beta} - \beta)$ are approximately the same, where $\hat{\beta}$ is some suitable classical estimator of β and the cumulative hazard function is given a Lévy process prior. This implies that the credible intervals for β obtained from the posterior distribution of it has asymptotically correct frequentist coverage. In the analysis, the posterior distribution can be explicitly expressed in terms of a new Lévy process, which has a new Lévy measure depending on the data. The nice structural property of a Lévy process, namely its independent increments, makes it possible to represnt the posterior distribution in a manageable form. Shen (2002) obtained similar results for general semiparametric models, but it appears that the conditions are harder to verify; see also Castillo (2008) for some recent results. For many other semiparametric problems, especially if the prior only satisfies some general conditions, much less is known about the asymptotic properties of the posterior distribution. Except for one-dimensional linear regression with an unknown error distribution (which is assumed to be symmetric for identifiability reasons) and binary response regression with unknown link function, even consistency of semi-parametric Bayesian procedures has not been studied.

In this paper, we fill up this gap by studying posterior consistency for various semi-parametric models including multiple linear regression with an unknown error distribution, exponential frailty model, generalized linear model (GLM) with unknown link function, Cox proportional hazard model, accelerated failure time (AFT) models and partial linear regression model. We begin with a general posterior consistency theorem for semiparametric models much in line of results of a celebrated result of Schwartz (1965) involving conditions on existence of certain exponentially consistent tests for the complement of the neighborhoods of the "true" value of the parameter and the prior positivity of a Kullback-Leibler type of neighborhood of the true distribution of the observations. Our result applies also to the case of independent, non-identically distributed observations. In the examples, we then verify the existence of exponentially consistent tests and the prior positivity condition.

Consistency implies that one's knowledge about the parameters converges to the perfect knowledge, in the sense that a degenerate distribution for a parameter stands for the perfect knowledge about it. Some early results on posterior consistency are obtained by Doob (1948), Freedman (1963, 1965), Schwartz (1965), Diaconis and Freedman (1986) and Doss (1985). More recently, important techniques of proving posterior consistency have been developed in Barron *et al.* (1999), Ghosal *et al.* (1999a, 1999b), Amewou-Atisso *et al.* (2003), Walker (2004), Choudhuri *et al.* (2004), Tokdar (2006), Wu and Ghosal (2008) and many other contemporary papers. To review the literature, the reader may consult Ghosh and Ramamoorthi (2003), Choudhuri *et al.* (2005) and Ghosal (2009).

As in Amewou-Atisso *et al.* (2003), we use a variant of Schwartz's (1965) theorem for independent, non-identically distributed variables. Indeed, our result, which addresses a general semiparametric problem and allows the use of sieves, is an easy generalization of Theorem 2.1 of Amewou-Atisso *et al.* (2003). This general result is discussed in Section 2. Subsequently in Sections in 3–8, we give verifiable sufficient conditions under which the consistency holds for the multiple regression model, exponential frailty model, GLM with unknown link function, Cox proportional hazard model respectively, AFT model and partial linear regression model, respectively. In each of the examples, the main task is to show the existence of the exponentially consistent tests for the true value of the parameter against the complement of a neighborhood intersected with a sieve. Proofs of auxiliary lemmas are deferred to the appendix. We conclude the paper with a discussion section.

2 Consistency of Posterior

Let $(a(\cdot),\theta)$ denote the unknown parameters of the semi-parametric model, where $a(\cdot)$ denotes the infinite dimensional parameter of the model and θ denotes the finite dimensional parameter of the model. Let $\Xi = \mathscr{A} \times \Theta$ be the parameter space, where $\mathscr{A}$ denotes the space of $a(\cdot)$ and Θ denotes the space of θ. Assume that Θ is a subset of $\mathbb{R}^d$. We give prior $\tilde{\Pi}$ on $\mathscr{A}$ and μ on Θ. Let Π stand for $\tilde{\Pi} \times \mu$.

Let Y_i, $i = 1, 2, \ldots$ denote the observations of the random variable of primary interest. Let X_i's or x_i's denote the corresponding covariates, which are respectively independent and identically distributed (i.i.d.) with cumulative distribution function (c.d.f.) $Q(x)$, or given as fixed non-random constants. Throughout this paper, Y is always one-dimensional variable, while X and x are d-dimensional, with $d \geq 1$. For the cases with random covariates, the density of $(Y|X)$ is denoted by $f_{a,\theta}$, corresponding to the parameters $a(\cdot)$ and θ. For the cases with fixed covariates, the density of Y_i's are denoted by $f_{a,\theta,i}$, corresponding to the parameters $a(\cdot)$, θ and the given covariate x_i's. Let $a_0(\cdot)$ and θ_0 denote the true values of the parameters, and put $f_0 = f_{a_0,\theta_0}$ and $f_{0i} = f_{a_0,\theta_0,i}$.

The sequence of posterior distribution $\Pi(\cdot|Y_1,\ldots,Y_n)$ or $\Pi(\cdot|(X_1,Y_1),\ldots,(X_n,Y_n))$ for $(a(\cdot),\theta)$, corresponding to non-stochastic covariates and stochastic covariates cases respectively, is said to be *consistent* at $(a_0(\cdot),\theta_0)$ if $\Pi(\mathscr{N}|Y_1,\ldots,Y_n)$ or $\Pi(\mathscr{N}|(X_1,Y_1),\ldots,(X_n,Y_n))$ respectively, converges to 1 almost surely (a.s.) as $n \to \infty$ for any neighborhood $\mathscr{N}$ of $(a_0(\cdot),\theta_0)$, when the distribution governing $Y_1, Y_2, \ldots$ or $(X_1,Y_1),(X_2,Y_2),\ldots$ has the "true" parameter $(a_0(\cdot),\theta_0)$. With the weak topology on $\mathscr{A}$ and the Euclidean distance on Θ, let $\mathscr{U}_{a_0}$ be a weak neighborhood of a_0, we only need to consider every $\mathscr{N}$ of the form $\mathscr{U}_{a_0} \times \{\theta : \|\theta - \theta_0\| < r\}$, for any $r > 0$.

Now, we give the definition of exponentially consistent tests.

DEFINITION 2.1. Let $\mathscr{W} \subset \Xi$. A sequence of tests $\Phi_n(\cdot)$, for testing $\mathrm{H}_0 : (a(\cdot),\theta) = (a_0(\cdot),\theta_0)$ against $\mathrm{H}_1 : (a(\cdot),\theta) \in \mathscr{W}$, is said to be *exponentially consistent* if for non-stochastic covariates x_i's, there exist constants C_1, C_2, $C > 0$, such that $\mathrm{E}_{\prod_1^n f_{0i}}(\Phi_n) \leq C_1 e^{-nC}$, and $\inf_{(a(\cdot),\theta)\in\mathscr{W}} \mathrm{E}_{\prod_1^n f_{a,\theta,i}}(\Phi_n) \geq 1 - C_2 e^{-nC}$, or for i.i.d. X_i's, there exist constants C_1, C_2, $C > 0$, such that $\int[\mathrm{E}_{f_0^n}(\Phi_n)]Q^n(dx) \leq C_1 e^{-nC}$, and $\inf_{(a(\cdot),\theta)\in\mathscr{W}} \int[\mathrm{E}_{f_{a,\theta,}^n}(\Phi_n)]Q^n(dx) \geq 1 - C_2 e^{-nC}$.

For any two densities f and g, let $K(f,g) = \int f \log \frac{f}{g}$, $V_+(f,g) = \int f \log_+ (\frac{f}{g})^2$, where $\log_+ x = \max(0, \log x)$. Put $K_i(a,\theta) = K(f_{0i}, f_{a,\theta,i})$, $K(a,\theta) = K(f_0, f_{a,\theta})$ and $V_i(a,\theta) = V_+(f_{0i}, f_{a,\theta,i})$.

For stochastic covariates, we restrict our attention to the case where the covariate X_i's are i.i.d.

THEOREM 2.1. *Let $\mathscr{W} \subset \mathscr{A} \times \Theta$ and let X_i's be i.i.d. with distribution function Q. If for some sequences $\mathscr{A}_n \subset \mathscr{A}$ and $m_n \leq M_n$,*

(i) *there is an exponentially consistent sequence of tests for* $\mathrm{H}_0 : (a(\cdot), \theta) = (a_0(\cdot), \theta_0)$ *against* $\mathrm{H}_1 : (a(\cdot), \theta) \in \mathscr{W} \cap \Big(\mathscr{A}_n \times [m_n, M_n]^d\Big)$;

(ii) *for all* $\delta > 0$, $\Pi\Big\{(a(\cdot), \theta) : \int K(a,\theta) dQ < \delta\Big\} > 0$;

(iii) $\Pi[(\mathscr{A}_n \times [m_n, M_n]^d)^c] \leq c_1 e^{-nc_2}$ *for some constants* c_1, $c_2 > 0$;

then with $(P_{f_0}^{\infty})$*-probability* 1,

$$\begin{aligned} &\Pi(\mathscr{W} | (X_1, Y_1), \ldots, (X_n, Y_n)) \\ &= \frac{\int_{\mathscr{W}} \prod_{i=1}^n (f_{a,\theta}(X_i, Y_i)/f_0(X_i, Y_i)) d\Pi(a,\theta)}{\int_{\mathscr{A} \times \Theta} \prod_{i=1}^n (f_{a,\theta}(X_i, Y_i)/f_0(X_i, Y_i)) d\Pi(a,\theta)} \to 0. \end{aligned} \tag{2.1}$$

This is a plain extension of the original Schwartz Theorem. The following is a straightforward but useful generalization of Theorem 2.1 in Amewou-Atisso *et al.*(2003).

THEOREM 2.2. *Let $\mathscr{W} \subset \mathscr{A} \times \Theta$ and x_i's be non-stochastic. If for some sequences $\mathscr{A}_n \subset \mathscr{A}$ and $m_n \leq M_n$,*

(i) *there is an exponentially consistent sequence of tests for* $\mathrm{H}_0 : (a(\cdot), \theta) = (a_0(\cdot), \theta_0)$ *against* $\mathrm{H}_1 : (a(\cdot), \theta) \in \mathscr{W} \cap \Big(\mathscr{A}_n \times [m_n, M_n]^d\Big)$;

(ii) *for all* $\delta > 0$,

$$\Pi\Big\{(a(\cdot), \theta) : K_i(a,\theta) < \delta \text{ for all } i, \text{ and } \sum_{i=1}^{\infty} \frac{V_i(a,\theta)}{i^2} < \infty\Big\} > 0;$$

(iii) $\Pi[(\mathscr{A}_n \times [m_n, M_n]^d)^c] \leq c_1 e^{-nc_2}$ *for some constants* c_1, $c_2 > 0$;

then with $(\Pi_1^\infty P_{f_{0i}})$*-probability* 1,

$$\Pi(\mathscr{W}|Y_1,\ldots,Y_n) = \frac{\int_{\mathscr{W}} \prod_1^n (f_{a,\theta,i}(Y_i)/f_{0i}(Y_i))d\Pi(a,\theta)}{\int_{\mathscr{A}\times\Theta} \prod_1^n (f_{a,\theta,i}(Y_i)/f_{0i}(Y_i))d\Pi(a,\theta)} \to 0. \tag{2.2}$$

Note that in the above two theorems, to allow more flexibility, the exponentially consistent tests are assumed to exist on the sieve $\mathscr{A}_n \times [m_n, M_n]^d$ only. For many semi-parametric applications, the use of a sieve is not necessary. For example, for all the semi-parametric models discussed later, sieves are used only for the exponential frailty model and generalized linear models with unknown link functions.

3 Multiple Regression

Now we generalize the result about consistency for the semi-parametric regression model to higher dimensions with stochastic or non-stochastic regressors. Consider a multiple regression model $Y_i = \alpha + X_i^T\beta + \varepsilon_i, \quad i = 1,2,\ldots,$ where X_i's and β are d-dimensional, ε_i's are i.i.d. f, with f having the same value of some fixed quantile. The unknown parameters are f, α and β, so the parameter space can be taken as $\Xi = \mathscr{F} \times \mathbb{R} \times \mathbb{R}^d$, where, for identifiability, we work with $\mathscr{F} = \{f : \int_{-\infty}^0 f(x)dx = a\}$ for some $0 < a < 1$ in this section. The choice $a = 1/2$ is common but not mandatory. Assume that $X_1,\ldots,X_n$ are i.i.d. with probability measure Q. Also, assume that X_i and ϵ_i are independent. We start with a prior $\tilde{\Pi}$ for f on $\mathscr{F}$ and independent of f, a prior μ for (α,β). Let Π stand for $\tilde{\Pi}\times\mu$. For this model, let $f_{\alpha,\beta} = f(y-\alpha-x^T\beta)$. Let $f_0 = f_{0,\alpha_0,\beta_0}$. For a density g and $\theta \in \mathbb{R}$, let g_θ stand for the density $g_\theta = g(y-\theta)$. Let $\gamma \in \{-1,1\}^d$, we define quadrant $\Gamma_\gamma = \{z \in \mathbb{R}^d : z_j\gamma_j > 0 \text{ for all } j = 1,\ldots,d\}$. Through this paper, let Υ_i denote the d-dimensional vector with all the components being 0 except the i-th component being 1. The following theorem is a generalization of Theorem 4.1 in Amewou-Atisso *et al.* (2003).

THEOREM 3.1. *Suppose that*

(i) *the covariate* X *is compactly supported, and for any* γ, $Q(\Gamma_\gamma\backslash\{X : |X^T\Upsilon_i| < \zeta\}) > 0$ *for each* $i = 1,\ldots,d$, *and some* $\zeta > 0$;

(ii) f_0 *is continuous,* $f_0(0) > 0$, *and for* $\eta > 0$ *sufficiently small, there exists* $g_\eta \in \mathscr{F}$ *and constant* $C_\eta > 0$ *such that for* $|\eta'| < \eta$, $f_0(y-\eta') < C_\eta g_\eta(y)$ *for all* y *and* $C_\eta \to 1$ *as* $\eta \to 0$;

(iii) *for all sufficiently small* η *and all* $\xi > 0$, $\tilde{\Pi}\{K(g_\eta, f) < \xi\} > 0$ *and* (α_0, β_0) *is in the support of* μ.

Then for any weak neighborhood $\mathscr{U}$ *of* f_0,

$$\Pi\Big\{(f, \alpha, \beta) : f \in \mathscr{U}, |\alpha - \alpha_0| < \epsilon, \|\beta - \beta_0\| < \epsilon \Big| (X_1, Y_1), \ldots, (X_n, Y_n)\Big\} \to 1$$

a.s. $P_{f_0}^{\infty}$.

To prove this theorem, we use the following lemma to show the existence of the exponentially consistent tests.

Lemma 3.1. *If Condition* (i) *of Theorem* 3.1 *holds, then there exist exponentially consistent tests for testing* $\mathrm{H}_0 : (f, \alpha, \beta) = (f_0, \alpha_0, \beta_0)$ *against* $\mathrm{H}_1 : \{(f, \alpha, \beta) : f \in \mathscr{U}_{f_0}, |\alpha - \alpha_0| < \epsilon, \|\beta - \beta_0\| < \epsilon\}$.

The proof of this lemma is given in the appendix.

Proof of Theorem 3.1. In view of Lemma 3.1, now we need only to verify Condition (ii) of Theorem 1.

For this model, to verify Condition (ii) of Theorem 1 is to show

$$\Pi\{(f, \alpha, \beta) : \int K(f_0, f_{\alpha,\beta}) dQ < \delta\} > 0,$$

for all $\delta > 0$. By Condition (ii) and applying Lemma 4.1 of Amewou-Atisso *et al.* (2003), we have that for any $f \in \mathscr{F}$ and $|\theta| < \eta$,

$$K(f_0, f_\theta) \leq (C_\eta + 1)\log C_\eta + C_\eta[K(g_\eta, f) + \sqrt{K(g_\eta, f)}].$$

Hence

$$\begin{aligned}
&\iint f_0(y - \alpha_0 - x^T\beta_0)\log\frac{f_0(y - \alpha_0 - x^T\beta_0)}{f(y - \alpha - x^T\beta)} dy dQ(x) \\
&= \iint f_0(y)\log\frac{f_0(y)}{f(y - (\alpha - \alpha_0) - x^T(\beta - \beta_0))} dy dQ(x) \\
&\leq (C_\eta + 1)\log C_\eta + C_\eta[K(g_\eta, f) + \sqrt{K(g_\eta, f)}],
\end{aligned}$$

for any $f \in \mathscr{F}$ and all (α, β) such that $|(\alpha - \alpha_0) + (\beta - \beta_0)x| < \eta$ for all possible x.

Now, by Condition (iii), we have that $\Pi\{(f, \alpha, \beta) : \int K(f_0, f_{\alpha,\beta}) dQ < \delta\} > 0$. An application of Theorem 2.1 completes the proof. □

REMARK 3.1. Condition (i) is automatically satisfied if X has a compactly supported positive density. A Polya tree prior with appropriate parameters satisfies Conditions (ii) and (iii) required by this theorem; see Section 5 of Amewou-Atisso *et al.* (2003) for details.

REMARK 3.2. If the covariate is non-stochastic, then the above theorem need some modifications. We need to change Condition (i) to the following: (i′) for some $L > 0$, $\|x_i\| < L$ for all i, and there exits $\epsilon_0 > 0$ such that the covariate $X_i's$ satisfy

$$\liminf n^{-1}\#\{i : x_i \in (\Gamma_\gamma \backslash \{x_i : |x_i^T \Upsilon_j| < \zeta\}) > 0,$$

for some $\zeta > 0$ and each $j = 1, \ldots, d$, and γ.

Also, we need to add the requirement that $V(g_\eta, f) < \infty$ in the definition of the set appearing in Condition (iii).

Now, consider the case that the random density f is given the Dirichlet mixture of normal prior. For any probability P on $\mathbb{R}$, let $f_{h,P}$ stand for the density

$$f_{h,P}(y) = \frac{1}{2}\int \phi_h(y-t)dP(t) + \frac{1}{2}\int \phi_h(y+t)dP(t),$$

where ϕ_h is the normal density with mean 0 and standard deviation h. Let ν denote the prior for h and D_π denote a Dirichlet process prior for the mixing distribution P with π as its base measure. Let $\tilde{\Pi} = D_\pi \times \nu$ denote the prior for f. Note that $f_{h,P}$ is a symmetric density. For the case with i.i.d. covariate X_i, we have the following result:

THEOREM 3.2. *Suppose $\tilde{\Pi}$ is a normal mixture prior for f as described above. If*

(i) *the covariate X is compactly supported, and for any γ, $Q(\Gamma_\gamma \backslash \{X : |X^T \Upsilon_i| < \zeta\}) > 0$ for each $i = 1, \ldots, d$, and some $\zeta > 0$;*

(ii) *$\tilde{\Pi}\{f : \int K(f_0, f)dQ < \delta\} > 0$ for all $\delta > 0$;*

(iii) *$\int y^2 f_0(y)dy < \infty$, $\mathrm{E}_{f_0}|\log f_0| < \infty$;*

(iv) *$\int t^2 dP(t) < \infty$ a.s. D_π;*

then the posterior distribution $\Pi(\cdot|(X_1, Y_1), \ldots, (X_n, Y_n))$ for (f, α, β) is weakly consistent at (f_0, α_0, β_0) provided (α_0, β_0) is in the support of the prior for (α, β).

To prove this Theorem, we need the following Lemma.

LEMMA 3.2. *Let f_0 be a density such that Condition* (iii) *of Theorem* 3.2 *holds. If $f(y) = \int \phi_h(y-t)dP(t)$ and Condition* (iv) *of Theorem* 3.2 *holds, then*

$$\lim_{\theta \to 0} K(f_0, f_\theta) = K(f_0, f).$$

PROOF. Obviously, $f(y)$ is positive and continuous, and

$$|\log f_\theta(y)| \leq |\log(\sqrt{2\pi}h)| + \left|\log \int e^{-(y-\theta-t)^2/(2h^2)} dP(t)\right|.$$

Applying Jensen's inequality to $-\log x$, the above expression is bounded by

$$|\log(\sqrt{2\pi}h)| + \int \frac{(y-\theta-t)^2}{h^2} dP(t).$$

The dominated convergence theorem (DCT) now applies. □

PROOF OF THEOREM 3.2. The existence of exponentially consistent tests is shown by Lemma 3.1. We only need to verify the Condition (ii) of Theorem 2.1. The verification in the following is very similar to the proof of Theorem 6.1 in Amewou-Atisso it et al. (2003). By (iv), $\int t^2 dP(t) < \infty$ a.s. D_π. So we may assume that

$$\tilde{\Pi}\{\mathscr{U}\} > 0,$$

where $\mathscr{U} = \{f : f = f_P, \text{(ii) holds}, \int t^2 dP(t) < \infty\}$.

For every $f \in \mathscr{U}$, using Lemma 3.2, choose δ_f such that, for $\|\theta\| > \delta_f$,

$$K(f_\theta, f) < \delta.$$

Now choose ϵ_f such that $|(\alpha - \alpha_0) + x^T(\beta - \beta_0)| < \delta_f$, when $|\alpha - \alpha_0| < \epsilon_f$ and $\|\beta - \beta_0\| < \epsilon_f$.

Therefore, if $f \in \mathscr{U}$, $|\alpha - \alpha_0| < \epsilon_f$ and $\|\beta - \beta_0\| < \epsilon_f$, we have

$$K(f, \alpha, \beta) < 2\delta.$$

Since $\tilde{\Pi}\{(f, \alpha, \beta) : f \in \mathscr{U}, |\alpha - \alpha_0| < \epsilon_f \text{ and } \|\beta - \beta_0\| < \epsilon_f\} > 0$, the proof is completed. □

REMARK 3.3. In the above theorem, we gave conditions under which posterior consistency holds with the prior modeled by symmetrized normal mixture. Other kernels also give similar results; see Section 6 for AFT models, where a Weibull mixture is used to construct the prior.

Remark 3.4. If the covariate is non-stochastic, then the above theorem need some modifications. First, Condition (i) need to be changed to (i′) as in Remark 3.2. Second, we need to add the requirement that $V(g_\eta, f) < \infty$ in the definition of the set appearing in Condition (ii). Finally, Condition (iii) needs to be changed to (iii′) $\int y^4 f_0(y)dy < \infty$, $\mathrm{E}_{f_0}(\log f_0)^2 < \infty$.

Remark 3.5. If the base measure π of the Dirichlet process prior D_π for the mixing distribution has full weak support and the support of ν is $\mathbb{R}^+$, then Condition (ii) holds. Condition (iv) holds if $\int t^2 d\pi(t) < \infty$.

4 Exponential Frailty Model

To study a paired-data for lifetime distribution, Cantor *et al.* (1985) proposed a joint distribution to model paired survival times where the dependence within the pair was modeled by an unobserved random variable W, the frailty, which follows an unknown distribution and varies from pair to pair. Such a model has the advantage of preserving the loss-of-memory property and the continuity of the model at the same time compared to other alternatives; see Owen *et al.* (2000) for more details.

Let $(X_1, Y_1), (X_2, Y_2), \ldots$ be paired observations. We assume that X_i and Y_i follow independent exponential distributions with parameters W_i and λW_i, where W_i's are also random variables with unknown common distribution F. We start with a prior $\tilde{\Pi}$ for F and independently put a prior μ for λ. For given $W = w$, X and Y are conditionally independent. The resulting semi-parametric model for n independently distributed observations $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ is

$$X|W = w \sim \mathrm{Exp}(w), \quad Y|W = w \sim \mathrm{Exp}(\lambda w), \quad \lambda \sim \mu, \quad W \sim F, \quad F \sim \tilde{\Pi},$$

where $\mathrm{Exp}(w)$ is the exponential distribution with parameter w, that is, it has the density function is we^{-wx}.

Let $f_0(x, y) = \int \lambda_0 w^2 e^{-w(x+\lambda_0 y)} dF_0(w)$ denote the probability density function (p.d.f.) of (X, Y) corresponding to the frailty distribution F_0 and parameter λ_0, and $f(x, y) = \int \lambda w^2 e^{-w(x+\lambda y)} dF(w)$ be the density corresponding to F and λ.

Theorem 4.1. *Suppose that* $\log(f_0(x, y))$, x *and* y *are* f_0*-integrable,* F_0 *is in the weak support of* $\tilde{\Pi}$, λ_0 *is in the support of* μ *and* $w_E := \int w^2 dF_0(w) < \infty$. *Let* $\mathscr{W} = \{(F, \lambda) : |\lambda - \lambda_0| < \epsilon, F \in \mathscr{U}\}^c$, *where* $\mathscr{U}$ *is a weak neighborhood of* F_0. *If for any* $\delta > 0$ *and* $\beta > 0$ *there exists a sequence* r_n *and a constant*

β_0 *such that* $r_n^2 < n\beta$ *and* $\tilde{\Pi}\{F : F(r_n) - F(r_n^{-1}) < 1 - \delta\} < e^{-n\beta_0}$, *then* $\Pi\{\mathscr{W}|(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)\} \to 0$ *a.s.* $P_{f_0}^{\infty}$.

To prove this theorem, we first show the existence of the exponentially consistent tests by the following lemma.

Lemma 4.1. *Under the conditions of Theorem* 4.1, *Condition* (i) *of Theorem* 2.1 *holds.*

The proof of this lemma is given in the appendix.

Proof of Theorem 4.1. The existence of the exponentially consistent tests is shown by Lemma 4.1. In the proof of Lemma 4.1, we construct the tests on the sieve $\mathscr{A}_n \times [m_n, M_n]^d = \{F : F(r_n) - F(r_n^{-1}) \geq 1 - \delta\} \times \mathbb{R}^+$. Under the condition of Theorem 4.1, Condition (iii) of Theorem 2.1 obviously holds.

To verify Condition (ii) of Theorem 2.1, we split the expression for the Kullback Leibler divergence in three parts,

$$\begin{aligned}\iint f_0(x,y)\log\frac{f_0(x,y)}{f(x,y)}dxdy &= \iint f_0(x,y)\log\frac{f_0(x,y)}{f_r(x,y)}dxdy \\ &\quad + \iint f_0(x,y)\log\frac{f_r(x,y)}{f_{r,\lambda}(x,y)}dxdy \\ &\quad + \iint f_0(x,y)\log\frac{f_{r,\lambda}(x,y)}{f(x,y)}dxdy, \quad (4.1)\end{aligned}$$

where

$$f_r(x,y) = \int \lambda_0 w^2 e^{-w(x+\lambda_0 y)} dF_r(w), \quad f_{r,\lambda}(x,y) = \int \lambda w^2 e^{-w(x+\lambda y)} dF_r(w),$$

$F_r(w) = t_r \int_0^w f_0(v)\mathbb{1}(v \in [r^{-1}, r])dv$, $t_r = [F_0(r) - F_0(r^{-1})]^{-1}$ and $\mathbb{1}(\cdot)$ is the indicator function. Now, by Lemma 10.2, 10.3 and 10.4 in the appendix, Condition (ii) of Theorem 2.1 holds. □

Remark 4.1. Dirichlet process prior with a base measure that has light tail satisfies the condition required for this theorem.

5 Generalized Linear Models with Unknown Link Functions

Traditionally, in a generalized linear model (GLM), a known function of the expectation of the observations, called the link function, is modeled to be linear in the predictors. Recently, for more flexible modeling, the link function is treated as unknown or the linearity assumption is removed. In the Bayesian context, the former is studied by Gelfand and Kuo (1991), Newton *et al.* (1996), Mallick and Gelfand (1994), Basu and Mukhopadhyay (2000) and others. In this section we give some sufficient conditions, under which posterior consistency holds in a GLM with an unknown link function.

Consider a random response Y_i measured along with a vector valued predictors X_i, where X_i's are i.i.d. with cumulative probability distribution function $Q(x)$. Assume that X is compactly supported, say, $\|X\| \leq L$. Assume that Y_i follows a distribution belonging to an exponential family as $Y_i \sim \exp[\theta_i y - b(\theta_i)]a(y)$. By the property of the exponential family, we have that $\mu_i = \mathrm{E}(Y_i) = b'(\theta_i)$ and $\mathrm{var}(Y_i) = b''(\theta_i)$. We model the relationship between the expectation of the observations and the predictors as $\mu_i = g(X_i^T\beta)$, where $\|\beta\| = 1$ and $g(\cdot)$ is differentiable and strictly increasing. Let $\mathscr{G}$ stand for the space of all possible link functions, and $\mathscr{M} = \{\mu : \mu = g(x^T\beta), \|x\| \leq L, \|\beta\| = 1\}$. Note that $\mathscr{M}$ is compact and connected, so a closed interval. Given $a(\cdot)$ and $b(\cdot)$, let T be a chosen strictly increasing differentiable function from $\mathscr{M}$ to $[0,1]$ and let $F(X_i^T\beta) := T(g(X_i^T\beta))$. Thus, modeling link $g(\cdot)$ is equivalent to modeling $F(\cdot)$, where $F(\cdot)$ is a c.d.f. on $[-L,L]$. Let $\mathscr{L}([-L,L]) = \{T \circ g : g \in \mathscr{G}\}$. In the Bayesian approach, we assign a prior $\tilde{\Pi}$ for $F(\cdot)$ and a prior π for β. Assume that $F(\cdot)$ and β are independently distributed. We define $\eta_i = X_i^T\beta$. Let $B_n = \{\beta : \|\beta\| = 1\} \cap \{\beta : \|\beta - \beta_0\| < r_n \text{ or } \|\beta + \beta_0\| < r_n\}$ where $r_n > 0$, $n \geq 1$. Let $\Pi = \tilde{\Pi} \times \pi$.

Note that the GLM with unknown link function as described above is identifiable. To see this, we show that

$$\exp[b'^{-1}(\mu_{1x})y - b(b'^{-1}(\mu_{1x}))]a(y) = \exp[b'^{-1}(\mu_{2x})y - b(b'^{-1}(\mu_{2x}))]a(y), \tag{5.1}$$

where $\mu_{ix} = g_i(x^T\beta_i)$, $i = 1,2$, for all x and y, if and only if $\beta_1 = \beta_2$ and $g_1 = g_2$. The "if" part is obvious, so we now show that the "only-if" part.

If $\beta_1 \neq \beta_2$ and $g_1 = g_2 = g$, then there exist x such that $x^T\beta_1 \neq x^T\beta_2$. For such x, if $y \neq \frac{b(b'^{-1}(g(x^T\beta_1))) - b(b'^{-1}(g(x^T\beta_2)))}{b'^{-1}(g(x^T\beta_1)) - b'^{-1}(g(x^T\beta_2))}$, then (5.1) does not hold. If $\beta_1 = \beta_2 = \beta$ but $g_1 \neq g_2$, then if x is such that $b'^{-1}(g_1(x^T\beta)) \neq b'^{-1}(g_2(x^T\beta))$ and $y \neq \frac{b(b'^{-1}(g_1(x^T\beta))) - b(b'^{-1}(g_2(x^T\beta)))}{b'^{-1}(g_1(x^T\beta)) - b'^{-1}(g_2(x^T\beta))}$, then (5.1) does not hold for such (x,y). If $\beta_1 \neq \beta_2$ and $g_1 \neq g_2$, without loss of generality, there exists η such that $g_1(\eta) > g_2(\eta)$, and hence $g_1(\eta^*) > g_2(\eta)$ for $\eta^* > \eta$. Choosing x such that $x^T\beta_1 = \eta^*$ and $x^T\beta_2 = \eta$, then for such x and $y \neq$

$\frac{b(b'^{-1}(g_1(\eta^*)))-b(b'^{-1}(g_2(\eta)))}{b'^{-1}(g_1(\eta^*))-b'^{-1}(g_2(\eta))}$, (5.1) does not hold. Note that for one dimensional x, we can let $\eta < 0$ and $\eta^* = -\eta$.

THEOREM 5.1. *Suppose that the covariate X is compactly supported on $\{x : \|x\| \le L\}$ for some constant L, and for any γ, $Q(\Gamma_\gamma \backslash \{X : |X^T \Upsilon_i| < \zeta\}) > 0$ for each $i = 1, \ldots, d$, and some $\zeta > 0$. Assume that the weak support of $\tilde{\Pi}$ contains $\mathscr{L}([-L, L])$. If $\pi(B_n^c) \le c_1 e^{-nc_2}$ for some constants $c_1, c_2 > 0$, then with $P_{f_0}^\infty$-probability 1, the posterior probability*

$$\Pi(\mathscr{W} | (X_1, Y_1), \ldots, (X_n, Y_n)) \to 0,$$

where $\mathscr{W}$ is of the form $(\mathscr{U} \times \{\beta : \|\beta - \beta_0\| < \delta\})^c$ and $\mathscr{U}$ is a weak neighborhood of F_0.

To prove this Theorem, we use the following two lemmas to construct the required exponentially consistent tests.

LEMMA 5.1. *For any $\delta > 0$, there exists a weak neighborhood $\mathscr{U}$ of F_0, such that, for $\mathscr{W} = \mathscr{U} \times \{\beta : \|\beta - \beta_0\| > \delta\}$, there exists an exponentially consistent sequence of tests for testing $\mathrm{H}_0 : (F, \beta) = (F_0, \beta_0)$ against $\mathrm{H}_1 : (F, \beta) \in \mathscr{W}$.*

LEMMA 5.2. *Suppose that the covariate X is compactly supported on $\{x : \|x\| \le L\}$ for some constant L, and for any γ, $Q(\Gamma_\gamma \backslash \{X : |X^T \Upsilon_i| < \zeta\}) > 0$ for each $i = 1, \ldots, d$, and some $\zeta > 0$. For any weak neighborhood $\mathscr{U}$ of F_0 there exists an exponentially consistent sequence of tests for testing $\mathrm{H}_0 : F = F_0, \beta = \beta_0$ against $\mathrm{H}_1 : F \notin \mathscr{U}, \beta \in B_n$.*

The proofs of these lemmas are given in the appendix.

PROOF OF THEOREM 5.1. By Lemmas 5.1 and 5.2, Condition (i) of Theorem 2.1 is satisfied, since $\mathscr{W} \cap (\mathscr{M}(\mathbb{R}) \times B_n)$ is the union of sets considered in these two lemmas, and hence the required exponentially consistent tests exist.

Condition (iii) of Theorem 2.1 is obviously satisfied under the conditions of this theorem.

We shall show that the Kullback-Leibler property described by Condition (ii) of Theorem 2.1 is satisfied to complete the proof. Write

$$I(y) = \exp[b'^{-1}(g(x^T\beta))y - b(b'^{-1}(g(x^T\beta)))]a(y),$$

and

$$I_0(y) = \exp[b'^{-1}(g_0(x^T\beta_0))y - b(b'^{-1}(g_0(x^T\beta_0)))]a(y).$$

In this model

$$\begin{aligned} K(a,\theta) &= \int I_0(y)\log\frac{I_0(y)}{I(y)}dydQ(x) \\ &= \int I_0(y)\Big[b'^{-1}(g_0(x^T\beta_0))y - b'^{-1}(g(x^T\beta))y \\ &\qquad -\Big(b(b'^{-1}(g_0(x^T\beta_0))) - b(b'^{-1}(g(x^T\beta)))\Big)\Big]dydQ(x) \\ &\le \int I_0(y)(\epsilon_{1\delta}y - \epsilon_{2\delta})dydQ(x) \le \epsilon_{1\xi}\int g_0(\eta_i)dQ(x) + \epsilon_{2\xi}, \end{aligned}$$

where

$$\epsilon_{1\xi} = \sup_{\|x\|\le L}|b'^{-1}(g_0(x^T\beta_0)) - b'^{-1}(g(x^T\beta))|$$

and

$$\epsilon_{2\xi} = \sup_{\|x\|\le L}|b(b'^{-1}(g_0(x^T\beta_0))) - b(b'^{-1}(g(x^T\beta)))|.$$

Now for any $\delta > 0$, if

$$\epsilon_{1\xi} < \frac{\delta}{2g_0(L)} \quad \text{and} \quad \epsilon_{2\xi} < \delta/2, \tag{5.2}$$

then $K(a,\theta) < \delta$ holds. We shall show that there exist weak neighborhood $\mathscr{U}$ of F_0 and neighborhood V of β_0 such that for any $F \in \mathscr{U}$ and $\beta \in V$, (5.2) holds.

Since b'^{-1} and $b\circ b'^{-1}$ are continuous on $\mathscr{M}$, they are uniformly continuous on $[g_0(-L), g_0(L)]$. Hence, there exists $\xi > 0$ such that for all the (β, F) such that $\sup_{\|x\|\le L}|F(x^T\beta) - F_0(x^T\beta_0)| < \xi$, (5.2) holds.

Since g_0 is uniformly continuous on $[L, L]$, there exists $\zeta_1 > 0$ such that for any x and β with $|x^T\beta_0 - x^T\beta| < \zeta_1$,

$$|g_0(x^T\beta_0) - g_0(x^T\beta)| < \xi/2. \tag{5.3}$$

Let $V = \{\beta : \|\beta - \beta_0\| < \zeta_1/L\}$, then for any $\beta \in V$ and $\|x\| \le L$, expression (5.3) holds.

Since T^{-1} is clearly continuous and bounded, there exists $\zeta_2 > 0$ such that for each F that satisfies $\sup_{\|x\|\le L}|g(x^T\beta) - g_0(x^T\beta)| < \zeta_2$, we have

$$\sup_{\|x\|\le L}|g(x^T\beta) - g_0(x^T\beta)| = \sup_{\|x\|\le L}|T^{-1}(F(x^T\beta)) - T^{-1}(F_0(x^T\beta))| < \xi/2 \tag{5.4}$$

for all $\|\beta\| = 1$. Let $\mathscr{U}' = \{F : \sup_x |F(x^T\beta) - F_0(x^T\beta)| < \zeta/2\}$. By Polya's Theorem, there exists a weak neighborhood $\mathscr{U}$ of F_0, contained in $\mathscr{U}'$.

Note that $|g(x^T\beta) - g_0(x^T\beta_0)| \leq |g_0(x^T\beta_0) - g_0(x^T\beta)| + |g_0(x^T\beta) - g(x^T\beta)|$. Combining (5.3) and (5.4), we have that for $F \in \mathscr{U}$ and $\beta \in V$, (5.2) holds, and hence Condition (ii) of Theorem 2.1 is satisfied. □

Remark 5.1. A Dirichlet process prior with a base measure whose support contains $[-L, L]$ satisfies the requirement of this theorem.

Remark 5.2. If the covariates x_i's are fixed constants, we need to assume conditions to ensure $\Pi\{(a(\cdot), \theta) : K_i(a, \theta) < \delta$ for all i, and $\sum_{i=1}^{\infty} \frac{V_i(a,\theta)}{i^2} < \infty\} > 0$. The K_i part is almost the same as K for the stochastic case. For the V_i part, in this model, the relation is guaranteed by $\epsilon_{1\xi}^2 < \delta/(3g_0^2(L))$, $\epsilon_{1\xi}\epsilon_{2\xi} < \delta/(6g_0(L))$, and $\epsilon_{2\xi}^2 < \delta/3$, which automatically hold for sufficiently small δ. Hence, we only need to modify the conditions on X. Under the conditions as in Remark 3.2, the theorem applies to non-stochastic cases.

6 Cox Proportional Hazard Model

For survival data with covariates, the proportional hazards model, introduced by Cox (1972), is the most popular model. Let $Y_1, Y_2, \ldots$ be survival times with covariates $X_1, X_2, \ldots$, where $X_i \in \mathbb{R}^d$ are i.i.d. with compactly supported probability measure $Q(x)$. Assume that for any γ, $Q(\Gamma_\gamma \backslash \{X : |X^T\Upsilon_i| < \zeta\}) > 0$ for each $i = 1, \ldots, d$, and some $\zeta > 0$, and the random hazard rate of Y_i is of the form

$$h(t)\exp(X_i^T\beta), \tag{6.1}$$

where $\beta \in \mathbb{R}^d$ is the unknown regression coefficient and $h(t)$ is the unknown baseline hazard function. We give priors μ for β and $\tilde{\Pi}$ for $h(\cdot)$. Assume that $h(\cdot)$ and β are a prior independent and define $\Pi := \tilde{\Pi} \times \mu$. Let $h_0(\cdot)$ and β_0 denote the true value of the unknown parameters.

The cumulative baseline hazard is given by $H(t) = \int_0^t h(s)ds$, and provided that $H(t) \to \infty$ as $t \to \infty$ a.s., we can define a baseline random density function for life time Y by

$$f(t) = h(t)\exp(-H(t)) = h(t)\bar{F}(t), \tag{6.2}$$

where $\bar{F}(t) = 1 - F(t) := \exp(-H(t))$ is the survival function, and $F(t)$ is the baseline c.d.f. of life time Y. Note that the p.d.f. of Y_i with the corresponding covariate X_i is given by

$$e^{X_i^T\beta} f(t)[\bar{F}(t)]^{\exp(X_i^T\beta)-1}. \tag{6.3}$$

Let f_0, F_0 and $\bar{F}_0$ denote the probability density, cumulative distribution and survival functions corresponding to h_0 respectively. Throughout this section we assume that

$$\mathrm{E}[H(t)] < \infty \qquad \text{for any } t > 0. \tag{6.4}$$

In the presence of right censoring, the life time Y cannot be observed every time. We only observe (Z, Δ), where $Z = Y \wedge C$, $\Delta = \mathbb{1}(Y \leq C)$ for C a censoring time with distribution F_c and for simplicity, assume that it has density f_c with respect to Lebesgue measure. Also, let $\bar{F}_c(t) = 1 - F_c(t)$. Clearly, this leads us to consider a prior of the space $\mathscr{F} \times \mathscr{F}$ and the corresponding prior Π^* induced on the space of the distribution of the observations (Z_i, Δ_i)'s. Assume that the support of F_c is given by $\mathrm{supp}(F_c) = \mathbb{R}^+$.

THEOREM 6.1. *Let the p.d.f. of life time Y_i be defined by* (6.1), (6.2) *and* (6.3). *Suppose the distribution of censoring times F_c is absolutely continuous and* $\mathrm{supp}(F_c) = \mathbb{R}^+$. *Moreover, assume that for any $q > 1$, the following conditions hold:*

(i) $f_0(t)$ *is strictly positive on* $(0, \infty)$ *and* $\int_{\mathbb{R}^+} \max\{\mathrm{E}[H(t)^q], t^q\} f_0(t)dt < \infty$;

(ii) *there exists $r > 0$ such that* $\liminf_{t \downarrow 0} h(t)/t^r = \infty$, *a.s.;*

(iii) $\Pi\{h : \sup_{0<t<T} |h(t) - h_0(t)| < \delta, \int_T^\infty |H(t) - H_0(t)| f_0(t)dt < \delta\} > 0$ *for some finite T and positive δ;*

(iv) $\int_0^\infty (\log f_0(t))^q f_0(t)dt < \infty$.

Then, for $\mathscr{W} = \{(h, \beta) : |\beta - \beta_0| < \epsilon, h : 1 - \exp(-H(t)) \in \mathscr{U}_{F_0}\}^c$, *where* $\mathscr{U}_{F_0}$ *is a weak neighborhood of* F_0, *we have that*

$$\Pi(\mathscr{W} | (X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)) \to 0 \;\; a.s. \;\; [P^\infty_{\beta_0, h_0}].$$

To prove this theorem, we first use the following two lemmas to show the existence of the exponentially consistent test.

LEMMA 6.1. *Under the conditions of Theorem* 6.1, *there is an exponentially consistent sequence of tests for*

$\mathrm{H}_0 : (h, \beta) = (h_0, \beta_0)$,
$\mathrm{H}_1 : 1 - e^{-H(t)} \notin \mathscr{U}_{F_0(t)}$, *where* $\mathscr{U}_{F_0}$ *is a weak neighborhood of* $F_0(\cdot)$.

LEMMA 6.2. *Under the conditions of Theorem 6.1, there is an exponentially consistent sequence of tests for*

$$
\begin{aligned}
&H_0 : (h, \beta) = (h_0, \beta_0),\\
&H_1 : \sup_{t\in\mathbb{R}} |\exp(-H(t)) - \bar{F}_0(t)| < \Delta,\ \|\tilde{\beta} - \tilde{\beta}_0\| > \sqrt{d}\ \delta.
\end{aligned}
$$

The proofs of this two lemmas are given in the appendix.

PROOF OF THEOREM 6.1. First, we see that Condition (i) of Theorem 2.1 is satisfied by Lemmas 6.1 and 6.2, since $\mathscr{W}$ is the union of sets considered in these two lemmas.

Now, showing that the Kullback-Leibler property described by Condition (ii) of Theorem 2.1 holds will complete the proof.

For this model, $K(h, \beta)$ equals to

$$
\begin{aligned}
&\iint_0^\infty \exp(x^T\beta_0)[\bar{F}_0(t)]^{\exp(x^T\beta_0)-1} f_0(t)\bar{F}_c(t)\\
&\qquad\qquad \times \log \frac{e^{x^T\beta_0}[\bar{F}_0(t)]^{\exp(x^T\beta_0)-1} f_0(t)}{e^{x^T\beta}[\bar{F}(t)]^{\exp(x^T\beta)-1} f(t)} dtdQ(x)\\
&\quad + \iint_0^\infty \log \frac{[\bar{F}_0(t)]^{\exp(x^T\beta_0)}}{[\bar{F}(t)]^{\exp(x^T\beta)}} f_c(t)[\bar{F}_0(t)]^{\exp(x^T\beta_0)} dtdQ(x). \qquad (6.5)
\end{aligned}
$$

Define the set

$$
\mathscr{V}(\delta, T) = \left\{ h : \sup_{t\le T} |h(t) - h_0(t)| < \delta \wedge \int_T^\infty |H - H_0| f_0 < \delta \right\} \qquad (6.6)
$$

We have $\Pi(\mathscr{V}) > 0$ under the Condition (iii).

Now, we are going to show that for any $\delta > 0$, there exists T' such that, for any $T > T'$, $\Pi(\mathscr{V}) > 0$ is a sufficient condition for

$$
\begin{aligned}
\Pi\Big\{ &\iint_0^\infty e^{x^T\beta_0}[\bar{F}_0(t)]^{\exp(x^T\beta_0)-1} f_0(t)\bar{F}_c(t)\\
&\qquad\qquad \times \log \frac{e^{x^T\beta_0}[\bar{F}_0(t)]^{\exp(x^T\beta_0)-1} f_0(t)}{e^{x^T\beta}[\bar{F}(t)]^{\exp(x^T\beta)-1} f(t)} dtdQ(x)\\
&\quad + \iint_0^\infty \log \frac{[\bar{F}_0(t)]^{\exp(x^T\beta_0)}}{[\bar{F}(t)]^{\exp(x^T\beta)}} f_c(t)[\bar{F}_0(t)]^{\exp(x^T\beta_0)} dtdQ(x) < \epsilon \Big\} > 0,
\end{aligned}
$$

for any $\epsilon > 0$. It is sufficient to show that for any $h \in \mathscr{V}(\delta, T)$, the following two inequalities hold,

$$\iint_T^\infty e^{x^T\beta_0}[\bar{F}_0(t)]^{\exp(x^T\beta_0)-1} f_0(t)\bar{F}_c(t) \times \log \frac{e^{x^T\beta_0}[\bar{F}_0(t)]^{\exp(x^T\beta_0)-1} f_0(t)}{e^{x^T\beta}[\bar{F}(t)]^{\exp(x^T\beta)-1} f(t)} dtdQ(x) + \iint_T^\infty \log \frac{[\bar{F}_0(t)]^{\exp(x^T\beta_0)}}{[\bar{F}(t)]^{\exp(x^T\beta)}} f_c(t)[\bar{F}_0(t)]^{\exp(x^T\beta_0)} dtdQ(x) < \epsilon/2 \quad (6.7)$$

and

$$\iint_0^T e^{x^T\beta_0}[\bar{F}_0(t)]^{\exp(x^T\beta_0)-1} f_0(t)\bar{F}_c(t) \times \log \frac{e^{x^T\beta_0}[\bar{F}_0(t)]^{\exp(x^T\beta_0)-1} f_0(t)}{e^{x^T\beta}[\bar{F}(t)]^{\exp(x^T\beta)-1} f(t)} dtdQ(x) + \iint_0^T \log \frac{[\bar{F}_0(t)]^{\exp(x^T\beta_0)}}{[\bar{F}(t)]^{\exp(x^T\beta)}} f_c(t)[\bar{F}_0(t)]^{\exp(x^T\beta_0)} dtdQ(x) < \epsilon/2. \quad (6.8)$$

The left hand side (l.h.s.) of (6.7) is equal to

$$\begin{aligned} &\iint_T^\infty e^{x^T\beta_0}[\bar{F}_0(t)]^{\exp(x^T\beta_0)-1} f_0(t)\bar{F}_c(t)(x^T(\beta_0 - \beta))dtdQ(x) \\ &+ \iint_T^\infty e^{x^T\beta_0}[\bar{F}_0(t)]^{\exp(x^T\beta_0)-1} f_0(t)\bar{F}_c(t)\log[h_0(t)/h(t)]dtdQ(x) \\ &- \iint_T^\infty e^{x^T\beta_0}[\bar{F}_0(t)]^{\exp(x^T\beta_0)-1} f_0(t)\bar{F}_c(t)e^{x^T\beta_0}H_0(t))dtdQ(x) \\ &+ \iint_T^\infty e^{x^T\beta_0}[\bar{F}_0(t)]^{\exp(x^T\beta_0)-1} f_0(t)\bar{F}_c(t)e^{x^T\beta}H(t))dtdQ(x) \\ &+ \iint_T^\infty f_c(t)[\bar{F}_0(t)]^{\exp(x^T\beta_0)-1}[e^{x^T\beta}H(t) - e^{x^T\beta_0}H_0(t)]dtdQ(x). \quad (6.9) \end{aligned}$$

The sum of the last three integrals in (6.9) are equal to

$$\int_T^\infty \{e^{x^T\beta_0}[\bar{F}_0(t)]^{\exp(x^T\beta_0)-1} f_0(t)\bar{F}_c(t) + f_c(t)[\bar{F}_0(t)]^{\exp(x^T\beta_0)-1}\} \times [e^{x^T\beta}H(t) - e^{x^T\beta_0}H_0(t)]dt. \quad (6.10)$$

Let x be fixed. It is easy to see that the first term in (6.9) is less than $\epsilon/6$ for β close enough to β_0 enough, since Q is compactly supported, and all such

β's form a set of positive probability under the assumption on the prior for β.

For the second term in (6.9), note that by Hölder's inequality, one has

$$\begin{aligned}&\int_T^\infty e^{x^T\beta_0}[\bar{F}_0(t)]^{\exp(x^T\beta_0)-1}f_0(t)\bar{F}_c(t)\log[h_0(t)/h(t)]dt\\ &\leq \left\{\int_T^\infty \left|e^{x^T\beta_0}[\bar{F}_0(t)]^{\exp(x^T\beta_0)-1}\right|^p f_0(t)dt\right\}^{1/p}\\ &\qquad\times\left\{\int_T^\infty\left|\log\frac{h_0(t)}{h(t)}\right|^q f_0(t)dt\right\}^{1/q}. \qquad (6.11)\end{aligned}$$

Since X is compactly supported, assume $\Pr_Q(\|X\|<L)=1$ and $\|\beta_0\|<L_2$. The first term on the right hand side (r.h.s.) of (6.11) is bounded when $p<(1-e^{-LL_2})^{-1}$, which corresponding to that $q>e^{LL_2}$. We choose q to be a sufficiently large even integer for the rest of the paper. It is sufficient to show that (6.11) is less than $\epsilon/6$ by showing that for any $\xi>0$ there exists T, such that $\int_T^\infty(\log[h_0(t)])^q f_0(t)dt<\xi$ and $\lim_{T\to\infty}\Pi\{h:\int_T^\infty(\log h(t))^q f_0(t)dt<\xi\}=1$. The latter is satisfied by Condition (iii). Note that $\log[h_0(t)]=H_0(t)+\log[f_0(t)]$, so the former is satisfied under the Conditions (i) and (iv).

Note that the expression in (6.10) is $\int \mathrm{E}[e^{x^T\beta}H(Z)-e^{x^T\beta_0}H_0(Z)]dQ(x)$. As Y is stochastically larger than $Z=Y\wedge C$, and as argued in Blasi *et al.* (2009),

$$\mathrm{E}|e^{x^T\beta}H(Z)-e^{x^T\beta_0}H_0(Z)|\leq \mathrm{E}|e^{x^T\beta}H(Y)-e^{x^T\beta_0}H_0(Y)|. \qquad (6.12)$$

Under the situation that $\|\beta-\beta_0\|$ small, the r.h.s. of (6.12) is less than $\epsilon/6$, for any $h\in\mathscr{V}(\delta,T)$ with sufficient small δ and sufficient large T.

It remains to show that (6.8) holds. We now consider the expressions

$$\iint_0^T e^{x^T\beta_0}[\bar{F}_0(t)]^{\exp(x^T\beta_0)-1}f_0(t)\bar{F}_c(t)(x^T(\beta_0-\beta))dtdQ(x), \qquad (6.13)$$

$$\iint_0^T e^{x^T\beta_0}[\bar{F}_0(t)]^{\exp(x^T\beta_0)-1}f_0(t)\bar{F}_c(t)\log[h_0(t)/h(t)]dtdQ(x) \qquad (6.14)$$

and

$$\begin{aligned}&\iint_0^T\{e^{x^T\beta_0}[\bar{F}_0(t)]^{\exp(x^T\beta_0)-1}f_0(t)\bar{F}_c(t)+f_c(t)[\bar{F}_0(t)]^{\exp(x^T\beta_0)-1}\}\\ &\qquad\times[e^{x^T\beta}H(t)-e^{x^T\beta_0}H_0(t)]dtdQ(x). \qquad (6.15)\end{aligned}$$

Obviously, the expression in (6.13) will be small enough by choosing an appropriate β.

Assume first that $h_0(0) > 0$, and let $c := \inf_{t \le T} h_0(t) > 0$. For $\delta < c$ and $h \in \mathscr{V}(\delta, T)$, the expression in (6.14) is less than

$$\iint_0^T \left| \frac{h_0(t)}{h(t)} - 1 \right| e^{x^T\beta_0} [\bar{F}_0(t)]^{\exp(x^T\beta_0)-1} f_0(t) dt dQ(x)$$
$$\le \frac{\delta}{c-\delta} \iint_0^T e^{x^T\beta_0} [\bar{F}_0(t)]^{\exp(x^T\beta_0)-1} f_0(t) dt dQ(x) = \frac{\delta}{c-\delta},$$

and the expression in (6.15) is less than

$$\iint_0^T [\sup_{s\le T} |h(s) - h_0(s)|] \cdot t \cdot [e^{x^T\beta} H(t) - e^{x^T\beta_0} H_0(t)] dt dQ(x) + \xi$$
$$\le \delta \iint_0^\infty t[e^{x^T\beta} H(t) - e^{x^T\beta_0} H_0(t)] dt dQ(x) + \xi \le \delta E_0 + \xi,$$

where ξ is a small positive number and is dependent on β, and

$$E_0 := \iint_0^\infty t e^{x^T\beta_0} [\bar{F}_0(t)]^{\exp(x^T\beta_0)-1} f_0(t) dt dQ(x),$$

which is finite by the Condition (i).

Now consider the case that $h(0) = 0$. We need a different bound for (6.14) in this case. Split the expression in (6.14) into two parts:

$$I_1 + I_2 := \iint_0^\zeta e^{x^T\beta_0} [\bar{F}_0(t)]^{\exp(x^T\beta_0)-1} f_0(t) \bar{F}_c(t) \log[h_0(t)/h(t)] dt dQ(x)$$
$$+ \iint_\zeta^T e^{x^T\beta_0} [\bar{F}_0(t)]^{\exp(x^T\beta_0)-1} f_0(t) \bar{F}_c(t) \log[h_0(t)/h(t)] dt dQ(x).$$

The analysis is essentially the same as before. We can find that for any $\epsilon > 0$, there are δ and T such that for any $h \in \mathscr{V}(\delta, T)$ implies the sum of I_2 and the expressions in (6.15) and (6.13) is less than $\epsilon/4$ for any ζ.

Now consider I_1, we shall show that for the same ϵ, there exists a small enough $\zeta > 0$, such that

$$\iint_0^\zeta e^{x^T\beta_0} [\bar{F}_0(t)]^{\exp(x^T\beta_0)-1} f_0(t) \bar{F}_c(t) \log[h_0(t)/h(t)] dt dQ(x) < \epsilon/4, \text{ a.s. } [\Pi].$$

If $h(0) \neq 0$, the above inequality holds. Hence, it is enough to show that $(\log[h_0(t)/h(t)])^q f_0(t)$ is Lebesgue integrable in 0 for Π-almost all h such that $h(0) = 0$. Note that $h_0(t) \to f_0(0)$ as $t \to 0$. Hence, under Condition (iv), we only need to show that $(\log h(t))^q$ is f_0-integrable, which is true under Condition (ii). □

REMARK 6.1. Blasi *et al.* (2009) showed that Condition (iii) holds, when the hazard function is of the form $h(t) = \int k(t, x)\tilde{\mu}(dx)$, where $\tilde{\mu}$ is a completely random measure on some Polish space, and $k(\cdot)$ is the Dykstra–Laud kernel, the rectangular kernel, the Ornstein–Uhlenbeck kernel or the exponential kernel.

REMARK 6.2. If the covariates X_i's are fixed constants, we need a sufficient condition about V_i as mentioned in Remark 5.2. For this model, we split the V_i as

$$\begin{aligned}
&\int e^{x_i^T\beta_0}[\bar{F}_0(t)]^{\exp(x_i^T\beta_0)-1} f_0(t)\bar{F}_c(t)(x_i^T(\beta_0-\beta))^2 dt \\
&\quad + \int e^{x_i^T\beta_0}[\bar{F}_0(t)]^{\exp(x_i^T\beta_0)-1} f_0(t)\bar{F}_c(t)(\log[h_0(t)/h(t)])^2 dt \\
&\quad + \int e^{x_i^T\beta_0}[\bar{F}_0(t)]^{\exp(x_i^T\beta_0)-1} f_0(t)\bar{F}_c(t)[e^{x_i^T\beta}H(t) - e^{x_i^T\beta_0}H_0(t)]^2 dt \\
&\quad + \int f_c(t)[\bar{F}_0(t)]^{\exp(x_i^T\beta_0)-1}[e^{x_i^T\beta}H(t) - e^{x_i^T\beta_0}H_0(t)]^2 dt.
\end{aligned}$$

The first term will be made arbitrarily small by choosing β close enough to β_0 enough. The second term can be made small by Condition (i), for $q > e^{LL_2/2}$, which also implies a bound for the "K" part. For the last two terms, by similar argument as above, the additional requirement on "V_i" part can be satisfied under the following condition:
(v) $\Pi\{h : \sup_{0<t<T}(h(t) - h_0(t))^2 < \delta, \int_T^\infty (H(t) - H_0(t))^2 f_0(t)dt < \delta\} > 0$ for some finite T and positive δ.

7 Accelerated failure time model

To analyze survival data with covariates, the accelerated failure time models are quite popular, too. For each subject $i = 1, \ldots, n$, let Y_i denote the failure time and C_i denote the censoring time. The observed survival data are $Z_i = \min(Y_i, C_i)$ and $\Delta_i = \mathbb{1}(Y_i \leq C_i)$. Let X_i denote a d-dimensional vector of covariates associated with subject i.

We assume that $Y/e^{X^T\beta} \sim f(y)$, where β and f are unknown parameters. Let β_0 and f_0 denote the true value of β and f respectively. Observe that

$\epsilon := \log Y - X^T\beta \sim \tilde{f}$, where $\tilde{f}(t) := e^t f(e^t)$. Note that the correspondence $f \mapsto \tilde{f}$ is one to one. Let F_c be the distribution of C, and for simplicity, assume that it has density f_c with respect to Lebesgue measure, and let $\tilde{f}_c$ and $\tilde{F}_c$ be the density and distribution of $\log C$ respectively. Let $\tilde{f}_0$ be the corresponding true value for $\tilde{f}$. Also, we assume that the covariate X is random and with probability measure $Q(x)$. To make this model identifiable, we further assume that $\int x dQ(x) = 0$.

Assume that the Dirichlet mixture prior assigned to f using the Weibull density function as the kernel. The Weibull kernel is given by

$$w(x; \theta, \phi) = \phi^{-1}\theta(x/\phi)^{\theta-1}\exp[-(x/\phi)^{\theta}].$$

Let $P \sim \tilde{\Pi}$, and given P, let $\log\phi_i$ be i.i.d. P. Let the model consist a prior ν for θ. Then, $\nu \times \tilde{\Pi}$ induces a prior on the survival function $\bar{\tilde{F}}(t)$ of the random variable $\log Y - X^T\beta$ via the map

$$(P, \theta) \mapsto \int \exp(-e^{(t-\log\phi)\theta})dP,$$

and induces another prior on the density function $\tilde{f}$ of the random variable $\log Y - X^T\beta$ via the map

$$(P, \theta) \mapsto \int \theta\exp(-e^{(t-\log\phi)\theta} + (t - \log\phi)\theta)dP.$$

Assign a prior μ for β, and let Π stand for $\tilde{\Pi} \times \nu \times \mu$.

THEOREM 7.1. *Assign a Dirichlet mixture prior with Weibull density kernel to f as described above. Suppose that*

(i) *the covariate X is compactly supported, and for any γ, $Q(\Gamma_\gamma \backslash \{X : |X^T\Upsilon_i| < \zeta\}) > 0$ for each $i = 1, \ldots, d$, and some $\zeta > 0$;*

(ii) *for some $0 < M < \infty$, $0 < \tilde{f}_0(t) < M$ for all t;*

(iii) *$\int \tilde{f}_0(t)|\log\tilde{f}_0(t)|dt < \infty$ and $\int \tilde{f}_c(t+\xi)|\log\bar{\tilde{F}}_0(t)|dt < \infty$ for all $\xi \in \mathbb{R}$;*

(iv) *for some $\delta > 0$, $\int \tilde{f}_0(t)\log\frac{\tilde{f}_0(t)}{\phi_\delta(t)}dt < \infty$, where $\phi_\delta(t) := \inf_{\|s-t\|<\delta}\tilde{f}_0(s)$;*

(v) *there exists $\eta > 0$ such that $|\int e^{2(\log t)|^\eta} f_0(t)dt| < \infty$, and $\int e^{(\log t - a)/b} f_0(t)dt < \infty$ for any $a \in \mathbb{R}$ and $b \in (0, \infty)$;*

(vi) *$\int(e^{-t/h} + t/h)dP(t) < \infty$ for any given $h > 0$, a.s. $\tilde{\Pi}$;*

(vii) *the weak support of* $\tilde{\Pi}$ *is the space of all probability measures on* $\mathbb{R}$*;*

(viii) *for some* $\delta > 0$ *and any* $\xi \in \mathbb{R}$, $\int \tilde{f}_c(t+\xi) \log \frac{\bar{\tilde{F}}_{0,\beta_0}(t)}{\phi_\delta(t)} dt < \infty$*, where* $\phi_\delta(t) := \inf_{\|s-t\|<\delta} \tilde{f}_0(s)$*;*

(ix) *there exists* $\eta > 0$ *such that* $|\int f_c(t) e^{2(\log t)^{1+\eta}} dt| < \infty$ *and* $\int f_c(t) e^{(\log t - a)/b} dt < \infty$ *for all* $a \in \mathbb{R}$ *and* $b \in (0,\infty)$*;*

Then for any weak neighborhood $\mathscr{U}$ *of* F_0*,*

$$\Pi\left(\left\{(F,\beta) : \tilde{f} \in \mathscr{U}, \|\beta-\beta_0\| < \epsilon\right\} \Big| (Z_1, X_1, \Delta_1), \ldots, (Z_n, X_n, \Delta_n)\right) \to 1$$

a.s. $\left[P_{f_0}^\infty\right]$.

To prove this theorem, we first use the following two lemmas to show the existence of the exponentially consistent test.

Lemma 7.1. *Under the conditions of Theorem* 7.1*, there is an exponentially consistent sequence of tests for testing* $\mathrm{H}_0 : (F,\beta) = (F_0,\beta_0)$ *against* $\mathrm{H}_1 : F \notin \mathscr{U}_{F_0}$*, where* $\mathscr{U}_{F_0}$ *is a weak neighborhood of* $F_0(\cdot)$*.*

Lemma 7.2. *Under the conditions of Theorem* 7.1*, there is an exponentially consistent sequence of tests for testing* $\mathrm{H}_0 : (F,\beta) = (F_0,\beta_0)$ *against*

$$\mathrm{H}_1 : \sup_{t \in \mathbb{R}} |F(t) - F_0(t)| < \Delta, \ \|\beta - \beta_0\| > \delta.$$

The proofs of these two lemmas are given in the appendix.

Proof of Theorem 7.1. Since $\{(\tilde{f},\beta) : \tilde{f} \in \mathscr{U}, \|\beta-\beta_0\| < \epsilon\}^c$ is the union of sets considered in the Lemmas 7.1 and 7.2, Condition (i) of Theorem 2.1 is satisfied.

Note that there is no sieve used in the tests we constructed in the above lemmas. We only need to verify Condition (ii) of Theorem 2.1 to complete the proof.

For this model, the Kullback-Leibler support condition on Π at (f_0,β_0) is

$$\Pi\left\{\iint e^{-x^T\beta_0} f_0(e^{-x^T\beta_0}y) \bar{F}_c(y) \log \frac{e^{-x^T\beta_0} f_0(e^{-x^T\beta_0}y)}{e^{-x^T\beta} f(e^{-x^T\beta}y)} dy dQ(x) \right.$$
$$\left. + \iint \bar{F}_0(e^{-x^T\beta_0}y) f_c(y) \log \frac{\bar{F}_0(e^{-x^T\beta_0}y)}{\bar{F}(e^{-x^T\beta}y)} dy dQ(x) < \epsilon \right\} > 0 \tag{7.1}$$

for any $\epsilon > 0$. By transformation, the expression (7.1) is equivalent to

$$\Pi\Bigg\{\iint \tilde{f}_0(t - x^T\beta_0)\bar{\bar{\tilde{F}}}_c(t)\log\frac{\tilde{f}_0(t - x^T\beta_0)}{\tilde{f}(t - x^T\beta)}dtdQ(x) + \iint \bar{\bar{\tilde{F}}}_0(t - x^T\beta_0)\tilde{f}_c(t)\log\frac{\bar{\bar{\tilde{F}}}_0(t - x^T\beta_0)}{\bar{\bar{\tilde{F}}}(t - x^T\beta)}dtdQ(x) < \epsilon\Bigg\} > 0 \tag{7.2}$$

for any $\epsilon > 0$.

By Condition (iii), (v), (vi) and (ix), applying Lemma 10.6 in the appendix, we have

$$\int \tilde{f}_0(t)\log\frac{\tilde{f}_0(t)}{\tilde{f}(t-\gamma)}dt \to \int \tilde{f}_0(t)\log\frac{\tilde{f}_0(t)}{\tilde{f}(t)}dt \tag{7.3}$$

as $\gamma \to 0$.

By Condition (iii), (iv) and (v), an application of Theorem 3 of Wu and Ghosal (2008), gives us

$$(\nu \times \tilde{\Pi})\left\{\int \tilde{f}_0(t)\log\frac{\tilde{f}_0(t)}{\tilde{f}(t)} < \epsilon\right\} > 0 \tag{7.4}$$

for every $\epsilon > 0$. Combining (7.3) and (7.4), we have that

$$\Pi\left\{\iint \tilde{f}_0(t - x^T\beta_0)\bar{\bar{\tilde{F}}}_c(t)\log\frac{\tilde{f}_0(t - x^T\beta_0)}{\tilde{f}(t - x^T\beta)}dtdQ(x) < \epsilon\right\} > 0 \tag{7.5}$$

for every $\epsilon > 0$. Note that Conditions (iii), (iv) and (v) correspond to Conditions B5-B7 of Theorem 3 of Wu and Ghosal (2008), while other conditions of that theorem obviously satisfied.

Now, we show that

$$\Pi\left\{\iint \bar{\bar{\tilde{F}}}_0(t - x^T\beta_0)\tilde{f}_c(t)\log\frac{\bar{\bar{\tilde{F}}}_0(t - x^T\beta_0)}{\bar{\bar{\tilde{F}}}(t - x^T\beta)}dtdQ(x) < \epsilon\right\} > 0 \tag{7.6}$$

To this end, by Lemma 10.7, we have

$$(\nu \times \tilde{\Pi})\left\{\int \tilde{f}_c(t+\xi)\log\frac{\bar{\bar{\tilde{F}}}_0(t)}{\bar{\bar{\tilde{F}}}(t)}dt < \epsilon\right\} > 0, \tag{7.7}$$

holds for any $\xi \in \mathbb{R}$.

By Condition (iii), (v), (vi) and (ix), applying Lemma 10.6, we have that

$$\int \tilde{f}_c(t) \log \frac{\bar{\tilde{F}}_0(t)}{\bar{\tilde{F}}(t-\gamma)} dt \to \int \tilde{f}_c(t) \log \frac{\bar{\tilde{F}}_0(t)}{\bar{\tilde{F}}(t)} dt \tag{7.8}$$

as $\gamma \to 0$. By (7.7) and (7.8), expression (7.6) holds true.

Note that the intersection of the sets in the expressions (7.4) and (7.7) containing a common weak neighborhood of P_m, and hence (7.2) holds. □

8 Partial linear regression model

Consider a partial linear model

$$Y_i = X_i^T \beta + f(Z_i) + \epsilon_i, \tag{8.1}$$

where Y_i is the dependent variable, X_i is a vector of d-dimensional explanatory variable, Z_i is a scalar explanatory variable, ϵ_i's are i.i.d. $\mathrm{Norm}(0, \sigma^2)$ for $i = 1, \ldots, n$, and f is an odd function. Assume that X and Z are random variables having probability distribution $Q_X(x)$ and p.d.f. $q_Z(z)$ repectively. Let $q_{f,Z}$ denote the p.d.f. of $f(Z)$ and the convolution $q_{f,Z} * \phi(0, \sigma^2)$ is the p.d.f. of $f(Z) + \epsilon$. We assign a prior μ on β and a prior $\tilde{\Pi}$ on $q_{f,Z} * \phi(0, \sigma^2)$. Let Π stand for $\tilde{\Pi} \times \mu$. We have the following result:

THEOREM 8.1. *Suppose that*

(i) *the distribution of covariates X and Z are compactly supported, and for any γ, $Q(\Gamma_\gamma \backslash \{X : |X^T \Upsilon_i| < \zeta\}) > 0$ for each $i = 1, \ldots, d$, and some $\zeta > 0$;*

(ii) *there exists a sufficiently small $\eta > 0$, C_η and a density g_η, such that for $|\eta'| < \eta$, $q_{f_0,Z} * \phi(0, \sigma_0^2)(y - \eta') < C_\eta g_\eta(y)$;*

(iii) *for all sufficiently small η and for all $\delta > 0$, $\tilde{\Pi}\{K(g_\eta, f) < \delta\} > 0$.*

Then for any weak neighborhood $\mathscr{U}$ of f_0,

$$\Pi\{(f, \beta) : f \in \mathscr{U}, \|\beta - \beta_0\| < \delta | (X_1, Y_1, Z_1), \ldots, (X_n, Y_n, Z_n)\} \to 1$$

a.s. $\left[P_{f_0}^\infty\right]$.

Proof. Note that by considering $f(Z) + \epsilon$ as the error part in multiple regression model, we can treat this model as a multiple regression model. Note that in Theorem 3.1, Condition (ii) is about the p.d.f. of the error term in the linear regression model, while in this theorem the same condition is given to $q_{f_0,Z} * \phi(0, \sigma_0^2)(y - \eta')$, which is the probability density function of the error term in the partial linear regression model. Also, the other two conditions are the same as the rest of the other two conditions in Theorem 3.1. Applying the same argument proves this theorem. □

Remark 8.1. Due to the similarity to the multiple regression model discussed in previous section, the prior applies for that model will apply here.

Remark 8.2. If the covariates are fixed constants x_i's and z_i's, Condition (i) and (iii) will need to be changed accordingly as mentioned in Remark 3.2 to ensure consistency.

9 Discussion

In this paper, we studied posterior consistency for some semi-parametric models, which was initiated by Amewou-Atisso *et al.* (2003). In that paper, the authors studied the univariate regression model, with Polya tree or Dirichlet mixture of normals priors being given, and the binary response regression with unknown link. Both stochastic and non-stochastic covariate cases were studied. The basic idea is to consider a sufficiently weak topology on the non-parametric part so that uniformly exponentially consistent tests can be constructed for testing the true value against the complement of a neighborhood. In this paper, we used the same idea to study some more models, which are commonly used in practice.

By doing this, we provide a ready catalog of conditions required to ensure sensible Bayesian inference in large samples for most commonly encountered semi-parametric models. This may help practitioners to select useful prior distributions. The technical difficulty lies in the proof of Kullback-Leibler property or constructing the exponentially consistent tests. For example, for proportional hazard model, it is more difficult to prove the Kullback-Leibler property of the model, while for exponential frailty model, it is harder to construct the tests.

Of course, there are many other Bayesian semi-parametric models used in practice. For example, the proportional hazard model with frailty structure and with censoring data is a commonly encountered model in survival analysis. Other examples include biased sampling model and projection pursuit model. In this paper, we have initiated the process of cataloging, which, we hope will get richer in the future.

10 Appendix

The following lemma, due to Amewou-Atisso *et al.* (2003) Lemma 3.1, will be used many times in our proofs of the existence of the exponentially consistent tests for the models.

LEMMA 10.1. *For $i = 1, 2, \ldots,$ let g_{0i} and g_i be densities on $\mathbb{R}$. If for each i there exists a function Φ_i, $0 \le \Phi_i \le 1$, such that*

$$\mathrm{E}_{g_{0i}}(\Phi_i) = \alpha_i \le \gamma_i = \mathrm{E}_{g_i}(\phi_i),$$

and if

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (\gamma_i - \alpha_i) > 0$$

then there exist a constant C, sets $B_n \subset \mathbb{R}^n$, $n = 1, 2, \ldots,$ and n_0, all depending only on (γ_i, α_i), such that for $n > n_0$,

$$\left[\prod_{i=1}^{n} P_{g_{0i}}\right](B_n) < e^{-nC},$$

$$\left[\prod_{i=1}^{n} P_{g_i}\right](B_n) > 1 - e^{-nC}.$$

PROOF OF LEMMA 3.1. Let $\mathscr{U}_{f_0}$ be a weak neighborhood of f_0, $\Delta > 0$, $\gamma = \{\gamma_1, \ldots, \gamma_d\} \in \{-1, 1\}^d$ and $\iota = \pm 1$. Consider the following three groups of hypotheses,

$$\mathrm{H}_0 : (f, \alpha, \beta) = (f_0, \alpha_0, \beta_0) \text{ against } \mathrm{H}_1 : (f, \alpha, \beta) \in \mathscr{W}_1,$$

$$\mathrm{H}_0 : (f, \alpha, \beta) = (f_0, \alpha_0, \beta_0) \text{ against } \mathrm{H}_1 : (f, \alpha, \beta) \in \mathscr{W}_{2,\iota,\gamma,j},$$

for each ι, γ and $j = 1, \ldots, d$, and

$$\mathrm{H}_0 : (f, \alpha, \beta) = (f_0, \alpha_0, \beta_0) \text{ against } \mathrm{H}_1 : (f, \alpha, \beta) \in \mathscr{W}_{3,\iota,\gamma,j},$$

for all ι and γ, where $\mathscr{W}_1 = \mathscr{U}_{f_0}^c \times \{|\alpha - \alpha_0| < \Delta, \|\beta - \beta_0\| < \Delta\}$, $\mathscr{W}_{2,\iota,\gamma,j} = \mathscr{F} \times \{(\alpha, \beta) : \iota(\alpha - \alpha_0) > 0, (\beta - \beta_0) \in \Gamma_\gamma, \gamma_j(\beta_j - \beta_{0j}) > \Delta\}$, $\beta = (\beta_1, \ldots, \beta_d)$ and $\mathscr{W}_{3,\iota,\gamma} = \mathscr{F} \times \{\iota(\alpha - \alpha_0) > \Delta, (\beta - \beta_0) \in \Gamma_\gamma\}$.

For the first group of hypotheses, we take

$$\mathscr{U} = \left\{ f : \int \Phi(y) f(y) - \int \Phi(y) f_0(y) < 0 \right\},$$

where $0 \leq \Phi \leq 1$ is uniformly continuous. Since Φ is uniformly continuous, given $\epsilon > 0$, there exists $\delta > 0$ such that $|y_1 - y_2| < \delta$ implies $|\Phi(y_1) - \Phi(y_2)| < \epsilon/2$. Let Δ be such that $|(\alpha - \alpha_0) + (\beta - \beta_0)X_i| < \delta$ for all $\|X_i\| < L$ and $\alpha, \beta \in \mathscr{W}_1$. Now for any $f \in \mathscr{U}^c$, we have that

$$\begin{aligned}
&\iint \Phi(y - \alpha_0 - x^T\beta_0) f(y - \alpha - x^T\beta) dy Q(dx) \\
\geq\ & \int \Phi(y) f(y) dy - \\
& \iint |\Phi(y) - \Phi(y - (\alpha - \alpha_0) - x^T(\beta - \beta_0))| \\
& \quad \times f(y - (\alpha - \alpha_0) - x^T(\beta - \beta_0)) dy Q(dx) \\
\geq\ & \int \Phi(y) f(y) dy - \epsilon/2 \\
\geq\ & \mathrm{E}_{f_0}\Phi + \epsilon/2.
\end{aligned}$$

By Lemma 10.1, such tests are exponentially consistent.

For the second group of hypotheses, by Condition (i), there exist X's such that the j-the component of X, $X_j\gamma_j > \epsilon$ and X belongs to the proper Γ_γ such that $\alpha + X^T\beta - \alpha_0 - X^T\beta_0 > \Delta\epsilon$. Condition (i) also implies that there exists $L > 0$ such that $\|X\| \leq L$ and $\liminf n^{-1}\#\{i : \|X_{i,j}\| > \epsilon_0, X_i \in \Gamma_\gamma\} > 0$ a.s. for any Γ_γ, where $\#\{\cdot\}$ stands for the cardinality of the set. Let $K_{n,\gamma} = \{i : \|X_{i,j}\| > \epsilon_0, X_i \in \Gamma_\gamma\} > 0$, for each $i \in K_{n,\gamma}$, there exists a set A_i such that

$$\alpha_i := P_{f_{0i}}(A_i) < a - C(\eta \wedge \Delta\epsilon)$$

and

$$\gamma_i := \inf_{W_{2,\iota,\gamma,j}} P_{f_{\alpha,\beta,i}}(A_i) \geq a,$$

where η is such that $\inf_{|y|<\eta} f_0(y) = C > 0$. For $i \notin K_n$, set $A_i = \mathbb{R}$, then $\alpha_i = \gamma_i = 1$. Note that

$$\liminf_{n\to\infty} \left(n^{-1} \sum_{i=1}^{n} (\gamma_i - \alpha_i) \right) > C(\eta \wedge \Delta\epsilon_0) \liminf_{n\to\infty} \#K_n/n > 0$$

Lemma 10.1 now applies. Hence we have exponentially consistent test for the second group of hypotheses.

The construction of the exponentially consistent test for the third group of hypotheses is similar.

Note that by choosing Δ sufficiently small, the union of the sets in the alternative hypotheses contains $\{(f,\alpha,\beta) : f \in \mathscr{U}_{f_0}, |\alpha-\alpha_0| < \epsilon, \|\beta-\beta_0\| < \epsilon\}^c$. The required exponentially consistent tests therefore exist. $\square$

LEMMA 10.2. *If $|\log f_0(x,y)|$, x and y are f_0-integrable, then for any $\epsilon > 0$, there exists r such that $\iint f_0(x,y)\log\frac{f_0(x,y)}{f_r(x,y)}dxdy < \epsilon$.*

PROOF. Since $\lambda_0 w^2 e^{-w(x+\lambda_0)} < \lambda_0 w^2$ and $f_r(w) \le 2f_0(w)$, one has that

$$\int \lambda_0 w^2 e^{-w(x+\lambda_0 y)} dF_r(w) \le 2\lambda_0/3. \tag{10.1}$$

By the dominated convergence theorem (DCT), we have that $\log\frac{f_0(x,y)}{f_r(x,y)} \to 0$ as $r\to\infty$ point-wise.

Note that $f_r(x,y) = \frac{1}{F_0(r)-F_0(r^{-1})}\int_{r^{-1}}^{r}\lambda_0 w^2 e^{-w(x+\lambda_0 y)}dF_0(w)$ and $f_0(x,y) = \int_{\mathbb{R}^+}\lambda_0 w^2 e^{-w(x+\lambda_0 y)}dF_0(w)$. Thus we have $f_r(x,y) < 2f_0(x,y)$ for sufficiently large r.

Observe that for given (x,y), $\lambda_0 w^2 e^{-w(x+\lambda_0 y)}$ is increasing on $(0, 2(x+\lambda_0 y)^{-1})$ and decreasing on $[2(x+\lambda_0 y)^{-1}, \infty)$ as a function of w. Now let w_1, w_2 and w_3 be such that $F_0(w_i) = i/4$, $i = 1,2,3$. Choose r sufficiently large such that $r^{-1} < w_1$ and $r > w_3$. For $x+\lambda_0 y \ge 2w_2^{-1}$, one has $f_r(x,y) > \lambda_0 w_3^2 e^{-w_3(x+\lambda_0 y)}/4$. For $x+\lambda_0 y < 2w_2^{-1}$, one has $f_r(x,y) > \lambda_0 w_1^2 e^{-w_1(x+\lambda_0 y)}/4$. Therefore, for r large, we have

$$2f_0(x,y) > f_r(x,y) > \begin{cases} \lambda_0 w_3^2 e^{-w_3(x+\lambda_0 y)}/4, & \text{if} \quad x+\lambda_0 y \ge 2w_2^{-1}, \\ \lambda_0 w_1^2 e^{-w_1(x+\lambda_0 y)}/4, & \text{if} \quad x+\lambda_0 y < 2w_2^{-1}. \end{cases}$$

Hence

$$\begin{aligned}\left|\log\frac{f_0(x,y)}{f_r(x,y)}\right| < \quad & |\log f_0(x,y)| + \\ & \max\{\log 2 + |\log f_0(x,y)|, |\log \lambda_0 w_3/4| + |w_3(x+\lambda_0 y)|\}.\end{aligned}$$

Since $|\log f_0(x,y)|$ and $x+y$ are f_0-integrable, by the DCT, we have

$$\iint f_0(x,y)\log\frac{f_0(x,y)}{f_r(x,y)}dxdy \to 0 \text{ as } r\to\infty.$$

$\square$

LEMMA 10.3. *If x and y are f_0-integrable and $w_E := \int w^2 dF_0(w) < \infty$, then for any $\epsilon > 0$, there exists a neighborhood $\mathscr{V}$ of λ_0 such that for $\lambda \in \mathscr{V}$ and r sufficiently large,*

$$\iint f_0(x,y) \log \frac{\int \lambda_0 w^2 e^{-w(x+\lambda_0 y)} dF_r(w)}{\int \lambda w^2 e^{-w(x+\lambda y)} dF_r(w)} dxdy < \epsilon.$$

PROOF. By arguments similar to those used in the derivation of (10.1) and the DCT, one has that $\log \frac{\int \lambda_0 w^2 e^{-w(x+\lambda_0 y)} dF_r(w)}{\int \lambda w^2 e^{-w(x+\lambda y)} dF_r(w)} \to 0$, point-wise.

Observe that

$$\begin{aligned} &\left| \log \frac{\int \lambda_0 w^2 e^{-w(x+\lambda_0 y)} dF_r(w)}{\int \lambda w^2 e^{-w(x+\lambda y)} dF_r(w)} \right| \\ \leq\ & \left| \log\left\{ \int \lambda_0 w^2 e^{-w(x+\lambda_0 y)} dF_r(w) \right\} \right| + \left| \log\left\{ \int \lambda w^2 e^{-w(x+\lambda y)} dF_r(w) \right\} \right|. \end{aligned}$$

In order to show that

$$\iint f_0(x,y) \log \frac{\int \lambda_0 w^2 e^{-w(x+\lambda_0 y)} dF_r(w)}{\int \lambda\lambda w^2 e^{-w(x+\lambda y)} dF_r(w)} dxdy \to 0, \tag{10.2}$$

it now suffices to bound $|\log(\int \lambda w^2 e^{-w(x+\lambda y)} dF_r(w))|$ for any $\lambda \in [\lambda_0/2, 2\lambda_0]$ by an $f_0(x,y)$-integrable function $u(x,y)$. For given values of λ and (x,y), $\lambda w^2 e^{-w(x+\lambda y)}$ as a function of w, is increasing on $(0, \frac{2}{x+\lambda y})$ and decreasing on $(\frac{2}{x+\lambda y}, \infty)$. Let w_1, w_2 and w_3 be defined as in the proof of Lemma 10.2. For $r^{-1} < w_1$ and $r > w_3$, $f_r(w) \geq f_0(w)$ on (w_1, w_3). For given λ,

$$\begin{aligned} \int \lambda w^2 e^{-w(x+\lambda y)} dF_r(w) &> \frac{1}{8} \lambda_0 w_1^2 e^{-w_1(x+2\lambda_0 y)}, \quad \text{if } x + \lambda y > \frac{2}{w_2}, \\ \int \lambda w^2 e^{-w(x+\lambda y)} dF_r(w) &> \frac{1}{8} \lambda_0 w_3^2 e^{-w_3(x+2\lambda_0 y)}, \quad \text{if } x + \lambda y \leq \frac{2}{w_2}. \end{aligned}$$

Hence, we have

$$\begin{aligned} &\log\left\{ \int \lambda w^2 e^{-w(x+\lambda y)} dF_r(w) \right\} \\ &\quad \geq \min\left\{ \log(\lambda_0 w_1^2/8) - w_1(x + 2\lambda_0 y), \log(\lambda_0 w_3^2/8) - w_3(x + 2\lambda_0 y) \right\}. \end{aligned}$$

Therefore, $\log(\int \lambda w^2 e^{-w(x\lambda y)} dF_r(w)) \geq \log(\lambda_0 w_1^2/8) - 2w_3(x + \lambda_0 y)$. Since $f_r(w) \leq 2f_0(w)$, we also have $\log(\int \lambda w^2 e^{-w(x+\lambda y)} dF_r(w)) \leq 4\lambda_0 w_E$. Hence,

$$\left|\log(\int \lambda w^2 e^{-w(x+\lambda y)} dF_0(w))\right|$$
$$\leq \max\left\{4\lambda_0 w_E, |\log(\lambda_0 w_1^2/8)| + 2w_3(x+\lambda_0 y)\right\}.$$

By the condition that x and y are f_0-integrable, (10.2) holds by the DCT. □

Proof of Lemma 4.1. To show that there exist exponentially consistent test for testing $\mathrm{H}_0 : (\lambda, F) = (\lambda_0, F_0)$ against $\mathrm{H}_1 : (F, \lambda) \in \mathscr{W}$, we show that there exist exponentially consistent tests for testing

$$\mathrm{H}_0 : (\lambda, F) = (\lambda_0, F_0) \text{ against } \mathrm{H}_1 : |\lambda - \lambda_0| > \epsilon; \tag{10.3}$$

and

$$\mathrm{H}_0 : (\lambda, F) = (\lambda_0, F_0) \text{ against } \mathrm{H}_1 : F \in \mathscr{U}^c \cap \mathscr{A}_n, \tag{10.4}$$

where $\mathscr{A}_n = \{F : F(r_n) - F(r_n^{-1}) \geq 1 - \delta\}$ for some given δ. Note that

$$Z = \frac{X}{X + \lambda_0 Y} \sim \mathrm{Unif}(0,1), \text{ if } \lambda = \lambda_0. \tag{10.5}$$

We use Kolmogorov-Smirnov test (KS test) to test $Z \sim \mathrm{Unif}(0,1)$, which in turn tests (10.3).

Let D_n denote the maximum difference between the empirical c.d.f. of Z and the standard uniform c.d.f. $U(z) = z$, $z \in [0,1]$. Let test $\Phi_n(\cdot)$ be $\mathbb{1}(D_n > \delta)$. We have that, for large samples,

$$\begin{aligned}
\mathrm{E}_{\lambda_0}(\Phi_n) &= \Pr(D_n \geq \delta) \leq 2e^{-2\delta^2 n}, \\
\mathrm{E}_{\lambda}(\Phi_n) &\geq 1 - (2\pi)^{-1/2} \int_{2(\Delta_\lambda\sqrt{n}-\delta)}^{2(\Delta_\lambda\sqrt{n}+\delta)} e^{-\frac{t^2}{2}} dt, \quad |\lambda - \lambda_0| > \epsilon,
\end{aligned}$$

where $\Delta_\lambda = \sup_{0\leq z\leq 1}\left|P_\lambda\left(\frac{X}{X+\lambda_0 Y} \leq z\right) - z\right|$; see Massey (1950) for details. In the above expression, P_λ denotes the probability law for $\frac{X}{X+\lambda_0 Y}$, when $X|W = w \sim \mathrm{Exp}(w)$ and $Y|W = w \sim \mathrm{Exp}(\lambda w)$.

We show that $\Delta := \inf_{|\lambda-\lambda_0|>\epsilon} \Delta_\lambda > 0$. Then it will follow that

$$\inf_{|\lambda-\lambda_0|>\epsilon} \mathrm{E}_\lambda(\Phi_n) \geq 1 - \frac{4\delta}{\sqrt{2\pi}} e^{-\frac{4(\Delta\sqrt{n}-\delta)^2}{2}} \geq 1 - 2\sqrt{2}\frac{\delta}{\sqrt{\pi}} e^{-\frac{1}{2}\Delta^2 n},$$

for n sufficiently large, leading to the desired test for (10.3).

Note that

$$P_\lambda\Big(\frac{X}{X+\lambda_0 Y}\le\frac{1}{2}\Big)=P_\lambda^*\Big(\frac{X}{Y}\le\lambda_0\Big)=\frac{\lambda_0}{\lambda_0+\lambda},$$

where P_λ^* denotes the probability law for X/Y when $X|W=w\sim\text{Exp}(w)$ and $Y|W=w\sim\text{Exp}(\lambda w)$. Hence, $\Delta_\lambda\ge|\frac{1}{2}-\frac{\lambda_0}{\lambda_0+\lambda}|\ge\frac{\epsilon}{4\lambda_0+2\epsilon}>0$, for all λ such that $|\lambda-\lambda_0|>\epsilon$. Thus, the sequence of tests Φ_n are exponentially consistent.

Now we show that there exist exponentially consistent tests for testing (10.4). Let

$$g_0(x)=\int we^{-wx}dF_0(w),\quad g_F(x)=\int we^{-wx}dF(w). \tag{10.6}$$

Let $\mathscr{G}=\{g_F: g_F(x)=\int we^{-wx}dF(w), F\in\mathscr{M}(\mathbb{R}^+)\}$. Note that the correspondence $F\mapsto g_F$ is one to one by the unicity of the Laplace transform. Let $\|g_F-g_0\|=\int|g_F(x)-g_0(x)|dx$ be the L_1-norm on $\mathscr{G}$. We now show that for any sequence $g_n\in\mathscr{G}$,

$$g_n\xrightarrow{L_1}g_0 \text{ implies } F_n\xrightarrow{w}F_0, \tag{10.7}$$

where $g_n=\int we^{-wx}dF_n(w)$ and $\xrightarrow{w}$ stands for weak convergence.

If $g_n\xrightarrow{L_1}g_0$, which means $\int_0^\infty we^{-wx}dF_n(w)\xrightarrow{L_1}\int_0^\infty we^{-wx}dF_0(w)$, then

$$\sup_{x\in\mathbb{R}^+}\left|\int_0^x\int_0^\infty we^{-wt}dF_n(w)dt-\int_0^x\int_0^\infty we^{-wt}dF_0(w)dt\right|\to 0.$$

Now, by Fubini's theorem, we have

$$\sup_{x\in\mathbb{R}^+}\left|\int_0^\infty e^{-wx}dF_n(w)-\int_0^\infty e^{-wx}dF_0(w)\right|\to 0.$$

Hence, for any $x>0$, $\int_0^\infty e^{-wx}dF_n(w)\to\int_0^\infty e^{-wx}dF_0(w)$. Since $\int e^{-wx}dF(w)$ is the Laplace transform of F, by Theorem 2 in Feller (1957, Vol. II XIII.1), it follows that $F_n\xrightarrow{w}F_0$.

Observe that there exists a sequence of exponentially consistent tests for testing

$$\text{H}_0:\ g=g_0\quad\text{against}\quad\text{H}_1:\ g\in\mathscr{U}^{*c}\cap\Big\{\int we^{-wx}dF(w): F\in\mathscr{A}_n\Big\} \tag{10.8}$$

by Lemma 10.5 below, where $\mathscr{U}^*$ is an L_1-neighborhood of g_0. The existence of a sequence of exponentially consistent tests for (10.8) implies the existence of a sequence of exponentially consistent tests Φ_n^* for testing (10.4).

Now the test $\max\{\Phi_n(\cdot), \Phi_n^*(\cdot)\}$ satisfies Condition (i) of Theorem 2.1, with the $\mathscr{A}_n \times [m_n, M_n]^d = \{F : F(r_n) - F(r_n^{-1}) \geq 1 - \delta\} \times \mathbb{R}^+$. □

LEMMA 10.4. *If $|\log f_0(x,y)|$ is $f_0(x,y)$-integrable, then for any $\epsilon > 0$ there exists a weak open neighborhood $\mathscr{U}$ of F_r, such that*

$$\iint f_0(x,y) \log \frac{\int \lambda w^2 e^{-w(x+\lambda y)} dF_r(w)}{\int \lambda w^2 e^{-w(x+\lambda y)} dF(w)} dxdy < \epsilon$$

for every $F \in \mathscr{U}$ and $\lambda \in \mathscr{V} \subset [\lambda_0/2, 2\lambda_0]$.

PROOF. We verify Conditions A7–A9 of Lemma 3 in Wu and Ghosal (2008) to prove this lemma, that is,

A7. for any $\phi \in A$, $\int \log \frac{f_{P_\epsilon,\phi}(x)}{\inf_{\theta \in D} K(x,\theta,\phi)} f_0(x)dx < \infty$;

A8. $c := \inf_{x \in C} \inf_{\theta \in D} K(x;\theta,\phi) > 0$, for any compact $C \subset \mathfrak{X}$;

A9. for any given $\phi \in A$ and compact $C \subset \mathfrak{X}$, such that the family of maps $\{\theta \mapsto K(x;\theta,\phi), x \in C\}$ is uniformly equicontinuous on $D \subset \Theta$.

In this case, $K(x,y;w,\lambda) = \lambda w^2 e^{-w(x+\lambda y)}$, where w plays the role of θ in the conditions, λ for ϕ and P_ϵ is substituted by proper F_r. Choose r and $\mathscr{V}$, such that $\iint f_0(x,y) \log \frac{f_0(x,y)}{f_{r,\lambda}(x,y)} dxdy < \epsilon$ for every $\lambda \in \mathscr{V}$. From Lemma 10.3, one has that $\iint f_0(x,y) \log f_{r,\lambda}(x,y)dxdy < \infty$. Let $D = [r^{-1}, r]$,

$$|\log(\inf_{w \in D} \lambda w^2 e^{-w(x+\lambda y)})| \leq (r + r^{-1})(x + 2\lambda_0 y) + 2\log r + \log(2\lambda_0).$$

Now $|\log(\inf_{w \in D} \lambda w^2 e^{-w(x+\lambda y)})|$ is integrable. Hence Condition A7 is satisfied. Condition A8 obviously holds. For Condition A9, the uniform equicontinuity condition holds for any compact E. By straightforward calculations, we will see that Condition A9 holds. Thus Lemma 3 of Wu and Ghosal (2008) concludes the proof. □

LEMMA 10.5. *If for any $\delta > 0$ and $\beta > 0$ there exists a sequence r_n and a constant β_0 such that $r_n^2 < n\beta$ and $\tilde{\Pi}\{F : F(r_n) - F(r_n^{-1}) < 1 - \delta\} < e^{-n\beta_0}$, then there exists an exponentially consistent sequence of tests for test* (10.8).

PROOF. Let $\mathscr{F}$ denote the space of density functions on $\mathbb{R}^+$. For any $\epsilon > 0$, let $\mathscr{U}^* = \{f : \|f - f_0\| < \epsilon\}$, where $\|f - f_0\| = \int |f(x) - f_0(x)|dx$. Choose $\delta < \epsilon/4$, let $\mathscr{F}_n = \{\int we^{-wx}dF(w) : F(r_n) - F(r_n^{-1}) \geq 1 - \delta\}$. By the assumption of this lemma, we have that $\tilde{\Pi}(\mathscr{F}^c) < e^{-n\beta_0}$. Choose $g_1, g_2 \dots, g_k \in \mathscr{F}$ be such that $V_n \subset \cup_{i=1}^k G_i$ where $V_n = \mathscr{F}_n \cap \mathscr{W}^{*c}$ and $G_i = \{f : \|f - g_i\| < \delta\}$. Let $f_i \in V_n \cap G_i$. Let $A_i = \{x : f_0(x) < f_i(x)\}$ and $B_i = \{(x_1, x_2, \dots, x_n) : \frac{1}{n}\sum_{j=1}^n 1\!\!1_{A_i}(x_j) \geq \frac{1}{2}(P_{f_0}(A_i) + P_{f_i}(A_i))\}$. Consider the test

$$\phi_n(X_1, X_2, \dots, X_n) = \max_{1\leq i\leq k} 1\!\!1_{B_i}(X_1, X_2, \dots, X_n),$$

where X_j are the first components from the observations (X_j, Y_j)'s.

Let $J(\delta, \mathscr{F})$ stand for the L_1-metric entropy, that is, the logarithm of the smallest number k of functions $f_1, \dots, f_k$ such that for any $f \in \mathscr{F}$, we have $\|f - f_i\| < \delta$ for some $i = 1, \dots, k$.

Now, as shown in the proof of Theorem 2 in Ghosal *et al.* (1999), if $J(\delta, \mathscr{F}_n) < n\beta$ for some $\beta < \epsilon^2/8$, then the sequence of tests constructed above is exponentially consistent. Hence showing that for any $\delta > 0$ and $\beta > 0$, $J(\delta, \mathscr{F}_n) < n\beta$ for some suitable choice of r_n is sufficient to complete the proof.

For $w_1 > w_2$, we have $w_1e^{-w_1x} > w_2e^{-w_2x}$ on $[0, \log\frac{w_1}{w_2}/(w_1 - w_2))$ and $w_1e^{-w_1x} \leq w_2e^{-w_2x}$ on $[\log\frac{w_1}{w_2}/(w_1 - w_2), \infty)$ and hence

$$\begin{aligned}\|w_1e^{-w_1x} - w_2e^{-w_2x}\| &\leq 2\int_0^{\log\frac{w_1}{w_2}/(w_1-w_2)} (w_1e^{-w_1x} - w_2e^{-w_2x})dx \\ &\leq 2\sup_{x\in A}(w_1e^{-w_1x} - w_2e^{-w_2x}) \log\frac{w_1}{w_2}/(w_1 - w_2) \\ &= 2(\log w_1 - \log w_2),\end{aligned}$$

where $A = [0, \log\frac{w_1}{w_2}/(w_1 - w_2))$ in the above display.

Given δ, let N be the smallest integer greater than $\frac{r_n^2-1}{e^{\delta/2-1}}$. Divide $(r_n^{-1}, r_n]$ in to N intervals. Let

$$E_i = \left(\frac{1}{r_n} + \frac{(r_n - r_n{}^{-1})(i-1)}{N}, \frac{1}{r_n} + \frac{(r_n - r_n{}^{-1})i}{N}\right], \quad i = 1, 2, \dots, N.$$

Then for any $w, w' \in E_i$, we have $\|we^{-wx} - w'e^{-w'x}\| < \delta$. Now, along the same lines in the proof of Lemma 1 and Lemma 2 in Ghosal *et al.* (1999), we have that $J(\delta, \mathscr{F}_n) < n\beta$ if $r_n^2 < n\beta$. □

Proof of Lemma 5.1. Consider X and β are both d-dimensional. Let $B=\{\beta:\|\beta\|=1\}$, $B^*=B\setminus\{\beta:\|\beta-\beta_0\|\le\epsilon\}$ and $\mathbb{X}=\{x:\|x-\beta\|\le\delta,\ \text{for some } \beta\in B\}$.

For any $\beta_1\in B^*$, let Y_0 and Y_1 be random variables corresponding to random variable X_0 taking the value of β_0 and X_1 taking the value of β_1 respectively. Let

$$\begin{aligned}\theta_0 &:= b'^{-1}(g_0(X_0^T\beta_0))=b'^{-1}(g_0(1)),\\ \theta_1 &:= b'^{-1}(g_0(X_1^T\beta_0))=b'^{-1}(g(\beta_1^T\beta_0)),\\ \theta_0^* &:= b'^{-1}(g(X_0^T\beta_1))=b'^{-1}(g(\beta_0^T\beta_1)),\\ \theta_1^* &:= b'^{-1}(g(X_1^T\beta))=b'^{-1}(g(1)).\end{aligned}$$

Now we have that

$$R(\theta_0,\theta_1):=\mathrm{E}_{\beta_0,g_0}\mathbb{1}(Y_0<Y_1)=\frac{\int_{-\infty}^{\infty}\int_{t_0}^{\infty}e^{t_1\theta_1}a(t_1)e^{t_0\theta_0}a(t_0)dt_1dt_0}{\int_{-\infty}^{\infty}e^{t_1\theta_1}a(t_1)dt_1\int_{\infty}^{\infty}e^{t_0\theta_0}a(t_0)dt_0} \qquad (10.9)$$

and

$$R(\theta_0^*,\theta_1^*):=\mathrm{E}_{\beta_1,g}\mathbb{1}(Y_0<Y_1)=\frac{\int_{-\infty}^{\infty}\int_{t_0}^{\infty}e^{t_1\theta_0^*}a(t_1)e^{t_0\theta_1^*}a(t_0)dt_1dt_0}{\int_{-\infty}^{\infty}e^{t_1\theta_0^*}a(t_1)dt_1\int_{\infty}^{\infty}e^{t_0\theta_1^*}a(t_0)dt_0}. \qquad (10.10)$$

By the fact that $\theta_0>\theta_1$ and $t_0\theta_0+t_1\theta_1<t_0\theta_1+t_1\theta_0$ when $t_0<t_1$, we have that $R(\theta_0,\theta_1)<R(\theta_1,\theta_0)$. Now, let

$$\Delta=\inf_{\beta\in B^*}\Big\{R\Big(b'^{-1}(g_0(\beta^T\beta_0)),b'^{-1}(g_0(1))\Big)-R\Big(b'^{-1}(g_0(1)),b'^{-1}(g_0(\beta^T\beta_0))\Big)\Big\}. \qquad (10.11)$$

By the continuity of $R(\cdot)$, we can find $\xi_1>0$, such that

$$|R(\theta_2,\theta_3)-R(\theta_2^*,\theta_3^*)|<\Delta/8 \qquad (10.12)$$

for all $|\theta_2-\theta_2^*|<\xi_1$, $|\theta_3-\theta_3^*|<\xi_1$ and

$$b'^{-1}(g_0(-1-\epsilon)-\xi_1)\le\theta_2,\theta_3,\theta_2^*,\theta_3^*\le b'^{-1}(g_0(1+\epsilon)+\xi_1).$$

By the continuity of F and T^{-1}, there exists $\xi>0$, such that for all F with $\|F-F_0\|_{\sup}<\xi$,

$$\sup_{-1-\epsilon\le t\le 1+\epsilon}|T^{-1}\circ F(t)-g_0(t)|<\xi_1.$$

Now, we can find a weak neighborhood $\mathscr{U}$ of F_0 such that for all the $F \in \mathscr{U}$. Also, by continuity, there exists $\delta > 0$ such that

$$\left\| R(b'^{-1}(g_0(X_2^T\beta)), b'^{-1}(g_0(X_3^T\beta))) - R(b'^{-1}(g_0(X_2^{*T}\beta^*)), b'^{-1}(g_0(X_3^{*T}\beta^*))) \right\| < \Delta/8 \tag{10.13}$$

for all for $\|\beta^* - \beta\| \le \delta$, $\|X_2^* - X_2\| \le \delta$ and $\|X_3^* - X_3\| \le \delta$, where $\beta, \beta^* \in B^*$ and $X_2, X_3, X_2^*, X_3^* \in \mathbb{X}$.

Now, let X_i and Y_i denote the ith covariate and observation, $K_{n0} = \{i : 1 \le i \le n, \|X_i - \beta_0\| \le \delta\}$ and $K_n = \{i : 1 \le i \le n, \|X_i - \beta_1\| \le \delta\}$. Let $\kappa = \min\{\#K_n,\ \#K_{n0}\}$. We will construct a test using only those Y_i for which the corresponding i is in K_n or K_{n0}. Take the first κ i's from K_n and K_{n0}, and setting them in pairs as $(X_{0k}, X_{1k}), k = 1, 2, \ldots \kappa$. Now we have $(Y_{0k}, Y_{1k}), k = 1, 2, \ldots, \kappa$, which denote the observations corresponding to κ pairs of X's. Note that all the X_{0k} are from $\{X_i : i \in K_{n0}\}$ and X_{1k} are from $\{X_i : i \in K_{n1}\}$.

By (10.13), we have

$$\mathrm{E}_{\beta_0} 1\!\!1(Y_{0k} < Y_{1k}) \le R(b'^{-1}(g_0(1)), b'^{-1}(g_0(\beta_1^T\beta_0))) + \Delta/8, \tag{10.14}$$

for each k. Let $B_{(\beta_1,\delta)} = \{\|\beta - \beta_1\| \le \delta\}$. By (10.12) and (10.13), we have

$$\begin{aligned}
&\inf\{\mathrm{E}_\beta 1\!\!1(Y_{0k} < Y_{1k}) : \beta \in B_{(\beta_1,\delta)} \cap B^*, F \in \mathscr{U}\} \\
= \ & \inf\Big\{ R(b'^{-1}(g_0(X_{0k}^T\beta)), b'^{-1}(g_0(X_{1k}^T\beta))) - R(b'^{-1}(g_0(\beta_1^T\beta)), b'^{-1}(g_0(1))) \\
& + R(b'^{-1}(g(X_{0k}^T\beta)), b'^{-1}(g(X_{1k}^T\beta))) - R(b'^{-1}(g_0(X_{0k}^T\beta)), b'^{-1}(g_0(X_{1k}^T\beta))) \\
& + R(b'^{-1}(g_0(\beta_1^T\beta)), b'^{-1}(g_0(1))) - R(b'^{-1}(g(\beta_1^T\beta)), b'^{-1}(g(1))) \\
& + R(b'^{-1}(g(\beta_1^T\beta)), b'^{-1}(g(1))) - R(b'^{-1}(g_0(1)), b'^{-1}(g_0(\beta_1^T\beta))) \\
& + R(b'^{-1}(g_0(1)), b'^{-1}(g_0(\beta_1^T\beta))) : \beta \in B_{(\beta_1,\delta)} \cap B^*, F \in \mathscr{U}^* \Big\} \\
\ge \ & R(b'^{-1}(g_0(1)), b'^{-1}(g_0(\beta_1^T\beta_0))) + \Delta/2.
\end{aligned} \tag{10.15}$$

Therefore, by the Lemma 10.1, there exist exponentially consistant sequence of tests Φ_{1n} for testing

$$\mathrm{H}_0 : \beta = \beta_0 \qquad \text{against} \qquad \mathrm{H}_1 : \beta \in B_{(\beta_1,\delta)} \cap B^*.$$

Hence, there exists $C > 0$ such that $\inf_{\beta^* \in B_{(\beta_1,\delta)} \cap B^*} \mathrm{E}_\beta \Phi_{1n} > 1 - e^{-nC}$ and $\mathrm{E}_{\beta_0}\Phi_{1n} < e^{-nC}$. Also note that C depends only on Δ, and Δ depends on ϵ.

Choose $\beta_1, \beta_2, \ldots, \beta_l$, in B^* such that $B^* \subset \cup_{j=1}^{l} B_j$ where $B_j = B_{(\beta_j, \delta)}$. Now there exists exponentially consistent Φ_{jn} for each $1 \le j \le l$. Set

$$\Phi_n(Y_1, \ldots, Y_n) = \max_{1 \le j \le l} \Phi_{jn},$$

then $\mathrm{E}_{\beta_0}\Phi_n \le le^{-nC}$ and $\inf_{\beta \in B^*} \mathrm{E}_\beta \Phi_n \ge 1 - e^{-nC}$. The test Φ_n is exponentially consistent if $\log l < nc$ for some constant $c > 0$. For given ϵ, Δ is fixed and hence so is δ. Now, by choosing $l \le [2d\delta^{1-d}] + 1$, the proof is completed. □

Proof of Lemma 5.2. By Proposition 2.5.2 of Ghosh and Ramamoorthi (2003), for any given weak neighborhood $\mathscr{U}$ of F_0, there exist $-L \le a_1, \ldots, a_m \le L \in \mathbb{Q}$ and $\delta > 0$ such that,

$$\mathscr{V} = \{P : |P[a_i, a_{i+1}) - P[a_i, a_{i+1})| < \delta \text{ for } 1 \le i \le m\} \subset \mathscr{U}.$$

Now for any $F \notin \mathscr{U}$, there exists i, such that $|F(a_{i+1}) - F(a_i) - F_0(a_{i+1}) + F_0(a_i)| \ge \delta$. Hence $|F(a_{i+1}) - F_0(a_{i+1})| + |F(a_i) - F_0(a_i)| \ge \delta$, and this implies that $|F(a_{i+1}) - F_0(a_{i+1})| \ge \delta/2$ or $|F(a_i) - F_0(a_i)| \ge \delta/2$ or both. Let

$$\mathscr{V}_{i+} = \{F : F(a_i) - F_0(a_i) \ge \delta/2\}, \tag{10.16}$$
$$\mathscr{V}_{i-} = \{F : F(a_i) - F_0(a_i) \le -\delta/2\}, \tag{10.17}$$

for $1 \le i \le m$. We see that $\cup_{i=1}^{m}(\mathscr{V}_{i+} \cup \mathscr{V}_{i-}) \supset \mathscr{U}^c$.

Now, if the tests for

$$\mathrm{H}_0 : F = F_0,\ \beta = \beta_0 \quad \text{against} \quad \mathrm{H}_1 : F \in \mathscr{V}_{i+},\ \beta \in B_n \tag{10.18}$$

and

$$\mathrm{H}_0 : F = F_0,\ \beta = \beta_0 \quad \text{against} \quad \mathrm{H}_1 : F \in \mathscr{V}_{i-},\ \beta \in B_n \tag{10.19}$$

are both exponentially consistent for each $i = 1, 2, \ldots, m$, then there exists an exponentially consistent sequence of tests.

For any given i, we consider the test for (10.18) in the following. Let $\mathbb{X}_{i0} = \{x : x^T\beta_0 = a_i\}$ and $\mathbb{X}_{i+} = \{x : a_i \le x^T\beta_0 \le a_i + \delta_i\}$, where $0 < \delta_i < \frac{1}{3}$ is chosen such that $T(F_0(a_i + \delta_i)) - T(F_0(a_i)) < \delta/4$.

For any $\beta \in B_n$, we know that there exists $x_\beta \in \mathbb{X}_{i0}$ such that

$$\|x_\beta - a_i\beta_0\| < \frac{1 + 3a_i}{r_n} \text{ for small } r_n \tag{10.20}$$

and $x_\beta^T\beta = a_i + 1$. Now for any $x \in \{x : \|x - x_\beta\| \le \frac{1}{3}\}$ we have

$$x^T\beta \ge x^T\beta_0 + \frac{1}{3}. \tag{10.21}$$

Therefore, there exist $x_j \in \mathbb{X}_{i+}$, $j = 1, \dots, l_{in}$, such that

$$\cup_{j=1}^{l_{in}} K_{ij+}^* \supset \mathbb{X}_{i+} \cap \{x : x \text{ satisfies } (10.20)\},$$

where $K_{ij+}^* = \{x : \|x - x_j\| \le \frac{1}{6}\}$. Let $K_{ij+} = K_{ij+}^* \cap \mathbb{X}_{i+}$. Let $B_{nj+} = \{\beta : x^T\beta \ge x^T\beta_0 \text{ for each } x \in K_{ij+}\}$.

Since $\cup_{j=1}^{l_{in}} K_{ij+}^* \supset \mathbb{X}_{i+} \cap \{x : x \text{ satisfies } (10.20)\}$, for any $\beta \in B_n$, the corresponding $x_\beta \in K_{ij+}$ for some j. For such j, $\|x_\beta - x_j\| \le \frac{1}{6}$ and $\sup_{x \in K_{ij+}} \|x_\beta - x\| \le \frac{1}{3}$, we see that (10.21) is satisfied for each $x \in K_{ij+}$, and hence $\beta \in B_{nj+}$. Therefore, $\cup_{j=1}^{l_{in}} B_{nj+} \supset B_n$.

For testing

$$\text{H}_0 : F = F_0,\ \beta = \beta_0 \quad \text{against} \quad \text{H}_1 : F \in \mathscr{V}_{i+},\ \beta \in B_{nj+}, \tag{10.22}$$

consider observations Y_k's that corresponding to $x_k \in K_{ij+}$ only. We have

$$\begin{aligned} & g(x^T\beta) - g_0(x^T\beta_0) \\ \ge\ & g(x^T\beta_0) - g_0(x^T\beta_0) = T(F(x^T\beta_0)) - T(F_0(x^T\beta_0)) \\ =\ & [T(F(x^T\beta_0)) - T(F(a_i))] + [T(F(a_i)) - T(F_0(a_i))] \\ & + [T(F_0(a_i)) - T(F_0(x^T\beta_0))] \\ >\ & \frac{\delta}{4}. \end{aligned}$$

Let $\zeta = \inf_{x \in K_{ij+}} \{b'^{-1}(g_0(x^T\beta_0) + \delta/4) - b'^{-1}(g_0(x^T\beta_0))\}$. If the k-th observation is Y_k with corresponding $x_k \in K_{ij+}$, then letting $\theta_{0k} = b'^{-1}(g_0(x_k^T\beta_0))$, and $\theta_k = b'^{-1}(g(x_k^T\beta))$, we have that $\theta_k > \theta_{0k} + \zeta$, for any $F \in \mathscr{V}_{i+}$ and $\beta \in B_{nj+}$.

Let $\Phi_k = 1 - 1\!\!1(Y_k < t)$, where $t = (b(\theta_{0k} + \zeta) - b(\theta_{0k}) - \log 2)/\zeta$. Note that $b''(\theta) > 0$, $\frac{b(\theta_{0k}+\zeta) - b(\theta_{0k})}{\zeta} < \frac{b(\theta) - b(\theta_{0k})}{\theta - \theta_{0k}}$ for all $\theta > \theta_{0k} + \zeta$.

Let $f_0(y)$ denote the density function for Y corresponding to $\beta = \beta_0, F = F_0$ and let $f(y)$ denote the density function corresponding to $\beta \in B_{nj+}, F \in \mathscr{V}_{i+}$. For $y < t$, we have that $f_0(y) > 2f(y)$. Hence, $\text{E}_{F_0,\beta_0}(\Phi_k) = 1 - \alpha_k$ and $\text{E}_{F \in \mathscr{V}_{i+}, \beta \in B_{nj+}}(\Phi_k) > 1 - \alpha_k/2$. Let

$$\underline{\alpha}_k = \inf_{x_k \in K_{ij+}} \text{E}_{F_0,\beta_0}(1\!\!1(Y_k < t)). \tag{10.23}$$

Since $t > b'(\theta_0) = \mathrm{E}_{F_0,\beta_0}(Y_k)$, we see that $\underline{\alpha}_k > 0$.

Let $\alpha_k = 1$ and $\Phi_k = 0$ when the corresponding $x_k \notin K_{ij+}$. Now, let

$$D_{nij+} = \left\{ (Y_1, \ldots, Y_n) : \frac{1}{n} \sum_{k=1}^{n} \Phi(Y_k) \geq \sum_{k=1}^{n} (1 - \frac{3}{4}\alpha_k) \right\}.$$

A straightforward application of Hoeffding's inequality (see Hoeffding (1963), Theorem 2) shows that

$$\mathrm{Pr}_{F_0,\beta_0}(D_{nij+}) \leq \exp\left[-2n\Big(\frac{\sum_{k=1}^{n} \alpha_k}{4n} \Big)^2 \right] \leq \exp\Big(-n\underline{\alpha}_n^2/8 \Big).$$

Applying Hoeffding's inequality to $1 - \Phi_k$, we have

$$\mathrm{Pr}_{F \in \mathscr{V}_{i+}, \beta \in B_{nj+}}(D_{nij+}^c) \leq \exp\Big(-n\frac{\underline{\alpha}_n^2}{8} \Big).$$

Now, we see that the tests described as above for (10.22) is exponentially consistent.

To construct the test for (10.18), we set $\Phi_n(Y_1, \ldots, Y_n) = \max_{1 \leq j \leq l_{in}} \Phi_k$. For such Φ_n, we have $\mathrm{E}_{F_0,\beta_0}\Phi_n \leq l_{in}\exp(-n\underline{\alpha}_n^2/8)$, and $\inf\{\mathrm{E}\Phi_n : F \in \mathscr{V}_{i+}, \beta \in B_n\} \geq 1 - \exp(-n\underline{\alpha}_n^2/8)$, by noticing that $\underline{\alpha}_n$ is only depended on the subscript of $\mathscr{V}_{\cdot}$, say $i+$, but does not depend on n or j or the choice of x_j.

Since $\log l_{in} < n\underline{\alpha}_n^2/8$, we showed that Φ_n is a sequence of exponentially consistent tests for (10.18). Similarly, we can show that a sequence of exponentially consistent tests for (10.19) exists. Along the same line that we constructed Φ_n from Φ_k, we will have a sequence of exponentially consistent tests. □

Proof of Lemma 6.1. For any given weak neighborhood $\mathscr{U}_{F_0}$ of F_0, there exist $a_1, \ldots, a_m$ such that $\{F : |F(a_k) - F_0(a_k)| < \xi\} \subset \mathscr{U}_{F_0}$ for all $k = 1, \ldots, m$. Let Q_b denote the quadrant indexed by b, where $b \in \{-1, 1\}^d$. Divide one test into $2^{d+1}m$ tests as $\mathrm{H}_0 : (h, \beta) = (h_0, \beta_0)$, against $\mathrm{H}_1 : 1 - \exp[-H(a_k)] > F_0(a_k) + \xi$, $\beta - \beta_0 \in Q_b$, and $\mathrm{H}_0 : (h, \beta) = (h_0, \beta_0)$, against H_1 : $1 - \exp[-H(a_k)] < F_0(a_k) - \xi$, $\beta_0 - \beta \in Q_b$ for $k = 1, \ldots, m$, and every b. Like before, for instance, for tests corresponding to $1 - \exp[-H(a_k)] > F_0(a_k) + \xi$, , by choosing only $X_i \in Q_b$ and corresponding Z_i's for the tests, we have that $\mathrm{E}_{F_0,b,k}(Z_i \leq a_k) = F_0(a_k)^{e^{X_i^T \beta_0}} \bar{F}_c(a_k) + F_c(a_k)$ and

$E_{F,b,k}(Z_i \leq a_k) \geq (F_0(a_k)+\xi)^{e^{X_i^T\beta}}\bar{F}_c(a_k)+F_c(a_k)$. For tests corresponding to $F_0(a_k) > F(a_k)+\xi$, just let the testing function corresponding to $Z_i \geq a_k$. Now, using Lemma 10.1 and constructing a test by taking maximum value of all these tests, the lemma follows. □

PROOF OF LEMMA 6.2. Note that β_l denotes the l-th component of the vector β and $\beta-\beta_0$ belongs to Q_b, one of the quadrants. Now, consider $d2^d$ tests such as

$$\begin{aligned} \mathrm{H}_0 &: (h,\beta)=(h_0,\beta_0) \text{ against} \\ \mathrm{H}_{1l} &: \sup_{t\in\mathbb{R}^+} |\exp(-H(t))-\bar{F}_0(t)|<\Delta,\ |\beta_l-\beta_{0l}|>\delta, \beta-\beta_0 \in Q_b, \end{aligned} \tag{10.24}$$

for every $l=1,\ldots,d$, and $b\in\{-1,1\}^d$. Find t such that $F_0(t)=1/2$. For given δ, choose Δ small, such that $-\delta\log 2/[\log(\frac{1}{2})-\log(\frac{1}{2}-\Delta)]>2$. Now choosing $X_i \in Q_b$, such that $(1/2-\Delta)^{\exp(X_i^T\beta)} > (1/2)^{\exp(X_i^T\beta_0)}+\xi$ for some $\xi>0$ and using only the corresponding Z_i's to construct the test (10.24), we have that

$$E_\beta(Z>t) \geq (\frac{1}{2}-\Delta)^{\exp(X_i^T\beta)}\bar{F}_c(t)=\gamma_i \geq \alpha_i = (\frac{1}{2})^{\exp(X_i^T\beta_0)}\bar{F}_c(t),$$

where the last expression is equal to $E_{\beta_0}(Z>t)$.

Using Lemma 10.1 and constructing a test by taking maximum value of all these tests prove the lemma. □

PROOF OF LEMMA 7.1. For any given weak neighborhood $\mathscr{U}_{F_0}$ of F_0, there exist $a_1,\ldots,a_m$ such that $\{F: |F(a_k)-F_0(a_k)|<\xi, k=1,\ldots,m\} \subset \mathscr{U}_{F_0}$. Let Q_b denote the quadrant indexed by b, where $b\in\{-1,1\}^d$. Divide one test into $2^{d+1}m$ tests as $\mathrm{H}_0 : (F,\beta)=(F_0,\beta_0)$, against $\mathrm{H}_1 : F(a_k)>F_0(a_k)+\xi,\ \beta-\beta_0\in Q_b$, and $\mathrm{H}_0 : (F,\beta)=(F_0,\beta_0)$, against $\mathrm{H}_1 : F(a_k)<F_0(a_k)-\xi,\ \beta_0-\beta\in Q_b$ for $k=1,\ldots,m$, and every b. Like before, for instance, for tests corresponding to $F(a_k)>F_0(a_k)+\xi$, , by choosing only $X_i\in Q_b$ and corresponding Z_i's for the tests, we have that $E_{F_0,b,k}(\log Z_i - X^T\beta_0 \leq a_k) = F_0(a_k)\bar{F}_c(a_k+X^T\beta_0)+F_c(a_k+X^T\beta_0)$ and $E_{F,b,k}(\log Z_i - X^T\beta_0 \leq a_k) \geq (F_0(a_k)+\xi)\bar{F}_c(a_k+X^T\beta_0)+F_c(a_k+X^T\beta_0)$. For tests corresponding to $F_0(a_k)>F(a_k)+\xi$, just let the testing function corresponding to $Z_i\geq a_k$. Now, using Lemma 10.1 and constructing a test by taking maximum value of all these tests, the proof is complete. □

Proof of Lemma 7.2. let β_l denote the l-th component of the vector β and $\beta - \beta_0$ belongs to Q_b, one of the quadrants. Now, consider $d2^d$ tests such as

$$\begin{aligned} \mathrm{H}_0 &: (F,\beta) = (F_0,\beta_0) \text{ against} \\ \mathrm{H}_{1l} &: \sup_{t\in\mathbb{R}^+} |F(t) - F_0(t)| < \Delta,\ |\beta_l - \beta_{0l}| > \delta, \beta_0 - \beta \in Q_b, \end{aligned} \tag{10.25}$$

for every $l = 1,\ldots,d$, and $b \in \{-1,1\}^d$. Find m such that $F_0(m) = 1/2$. For given δ, let $\delta_{F_c} = \min_{s\in[-L,L]}(F_c(s) - F_c(s-\delta))$, then find Δ such that

$$\frac{\bar{F}_c(m)}{\bar{F}_c(m) + \delta_{F_c}} \frac{1/2}{1/2 - \Delta} > 1 + \xi,$$

for some small $\xi > 0$. Now, choose i such that $X_i^T\beta_0 > 0$ and the l-th component of X_i, $X_{il} > \zeta > 0$. By Condition (i) of Theorem 7.1, we have that $\liminf n^{-1}\#\{i : X_i^T\beta_0 > 0 \text{ and } X_{il} > \zeta > 0\} > 0$. Using only the Z_i's with the selceted i's to construct the test (10.25), we have that

$$\mathrm{E}_\beta(\log Z > m) \geq (\frac{1}{2} - \Delta)\bar{F}_c(m) = \gamma_i \geq \alpha_i = (\frac{1}{2})\bar{F}_c(m+\delta) \geq \mathrm{E}_{\beta_0}(\log Z > m).$$

Using Lemma 10.1 and constructing a test by taking maximum value of all these tests, we finish the proof of the lemma. □

Lemma 10.6. *Let $\tilde{f}_0$ and $\tilde{f}_c$ be density functions and $\bar{\tilde{F}}_0$ be a survival function such that Conditions* (iii), (v), (vi) *and* (ix) *in Theorem* 7.1 *holds, then*

(i) $\lim_{\gamma\to 0} K(\tilde{f}_0, \tilde{f}(\cdot - \gamma)) = K(\tilde{f}_0, \tilde{f})$

(ii) $\lim_{\gamma\to 0} \int \tilde{f}_c(t) \log \frac{\bar{\tilde{F}}_0(t)}{\bar{\tilde{F}}(t-\gamma)} dt = \int \tilde{f}_c(t) \log \frac{\bar{\tilde{F}}_0(t)}{\bar{\tilde{F}}(t)} dt,$

Proof. Note that by Condition (v) and (ix), we have $\int e^{t/h}\tilde{f}_0(t)dt < \infty$ and $\int e^{t/h}\tilde{f}_c(t)dt < \infty$.

Since $\tilde{f}(t)$ is postive and continuous, and

$$|\log \tilde{f}(t-\gamma)| \leq |\log h| + \left|\log \int \exp(-e^{(t-\lambda-\gamma)/h} + (t-\lambda-\gamma)/h) dP(\lambda)\right|.$$

By Jensen's inequality , the last term in above expression is less than

$$\int [e^{(t-\lambda-\gamma)/h} - (t-\lambda-\gamma)/h] dP(\lambda).$$

The DCT now applies. For $\bar{\tilde{F}}$, the proof is similar. □

LEMMA 10.7. *Under the conditions of Theorem* 7.1, *assertion* (7.7) *holds.*

PROOF. To show that (7.7) holds, our first step is to show that there exist probability measures P_m such that

$$\int_{-\infty}^{\infty} \tilde{f}_c(t+\xi)\log\frac{\bar{\bar{F}}_0(t)}{\bar{\bar{F}}_{P_m}(t)}dt \to 0, \tag{10.26}$$

as $m\to\infty$ for any $\xi\in\mathbb{R}$. To this end, we define

$$\tilde{f}_m(x) = \begin{cases} t_m\tilde{f}_0(x), & \|x\|<m, \\ 0, & \text{otherwise}, \end{cases} \qquad m\geq 1,$$

where $t_m^{-1} = \int_{\|x\|<m}\tilde{f}_0(x)dx$, $h_m = m^{-\eta}$, $\tilde{F}_m$ is the probability measure corresponding to $\tilde{f}_m$, $P_m = \tilde{F}_m\times\delta(h_m)$, where $\delta(\cdot)$ is the degenerate distribution. Obviously, P_m is compactly supported.

Now let

$$\bar{\bar{F}}_m := t_m\int_{-m}^{m}\exp(-e^{(t-\lambda)/h_m})\tilde{f}_0(\lambda)d\lambda, \tag{10.27}$$

and $\chi(t) := \exp(-e^t)$.

Using the transformation $a = (x-\theta)/h_m$,

$$\begin{aligned}\bar{\bar{F}}_{P_m}(x) = \int\frac{1}{h_m^d}\chi\left(\frac{x-\theta}{h_m}\right)d\tilde{F}_m(\theta) &= t_m\int_{-m}^{m}\frac{1}{h_m^d}\chi\left(\frac{x-\theta}{h_m}\right)f_0(\theta)d\theta \\ &= \int_{(x-m/h_m)}^{(x+m/h_m)}\chi(a)f_0(x-ah_m)da.\end{aligned}$$

Since for any given a, $\chi(a)f_0(x-ah_m)\to\chi(a)f_0(x)$ as $h_m\to 0$ and f_0 is bounded, by the DCT, we obtain $\bar{\bar{F}}_{P_m}(t)\to\bar{\bar{F}}_0(t)$.

Now we need to bound $\log\frac{\bar{\bar{F}}_0(t)}{\bar{\bar{F}}_{P_m}(t)}$ and apply the DCT to show that (10.26) holds. Observe that

$$\begin{aligned}\bar{\bar{F}}_{P_m}(x) &= t_m\int_{\|\theta\|<m}\frac{1}{h_m^d}\chi\left(\frac{x-\theta}{h_m}\right)\tilde{f}_0(\theta)d\theta \\ &\leq Mt_m\int_{\|\theta\|<m}\frac{1}{h_m^d}\chi\left(\frac{x-\theta}{h_m}\right)d\theta \\ &\leq Mt_m\leq Mt_1.\end{aligned}$$

Hence, as $\log \frac{\bar{\tilde{F}}_0(x)}{Mt_1} < 0$,

$$\log \frac{\bar{\tilde{F}}_0(x)}{\bar{\tilde{F}}_{P_m}(x)} \geq \log \frac{\bar{\tilde{F}}_0(x)}{Mt_1}. \tag{10.28}$$

Also

$$\begin{aligned}
&\int_{-\infty}^{\infty} \tilde{f}_c(x) \log \frac{\bar{\tilde{F}}_0(x)}{\bar{\tilde{F}}_{P_m}(x)} dx \\
&= \int_{|x|<m} \tilde{f}_c(x) \log \frac{\bar{\tilde{F}}_0(x)}{\bar{\tilde{F}}_{P_m}(x)} dx + \int_{|x|\geq m} \tilde{f}_c(x) \log \frac{\bar{\tilde{F}}_0(x)}{\bar{\tilde{F}}_{P_m}(x)} dx.
\end{aligned}$$

Let $m > l_1$. Now, for $x > m$,

$$\begin{aligned}
\bar{\tilde{F}}_{P_m}(x) &= t_m \int_0^m \frac{1}{h_m} \chi\left(\frac{x-\theta}{h_m}\right) \tilde{f}_0(\theta) d\theta \\
&\geq t_m \int_0^m \frac{1}{h_m} \chi\left(\frac{x+m}{h_m}\right) \tilde{f}_0(\theta) d\theta \\
&= \frac{1}{h_m} \chi\left(\frac{x+m}{h_m}\right) t_m \int_0^m \tilde{f}_0(\theta) d\theta \\
&= \frac{1}{h_m} \chi\left(\frac{x+m}{h_m}\right) \\
&= m^\eta \chi\left(m^\eta x + m^{1+\eta}\right) \\
&\geq \|x\|^\eta \chi(2x^{1+\eta}).
\end{aligned} \tag{10.29}$$

The last inequality holds when $T \mapsto T^\eta \chi(T^\eta(x+T))$ is decreasing for $T > T_0 > 0$. This follows because, with $z = T^\eta x + T^{\eta+1}$,

$$\begin{aligned}
&\frac{d}{dT}\left\{\eta \log T + \log \chi\left(T^\eta x + T^{\eta+1}\right)\right\} \\
&= \frac{\eta}{T} + \frac{\chi'(z)}{\chi(z)}\left(\frac{\eta}{T} z + T^\eta\right) \\
&= \frac{\eta}{T}\left\{1 - e^z(z + \frac{T^{1+\eta}}{\eta})\right\} \leq 0.
\end{aligned}$$

For $x < -m$, we will have similar result.

For $|x| \le m$, let $\delta > 0$ be fixed, and $\phi_m^*(x) = \inf_{\|t-x\|<\delta h_m} \tilde{f}_0(t)$,

$$\begin{aligned}
\bar{\tilde{F}}_{P_m}(x) &= t_m \int_{-m}^{m} \frac{1}{h_m}\chi\left(\frac{x-\theta}{h_m}\right)\tilde{f}_0(\theta)d\theta \\
&\ge t_m \int_{\{|\theta|<m\}\cap\{|\theta-x|<\delta h_m\}} \frac{1}{h_m}\chi\left(\frac{x-\theta}{h_m}\right)\tilde{f}_0(\theta)d\theta \\
&\ge t_m\phi_m^*(x) \int_{\{|\theta|<m\}\cap\{|\theta-x|<\delta h_m\}} \frac{1}{h_m}\chi\left(\frac{x-\theta}{h_m}\right)d\theta \\
&= t_m\phi_m^*(x) \int_{\{|x-uh_m|\le m\}\cap\{|u|\le\delta\}} \chi(u)du \\
&\ge t_m\phi_m^*(x) \int_0^{\delta} \chi(u)du,
\end{aligned}$$

with the convention that $[a, b] = [b, a]$ if $b < a$. The last inequality holds because when $|x| \le m$,

$$\Big\{u : u \in [0, \delta]\Big\} \subset \Big\{u : |x/h_m - u| \le m/h_m \text{ and } |u| \le \delta\Big\}.$$

We have $t_m \ge 1$, $\phi_m^*(x) \ge \phi_1(x)$. Let

$$c := \min_{x\in\{-\delta,\,\delta\}} \Big(\int_{[0,x]} \chi(u)du\Big).$$

Then, $\bar{\tilde{F}}_{P_m}(x) \ge c\phi_1(x)$, for all $|x| < m$. For $0 < R < m$,

$$\bar{\tilde{F}}_{P_m}(x) \ge \begin{cases} c\phi_1(x), & |x| < R, \\ \min\Big\{|x|^{\eta}\chi(2|x|^{1+\eta}\frac{x}{|x|}), c\phi_1(x)\Big\}, & |x| \ge R. \end{cases}$$

$$\log\frac{\bar{\tilde{F}}_0(x)}{\bar{\tilde{F}}_{P_m}(x)} \le \xi(x) := \begin{cases} \log\frac{\bar{\tilde{F}}_0(x)}{c\phi_1(x)}, & |x| < R, \\ \max\left(\log\frac{\bar{\tilde{F}}_0(x)}{|x|^{\eta}\chi(2|x|^{1+\eta}\frac{x}{|x|})}, \log\frac{\bar{\tilde{F}}_0(x)}{c\phi_1(x)}\right), & |x| \ge R. \end{cases} \tag{10.30}$$

Combining (10.28) and (10.30), we obtain

$$\Big|\log\frac{\bar{\tilde{F}}_0(x)}{\bar{\tilde{F}}_{P_m}(x)}\Big| \le \max\Big(\xi(x), \Big|\log\frac{\bar{\tilde{F}}_0(x)}{Mt_1}\Big|\Big).$$

From Condition (iii) of Theorem 7.1,

$$\int \Big| \log \frac{\bar{\tilde{F}}_0(x)}{Mt_1} \Big| \tilde{f}_c(x)dx = \log Mt_1 - \int f_0(x) \log f_0(x) dx < \infty.$$

Now

$$\begin{aligned} \int \xi(x) \tilde{f}_c(x) dx &= \int_{|x|<R} \tilde{f}_c(x) \log \frac{\bar{\tilde{F}}_0(x)}{c\phi_1(x)} dx \\ &\quad + \int_{\|x\| \geq R} \tilde{f}_c(x) \max \Big(\log \frac{\bar{\tilde{F}}_0(x)}{|x|^\eta \chi(2|x|^{1+\eta})}, \log \frac{\bar{\tilde{F}}_0(x)}{c\phi_1(x)} \Big). \end{aligned}$$

Hence,

$$\begin{aligned} \int \xi(x) \tilde{f}_c(x) dx \leq \int \tilde{f}_c(x) \log \frac{\bar{\tilde{F}}_0(x)}{c\phi_1(x)} dx \\ + \int_{|x| \geq R, f_0(x) > |x|^\eta \chi(2|x|^\eta x)} \tilde{f}_c(x) \log \frac{\bar{\tilde{F}}_0(x)}{|x|^\eta \chi(2|x|^\eta x)} dx, \end{aligned}$$

since $\max(x_1, x_2) \leq x_1 + x_2^+$ if $x_1 \geq 0$. The first term on the r.h.s. of the above inequality is finite, by Condition (iv) of Theorem 7.1. By Conditions (iii), (v) and (ix) of Theorem 7.1, the second term is also finite. Thus $\int \tilde{f}_c(x) \log \frac{\bar{\tilde{F}}_0(x)}{\bar{\tilde{F}}_{P_m}(x)} dx \to 0$ as $m \to \infty$, i.e., (10.26) holds.

Now, With the continuity of $\chi((x-\theta)/h)$ as a function of h, the uniformly eqicontinuity of $\chi((x-\theta)/h)$ as a function of θ and the fact that $0 < \chi(\cdot) < 1$, (10.26) implies (7.7) by using the DCT. □

References

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

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

Basu, S. and Mukhopadhyay, S. (2000). Binary response regression with normal scale mixtures links, In *Generalized Linear Models: A Bayesian Perspective,* (D. K. Dey, S. K. Ghosh, and B. K. Mallick, eds.) New York, Marcel Dekker.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and adaptive estimation for semiparametric models.* Johns Hopkins University Press, Baltimore.

BLASI, P. D., PECCATI, G. and PRÜNSTER, I. (2009). Asymptiotics for posterior hazards. *Ann. Statist.* (to appear).

CANTOR, A. B. and KNAPP, R. G. (1985). A test of the equality of survival distributions based on paired observations from conditionally independent exponential distributions. *IEEE Transactions on Reliability*, **34** 342–346.

CASTILLO, I. (2008). A semiparametric Bernstein-von Mises theorem. Preprint.

CHOUDHURI, N., GHOSAL, S. and ROY, A. (2004). Bayesian estimation of the spectral density of a time series. *J. Amer. Statist. Assoc.* **99**, 1050-1059.

CHOUDHURI, N., GHOSAL, S. and ROY, A. (2005). Bayesian methods for function estimation. *Handbook of Statistics*, **25**, (D. K. Dey and C. R. Rao, eds.) 373–414.

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

DIACONIS, P. and FREEDMAN, D. (1986). On the consistency of Bayes estimates (with discussion). *Ann. Statist.* **14**, 1-67.

DOOB, J. L. (1948). Application of the theory of martingales. *Coll. Int. du CNRS*, Paris, 22–28.

DOSS, H. (1985). Bayesian nonparametric estimation of the median. II: Asymptotic properties of the estimates. *Ann. Statist.* **13**, 1445–1464.

FREEDMAN, D. (1963). On the asymptotic distribution of Bayes estimates in the discrete case I. *Ann. Math. Statist.* **34**, 1386–1403.

FREEDMAN, D. (1965). On the asymptotic distribution of Bayes estimates in the discrete case II. *Ann. Math. Statist.* **36**, 454–456.

GELFAND, A. E. and KUO, L. (1991). Nonparametric Bayesian bioassay including ordered polytomous response. *Biometrika* **78**, 657–666.

GHOSAL, S., GHOSH, J. K. and RAMAMOORTHI, R. V. (1999a). Consistency issues in Bayesian nonparametrics. In *Asymptotics, Nonparametrics and Time Series: A Tribute to Madan Lal Puri.* (S. Ghosh ed.,) Marcel Dekker, New York, 639–668.

GHOSAL, S., GHOSH, J. K. and RAMAMOORTHI, R. V. (1999b). Posterior consistency of Dirichlet mixtures in density estimation. *Ann. Statist.* **27** 143–158.

GHOSAL, S. (2009). Dirichlet process, related priors and posterior asymptotics. In *Bayesian Nonparametrics in Practice* (Hjort, N. L., *et al.*, eds.) Cambridge University Press (to appear).

GHOSH, J. K. and RAMAMOORTHI, R. V. (2003). *Bayesian Nonparametrics.* Springer–Verlag, New York.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13–30.

KIM, Y. (2006). The Bernstein-von Mises theorem for the proportional hazard model. *Ann. Statist.* **34**, 1678–1700.

MALLICK, B. K. and GELFAND, A. E. (1994). Generalized linear models with unknown link functions. *Biometrika* **81** 237–245.

MASSEY, F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *J. Amer. Statist. Assoc.*, **46** 68–78.

NEWTON, M. A., CZADO, C. and CHAPPELL, R. (1996). Bayesian inference for semiparametric binary regression. *J. Amer. Statist. Assoc.* **91**, 142–153.

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

Shen, X. (2002). Asymptotic normality of semiparametric and nonparametric posterior distributions. *J. Amer. Statist. Assoc.* **97**, 222–235.

Tokdar, S. (2006). Posterior consistency of Dirichlet location-scale mixture of normals in density estimation and regression. *Sankhya* **67**, 90–110.

van der Vaart, A. W. and Wellner, J. A. (1996). Weak Convergence and Empirical Processes with Applications to Statistics. Springer–Verlag, New York.

Walker, S. G. (2004). New approaches to Bayesian consistency. *Ann. Statist.* **32**, 2028–2043.

Wu, Y. and Ghosal, S. (2008). Kullback Leibler property of kernel mixture priors in Bayesian density estimation. *Electronic J. Statist.* **2**, 298–331.

Yuefeng Wu
Department of Statistics
North Carolina State University
USA
E-mail: ywu9@ncsu.edu

Subhashis Ghosal
Department of Statistics
North Carolina State University
USA
E-mail: sghoshal@stat.ncsu.edu

Paper received ; revised June 2009.