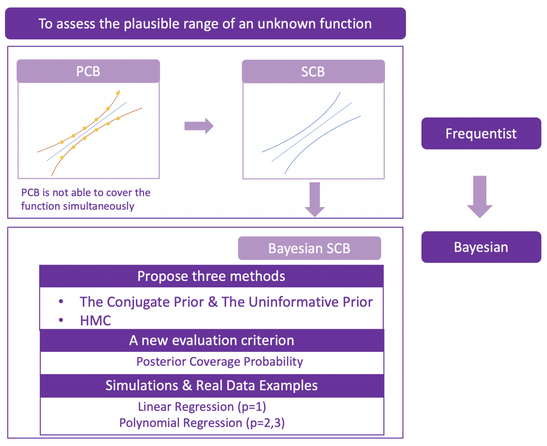
Bayesian Simultaneous Credible Bands for Polynomial Regression

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

[1]Department of Mathematics, The University of Manchester
[2]School of Mathematical Sciences, University of Southampton

1. Motivation

- Regression analysis requires assessing the plausible range of an unknown function.
- Traditional Pointwise Confidence Bands (PCB) fail to cover the function simultaneously at the nominal level, motivating the use of Simultaneous Confidence Bands (SCB).
- SCBs have been widely developed in the Frequentist framework. However, in scenarios with limited data or with domain information, a Bayesian approach is a better way.
- We propose three Bayesian SCB methods using a conjugate prior, Jeffreys' non-informative prior, and HMC separately. Simulations and real analysis in a dose-response study for the polynomial models with $p = 1, 2, 3$ show the effectiveness of the methods.



2. Methodology

2.1 Notations

Consider a standard univariate polynomial model:

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

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

The Key Procedure

For any given $x \in (a, b)$, denote $\boldsymbol{x} = (1, x, \ldots, x^p)$, we consider the construction of $1-\alpha$ level Bayesian SCB for the whole regression curve $\mathbf{x^T\boldsymbol{\theta}}$,

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

$P\{\cdot\}$, $E(\boldsymbol{\theta})$, 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.

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

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

$$\lambda = \sup_{x\in(a,b)} \frac{\left|\boldsymbol{x}^T(\boldsymbol{\theta} - E(\boldsymbol{\theta}))\right|}{\sqrt{\boldsymbol{x}^T \mathrm{Var}(\boldsymbol{\theta})\boldsymbol{x}}}. \tag{1}$$

2.2 The Proposed Methods

(1) The Normal-Gamma Conjugate Prior

Assume the prior $\xi(\boldsymbol{\theta}, \tau)$ is a normal-gamma prior, where $\boldsymbol{\theta}|\tau \sim \mathcal{N}(\mu, \tau^{-1}\mathcal{P})$, and $\tau \sim \mathrm{Gamma}(\alpha_0, \beta_0)$. The posterior $(\boldsymbol{\theta}|\boldsymbol{Y})$ is a $p-$dimensional t distribution with the location vector μ^* and the precision matrix D^* in analytical forms:

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

After we derive $E(\boldsymbol{\theta})$ and $\mathrm{Var}(\boldsymbol{\theta})$, the only numerical optimisation step for **Function (1)** is the monotonic, one-dimensional root finding problem for $\hat{\lambda}$. Then we use a simulation-based method to find it.

Algorithm 1 To compute $\hat{\lambda}_{\mathrm{BSCB}}$ based on a normal-gamma conjugate prior setting for the construction of the Bayesian SCB 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: $\hat{\lambda}_{\mathrm{BSCB}}$

- Step 1: Compute $E(\boldsymbol{\theta}) = \mu^*$, and $\mathrm{Var}(\boldsymbol{\theta}) = \frac{\nu}{\nu-2}(D^*)^{-1}$.
- Step 2: For $l = 1, 2, \ldots, L$, repeat the following:
  - a. Generate one value of $\boldsymbol{\theta}^{(l)}$ from the posterior distribution.
  - b. Compute $S^{(l)}$ which is given by

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

- Step 3: Order these $S^{(l)}$ values as $S_{[1]} \le \cdots \le S_{[L]}$ and use $S_{[<(1-\alpha)L>]}$ as the $\hat{\lambda}_{\mathrm{BSCB}}$ 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:

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

return $\hat{\lambda}_{\mathrm{BSCB}}$

(2) The Jeffery's Non-informative Prior

Additionally, we can use Jeffery's non-informative prior to produce the posterior, in which $(\boldsymbol{\theta}|Y)$ is also a multivariate t-distribution:

$$(\boldsymbol{\theta}|\boldsymbol{Y}) \sim t_n(\hat{\boldsymbol{\theta}}, \frac{\mathrm{SSE}}{n}\Sigma_0).$$

where $\hat{\boldsymbol{\theta}} = (\boldsymbol{X}^T V^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T V^{-1}\boldsymbol{Y}, \Sigma_0 = (\boldsymbol{X}^T V^{-1}\boldsymbol{X})^{-1}, \mathrm{SSE} = (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\theta}})^T V^{-1}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\theta}})$.

(3) The Hamiltonian Monte Carlo Sampler Method

The third algorithm employs a sampling method, the HMC method, to approximate the posterior distribution. We use four independent chains of 8000 iterations each, with the first 4000 as warmup.

2.3 Comparisons

The following methods are selected to compare with the proposed one:

1. The Frequentist SCB
   - The exact SCB of Liu et al.(2013)[1]
   - The conservative SCB of Naiman (1986)[2]
2. The Bayesian PCB
3. The Frequentist PCB

Differences of these methods lie in three parts: (1) μ^* or $\hat{\boldsymbol{\theta}}$; (2) The critical constant; (3) $\mathrm{Var}(\boldsymbol{x}^T\boldsymbol{\theta})$.

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

3. Simulation Study

3.1 Settings

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

$x_i \sim U(-5, 5)$, $e_i \sim N(0, \sigma^2), i = 1, \ldots, n$. For the polynomial model, we use the D-optimal design to construct the design matrix.

	$\boldsymbol{\theta}$	n	σ
Linear Setting	$(1,2)^T$	100	0.25
Quadratic Setting	$(-6, -3, 0.25)^T$	20, 50, 100, 200	0.2, 0.5, 1
Cubic Setting	$(1, 2, -1, 0.5)^T$	200	1

Table 1. Settings for the simulation study

3.2 Evaluation Criterion

The Empirical Coverage Rate (ECR, Frequentist Criterion)

The ECR is the most commonly used evaluation criterion in the existing literature for the Frequentist SCB, which is defined as the proportion of successful catches among experiments. Yet it is **not an appropriate criterion** to use in the Bayesian context.

The Posterior Coverage Probability (PCP, Bayesian Criterion)

Therefore, we change to a new criterion called: The Posterior Coverage Probability, which reflects the probability that the band contains the true regression function under the posterior distribution.

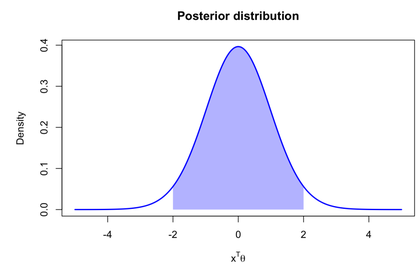


Figure 1. The definition of posterior coverage probability

3.3 Results

p	σ	n	Average Posterior Coverage Probability			
			95% Bayesian SCB Conjugate Prior	95% Bayesian SCB Non-informative Prior	95% Bayesian SCB HMC	95% Bayesian PCB
3	0.5	20	0.984	0.982	0.987	0.95
		50	0.987	0.987	0.988	0.95
		100	0.988	0.988	0.989	0.95
		200	0.988	0.988	0.989	0.95

Table 2. Average Posterior Coverage Probability (APCP) in 1000 repetitions for the Non-informative Prior method

The non-informative prior method is better than the conjugate prior method in terms of the ECR and APCP. The HMC method is less sensitive to the choice of initial hyperparameters, making it more flexible and easier to use than the two prior approaches.

4. Real Data Examples

Background: Identifying the right dose is a critical step in pharmaceutical drug development. An insufficient dose may lead to inadequate efficacy, while an excessive dose can result in safety or tolerability issues.

Goal: We use the dataset from Bretz et al. (2005)[3] to find the dose-response relationship in 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 $x = 0.05, 0.2, 0.6, 1$.

Dataset: Y is the response to the doses of treatment, while x is the doeses 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

Table 3. Summary of clinical dose response data from Bretz et al. (2005).

- Following Liu et al. (2013) [1], the best-fitted model is:

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

- For $x \in (a, b) = (0, 1)$ and $1-\alpha = 0.95$, we obtain $\lambda = 2.442347$ by the HMC method, which is the same as the one when covariates are centered.
- The Bayesian SCB, which combines more information from the prior knowledge, is very close to the exact Frequentist SCB of Liu et al.(2013)[1]. While the conservative Frequentist SCB of Naiman (1986)[2] is on the edge, and the Bayesian PCB and the Frequentist SCB in the inner side.

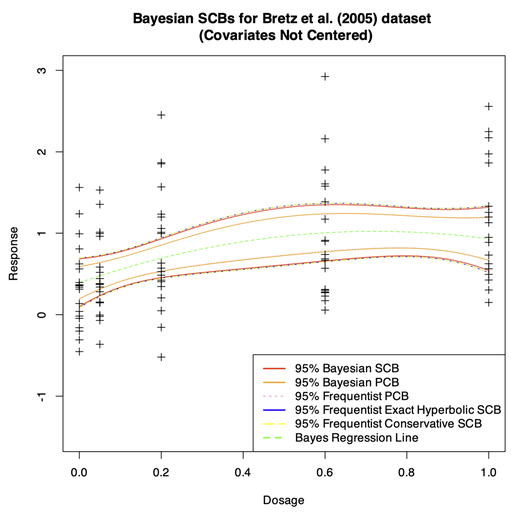


Figure 2. The 95% Bayesian SCB, the 95% Bayesian PCB, and the 95% Frequentist PCB for Bretz et al.(2005) dataset

5. Conclusion

5.1 Summary

- To assess where lies the true regression function $\boldsymbol{x}^T\boldsymbol{\theta}$, we propose three Bayesian 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.
- The conjugate prior method, the non-informative prior method, and the HMC method are computationally **convenient**. The HMC method is **more generally applicable** than the prior methods, as it is **less sensitive** to the hyperparameters.

5.2 Future Work

1. Extend the Bayesian approach into other models: GLM, random effects linear model, or the quantile regression model.
2. Combine the framework with the machine learning algorithms for uncertainty quantification, e.g., model averaging.

References

[1] 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.

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

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