Posterior consistency for some Bayesian Semi-parametric Problems

Yuefeng Wu

Department of Biological Statistics and Computational Biology
Cornell University
Joint work with Dr. Subhashis Ghosal

July 31, 2009

Bayesian Nonparametric and Semi-parametric Models

- Characteristics of the models

Bayesian Nonparametric and Semi-parametric Models

- Characteristics of the models
  1. Having at least one parameter in the form of function.

Bayesian Nonparametric and Semi-parametric Models

- Characteristics of the models
  1. Having at least one parameter in the form of function.
  2. Prior distributions on parameter space including the function spaces.

Bayesian Nonparametric and Semi-parametric Models

- Characteristics of the models
  1. Having at least one parameter in the form of function.
  2. Prior distributions on parameter space including the function spaces.
- Advantage of the models

Bayesian Nonparametric and Semi-parametric Models

- Characteristics of the models
  1. Having at least one parameter in the form of function.
  2. Prior distributions on parameter space including the function spaces.
- Advantage of the models
  1. More flexible than parametric models.

Bayesian Nonparametric and Semi-parametric Models

- Characteristics of the models
  1. Having at least one parameter in the form of function.
  2. Prior distributions on parameter space including the function spaces.
- Advantage of the models
  1. More flexible than parametric models.
  2. Advantage of the Bayesian approach

Bayesian Nonparametric and Semi-parametric Models

- Characteristics of the models
  1. Having at least one parameter in the form of function.
  2. Prior distributions on parameter space including the function spaces.
- Advantage of the models
  1. More flexible than parametric models.
  2. Advantage of the Bayesian approach
     1. Reflecting one's prior beliefs into the analysis.

Bayesian Nonparametric and Semi-parametric Models

- Characteristics of the models
  1. Having at least one parameter in the form of function.
  2. Prior distributions on parameter space including the function spaces.
- Advantage of the models
  1. More flexible than parametric models.
  2. Advantage of the Bayesian approach
     1. Reflecting one's prior beliefs into the analysis.
     2. Straightforward in principle where inference is based on the posterior distribution only.

Priors on Infinite-dimensional Spaces

- Dirichlet process prior: $\mathscr{D}_\alpha$. (Ferguson (1973))

Priors on Infinite-dimensional Spaces

- Dirichlet process prior: $\mathscr{D}_\alpha$. (Ferguson (1973))
  - Let $\mathscr{X}$ be a sample space, $\mathscr{B}$ be the collection of all Borel sets on $\mathscr{X}$. For any $B \in \mathscr{B}$, $P(B)$ is a measurable random variable taking value in $[0, 1]$;

Priors on Infinite-dimensional Spaces

- Dirichlet process prior: $\mathscr{D}_\alpha$. (Ferguson (1973))
  - Let $\mathscr{X}$ be a sample space, $\mathscr{B}$ be the collection of all Borel sets on $\mathscr{X}$. For any $B \in \mathscr{B}$, $P(B)$ is a measurable random variable taking value in $[0, 1]$;
  - Each realization of P is a probability measure on $(\mathscr{X}, \mathscr{B})$;

Priors on Infinite-dimensional Spaces

- Dirichlet process prior: $\mathscr{D}_\alpha$. (Ferguson (1973))
  - Let $\mathscr{X}$ be a sample space, $\mathscr{B}$ be the collection of all Borel sets on $\mathscr{X}$. For any $B \in \mathscr{B}$, $P(B)$ is a measurable random variable taking value in $[0, 1]$;
  - Each realization of P is a probability measure on $(\mathscr{X}, \mathscr{B})$;
  - For each measurable finite partition $\{B_1, \ldots, B_k\}$ of $\mathscr{X}$, the joint distribution of the vector $(P(B_1), \ldots, P(B_k))$ has Dirichlet distribution on the k-dimensional uni simplex with parameters $(k; \alpha(B_1), \ldots, \alpha(B_k))$.

Priors on Infinite-dimensional Spaces

- Dirichlet mixture prior (Ferguson (1983), Lo (1984)).

Priors on Infinite-dimensional Spaces

- Dirichlet mixture prior (Ferguson (1983), Lo (1984)).
  - $K(x,\theta)$ be the kernel function on $\mathscr{X} \times \Theta$, $\mathscr{X}$: sample space, Θ: parameter space;

Priors on Infinite-dimensional Spaces

- Dirichlet mixture prior (Ferguson (1983), Lo (1984)).
  - $K(x,\theta)$ be the kernel function on $\mathscr{X} \times \Theta$, $\mathscr{X}$: sample space, Θ: parameter space;
  - $P \sim \mathscr{D}_\alpha$: Random mixing distribution on Θ;

Priors on Infinite-dimensional Spaces

- Dirichlet mixture prior (Ferguson (1983), Lo (1984)).
  - $K(x, \theta)$ be the kernel function on $\mathscr{X} \times \Theta$, $\mathscr{X}$: sample space, Θ: parameter space;
  - $P \sim \mathscr{D}_\alpha$: Random mixing distribution on Θ;
  - $\mathscr{D}_\alpha$: a prior on $\mathscr{M}(\Theta)$, the space of probability measures on Θ

Priors on Infinite-dimensional Spaces

- Dirichlet mixture prior (Ferguson (1983), Lo (1984)).
  - $K(x, \theta)$ be the kernel function on $\mathscr{X} \times \Theta$, $\mathscr{X}$: sample space, Θ: parameter space;
  - $P \sim \mathscr{D}_\alpha$: Random mixing distribution on Θ;
  - $\mathscr{D}_\alpha$: a prior on $\mathscr{M}(\Theta)$, the space of probability measures on Θ
  - Via the map $P \mapsto f_P(x) := \int K(x; \theta) dP(\theta)$, there is a prior on $\mathscr{D}(\mathfrak{X})$, the space of densities on $\mathfrak{X}$.

Priors on Infinite-dimensional Spaces

- Mixture of Dirichlet process (Antoniak (1974)).

Priors on Infinite-dimensional Spaces

- Mixture of Dirichlet process (Antoniak (1974)).
- Invariant Dirichlet process, Pinned-down Dirichlet (Dalal (1979), Freedman (1986a, b)).

Priors on Infinite-dimensional Spaces

- Mixture of Dirichlet process (Antoniak (1974)).
- Invariant Dirichlet process, Pinned-down Dirichlet (Dalal (1979), Freedman (1986a, b)).
- Generalization of the Dirichlet process,

Priors on Infinite-dimensional Spaces

- Mixture of Dirichlet process (Antoniak (1974)).
- Invariant Dirichlet process, Pinned-down Dirichlet (Dalal (1979), Freedman (1986a, b)).
- Generalization of the Dirichlet process,
  - Tail-free and neutral to the right process (Freedman (1963))

Priors on Infinite-dimensional Spaces

- Mixture of Dirichlet process (Antoniak (1974)).
- Invariant Dirichlet process, Pinned-down Dirichlet (Dalal (1979), Freedman (1986a, b)).
- Generalization of the Dirichlet process,
  - Tail-free and neutral to the right process (Freedman (1963))
  - Polya tree process (Ferguson (1974), Blackwell and MacQueen (1973), Mauldin *et al.* (1992) and Lavine (1992, 1994))

Priors on Infinite-dimensional Spaces

- Mixture of Dirichlet process (Antoniak (1974)).
- Invariant Dirichlet process, Pinned-down Dirichlet (Dalal (1979), Freedman (1986a, b)).
- Generalization of the Dirichlet process,
  - Tail-free and neutral to the right process (Freedman (1963))
  - Polya tree process (Ferguson (1974), Blackwell and MacQueen (1973), Mauldin *et al.* (1992) and Lavine (1992, 1994))
  - Generalized Dirichlet procesess (k-dimensional) (Hjort (1996)).

Priors on Infinite-dimensional Spaces

- Mixture of Dirichlet process (Antoniak (1974)).
- Invariant Dirichlet process, Pinned-down Dirichlet (Dalal (1979), Freedman (1986a, b)).
- Generalization of the Dirichlet process,
  - Tail-free and neutral to the right process (Freedman (1963))
  - Polya tree process (Ferguson (1974), Blackwell and MacQueen (1973), Mauldin *et al.* (1992) and Lavine (1992, 1994))
  - Generalized Dirichlet procesess (k-dimensional) (Hjort (1996)).
  - Priors obtained from random series representation (Sethuraman (1994)).

Priors on Infinite-dimensional Spaces

- Mixture of Dirichlet process (Antoniak (1974)).
- Invariant Dirichlet process, Pinned-down Dirichlet (Dalal (1979), Freedman (1986a, b)).
- Generalization of the Dirichlet process,
  - Tail-free and neutral to the right process (Freedman (1963))
  - Polya tree process (Ferguson (1974), Blackwell and MacQueen (1973), Mauldin *et al.* (1992) and Lavine (1992, 1994))
  - Generalized Dirichlet procesess (k-dimensional) (Hjort (1996)).
  - Priors obtained from random series representation (Sethuraman (1994)).
- Gaussian process Lèvy process (Leonard (1978), Lenk (1988,1991), Ferguson and Phadia (1979) and Lo (1982)).

Consistency

- Purpose

Consistency

- Purpose
  - Check the validation of the Bayesian models.

Consistency

- Purpose
  - Check the validation of the Bayesian models.
- Meaning

Consistency

- Purpose
  - Check the validation of the Bayesian models.
- Meaning
  - Bayesian's assessment to the unknown parameter should be close to the true value.

Consistency

- Purpose
  - Check the validation of the Bayesian models.
- Meaning
  - Bayesian's assessment to the unknown parameter should be close to the true value.
  - Robustness with respect to the choice of the prior.

Consistency

- Purpose
  - Check the validation of the Bayesian models.
- Meaning
  - Bayesian's assessment to the unknown parameter should be close to the true value.
  - Robustness with respect to the choice of the prior.

Definition

For each n, let $\Pi(\cdot|\mathbb{X}_n)$ be a posterior given $X_1, \ldots, X_n$. The sequence $\{\Pi(\cdot|\mathbb{X}_n)\}$ is said to be consistent at θ_0 if there is a $\Omega_0 \subset \Omega$ with $P_{\theta_0}^{\infty}(\Omega_0) = 1$ such that if $\omega \in \Omega_0$, then for every neighborhood U of θ_0,

$$\Pi(U|\mathbb{X}_n(\omega)) \to 1. \tag{1}$$

Some Basic Results about Posterior Consistency

- For finite dimensional cases, having positive prior probability in a neighborhood of the true value of the parameter implies consistency at that value.

Some Basic Results about Posterior Consistency

- For finite dimensional cases, having positive prior probability in a neighborhood of the true value of the parameter implies consistency at that value.
- Doob (1948): The prior probability of the set of all $\theta \in \Theta$ at which consistency does not hold, is 0. This result does not imply consistency at a given point if there is no prior point mass there.

Some Basic Results about Posterior Consistency

- For finite dimensional cases, having positive prior probability in a neighborhood of the true value of the parameter implies consistency at that value.
- Doob (1948): The prior probability of the set of all $\theta \in \Theta$ at which consistency does not hold, is 0. This result does not imply consistency at a given point if there is no prior point mass there.
- Freedman (1963, 1986): Only having positive prior probability in a neighborhood of the true value of the parameter does not imply consistency at that value.

Basic Theorem about Posterior Consistency

- Schwartz (1965)

Theorem

Let $\theta_0 \in U \subset \Theta$. *If there exists* $m \geq 1$, *a test function* $\phi(\mathbb{X}_m)$ *for testing* $H_0 : \theta = \theta_0$ *against* $H_a : \theta \in U^c$ *with the property that* $\inf\{E_\theta \phi(\mathbb{X}_m) : \theta \in U^c\} > E_{\theta_0} \phi(\mathbb{X}_m)$ *and* $\theta_0 \in KL(\Pi)$, *then* $\Pi\{\theta \in U^c | \mathbb{X}_n\} \to 0$ *a.s.* $[P_{\theta_0}^\infty]$.

Basic Theorem about Posterior Consistency

- Schwartz (1965)

Theorem

Let $\theta_0 \in U \subset \Theta$. If there exists $m \geq 1$, a test function $\phi(\mathbb{X}_m)$ for testing $H_0 : \theta = \theta_0$ against $H_a : \theta \in U^c$ with the property that $\inf\{E_\theta \phi(\mathbb{X}_m) : \theta \in U^c\} > E_{\theta_0}\phi(\mathbb{X}_m)$ and $\theta_0 \in KL(\Pi)$, then $\Pi\{\theta \in U^c | \mathbb{X}_n\} \to 0$ a.s. $[P_{\theta_0}^\infty]$.

- Test condition

Basic Theorem about Posterior Consistency

- Schwartz (1965)

Theorem

Let $\theta_0 \in U \subset \Theta$. If there exists $m \geq 1$, a test function $\phi(\mathbb{X}_m)$ for testing $H_0 : \theta = \theta_0$ against $H_a : \theta \in U^c$ with the property that $\inf\{E_\theta \phi(\mathbb{X}_m) : \theta \in U^c\} > E_{\theta_0}\phi(\mathbb{X}_m)$ and $\theta_0 \in KL(\Pi)$, then $\Pi\{\theta \in U^c | \mathbb{X}_n\} \to 0$ a.s. $[P_{\theta_0}^{\infty}]$.

- Test condition
- KL property

Kullback Leibler Property

- Let f be given a prior Π^*.

Kullback Leibler Property

- Let f be given a prior Π^*.
- Kullback-Leibler divergence between f and g: $\mathcal{K}(f, g) = \int f \log(f/g)$.

Kullback Leibler Property

- Let f be given a prior Π^*.
- Kullback-Leibler divergence between f and g: $\mathcal{K}(f, g) = \int f \log(f/g)$.
- Kullback-leibler neighborhood of f of size ϵ: $\mathscr{K}_\epsilon(f) = \{g : \mathcal{K}(f, g) < \epsilon\}$.

Kullback Leibler Property

- Let f be given a prior Π^*.
- Kullback-Leibler divergence between f and g: $\mathcal{K}(f, g) = \int f \log(f/g)$.
- Kullback-leibler neighborhood of f of size ϵ: $\mathscr{K}_\epsilon(f) = \{g : \mathcal{K}(f, g) < \epsilon\}$.
- We say that the KL property holds at f_0 or f_0 is in the Kullback-Leibler support of Π^* and write $f_0 \in KL(\Pi^*)$, if $\Pi^*(\mathscr{K}_\epsilon(f_0)) > 0$ for every $\epsilon > 0$.

Studies on Semi-parametric Models

- Examples from literature

Studies on Semi-parametric Models

- Examples from literature
  - Bickel *et al.* (1988) and van der Vaart and Wellner (1996): Consistent and efficient methods of estimation of the parameters, estimation of the finite dimensional part at the $\sqrt{n}$ rate.

Studies on Semi-parametric Models

- Examples from literature
  - Bickel *et al.* (1988) and van der Vaart and Wellner (1996): Consistent and efficient methods of estimation of the parameters, estimation of the finite dimensional part at the $\sqrt{n}$ rate.
  - Kim (2006): For the Cox proportional hazard model, 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.

Studies on Semi-parametric Models

- Examples from literature
  - Bickel *et al.* (1988) and van der Vaart and Wellner (1996): Consistent and efficient methods of estimation of the parameters, estimation of the finite dimensional part at the $\sqrt{n}$ rate.
  - Kim (2006): For the Cox proportional hazard model, 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.
- My work:
  Studying posterior consistency for various semi-parametric models.

General Model Notations

- $a(\cdot)$: the infinite dimensional parameter; θ: the finite dimensional parameter; $\Xi = \mathscr{A} \times \Theta$: the parameter space.

General Model Notations

- $a(\cdot)$: the infinite dimensional parameter; θ: the finite dimensional parameter; $\Xi = \mathscr{A} \times \Theta$: the parameter space.
- Prior: $\tilde{\Pi}$ on $\mathscr{A}$, μ on Θ and $\Pi = \tilde{\Pi} \times \mu$.

General Model Notations

- $a(\cdot)$: the infinite dimensional parameter; θ: the finite dimensional parameter; $\Xi = \mathscr{A} \times \Theta$: the parameter space.
- Prior: $\tilde{\Pi}$ on $\mathscr{A}$, μ on Θ and $\Pi = \tilde{\Pi} \times \mu$.
- Y_i, $i = 1, 2, \ldots$: the observations of the random variable of primary interest; X_i's (or x_i's): the corresponding covariates.

General Model Notations

- $a(\cdot)$: the infinite dimensional parameter; θ: the finite dimensional parameter; $\Xi = \mathscr{A} \times \Theta$: the parameter space.
- Prior: $\tilde{\Pi}$ on $\mathscr{A}$, μ on Θ and $\Pi = \tilde{\Pi} \times \mu$.
- Y_i, $i = 1, 2, \ldots$: the observations of the random variable of primary interest; X_i's (or x_i's): the corresponding covariates.
- $f_{a,\theta}$ (or $f_{a,\theta,i}$): the density of $(Y|X)$.

General Model Notations

- $a(\cdot)$: the infinite dimensional parameter; θ: the finite dimensional parameter; $\Xi = \mathscr{A} \times \Theta$: the parameter space.
- Prior: $\tilde{\Pi}$ on $\mathscr{A}$, μ on Θ and $\Pi = \tilde{\Pi} \times \mu$.
- Y_i, $i = 1, 2, \ldots$: the observations of the random variable of primary interest; X_i's (or x_i's): the corresponding covariates.
- $f_{a,\theta}$ (or $f_{a,\theta,i}$): the density of $(Y|X)$.
- $a_0(\cdot)$ and θ_0: the true values of the parameters. Let $f_0 = f_{a_0,\theta_0}$ and $f_{0i} = f_{a_0,\theta_0,i}$.

Main Theorem

Theorem

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 testing: $(a_0(\cdot), \theta_0)$ v.s. $\mathscr{W} \cap \left(\mathscr{A}_n \times [m_n, M_n]^d\right)$;*

(ii) *Kullback Leibler property holds;*

(iii) *the prior measure of the outside of the sieve is exponentially small;*

then with $(P_{f_0}^{\infty})$-probability 1, the model is consist on $(a_0(\cdot), \theta_0)$, i.e.

$$\Pi(\mathscr{W} | (X_1, Y_1), \ldots, (X_n, Y_n)) \to 0. \quad (2)$$

Exponential consistency:
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
$\inf_{(a(\cdot),\theta)\in \mathscr{W}} \int [\mathrm{E}_{f^n_{a,\theta,}}(\Phi_n)] Q^n(dx) \geq 1 - C_2 e^{-nC}$.

Consistency of Multiple Regression

Generalization of Theorem 4.1 in Amewou-Atiss *et al.* (2003).

Theorem

Suppose that

(i) *the covariate* X *is compactly supported, and for any quadrant* Γ, $Q(\Gamma \backslash \{|X_i| < \zeta\}) > 0$ *for each* $i = 1, \ldots, d$, *and some* $\zeta > 0$;

(ii) f_0 *is continuous,* $f_0(0) > 0$, *and* f_0 *is covered by some other function: for* $|\eta'| < \eta$, $f_0(y - \eta') < C_\eta g_\eta(y)$ *for all* y;

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

Then for any weak neighborhood $\mathscr{U}$ *of* f_0, $\Pi\Big\{f \in \mathscr{U}, |\alpha - \alpha_0| < \epsilon, \|\beta - \beta_0\| < \epsilon \Big| (X_1, Y_1), \ldots, (X_n, Y_n)\Big\} \to 1$ *a.s.* $P_{f_0}^{\infty}$.

Exponential Frailty Model

- Let $(X_1, Y_1), (X_2, Y_2), \ldots$ be paired observations.

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

where $\mathrm{Exp}(w)$ is corresponding to p.d.f. we^{-wx}.
For given $W = w$, X and Y are conditionally independent, $\tilde{\Pi}$ and μ are independent.

Exponential Frailty Model

- Let $(X_1, Y_1), (X_2, Y_2), \ldots$ be paired observations.

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

  where $\mathrm{Exp}(w)$ is corresponding to p.d.f. we^{-wx}.
  For given $W = w$, X and Y are conditionally independent, $\tilde{\Pi}$ and μ are independent.

- The likelihood function of observations:
  $f(x, y) = \int \lambda w^2 e^{-w(x+\lambda y)} dF(w)$

Consistency of Exponential Frailty Model

Theorem

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.* $\prod_{i=1}^{\infty} P_{f_{0i}}$.

Other studies about this model: Cantor *et al.* (1985) and Owen *et al.* (2000).

Cox Proportional Hazard Model

- Y_i's: survival times with i.i.d. covariates $X_i \sim Q$.

Cox Proportional Hazard Model

- Y_i's: survival times with i.i.d. covariates $X_i \sim Q$.
- $h(t)\exp(X_i^T\beta)$: the random hazard rate function of Y.

Cox Proportional Hazard Model

- Y_i's: survival times with i.i.d. covariates $X_i \sim Q$.
- $h(t)\exp(X_i^T\beta)$: the random hazard rate function of Y.
- Priors: $\beta \sim \mu$, $h \sim \tilde{\Pi}$, $\Pi = \mu \times \tilde{\Pi}$.

Cox Proportional Hazard Model

- Y_i's: survival times with i.i.d. covariates $X_i \sim Q$.
- $h(t)\exp(X_i^T\beta)$: the random hazard rate function of Y.
- Priors: $\beta \sim \mu$, $h \sim \tilde{\Pi}$, $\Pi = \mu \times \tilde{\Pi}$.
- $f(t) = h(t)\exp(-H(t)) = h(t)\bar{F}(t)$: baseline random density function for life time Y.

Cox Proportional Hazard Model

- Y_i's: survival times with i.i.d. covariates $X_i \sim Q$.
- $h(t)\exp(X_i^T\beta)$: the random hazard rate function of Y.
- Priors: $\beta \sim \mu$, $h \sim \tilde{\Pi}$, $\Pi = \mu \times \tilde{\Pi}$.
- $f(t) = h(t)\exp(-H(t)) = h(t)\bar{F}(t)$: baseline random density function for life time Y.
- Censoring: (Z, Δ), where $Z = Y \wedge C$, $\Delta = 1\!\!1(Y \leq C)$ for C a censoring time with distribution F_c.

Consistency of Cox Proportional Hazard Model

Theorem

If

(a1) $f_0(t)$ *is strictly positive on* $(0, \infty)$ *and* $\int_{\mathbb{R}^+} \max\{\mathrm{E}[H(t)^q], t^q\} f_0(t)dt < \infty$ *for any* $q > 1$*;*

(a2) *there exists* $r > 0$ *such that* $\liminf_{t \downarrow 0} h(t)/t^r = \infty$*, a.s.;*

(a3) $\{h : \sup_{0<t<T} |h(t) - h_0(t)| < \delta, \int_T^\infty |H(t) - H_0(t)| f_0(t)dt < \delta\}$ *has positive prior probability for some finite* T *and* $\delta > 0$*; (Blasi et al. (2009))*

(a4) $\int_0^\infty \log(f_0(t))^q f_0(t)dt < \infty$*.*

Then, for $\mathscr{W} = \{(h, \beta) : |\beta - \beta_0| < \epsilon, h : 1 - \exp(-H(t)) \in \mathscr{U}_{F_0}\}^c$*, we have that* $\Pi(\mathscr{W} | (X_1, Y_1), \ldots, (X_n, Y_n)) \to 0$ *a.s.* $[P^\infty_{\beta_0, h_0}]$*.*

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.

The Model

- Y_i's: observations; $X_i \sim Q$, i.i.d.: predictors, compactly supported;

The Model

- Y_i's: observations; $X_i \sim Q$, i.i.d.: predictors, compactly supported;
- Relation: $\mu_i = g(X_i^T \beta)$, g is random;

The Model

- Y_i's: observations; $X_i \sim Q$, i.i.d.: predictors, compactly supported;
- Relation: $\mu_i = g(X_i^T \beta)$, g is random;
- T: known, strictly increasing differentiable function from $\mathscr{M}$ (the space of *mu*) to $[0, 1]$;

The Model

- Y_i's: observations; $X_i \sim Q$, i.i.d.: predictors, compactly supported;
- Relation: $\mu_i = g(X_i^T \beta)$, g is random;
- T: known, strictly increasing differentiable function from $\mathscr{M}$ (the space of mu) to $[0, 1]$;
- $F(X_i^T \beta) := T(g(X_i^T \beta))$;

The Model

- Y_i's: observations; $X_i \sim Q$, i.i.d.: predictors, compactly supported;
- Relation: $\mu_i = g(X_i^T \beta)$, g is random;
- T: known, strictly increasing differentiable function from $\mathscr{M}$ (the space of mu) to $[0, 1]$;
- $F(X_i^T \beta) := T(g(X_i^T \beta))$;
- Prior: $F \sim \tilde{\Pi}$, $\beta \sim \mu$;

The Model

- Y_i's: observations; $X_i \sim Q$, i.i.d.: predictors, compactly supported;
- Relation: $\mu_i = g(X_i^T \beta)$, g is random;
- T: known, strictly increasing differentiable function from $\mathscr{M}$ (the space of mu) to $[0, 1]$;
- $F(X_i^T \beta) := T(g(X_i^T \beta))$;
- Prior: $F \sim \tilde{\Pi}$, $\beta \sim \mu$;
- $\|\beta\| = 1$.

Consistency of Generalized Linear Models with Unknown Link Functions

Theorem

The covariate X is compactly supported, and for any quadrant Γ, $Q(\Gamma \backslash \{X : |X_i| < \zeta\}) > 0$ for each $i = 1, \ldots, d$, and some $\zeta > 0$; Assume that the weak support of $\tilde{\Pi}$ covers $\mathscr{M}([-L, L])$. If $\pi(B_n^c) \leq c_1 e^{-nc_2}$ 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 $(\mathscr{U}_{F_0} \times \{\beta : \|\beta - \beta_0\| < \delta\})^c$ and $\mathscr{U}_{F_0}$ is the weak neighborhood of F_0.

Accelerated Failure Time Model

- For each subject $i = 1, \dots, n$, 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 $\Delta_i = 1\!\!1(Y_i \leq C_i)$. Let X_i denote a d-dimensional vector of covariates associated with subject i.

Accelerated Failure Time Model

- 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.
- Assume $Y/e^{X^T\beta} \sim f(y)$.

Accelerated Failure Time Model

- 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.
- Assume $Y/e^{X^T\beta} \sim f(y)$.
- Observe that $\epsilon := \log Y - X^T\beta \sim \tilde{f}$, where $\tilde{f}(t) := e^t f(e^t)$. Note that $f \mapsto \tilde{f}$ is one to one.

Accelerated Failure Time Model

- 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 $\Delta_i = 1\!\!1(Y_i \leq C_i)$. Let X_i denote a d-dimensional vector of covariates associated with subject i.
- Assume $Y/e^{X^T\beta} \sim f(y)$.
- Observe that $\epsilon := \log Y - X^T\beta \sim \tilde{f}$, where $\tilde{f}(t) := e^t f(e^t)$. Note that $f \mapsto \tilde{f}$ is one to one.
- $\int x dQ(x) = 0$.

Prior for AFT

- f is given a weibull kernel Dirichlet mixture prior.

Prior for AFT

- f is given a weibull kernel Dirichlet mixture prior.
- Weibull: $w(x;\theta,\phi)=\phi^{-1}\theta(x/\phi)^{\theta-1}\exp[-(x/\phi)^{\theta}]$.

Prior for AFT

- f is given a weibull kernel Dirichlet mixture prior.
- Weibull: $w(x;\theta,\phi)=\phi^{-1}\theta(x/\phi)^{\theta-1}\exp[-(x/\phi)^{\theta}]$.
- $\log\phi|P\sim P$, $P\sim\tilde{\Pi}$, and $\theta\sim\nu$.

Prior for AFT

- f is given a weibull kernel Dirichlet mixture prior.
- Weibull: $w(x;\theta,\phi) = \phi^{-1}\theta(x/\phi)^{\theta-1}\exp[-(x/\phi)^{\theta}]$.
- $\log\phi|P \sim P$, $P \sim \tilde{\Pi}$, and $\theta \sim \nu$.
- $\beta \sim \mu$.

Prior for AFT

- f is given a weibull kernel Dirichlet mixture prior.
- Weibull: $w(x;\theta,\phi) = \phi^{-1}\theta(x/\phi)^{\theta-1}\exp[-(x/\phi)^{\theta}]$.
- $\log\phi|P \sim P$, $P \sim \tilde{\Pi}$, and $\theta \sim \nu$.
- $\beta \sim \mu$.
- $\tilde{\Pi} \times \nu \times \mu$ nduces a prior on the density function $\tilde{f}$ of $\log Y - X^T\beta$ via the map

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

Sufficient Conditions for Consistency of AFT

1. the covariate X is compactly supported, and for any quadrant Γ, $Q(\Gamma \backslash \{\|x_i\| < \zeta\}) > 0$, for all $i = 1, \ldots, d$ and some $\zeta > 0$;
2. for some $0 < M < \infty$, $0 < \tilde{f}_0(t) < M$ for all t;
3. $\int \tilde{f}_0(t) |\log \tilde{f}_0(t)| dt < \infty$ and $\int \tilde{f}_c(t+\xi) |\log \bar{\tilde{F}}_0(t)| dt < \infty$ for all $\xi \in \mathbb{R}$;
4. for some $\delta > 0$, $\int \tilde{f}_0(t) \log \frac{\tilde{f}_0(t)}{\phi_\delta(t)} dt < \infty$;
5. there exists $\eta > 0$ such that $|\int e^{2t|t|^\eta} \tilde{f}_0(t) dt| < \infty$, and $\int \tilde{f}_0(t) e^{(t-a)/b} dt < \infty$ for all $a \in \mathbb{R}$ and $b \in (0, \infty)$;
6. $\int (e^{-t/h} + t/h) dP(t) < \infty$ for any given $h > 0$, a.s. $\tilde{\Pi}$;
7. the weak support of $\tilde{\Pi}$ is the space of all probability measures on $\mathbb{R}$;
8. for some $\delta > 0$ and any $\xi \in \mathbb{R}$, $\int \tilde{f}_c(t+\xi) \log \frac{\bar{\tilde{F}}_{0,\beta_0}(t)}{\phi_\delta(t)} dt < \infty$;
9. there exists $\eta > 0$ such that $|\int \tilde{f}_c(t+\xi) e^{2t|t|^\eta} dt| < \infty$ for any $\xi \in \mathbb{R}$, and $\int \tilde{f}_c(t) e^{(t-a)/b} dt < \infty$ for all $a \in \mathbb{R}$ and $b \in (0, \infty)$;
10. $\int e^{t/h} \tilde{f}_0(t) dt < \infty$ and $\int e^{t/h} \tilde{f}_c(t) dt < \infty$ for any given $h > 0$.

Partial Linear Regression

- Model: $Y_i = X_i^T\beta + f(Z_i) + \epsilon_i$,
- ϵ_i's are i.i.d. $\mathrm{Norm}(0, \sigma^2)$,
- f is an odd function, random.
- Prior: $\beta \sim \mu$ $q_{f,z} * \phi(0, \sigma^2) \sim \tilde{\Pi}$, $\Pi = \tilde{\Pi} \times \mu$.

Theorem

Suppose that

(1) *X and Z are compactly supported, and distrbuted well;*

(2) *Cover: $q_{f_0,z} * \phi(0, \sigma_0^2)(y - \eta') < C_\eta g_\eta(y)$;*

(3) *KL property: $\tilde{\Pi}\{K(g_\eta, f) < \delta\} > 0$.*

Then $\Pi\{(f, \beta) : f \in \mathscr{U}, \|\beta - \beta_0\| < \delta | (X_1, Y_1, Z_1), \ldots (X_n, Y_n, Z_n)\} \to 1$ a.s. $\left[P_{f_0}^\infty\right]$, for any weak neighborhood $\mathscr{U}$ of f_0,.

The End