A Theoretical Framework for Bayesian Nonparametric Regression: Orthonormal Random Series and Rates of Contraction

Fangzheng Xie (fxie5@jhu.edu), Wei Jin (wjin@jhu.edu), and Yanxun Xu (yanxun.xu@jhu.edu)
Department of Applied Mathematics and Statistics, Johns Hopkins University

1. Overview

- A general framework to study rates of contraction w.r.t. to $\|\cdot\|_2$ for Bayesian nonparametric regression
- Key features:
  - Flexibility: Orthonormal series;
  - Convenience: Drop L_∞-bound.
- Applications:
  - Finite random series prior;
  - Block prior w/o truncation;
  - SE-GP w/ fixed design.
- Extension: sparse additive models in high dimensions.

2. Background: G-VDV Method

Sufficient conditions for $\Pi(d(\theta,\theta_0) > M\epsilon_n \mid \text{data}) = o_{\mathbb{P}_0}(1)$:

1. The prior concentration condition: $\Pi\left(B_{\mathrm{KL}}(p_0,\epsilon_n)\right) \geq \mathrm{e}^{-Dn\epsilon_n^2}$.
2. Existence of sieves $(\Theta_n)_{n=1}^\infty$ s.t. $\Pi(\Theta_n^c) \leq \mathrm{e}^{-(D+4)n\epsilon_n^2}$.
3. Existence of tests $(\phi_n)_{n=1}^\infty$ s.t.

$$\mathbb{E}_0\phi_n \to 0,$$
$$\sup_{\{d(\theta,\theta_0)>M\epsilon_n\}} \mathbb{E}_\theta(1-\phi_n) \leq \mathrm{e}^{-\bar{D}Mn\epsilon_n^2}.$$

3. The Framework and Main Results

- Sampling model: $y_i = f(\mathbf{x}_i) + e_i$.
- Tool: orthonormal basis $(\psi_k)_k$
- Prior: $f = \sum_k \beta_k\psi_k$, $(\beta_k)_{k=1}^\infty \sim \Pi$.

Sufficient conditions for $\Pi(\|f - f_0\|_2 > M\epsilon_n \mid \text{data}) = o_{\mathbb{P}_0}(1)$:

1. The prior concentration condition: $\Pi(B(f_0,\epsilon_n)) \geq \mathrm{e}^{-Dn\epsilon_n^2}$, where

$$B(f_0,\epsilon) = B_2(f_0,\epsilon) \cap \left\{\sum_{k>k_n} |\beta_k - \beta_{0k}| = \omega\right\},$$

   and $k_n\epsilon_n^2 = O(1)$.
2. Existence of sieves $(\mathcal{F}_n)_{n=1}^\infty$ s.t. $\Pi(\mathcal{F}_n^c) \leq \mathrm{e}^{-(2D+\sigma^{-2})n\epsilon_n^2}$,, where

$$\mathcal{F}_n \subset \left\{\sum_{k>m_n}^\infty |\beta_k - \beta_{0k}| = \delta\right\},$$

   and $m_n\epsilon_n^2 \to 0$.
3. Existence of tests $(\phi_n)_{n=1}^\infty$ guaranteed by metric entropy:

$$\mathcal{N}\left(\xi j\epsilon_n, \mathcal{F}_n \cap B_2(f_0, 2j\epsilon_n), \|\cdot\|_2\right) \leq \exp\left(Dn\epsilon_n^2/2\right).$$

4. Applications

- Finite Random Series prior
  - Truth f_0 is α-Hölder, $\alpha > 1/2$.
  - Assume α unknown, $\gamma \in (1/2,\alpha)$.
  - Prior: $(f \mid N = m) = \sum_{k=1}^m \beta_k\psi_k$, $((k^\gamma\beta_k)_{k=1}^m \mid N = m) \overset{\text{i.i.d.}}{\sim} g$, $g(\beta) \propto \mathrm{e}^{-\tau_0|\beta|^\tau}$, $N \sim \mathrm{ZTP}(\lambda)$.
  - Adaptive contraction: $\Pi\left(\|f - f_0\|_2 > M\epsilon_n \mid \text{data}\right) \overset{\mathbb{P}_0}{\to} 0$, where $\epsilon_n = n^{-\alpha/(2\alpha+1)}(\log n)^t$, $t > \alpha/(2\alpha+1)$.
- Block prior w/o truncation
  - Truth f_0 is α-Sobolev, $\alpha > 1/2$.
  - Prior: $[\beta_{k_\ell},\cdots,\beta_{k_{\ell+1}-1}] \mid A_\ell \sim \mathrm{N}(\mathbf{0}, A_\ell\mathbf{I}_{n_\ell})$, $k_\ell = \lceil \mathrm{e}^\ell\rceil$, $A_\ell \sim g_\ell$, and g_ℓ shrinks toward 0.
  - Exact minimax-optimal contraction: $\Pi\left(\|f - f_0\|_2 > M\epsilon_n \mid \text{data}\right) \overset{\mathbb{P}_0}{\to} 0$, where $\epsilon_n = n^{-\alpha/(2\alpha+1)}$.
- SE-GP prior ($K(x,x') = \mathrm{e}^{-(x-x')^2}$)
  - Truth f_0 is supersmooth.
  - Prior: $\beta_k \sim \mathrm{N}(0,\lambda_k)$, $\lambda_k \asymp \mathrm{e}^{-k^2/4}$
  - Design points $(\mathbf{x}_i)_{i=1}^n$ are fixed
  - Near-parametric contraction: $\Pi\left(\|f - f_0\|_2 > M\epsilon_n \mid \text{data}\right) \overset{\mathbb{P}_0}{\to} 0$, where $\epsilon_n = (\log n)/\sqrt{n}$.

5. Extension: Sparse Additive Model

- Sparse additive model: $f(\mathbf{x}) = \mu + \sum_{r=1}^q f_{j_r}(x_{j_r})$, $p \gg n$
- Prior: $f(\mathbf{x}) = \mu + \sum_{jk} z_j\beta_{jk}\psi_k(x_j)$, $z_j \sim \mathrm{Bern}(1/p)$, $(\beta_{jk})_{jk} \sim \Pi$

Sufficient conditions for $\Pi(\|f - f_0\|_2 > M\epsilon_n \mid \text{data}) \overset{\mathbb{P}_0}{\to} 0$:

1. Prior: $\Pi(\widetilde{B}(f_0,\epsilon_n)) \geq \mathrm{e}^{-Dn\epsilon_n^2}$, where

$$\widetilde{B}(f_0,\epsilon) = B_2(f_0,\epsilon) \cap \{\|\mathbf{z}\|_1 \leq 2q\} \cap \left\{\sum_{j,k>k_n} |z_j\beta_{jk} - \beta_{0jk}| \leq \omega\right\}$$

   and $k_n\epsilon_n^2 = O(1)$.
2. Sieves $\mathcal{G}_n$: $\Pi(\mathcal{G}_n^c) \leq \mathrm{e}^{-(2D+\sigma^{-2})n\epsilon_n^2}$, $A_nm_n\epsilon_n^2 \to 0$, $\mathcal{G}_n = \bigcup_{\|\mathbf{z}\|_1 \leq A_nq} \mathcal{G}_n^{\mathbf{z}}$,

$$\mathcal{G}_n^{\mathbf{z}} \subset \left\{\sum_{j,k>m_n} |z_j\beta_{jk} - \beta_{0jk}| \leq \delta\right\}.$$

3. Tests $(\phi_n)_{n=1}^\infty$ guaranteed by metric entropy:

$$\mathcal{N}\left(\xi j\epsilon_n, \mathcal{G}_n \cap B_2(f_0, 2j\epsilon_n), \|\cdot\|_2\right) \leq \exp\left(Dn\epsilon_n^2/2\right).$$