Bayesian simultaneous credible bands for polynomial regression

Fei Yang[1] Yang Han[1] Wei Liu[2]

[1]University of Manchester
[2]University of Southampton

fei.yang@manchester.ac.uk

13 July 2025

MANCHESTER 1824
The University of Manchester

Table of Contents

Bayesian simultaneous credible bands for polynomial regression

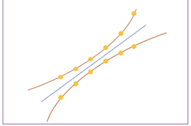
Statistics, Dept of Math, Manchester

Fei Yang

Motivation

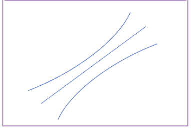
To assess the plausible range of an unknown function

PCB



PCB is not able to cover the function simultaneously

SCB



Bayesian SCB

Propose two methods

- The Conjugate Prior
- HMC

A new evaluation criterion

Posterior Coverage Probability

Simulations & Real Data Examples

Linear Regression (p=1)
Polynomial Regression (p=2,3)

Frequentist

Bayesian

Methodology

Notations

Methodology

Notations

$$Y = X\theta + e, \tag{1}$$

- $\boldsymbol{Y} = (y_1, \ldots, y_n)^T$
- $\boldsymbol{X}$ is a $n \times (p+1)$ full column-rank design matrix with the lth $(1 \leq l \leq n)$ row given by $(1, x_l, \ldots, x_l^p)$. x has been mean-centered.
- $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_p)^T$.
- $\boldsymbol{e} = (e_1, \ldots, e_n)^T$ with $e_i \sim N(0, \sigma^2 V)$ where the covariance matrix V is assumed to be a known positive-definite matrix.
- $\boldsymbol{\theta}$ and σ^2 are unknown.

Key Property

Methodology

Key Property

For a newly given $x \in (a, b)$, denote $\boldsymbol{x} = (1, x, \ldots, x^p)$, we consider the construction of $1 - \alpha$ level Bayesian simultaneous credible bands for $\mathbf{x^T}\boldsymbol{\theta}$,

$$P\{\boldsymbol{x}^T\boldsymbol{\theta} \in \boldsymbol{x}^T E(\boldsymbol{\theta}) \pm \lambda\sqrt{Var(\mathbf{x^T}\boldsymbol{\theta})} \quad \forall x \in (a, b)\} = 1 - \alpha. \qquad (2)$$

- $P\{\cdot\}$, $E(\theta)$, and $\mathrm{Var}(\boldsymbol{x}^T\boldsymbol{\theta})$ are with respect to the posterior distribution of $\boldsymbol{\theta}|\boldsymbol{Y}$
- λ is the critical constant.

Methodology

Key Property

For a newly given $x \in (a,b)$, denote $\boldsymbol{x} = (1, x, \ldots, x^p)$, we consider the construction of $1-\alpha$ level Bayesian simultaneous credible bands for $\mathbf{x^T}\boldsymbol{\theta}$,

$$P\{\boldsymbol{x}^T\boldsymbol{\theta} \in \boldsymbol{x}^T E(\boldsymbol{\theta}) \pm \lambda\sqrt{Var(\mathbf{x^T}\boldsymbol{\theta})} \quad \forall x \in (a,b)\} = 1-\alpha. \tag{2}$$

$$P\{-\lambda \le \frac{\boldsymbol{x}^T(\boldsymbol{\theta} - E(\boldsymbol{\theta}))}{\sqrt{\text{Var}(\boldsymbol{x}^T\boldsymbol{\theta})}} \le \lambda, \quad \forall x \in (a,b)\} = 1-\alpha, \tag{3}$$

Methodology

Key Property

For a newly given $x \in (a,b)$, denote $\boldsymbol{x} = (1, x, \ldots, x^p)$, we consider the construction of $1-\alpha$ level Bayesian simultaneous credible bands for $\mathbf{x}^{\mathbf{T}}\boldsymbol{\theta}$,

$$P\{\boldsymbol{x}^T\boldsymbol{\theta} \in \boldsymbol{x}^T E(\boldsymbol{\theta}) \pm \lambda\sqrt{Var(\mathbf{x}^{\mathbf{T}}\boldsymbol{\theta})} \quad \forall x \in (a,b)\} = 1-\alpha. \tag{2}$$

$$P\{-\lambda \leq \frac{\boldsymbol{x}^T(\boldsymbol{\theta} - E(\boldsymbol{\theta}))}{\sqrt{\mathrm{Var}(\boldsymbol{x}^T\boldsymbol{\theta})}} \leq \lambda, \quad \forall x \in (a,b)\} = 1-\alpha, \tag{3}$$

The Key Procedure

Ultimately, we need to evaluate the value of the critical constant λ:

$$\lambda = \sup_{x \in (a,b)} \frac{\left|\boldsymbol{x}^T(\boldsymbol{\theta} - E(\boldsymbol{\theta}))\right|}{\sqrt{\boldsymbol{x}^T \mathrm{Var}(\boldsymbol{\theta})\boldsymbol{x}}}$$

Methodology

The Conjugate Prior Method

The normal-gamma conjugate prior

- Denote $\tau = \frac{1}{\sigma^2}$, τ is the precision matrix of the random errors $\boldsymbol{e}$. Now the parameters of interest are $\boldsymbol{\theta}$ and τ.
- Here we assume the prior $\xi(\boldsymbol{\theta}, \tau)$ is a normal-gamma prior density,
  - $\xi(\boldsymbol{\theta}, \tau) = \xi_1(\boldsymbol{\theta}|\tau) \cdot \xi_2(\tau), \quad \boldsymbol{\theta} \in \mathbb{R}^p, \tau > 0,$
  - $\boldsymbol{\theta}|\tau \sim \mathcal{N}(\mu, \tau^{-1}\mathcal{P})$
  - $\tau \sim \text{Gamma}(\alpha_0, \beta_0)$
- Use a data-driven approach for the hyperparameters in the priors:
  - $\mu = \hat{\boldsymbol{\theta}}_{GLS} = (\boldsymbol{X}^T V^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^T V^{-1} \boldsymbol{Y},$
  - $\mathcal{P} = c \cdot I_p,$
  - $\alpha_0 = 1,$
  - $\beta_0 = \frac{\hat{\boldsymbol{e}}_{GLS}^T \hat{\boldsymbol{e}}_{GLS}}{n-p}$

The multivariate t distribution

It can be proved that $\boldsymbol{\theta}|\boldsymbol{Y}$ follows a $p-$dimensional t distribution with $(n+2\alpha_0)$ degrees of freedom:

$$(\boldsymbol{\theta}|\boldsymbol{Y}) \sim t_{(n+2\alpha_0)}(\mu^*, (D^*)^{-1}), \tag{4}$$

with location vector

$$\mu^*(\boldsymbol{\theta}|\boldsymbol{Y}) = (\boldsymbol{X}^T V^{-1}\boldsymbol{X} + \mathcal{P})^{-1}(\boldsymbol{X}^T V^{-1}\boldsymbol{Y} + \mathcal{P}\mu), \tag{5}$$

and precision matrix

$$D^*(\boldsymbol{\theta}|\boldsymbol{Y}) = (n+2\alpha_0)(\boldsymbol{X}^T V^{-1}\boldsymbol{X} + \mathcal{P})\cdot$$

$$\left[2\beta_0 + \boldsymbol{Y}^T V^{-1}\boldsymbol{Y} + \boldsymbol{\mu}^T\mathcal{P}\boldsymbol{\mu} - (\boldsymbol{X}^T V^{-1}\boldsymbol{Y} + \mathcal{P}\mu)^T(\boldsymbol{X}^T V^{-1}\boldsymbol{X} + \mathcal{P})^{-1}(\boldsymbol{X}^T V^{-1}\boldsymbol{Y} + \mathcal{P}\mu)\right]^{-1} \tag{6}$$

The multivariate t distribution

The Key Procedure

$$\lambda = \sup_{x \in (a,b)} \frac{\left| \boldsymbol{x}^T (\boldsymbol{\theta} - E(\boldsymbol{\theta})) \right|}{\sqrt{\boldsymbol{x}^T \mathrm{Var}(\boldsymbol{\theta}) \boldsymbol{x}}}$$

- $E(\boldsymbol{\theta}|\boldsymbol{Y}) = \mu^*$,
- $\mathrm{Var}(\boldsymbol{\theta}|\boldsymbol{Y}) = \frac{n+2\alpha_0}{n+2\alpha_0-2}(D^*)^{-1}$, for $n + 2\alpha_0 > 2$,
- The only numerical optimization step is the monotonic, one-dimensional root finding problem for $\hat{\lambda}$.
- Use a simulation-based method to find $\hat{\lambda}$.

Methodology

Algorithm Use the conjugate prior method to compute λ by simulation for $\boldsymbol{x}^T\boldsymbol{\beta}$ where $x \in (a, b)$

Input: $\boldsymbol{X}, \boldsymbol{Y}, n, \mu, \mathcal{P}, \alpha_0, \beta_0, x \in (a, b), 1-\alpha,$

Output: $\hat{\lambda}$

- Step 1: Compute $E(\boldsymbol{\theta}) = \mu^*$, and $\text{Var}(\boldsymbol{\theta}) = \frac{n+2\alpha_0}{n+2\alpha_0-2}(D^*)^{-1}$.
- Step 2: For $l = 1, 2, \ldots, L$, repeat the following:
  - a. Generate one value of $\boldsymbol{\theta}^{(l)}$ from the posterior (4). That is, generate $\boldsymbol{\theta}^{(l)} \sim t_{(n+2\alpha_0)}(\mu^*, D^*)$, where μ^* and D^* is given by (5) and (11).
  - b. Compute $\hat{\lambda}^{(l)}$ which is given by

$$\hat{\lambda}^{(l)} = \sup_{x \in (a,b)} \frac{\left|\boldsymbol{x}^T(\boldsymbol{\theta}^{(l)} - \mu^*)\right|}{\sqrt{\boldsymbol{x}^T \text{Var}(\boldsymbol{\theta}^{(l)})\boldsymbol{x}}},$$

Methodology

Algorithm Use the conjugate prior method to compute λ by simulation for $\boldsymbol{x}^T\boldsymbol{\beta}$ where $x \in (a,b)$

Input: $\boldsymbol{X}, \boldsymbol{Y}, n, \mu, \mathcal{P}, \alpha_0, \beta_0, x \in (a,b), 1-\alpha,$

Output: $\hat{\lambda}$

- Step 1: Compute $E(\boldsymbol{\theta}) = \mu^*$, and $\mathrm{Var}(\boldsymbol{\theta}) = \frac{n+2\alpha_0}{n+2\alpha_0-2}(D^*)^{-1}$.
- Step 2: For $l = 1, 2, \ldots, L$, repeat the following:
  - a. Generate one value of $\boldsymbol{\theta}^{(l)}$ from the posterior (4). That is, generate $\boldsymbol{\theta}^{(l)} \sim t_{(n+2\alpha_0)}(\mu^*, D^*)$, where μ^* and D^* is given by (5) and (11).
  - b. Compute $\hat{\lambda}^{(l)}$ which is given by

$$\hat{\lambda}^{(l)} = \sup_{x\in(a,b)} \frac{\left|\boldsymbol{x}^T(\boldsymbol{\theta}^{(l)} - \mu^*)\right|}{\sqrt{\boldsymbol{x}^T \mathrm{Var}(\boldsymbol{\theta}^{(l)})\boldsymbol{x}}},$$

- Step 3: Order these $\hat{\lambda}^{(l)}$ values as $\lambda_{[1]} \leq \cdots \leq \lambda_{[L]}$ and use $\lambda_{[<(1-\alpha)L>]}$ as the $\hat{\lambda}$ we want. Here $<(1-\alpha)L>$ denotes the integer part of $(1-\alpha)L$. $L_{<(1-\alpha)L>}$ converges to λ with probability one as $L \to \infty$ (Serfling, 1980 [1]). The simultaneous credible bands for $\boldsymbol{x}^T\boldsymbol{\theta}$ is given by:

$$\left[\boldsymbol{x}^T E(\boldsymbol{\theta}) - \lambda_{[<(1-\alpha)L>]}\sqrt{\mathbf{x}^{\mathbf{T}}\mathrm{Var}(\boldsymbol{\theta})\mathbf{x}}, \quad \boldsymbol{x}^T E(\boldsymbol{\theta}) + \lambda_{[<(1-\alpha)L>]}\sqrt{\mathbf{x}^{\mathbf{T}}\mathrm{Var}(\boldsymbol{\theta})\mathbf{x}}\right]$$

The HMC Method

Methodology

The HMC Method

- The Hamiltonian Monte Carlo sampler method is originated in the physics literature as an approach uniting MCMC and molecular dynamics approaches
  - A special case of the Metropolis-Hastings method.
    - We generate a proposal for a new sample. The sampling trajectory is treated as a process where a frictionless object sliding over a surface of varying height.
    - Then accept or reject it based on an MH acceptance probability.
  - Enables more efficient exploration of the state space than standard random-walk proposals.
- Hyperparameters:
  - Chains: 4
  - Iterations per chain: 8000
  - Warmup: 4000
- Outputs:
  - Posterior distribution of $\boldsymbol{\theta}$
  - $E(\boldsymbol{\theta})$, $\mathrm{Var}(\boldsymbol{\theta})$

Algorithm Use the HMC method to compute λ by simulation for $\boldsymbol{x}^T\boldsymbol{\beta}$ where $x \in (a, b)$

Input: $\boldsymbol{X}, \boldsymbol{Y}, V, n, \mu, \mathcal{P}, \alpha_0, \beta_0, x \in (a, b), 1-\alpha$

Output: $\hat{\lambda}$

- Step 1: Derive the $E(\boldsymbol{\theta}) = \mu^*$, and Var$(\boldsymbol{\theta})$ according to the posterior distribution.
- Step 2: For $l = 1, 2, \ldots, L$, repeat the following:
  - a. Using the HMC to produce the posterior distribution of $\boldsymbol{\theta}$. Draw one value of $\boldsymbol{\theta}^{(l)}$.
  - b. Compute $\hat{\lambda}^{(l)}$ which is given by

$$\hat{\lambda}^{(l)} = \sup_{x\in(a,b)} \frac{\left|\boldsymbol{x}^T(\boldsymbol{\theta}^{(l)} - \mu^*)\right|}{\sqrt{\boldsymbol{x}^T \mathrm{Var}(\boldsymbol{\theta}^{(l)})\boldsymbol{x}}},$$

Algorithm Use the HMC method to compute λ by simulation for $\boldsymbol{x}^T\boldsymbol{\beta}$ where $x \in (a,b)$

Input: $\boldsymbol{X}, \boldsymbol{Y}, V, n, \mu, \mathcal{P}, \alpha_0, \beta_0, x \in (a,b), 1-\alpha$

Output: $\hat{\lambda}$

- Step 1: Derive the $E(\boldsymbol{\theta}) = \mu^*$, and Var$(\boldsymbol{\theta})$ according to the posterior distribution.
- Step 2: For $l = 1, 2, \ldots, L$, repeat the following:
  - a. Using the HMC to produce the posterior distribution of $\boldsymbol{\theta}$. Draw one value of $\boldsymbol{\theta}^{(l)}$.
  - b. Compute $\hat{\lambda}^{(l)}$ which is given by

$$\hat{\lambda}^{(l)} = \sup_{x\in(a,b)} \frac{\left|\boldsymbol{x}^T(\boldsymbol{\theta}^{(l)} - \mu^*)\right|}{\sqrt{\boldsymbol{x}^T \mathrm{Var}(\boldsymbol{\theta}^{(l)})\boldsymbol{x}}},$$

- Step 3: Order these $\hat{\lambda}^{(l)}$ values as $\lambda_{[1]} \leq \cdots \leq \lambda_{[L]}$ and use $\lambda_{[<(1-\alpha)L>]}$ as the $\hat{\lambda}$ we want. Here $<(1-\alpha)L>$ denotes the integer part of $(1-\alpha)L$. The simultaneous credible bands for $\boldsymbol{x}^T\boldsymbol{\theta}$ is given by:

$$\left[\boldsymbol{x}^T E(\boldsymbol{\theta}) - \lambda_{[<(1-\alpha)L>]}\sqrt{\mathbf{x}^{\mathbf{T}}\mathrm{Var}(\boldsymbol{\theta})\mathbf{x}}, \quad \boldsymbol{x}^T E(\boldsymbol{\theta}) + \lambda_{[<(1-\alpha)L>]}\sqrt{\mathbf{x}^{\mathbf{T}}\mathrm{Var}(\boldsymbol{\theta})\mathbf{x}}\right]$$

Comparison Methods & Evaluation Criterion

We select the following methods to compare with the proposed method,

1. The Frequentist simultaneous confidence band (Frequentist SCB)
   - The exact SCB of Liu et al.(2013)[2];
   - The conservative SCB of Naiman (1986)[3]
2. The Bayesian pointwise credible band (Bayesian PCB)
3. The Frequentist pointwise confidence band (Frequentist PCB)

Methodology

Comparison Methods & Evaluation Criterion

Difference

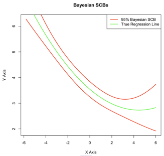
Method	Construction
Bayes SCB	$P\{\boldsymbol{x}^T\boldsymbol{\theta} \in \boldsymbol{x}^T\mu^* \pm \hat{\lambda}_{\text{Bayes}}\sqrt{\mathbf{x^T}\text{Var}(\boldsymbol{\theta})\mathbf{x}} \quad \forall x \in (a,b)\} = 1-\alpha.$
exact Freq SCB	$P\left\{\boldsymbol{x}^T\boldsymbol{\theta} \in \boldsymbol{x}^T\hat{\boldsymbol{\theta}} \pm \hat{\lambda}_{\text{exact}}\hat{\sigma}\sqrt{\boldsymbol{x}^T(\boldsymbol{X}V^{-1}\boldsymbol{X})^{-1}\boldsymbol{x}} \quad \forall x \in (a,b)\right\} = 1-\alpha,$
Freq PCB	$P\left\{\boldsymbol{x}^T\boldsymbol{\theta} \in \boldsymbol{x}^T\hat{\boldsymbol{\theta}} \pm t_{n-p-1}^{\alpha/2}\hat{\sigma}\sqrt{\boldsymbol{x}^T(\boldsymbol{X}V^{-1}\boldsymbol{X})^{-1}\boldsymbol{x}}\right\} = 1-\alpha,$
Bayes PCB	$P\{\boldsymbol{x}^T\boldsymbol{\theta} \in \boldsymbol{x}^T\mu^* \pm t_{\nu,\alpha/2}\sqrt{\boldsymbol{x}^T\text{Var}(\boldsymbol{\theta})\boldsymbol{x}}\} = 1-\alpha,$

- μ^* or $\hat{\boldsymbol{\theta}}$
- The critical constant
- $\text{Var}(\boldsymbol{x}^T\boldsymbol{\theta})$

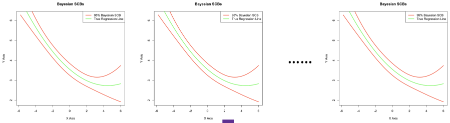
Methodology

The Empirical Coverage Rate

If the band covers the true regression line in the given range, we call it as one successful catch



Repeat testing for K times



Empirical Coverage Rate = $\frac{\#\, Successful\ Catch}{K}$

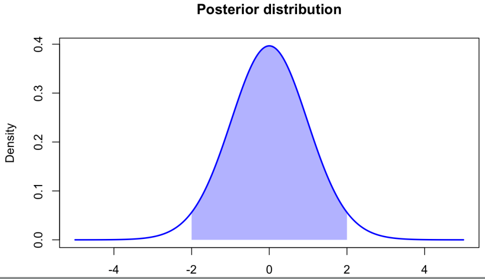
Yet, it is **not an appropriate criterion** to use in the Bayesian context.

MANCHESTER 1824
The University of Manchester

The Posterior Coverage Probability

It reflects the probability that the band contains the true regression function under the posterior distribution. Denote the Bayesian SCB for $\boldsymbol{x}^T\boldsymbol{\theta}$, $x \in (a, b)$, as $\mathcal{I}_A$: the posterior coverage probability is defined as

$$\inf P_{\boldsymbol{\theta}|\mathbf{Y}}\{\{\boldsymbol{x}^T\boldsymbol{\theta} \in \mathcal{I}_A\} \quad \forall x \in (a, b)\}, \tag{7}$$



Methodology

Comparison Methods & Evaluation Criterion

Denote $T = \boldsymbol{x}^T\boldsymbol{\theta}$, then $T \sim t_\nu(\mu_t, \sigma_t)$,$\mu_t = \boldsymbol{x}^T\mu^*, \sigma_t = \boldsymbol{x}^T\text{Cov}(\theta)\boldsymbol{x}, \nu = n + 2\alpha$. Thus $T_\nu := \frac{T-\mu_t}{\sigma_t} \sim t_\nu$.

$$
\begin{aligned}
& P_{\boldsymbol{x}^T\boldsymbol{\theta}|\mathbf{Y}}\{\boldsymbol{x}^T\boldsymbol{\theta} \in \mathcal{I}_A \quad \forall x \in (a,b)\} \\
&= P_{\boldsymbol{x}^T\boldsymbol{\theta}|\mathbf{Y}}\{L(x) \leq \boldsymbol{x}^T\boldsymbol{\theta} \leq U(x) \quad \forall x \in (a,b)\}, \\
&= P_{\boldsymbol{x}^T\boldsymbol{\theta}|\mathbf{Y}}\{\frac{L(x)-\mu_t}{\sigma_t} \leq T_\nu \leq \frac{U(x)-\mu_t}{\sigma_t} \quad \forall x \in (a,b)\} \\
&= \int_a^b \left[F_\nu(\frac{U(x)-\mu_t}{\sigma_t}) - F_\nu(\frac{L(x)-\mu_t}{\sigma_t})\right] dF(x), \\
&= \int_a^b \left[F_\nu(\frac{U(x)-\mu_t}{\sigma_t}) - F_\nu(\frac{L(x)-\mu_t}{\sigma_t})\right] f(x|a \leq x \leq b)dx, \\
&= \int_a^b \left[F_\nu(\frac{U(x)-\mu_t}{\sigma_t}) - F_\nu(\frac{L(x)-\mu_t}{\sigma_t})\right] \frac{1}{\sqrt{2\pi}(\Phi(b)-\Phi(a))} \exp{(-x^2/2)}dx.
\end{aligned}
$$

Methodology

Comparison Methods & Evaluation Criterion

Obtain large discreted samples $x^{(i)} \in [a,b], i = 1,\ldots,n_x$. For each $x^{(i)}$, we have large number of $\boldsymbol{\theta}_j, j = 1,\ldots,n_\theta$. Denote $f_j^{(i)} = \boldsymbol{x}^{(i)^T}\boldsymbol{\theta}_j$, which is a random variable, $\boldsymbol{f}^{(i)}$ as the distribution of $f_j^{(i)}$, $F^{(i)}$ as the CDF. The posterior coverage probability is approximated by the following equation:

$$
\begin{aligned}
&P_{\boldsymbol{x}^T\boldsymbol{\theta}|\mathbf{Y}}\{\boldsymbol{x}^T\boldsymbol{\theta} \in \mathcal{I}_A \quad \forall x \in (a,b)\} \\
&= P_{\boldsymbol{x}^T\boldsymbol{\theta}|\mathbf{Y}}\{L(x) \le \boldsymbol{x}^T\boldsymbol{\theta} \le U(x) \quad \forall x \in (a,b)\}, \\
&= \int_a^b \{F(U(x)) - F(L(x))\}\, dF(x), \\
&\approx \sum_{i=1}^{n_x} \left\{F^{(i)}(U(x^{(i)})) - F^{(i)}(L(x^{(i)}))\right\}
\end{aligned}
$$

Simulation

A Toy Example

Simulation

A Toy Example

$$\boldsymbol{Y} = \boldsymbol{X}^T\boldsymbol{\theta} + \boldsymbol{e},$$

- $x_i \sim U(-5,5)$, $e_i \sim N(0,\sigma^2), i = 1,\ldots,n,$
  - Linear Setting: $\boldsymbol{\theta} = (1,2)^T$
    - n=100
    - $\sigma = 0.25$
  - Quadratic Setting: $\boldsymbol{\theta} = (-6,-3,0.25)^T$
    - $n = 20, 50, 100, 200$
    - $\sigma = 0.2, 0.5, 1$
  - Cubic Setting: $\boldsymbol{\theta} = (1,2,-1,0.5)^T$
    - $n = 200$
    - $\sigma = 1$
- For the polynomial model, we use D-optimal design to construct the design matrix.
  - e.g. The optimal design for $x \in [-5,5]$ is: $x = -5, -2.237447, -2.235447, 2.235447, 2.237447, 5.$

Simulation

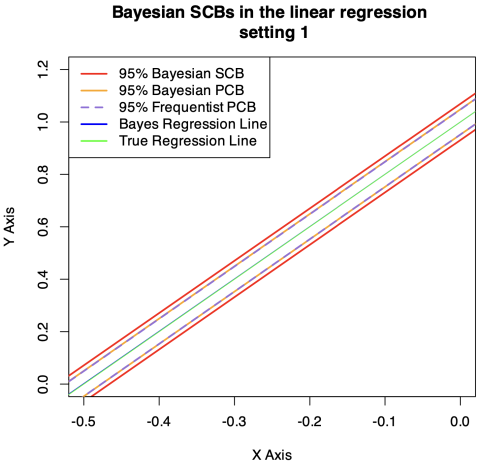


Figure: The 95% Bayesian SCB, the 95% Bayesian PCB, and the 95% Frequentist PCB for the linear regression line

Simulation

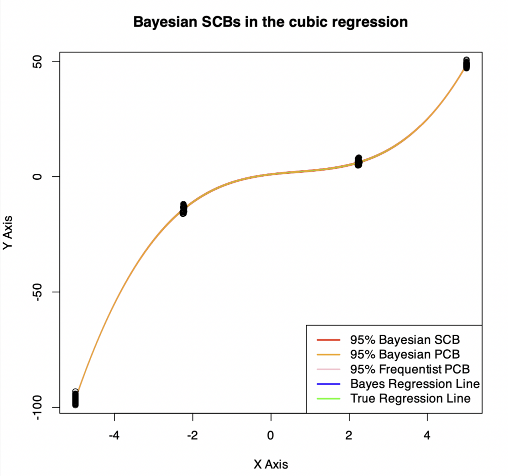


Figure: The 95% Bayesian SCB, the 95% Bayesian PCB, and the 95% frequentist PCB for the regression curve in the cubic regression example with a D-optimal design matrix

Simulation

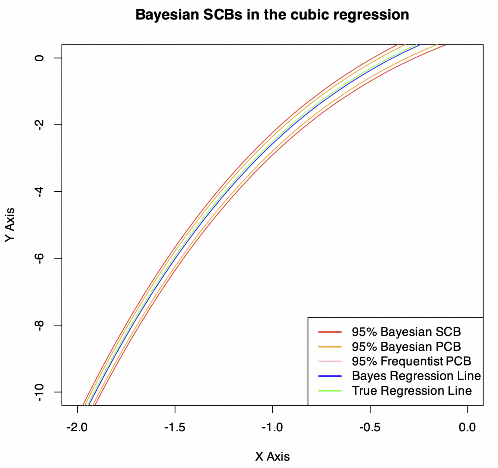


Figure: A zoomed-in look when $x \in [-2, 2]$ for the 95% Bayesian SCB, the 95% Bayesian PCB, and the 95% frequentist PCB in the cubic regression example

MANCHESTER 1824
The University of Manchester

The Posterior Coverage Probability

p	σ	n	Average Posterior Coverage Probability: 95% Bayesian SCB Conjugate Prior	Average Posterior Coverage Probability: 95% Bayesian PCB
2	0.2	20	0.981	0.950
		50	0.984	0.950
		100	0.985	0.950
		200	0.986	0.950
	0.5	20	0.981	0.950
		50	0.984	0.950
		100	0.985	0.950
		200	0.986	0.950
	1	20	0.981	0.950
		50	0.984	0.950
		100	0.985	0.950
		200	0.986	0.950
3	1	20	0.984	0.950
		50	0.987	0.950
		100	0.988	0.950
		200	0.988	0.950

Figure: The posterior coverage probability of the 95 % Bayesian SCB, and the 95 % Bayesian PCB for the quadratic and cubic regression

- The average posterior coverage probability (APCP) is the mean average under 1000 repetitions.
- APCP increases as n increases
- APCP is not sensitive to the changes of the noise level σ.

The HMC Method

p	σ	n	Empirical Coverage Rate				Average Posterior Coverage Probability		
			95% Bayesian SCB Conjugate Prior	95% Bayesian SCB HMC	95% Bayesian PCB	95% Frequentist PCB	95% Bayesian SCB Conjugate Prior	95% Bayesian SCB HMC	95% Bayesian PCB
2	0.2	20	0.901	0.949	0.766	0.814	0.981	0.985	0.950
2	0.2	50	0.917	0.932	0.787	0.799	0.984	0.986	0.950
2	0.2	100	0.937	0.941	0.804	0.810	0.985	0.986	0.950
2	0.2	200	0.933	0.938	0.811	0.814	0.986	0.986	0.950

Figure: The posterior coverage probability of the 95 % Bayesian SCB, and the 95 % Bayesian PCB for the HMC method in the quadratic regression

The HMC method is less sensitive to the choice of initial hyperparameters, making it more flexible and easier to use than the conjugate prior approach.

Real Data Example

Dose-response Dataset in a Phase II study

Real Data Example

Dose-response Dataset in a Phase II study (Bretz et al., 2005)[4]

- Goal: To accurately finding the dose-response relationship in a a randomized double-blind parallel group trial involving 100 patients who were randomly assigned, with equal probability, to receive either placebo or one of four active doses, coded as $x = 0.05, 0.2, 0.6, 1$.
- Y: The response to the doses of treatment.
- x: The doses of the drug.

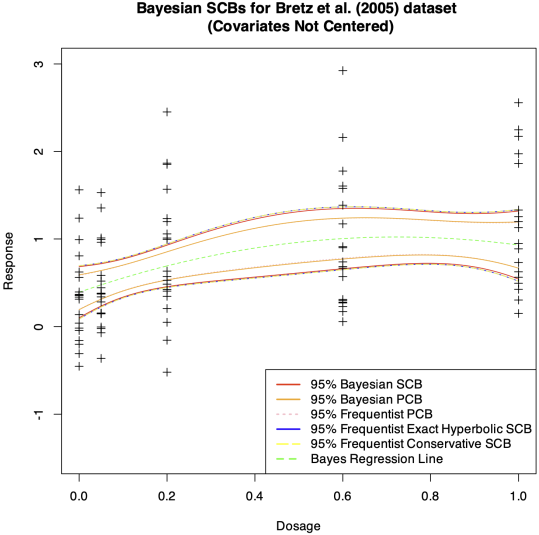
Dose	Sample size	Sample mean	Sample SD
0	20	0.34	0.52
0.05	20	0.46	0.49
0.2	20	0.81	0.74
0.6	20	0.93	0.76
1	20	0.95	0.95

Real Data Example

- When the covariates are not centered:

  $$Y = 0.392 + 1.743x - 1.205x^2,$$

- $\lambda : 2.442347$, the same as the one when covariates are centered.
- Compared with:
  - The exact Frequentist SCB of Liu et al.(2013)[2];
  - The conservative Frequentist SCB of Naiman (1986)[3]



MANCHESTER 1824
The University of Manchester

Statistics, Dept of Math, Manchester

Fei Yang

Discussion

Summary & Future Work

Discussion

Summary

- To assess where lies the true regression function $\boldsymbol{x}^T\boldsymbol{\theta}$, we propose two methods for constructing two-sided hyperbolic **Bayesain SCBs** over a finite interval on the covariates for the **polynomial regression**.
- Compared to the Frequentist approach, Bayesian methods are more suitable when data are **limited** or when **domain knowledge** needs to be incorporated.
- Both the conjugate method and the HMC method are computationally **convenient**. The HMC method is **more generally applicable** than the conjugate method, as it is **less sensitive** to the hyperparameters.

Discussion

Future Work

1. Extend the Bayesian approach into other models:
    1. The GLM,
    2. Random effects linear model,
    3. Quantile regression model
2. Combine with the machine learning algorithms (Sluijterman et al., 2024 [5]).

References I

[1] Robert J Serfling. *Approximation theorems of mathematical statistics*. John Wiley & Sons, 1980.

[2] Wei Liu, Sanyu Zhou, and Frank Bretz. Exact simultaneous confidence bands for quadratic and cubic polynomial regression with applications in dose response study. *Australian & New Zealand Journal of Statistics*, 55(4):421–434, 2013.

[3] Naiman Daniel Q. Conservative confidence bands in curvilinear regression. *The Annals of Statistics*, pages 896–906, 1986.

[4] Frank Bretz, José C Pinheiro, and Michael Branson. Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrics*, 61(3):738–748, 2005.

[5] Laurens Sluijterman, Eric Cator, and Tom Heskes. How to evaluate uncertainty estimates in machine learning for regression? *Neural Networks*, 173:106203, 2024.

This is the end of my presentation.
Thank you for your attention.

Appendix

The HMC Method

Denote $H(\cdot)$ as the Hamiltonial function that represents the total energy in the system, such that

$$H(q,p) = U(q) + K(p),$$

where q is the position of the object and $U(q)$ represents the potential energy; p is the momentum of the and $K(p)$ represents the kinetic energy. p and q are changed with time t, and can be expressed as the following set of $2d$ first-order differential equations:

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}$$
$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}$$

for $i = 1, \ldots, d$.

We can outline the HMC algorithm, as follows:

1. Initialise with $\boldsymbol{\theta} = \boldsymbol{\theta}^{(0)}$.
2. For $i = 1, \ldots, N$:
   1. Draw $p^{(i)} \sim \mathcal{N}_d(0, M)$.
   2. Use $\boldsymbol{\theta}^{(i-1)}$ and $p^{(i)}$ to simulate the Hamiltonian dynamics and propose $(\boldsymbol{\theta}^*, p^*)$.
   3. Calculate the acceptance probability, given by

$$\alpha(\boldsymbol{\theta}^{(i-1)}, \boldsymbol{\theta}^*) = \min\{\exp\{H(\boldsymbol{\theta}^{(i-1)}, p^{(i)}) - H(\boldsymbol{\theta}^*, p^*)\}, 1\}$$

   4. With probability $\alpha(\boldsymbol{\theta}^{(i-1)}, \boldsymbol{\theta}^*)$, accept the candidate value and set $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^*$; otherwise reject the candidate value and set $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^{(i-1)}$.
3. Repeat until a sample of the desired size is obtained.