Coverage of credible intervals under multivariate monotonicity

Subhashis Ghoshal, North Carolina State University

Bayesian Networking Workskshop 2023, Monash University, Melbourne, Australia

- Based on joint work with **Kang Wang**

- Based on joint work with **Kang Wang**
- Reference: Coverage of credible intervals in Bayesian multivariate isotonic regression, *The Annals of Statistics*, Vol. 51, No 3, pages 1376–1400 (2023).

- Based on joint work with **Kang Wang**
- Reference: Coverage of credible intervals in Bayesian multivariate isotonic regression, *The Annals of Statistics*, Vol. 51, No 3, pages 1376–1400 (2023).
- Fund provided by **National Science Foundation**

- Based on joint work with **Kang Wang**
- Reference: Coverage of credible intervals in Bayesian multivariate isotonic regression, *The Annals of Statistics*, Vol. 51, No 3, pages 1376–1400 (2023).

- Fund provided by **National Science Foundation**

- Travel fund from NSF, Army Research Office, and Goodnight Professorship.

Preliminaries

Monotone Regression

$Y = f(X) + \varepsilon$, X univariate or multivariate, f increasing in (each component of) x.

Monotone Regression

$Y = f(X) + \varepsilon$, X univariate or multivariate, f increasing in (each component of) x.

- The parameter space of monotone regression functions Θ_0 can be thought of as a subset of a bigger parameter space of all regression functions Θ where the model still makes sense.

Monotone Regression

$Y = f(X) + \varepsilon$, X univariate or multivariate, f increasing in (each component of) x.

- The parameter space of monotone regression functions Θ_0 can be thought of as a subset of a bigger parameter space of all regression functions Θ where the model still makes sense.
- Posterior analysis is simpler on Θ aided by conjugacy.

Monotone Regression

$Y = f(X) + \varepsilon$, X univariate or multivariate, f increasing in (each component of) x.

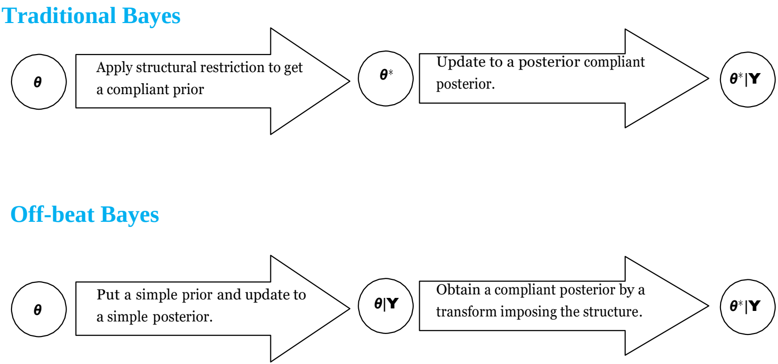
- The parameter space of monotone regression functions Θ_0 can be thought of as a subset of a bigger parameter space of all regression functions Θ where the model still makes sense.
- Posterior analysis is simpler on Θ aided by conjugacy.
- Consider a map that takes an element of Θ to an element of Θ_0. This may be used to induce a prior distribution on Θ_0, but the corresponding posterior distribution is complicated.

Monotone Regression

$Y = f(X) + \varepsilon$, X univariate or multivariate, f increasing in (each component of) x.

- The parameter space of monotone regression functions Θ_0 can be thought of as a subset of a bigger parameter space of all regression functions Θ where the model still makes sense.
- Posterior analysis is simpler on Θ aided by conjugacy.
- Consider a map that takes an element of Θ to an element of Θ_0. This may be used to induce a prior distribution on Θ_0, but the corresponding posterior distribution is complicated.
- The key idea is to switch the order of a map $\iota : \theta \mapsto \theta^*$ enforcing the desirable restriction, and the posterior updating, and use the induced posterior distribution to make an inference.

Traditional Bayes vs the Off-beat Bayesian Idea



Immersion-posterior

- The map ι complying with the restriction will be called an immersion map, and the resulting induced posterior is called the immersion posterior.

Immersion-posterior

- The map ι complying with the restriction will be called an immersion map, and the resulting induced posterior is called the immersion posterior.
- A projection map is a natural choice for ι, and in that case, the induced posterior is called the projection-posterior. A simple concentration inequality will show that the projection-posterior inherits the posterior contraction rate from the unrestricted posterior.

Immersion-posterior

- The map ι complying with the restriction will be called an immersion map, and the resulting induced posterior is called the immersion posterior.
- A projection map is a natural choice for ι, and in that case, the induced posterior is called the projection-posterior. A simple concentration inequality will show that the projection-posterior inherits the posterior contraction rate from the unrestricted posterior.
- An immersion map only needs to be defined on the support of the prior/posterior. Oftentimes, priors are put using basis expansion. Then the immersion map will typically be a finite-dimensional optimization.

Univariate Monotone Regression

Chakraborty and G. (2021, EJS; 2021, AoS)

- A finite random series of step functions: $f(x) = \sum_{j=1}^{J} \theta_j \mathbb{1}((j-1)/J < x \leq j/J]$, $\theta_1, \ldots, \theta_J$ are the coefficients given a prior distribution, J is the number of terms.

Univariate Monotone Regression

Chakraborty and G. (2021, EJS; 2021, AoS)

- A finite random series of step functions: $f(x) = \sum_{j=1}^{J} \theta_j \mathbb{1}((j-1)/J < x \leq j/J]$, $\theta_1, \ldots, \theta_J$ are the coefficients given a prior distribution, J is the number of terms.
- Monotone projection reduces to isotonization of the θ-coefficients, for which efficient algorithms like the PAVA exist.

Univariate Monotone Regression

Chakraborty and G. (2021, EJS; 2021, AoS)

- A finite random series of step functions: $f(x) = \sum_{j=1}^{J} \theta_j \mathbb{1}((j-1)/J < x \leq j/J]$, $\theta_1, \ldots, \theta_J$ are the coefficients given a prior distribution, J is the number of terms.
- Monotone projection reduces to isotonization of the θ-coefficients, for which efficient algorithms like the PAVA exist.
- Work with a deterministic choice of J, depending on the sample size n.

The coverage problem

- Consider an equal tailed $(1-\alpha)$-credible interval for $f(x_0)$ at an interior point x_0. Does its coverage go to $1-\alpha$?

The coverage problem

- Consider an equal tailed $(1-\alpha)$-credible interval for $f(x_0)$ at an interior point x_0. Does its coverage go to $1-\alpha$?
- **Cox phenomenon:** Agreement of credibility and coverage happens in parametric problems by the Bernstein-von Mises theorem, but for smoothing problems, the coverage may be arbitrarily low (Cox (1993, Ann Stat)).

The coverage problem

- Consider an equal tailed $(1-\alpha)$-credible interval for $f(x_0)$ at an interior point x_0. Does its coverage go to $1-\alpha$?
- **Cox phenomenon:** Agreement of credibility and coverage happens in parametric problems by the Bernstein-von Mises theorem, but for smoothing problems, the coverage may be arbitrarily low (Cox (1993, Ann Stat)).
- We shall obtain coverage of the monotone regression problem with the empirical $\mathbb{L}_2$-projection-posterior, which corresponds to the weighted isotonization problem of minimizing $\sum_{j=1}^{J} N_j(\theta_j - \theta_j^*)^2$ subject to $\theta_1^* \leq \cdots \leq \theta_J^*$, where $N_j = \sum_{i=1}^{n} \mathbb{1}\{(j-1)/J < X_i \leq j/J]\}$.

Limiting Coverage of a Credible interval

Let $\Delta_f^* = \arg\min\{f(t) + t^2 : t \in \mathbb{R}\}$, W_1, W_2 be independent two-sided Brownian motions, $C_0 = 2b\,(a/b)^{2/3}$ with $a = \sqrt{\sigma_0^2/g(x_0)}$, $b = f_0'(x_0)/2$, g the density of X. Let $\hat{f}_n$ be the sieve-MLE.
Let $F^*_{a,b}(z|w) = \mathrm{P}(C_0\Delta^*_{w+W_2} \le z)$.

Limiting Coverage of a Credible interval

Let $\Delta_f^* = \arg\min\{f(t) + t^2 : t \in \mathbb{R}\}$, W_1, W_2 be independent two-sided Brownian motions, $C_0 = 2b\,(a/b)^{2/3}$ with $a = \sqrt{\sigma_0^2/g(x_0)}$, $b = f_0'(x_0)/2$, g the density of X. Let $\hat{f}_n$ be the sieve-MLE.
Let $F^*_{a,b}(z|w) = \mathrm{P}(C_0 \Delta^*_{w+W_2} \leq z)$.

Theorem

(a) *for every* $z \in \mathbb{R}$, $\Pi(n^{1/3}(f^*(x_0) - f_0(x_0)) \leq z|\mathbb{D}_n) \rightsquigarrow F^*_{a,b}(z|W_1)$;

(b) *Limiting coverage of* $I_{n,\gamma}$*: with* $Z_B = \mathrm{P}(\Delta^*_{W_1+W_2} \geq 0 \big| W_1)$,
$\mathrm{P}_0\left(f_0(x_0) \in I_{n,\gamma}\right) \to \mathrm{P}(\gamma/2 \leq Z_B \leq 1 - \gamma/2)$.

Note that the nuisance parameters a, b magically vanish from the limit.

Bayes-Chernoff Distribution

- We say that the random variable $Z_B = \mathrm{P}(\Delta^*_{W_1+W_2} \geq 0 \big| W_1)$ has the **Bayes-Chernoff distribution**.

Bayes-Chernoff Distribution

- We say that the random variable $Z_B = \mathrm{P}(\Delta^*_{W_1+W_2} \geq 0 \big| W_1)$ has the **Bayes-Chernoff distribution**.
- The Bayes-Chernoff distribution is slightly more concentrated about the center $1/2$ compared with the uniform.

Bayes-Chernoff Distribution

- We say that the random variable $Z_B = \mathrm{P}(\Delta^*_{W_1+W_2} \geq 0 \big| W_1)$ has the **Bayes-Chernoff distribution**.
- The Bayes-Chernoff distribution is slightly more concentrated about the center $1/2$ compared with the uniform.
- This means that the limiting coverage is slightly higher than the credibility — the opposite of the Cox phenomenon.

Bayes-Chernoff Distribution

- We say that the random variable $Z_B = \mathrm{P}(\Delta^*_{W_1+W_2} \geq 0 \big| W_1)$ has the **Bayes-Chernoff distribution**.
- The Bayes-Chernoff distribution is slightly more concentrated about the center $1/2$ compared with the uniform.
- This means that the limiting coverage is slightly higher than the credibility — the opposite of the Cox phenomenon.
- For instance, the 95% equal-tailed credibility interval gives 96.4% limiting coverage.

Bayes-Chernoff Distribution

- We say that the random variable $Z_B = \mathrm{P}(\Delta^*_{W_1+W_2} \geq 0 \big| W_1)$ has the **Bayes-Chernoff distribution**.
- The Bayes-Chernoff distribution is slightly more concentrated about the center $1/2$ compared with the uniform.
- This means that the limiting coverage is slightly higher than the credibility — the opposite of the Cox phenomenon.
- For instance, the 95% equal-tailed credibility interval gives 96.4% limiting coverage.
- If the target coverage is $(1-\alpha)$, starting with a $(1-\gamma)$-credible interval, where $A(\gamma/2) = \alpha/2$, the **limiting coverage** $(1-\alpha)$ **is attained exactly.**

Bayes-Chernoff Distribution

- We say that the random variable $Z_B = \mathrm{P}(\Delta^*_{W_1+W_2} \geq 0 \big| W_1)$ has the **Bayes-Chernoff distribution**.
- The Bayes-Chernoff distribution is slightly more concentrated about the center $1/2$ compared with the uniform.
- This means that the limiting coverage is slightly higher than the credibility — the opposite of the Cox phenomenon.
- For instance, the 95% equal-tailed credibility interval gives 96.4% limiting coverage.
- If the target coverage is $(1-\alpha)$, starting with a $(1-\gamma)$-credible interval, where $A(\gamma/2) = \alpha/2$, the **limiting coverage** $(1-\alpha)$ **is attained exactly.**
- For instance, 93.2% equal-tailed credible interval has 95% limiting coverage.

Multivariate Monotone Regression

Prior and projection-posterior

Wang and G. (2023, AoS).

- $Y = f(\boldsymbol{X}) + \varepsilon$ with $\boldsymbol{X} \sim G$ on $[0,1]^d$, f multivariate monotone, that is, $f(\boldsymbol{x}_1) \leq f(\boldsymbol{x}_2)$ if $\boldsymbol{x}_1 \preceq \boldsymbol{x}_2$, where $\boldsymbol{x}_1 \preceq \boldsymbol{x}_2$ means that $x_{1,k} \leq x_{2,k}$, $k = 1, \ldots, d$.

Prior and projection-posterior

Wang and G. (2023, AoS).

- $Y = f(\boldsymbol{X}) + \varepsilon$ with $\boldsymbol{X} \sim G$ on $[0,1]^d$, f multivariate monotone, that is, $f(\boldsymbol{x}_1) \leq f(\boldsymbol{x}_2)$ if $\boldsymbol{x}_1 \preceq \boldsymbol{x}_2$, where $\boldsymbol{x}_1 \preceq \boldsymbol{x}_2$ means that $x_{1,k} \leq x_{2,k}$, $k = 1, \ldots, d$.
- Prior: On piecewise constant functions $f = \sum_{\boldsymbol{j}} \theta_{\boldsymbol{j}} \mathbb{1}_{I_{\boldsymbol{j}}}$ on the blocks $I_{\boldsymbol{j}} = \prod_{k=1}^{d} ((j_k - 1)/J, j_k/J]$ with a prior on the random heights $\theta_{\boldsymbol{j}}$.

Prior and projection-posterior

Wang and G. (2023, AoS).

- $Y = f(\boldsymbol{X}) + \varepsilon$ with $\boldsymbol{X} \sim G$ on $[0,1]^d$, f multivariate monotone, that is, $f(\boldsymbol{x}_1) \leq f(\boldsymbol{x}_2)$ if $\boldsymbol{x}_1 \preceq \boldsymbol{x}_2$, where $\boldsymbol{x}_1 \preceq \boldsymbol{x}_2$ means that $x_{1,k} \leq x_{2,k}$, $k = 1, \ldots, d$.
- Prior: On piecewise constant functions $f = \sum_{\boldsymbol{j}} \theta_{\boldsymbol{j}} \mathbb{1}_{I_{\boldsymbol{j}}}$ on the blocks $I_{\boldsymbol{j}} = \prod_{k=1}^{d} ((j_k - 1)/J, j_k/J]$ with a prior on the random heights $\theta_{\boldsymbol{j}}$.
- Monotonicity will be ensured if the array of coefficients $\boldsymbol{\theta}$ belongs to the cone $\mathcal{C} := \{\boldsymbol{\theta} \in \mathbb{R}^{J^d} : \theta_{\boldsymbol{j}_1} \leq \theta_{\boldsymbol{j}_2} \text{ if } \boldsymbol{j}_1 \preceq \boldsymbol{j}_2\}$.

Prior and projection-posterior

Wang and G. (2023, AoS).

- $Y = f(\boldsymbol{X}) + \varepsilon$ with $\boldsymbol{X} \sim G$ on $[0,1]^d$, f multivariate monotone, that is, $f(\boldsymbol{x}_1) \leq f(\boldsymbol{x}_2)$ if $\boldsymbol{x}_1 \preceq \boldsymbol{x}_2$, where $\boldsymbol{x}_1 \preceq \boldsymbol{x}_2$ means that $x_{1,k} \leq x_{2,k}$, $k = 1, \ldots, d$.
- Prior: On piecewise constant functions $f = \sum_{\boldsymbol{j}} \theta_{\boldsymbol{j}} \mathbb{1}_{I_{\boldsymbol{j}}}$ on the blocks $I_{\boldsymbol{j}} = \prod_{k=1}^{d}((j_k - 1)/J, j_k/J]$ with a prior on the random heights $\theta_{\boldsymbol{j}}$.
- Monotonicity will be ensured if the array of coefficients $\boldsymbol{\theta}$ belongs to the cone $\mathcal{C} := \{\boldsymbol{\theta} \in \mathbb{R}^{J^d} : \theta_{\boldsymbol{j}_1} \leq \theta_{\boldsymbol{j}_2} \text{ if } \boldsymbol{j}_1 \preceq \boldsymbol{j}_2\}$.
- $\mathbb{L}_1$-approximation rate is J^{-1} using a total of J^d basis functions; the $\mathbb{L}_2$-rate is $J^{-1/2}$.

Prior and projection-posterior

Wang and G. (2023, AoS).

- $Y = f(\boldsymbol{X}) + \varepsilon$ with $\boldsymbol{X} \sim G$ on $[0,1]^d$, f multivariate monotone, that is, $f(\boldsymbol{x}_1) \le f(\boldsymbol{x}_2)$ if $\boldsymbol{x}_1 \preceq \boldsymbol{x}_2$, where $\boldsymbol{x}_1 \preceq \boldsymbol{x}_2$ means that $x_{1,k} \le x_{2,k}$, $k = 1, \ldots, d$.
- Prior: On piecewise constant functions $f = \sum_{\boldsymbol{j}} \theta_{\boldsymbol{j}} \mathbb{1}_{I_{\boldsymbol{j}}}$ on the blocks $I_{\boldsymbol{j}} = \prod_{k=1}^{d} ((j_k - 1)/J, j_k/J]$ with a prior on the random heights $\theta_{\boldsymbol{j}}$.
- Monotonicity will be ensured if the array of coefficients $\boldsymbol{\theta}$ belongs to the cone $\mathcal{C} := \{\boldsymbol{\theta} \in \mathbb{R}^{J^d} : \theta_{\boldsymbol{j}_1} \le \theta_{\boldsymbol{j}_2} \text{ if } \boldsymbol{j}_1 \preceq \boldsymbol{j}_2\}$.
- $\mathbb{L}_1$-approximation rate is J^{-1} using a total of J^d basis functions; the $\mathbb{L}_2$-rate is $J^{-1/2}$.
- Optimal $\mathbb{L}_1$-contraction rate $n^{-1/(d+2)}$ using the $\mathbb{L}_1$ or $\mathbb{L}_2$-projection posterior based on the conjugate normal prior.

Prior and projection-posterior

Wang and G. (2023, AoS).

- $Y = f(\boldsymbol{X}) + \varepsilon$ with $\boldsymbol{X} \sim G$ on $[0,1]^d$, f multivariate monotone, that is, $f(\boldsymbol{x}_1) \leq f(\boldsymbol{x}_2)$ if $\boldsymbol{x}_1 \preceq \boldsymbol{x}_2$, where $\boldsymbol{x}_1 \preceq \boldsymbol{x}_2$ means that $x_{1,k} \leq x_{2,k}$, $k = 1, \ldots, d$.
- Prior: On piecewise constant functions $f = \sum_{\boldsymbol{j}} \theta_{\boldsymbol{j}} \mathbb{1}_{I_{\boldsymbol{j}}}$ on the blocks $I_{\boldsymbol{j}} = \prod_{k=1}^d ((j_k - 1)/J, j_k/J]$ with a prior on the random heights $\theta_{\boldsymbol{j}}$.
- Monotonicity will be ensured if the array of coefficients $\boldsymbol{\theta}$ belongs to the cone $\mathcal{C} := \{\boldsymbol{\theta} \in \mathbb{R}^{J^d} : \theta_{\boldsymbol{j}_1} \leq \theta_{\boldsymbol{j}_2} \text{ if } \boldsymbol{j}_1 \preceq \boldsymbol{j}_2\}$.
- $\mathbb{L}_1$-approximation rate is J^{-1} using a total of J^d basis functions; the $\mathbb{L}_2$-rate is $J^{-1/2}$.
- Optimal $\mathbb{L}_1$-contraction rate $n^{-1/(d+2)}$ using the $\mathbb{L}_1$ or $\mathbb{L}_2$-projection posterior based on the conjugate normal prior.
- This also leads to a consistent Bayesian test as before.

Coverage for $f(\boldsymbol{x}_0)$

- Unlike the univariate case, the projection-posterior approach does not lead to the optimal rate of contraction for $f(\boldsymbol{x}_0)$.

Coverage for $f(\boldsymbol{x}_0)$

- Unlike the univariate case, the projection-posterior approach does not lead to the optimal rate of contraction for $f(\boldsymbol{x}_0)$.
- We instead consider the 'maxi-min' immersion map

$$\underline{\iota}(f)(\boldsymbol{x}_0) = \max_{\boldsymbol{j}_1 \preceq \boldsymbol{j}_0(\boldsymbol{x}_0)} \min_{\substack{\boldsymbol{j}_0(\boldsymbol{x}_0) \preceq \boldsymbol{j}_2 \\ N_{[\boldsymbol{j}_1:\boldsymbol{j}_2]}>0}} \frac{\sum_{\boldsymbol{j}\in[\boldsymbol{j}_1:\boldsymbol{j}_2]} N_{\boldsymbol{j}}\theta_{\boldsymbol{j}}}{N_{[\boldsymbol{j}_1:\boldsymbol{j}_2]}}$$

Coverage for $f(\boldsymbol{x}_0)$

- Unlike the univariate case, the projection-posterior approach does not lead to the optimal rate of contraction for $f(\boldsymbol{x}_0)$.
- We instead consider the 'maxi-min' immersion map

$$\underline{\iota}(f)(\boldsymbol{x}_0) = \max_{\boldsymbol{j}_1 \preceq \boldsymbol{j}_0(\boldsymbol{x}_0)} \min_{\substack{\boldsymbol{j}_0(\boldsymbol{x}_0) \preceq \boldsymbol{j}_2 \\ N_{[\boldsymbol{j}_1:\boldsymbol{j}_2]} > 0}} \frac{\sum_{\boldsymbol{j} \in [\boldsymbol{j}_1:\boldsymbol{j}_2]} N_{\boldsymbol{j}} \theta_{\boldsymbol{j}}}{N_{[\boldsymbol{j}_1:\boldsymbol{j}_2]}}$$

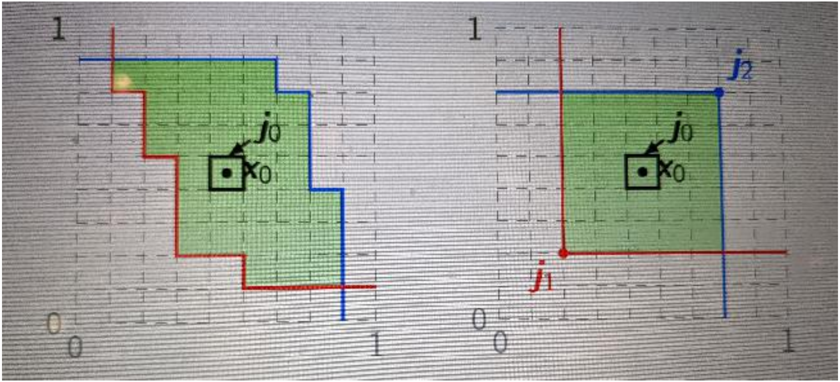
- The clue comes from the operation used in the construction of the maxi-min estimator of Han and Zhang (2020, Ann Stat).

Coverage for $f(\boldsymbol{x}_0)$

- Unlike the univariate case, the projection-posterior approach does not lead to the optimal rate of contraction for $f(\boldsymbol{x}_0)$.
- We instead consider the 'maxi-min' immersion map

$$\underline{\iota}(f)(\boldsymbol{x}_0) = \max_{\boldsymbol{j}_1 \preceq \boldsymbol{j}_0(\boldsymbol{x}_0)} \min_{\substack{\boldsymbol{j}_0(\boldsymbol{x}_0) \preceq \boldsymbol{j}_2 \\ N_{[\boldsymbol{j}_1:\boldsymbol{j}_2]}>0}} \frac{\sum_{\boldsymbol{j}\in[\boldsymbol{j}_1:\boldsymbol{j}_2]} N_{\boldsymbol{j}}\theta_{\boldsymbol{j}}}{N_{[\boldsymbol{j}_1:\boldsymbol{j}_2]}}$$

- The clue comes from the operation used in the construction of the maxi-min estimator of Han and Zhang (2020, Ann Stat).
- A major difference with theirs is that the operation is in a discrete domain, and hence is also simpler, because of the binning through the hypercubes used in constructing the basis of step functions.



1
j_0
x_0
0
0
1
1
j_2
j_0
x_0
j_1
0
0
1

- A natural alternative is obtained by switching the maximization and the minimization operations

$$\bar{\iota}(f)(\boldsymbol{x}_0) = \min_{\boldsymbol{j}_0(\boldsymbol{x}_0) \preceq \boldsymbol{j}_2} \max_{\substack{\boldsymbol{j}_1 \preceq \boldsymbol{j}_0(\boldsymbol{x}_0) \\ N_{[\boldsymbol{j}_1 : \boldsymbol{j}_2]} > 0}} \frac{\sum_{\boldsymbol{j} \in [\boldsymbol{j}_1 : \boldsymbol{j}_2]} N_{\boldsymbol{j}} \theta_{\boldsymbol{j}}}{N_{[\boldsymbol{j}_1 : \boldsymbol{j}_2]}}.$$

- A natural alternative is obtained by switching the maximization and the minimization operations

$$\bar{\iota}(f)(\boldsymbol{x}_0) = \min_{\boldsymbol{j}_0(\boldsymbol{x}_0) \preceq \boldsymbol{j}_2} \max_{\substack{\boldsymbol{j}_1 \preceq \boldsymbol{j}_0(\boldsymbol{x}_0) \\ N_{[\boldsymbol{j}_1:\boldsymbol{j}_2]} > 0}} \frac{\sum_{\boldsymbol{j} \in [\boldsymbol{j}_1:\boldsymbol{j}_2]} N_{\boldsymbol{j}} \theta_{\boldsymbol{j}}}{N_{[\boldsymbol{j}_1:\boldsymbol{j}_2]}}.$$

- Both operations $\underline{\iota}$ and $\bar{\iota}$ are asymmetric in terms of the direction. A symmetric operation is obtained by averaging: $\iota(f)(\boldsymbol{x}_0) = (\underline{\iota}(f)(\boldsymbol{x}_0) + \bar{\iota}(f)(\boldsymbol{x}_0))/2$.

- A natural alternative is obtained by switching the maximization and the minimization operations

$$\bar{\iota}(f)(\boldsymbol{x}_0) = \min_{\boldsymbol{j}_0(\boldsymbol{x}_0) \preceq \boldsymbol{j}_2} \max_{\substack{\boldsymbol{j}_1 \preceq \boldsymbol{j}_0(\boldsymbol{x}_0) \\ N_{[\boldsymbol{j}_1:\boldsymbol{j}_2]} > 0}} \frac{\sum_{\boldsymbol{j} \in [\boldsymbol{j}_1:\boldsymbol{j}_2]} N_{\boldsymbol{j}} \theta_{\boldsymbol{j}}}{N_{[\boldsymbol{j}_1:\boldsymbol{j}_2]}}.$$

- Both operations $\underline{\iota}$ and $\bar{\iota}$ are asymmetric in terms of the direction. A symmetric operation is obtained by averaging: $\iota(f)(\boldsymbol{x}_0) = (\underline{\iota}(f)(\boldsymbol{x}_0) + \bar{\iota}(f)(\boldsymbol{x}_0))/2$.
- Denote the images under the corresponding immersion maps by $f_*(\boldsymbol{x}_0)$, $f^*(\boldsymbol{x}_0)$, and $\tilde{f}(x_0)$ respectively.

- A natural alternative is obtained by switching the maximization and the minimization operations

$$\bar{\iota}(f)(\boldsymbol{x}_0) = \min_{\boldsymbol{j}_0(\boldsymbol{x}_0) \preceq \boldsymbol{j}_2} \max_{\substack{\boldsymbol{j}_1 \preceq \boldsymbol{j}_0(\boldsymbol{x}_0) \\ N_{[\boldsymbol{j}_1:\boldsymbol{j}_2]}>0}} \frac{\sum_{\boldsymbol{j}\in[\boldsymbol{j}_1:\boldsymbol{j}_2]} N_{\boldsymbol{j}}\theta_{\boldsymbol{j}}}{N_{[\boldsymbol{j}_1:\boldsymbol{j}_2]}}.$$

- Both operations $\underline{\iota}$ and $\bar{\iota}$ are asymmetric in terms of the direction. A symmetric operation is obtained by averaging: $\iota(f)(\boldsymbol{x}_0) = (\underline{\iota}(f)(\boldsymbol{x}_0) + \bar{\iota}(f)(\boldsymbol{x}_0))/2$.
- Denote the images under the corresponding immersion maps by $f_*(\boldsymbol{x}_0)$, $f^*(\boldsymbol{x}_0)$, and $\tilde{f}(x_0)$ respectively.
- In the univariate case, all three operations coincide and reduce to the empirical $\mathbb{L}_2$-projection on monotone functions for stepwise functions given by the standard isotonization procedure for the step-heights.

Key assumptions

- Local smoothness: let β_k be the order of the first non-zero derivative of f at $\boldsymbol{x}_0$ along the kth coordinate, that is, $\beta_k = \min_{l \geq 1}\{l : \partial_k^l f(\boldsymbol{x}_0) \neq 0\}$.

Key assumptions

- Local smoothness: let β_k be the order of the first non-zero derivative of f at $\boldsymbol{x}_0$ along the kth coordinate, that is, $\beta_k = \min_{l\geq 1}\{l : \partial_k^l f(\boldsymbol{x}_0) \neq 0\}$.
- Assume $1 \leq \beta_1, \ldots, \beta_s < \infty$, and that $\beta_{s+1} = \ldots = \beta_d = \infty$ for some $0 \leq s \leq d$.

Key assumptions

- Local smoothness: let β_k be the order of the first non-zero derivative of f at $\boldsymbol{x}_0$ along the kth coordinate, that is, $\beta_k = \min_{l \geq 1}\{l : \partial_k^l f(\boldsymbol{x}_0) \neq 0\}$.
- Assume $1 \leq \beta_1, \dots, \beta_s < \infty$, and that $\beta_{s+1} = \dots = \beta_d = \infty$ for some $0 \leq s \leq d$.
- For a positive sequence $\omega_n \downarrow 0$, set $\boldsymbol{r}_n = (\omega_n^{1/\beta_1}, \dots, \omega_n^{1/\beta_s}, 1, \dots, 1)^{\mathrm{T}}$. For any $t > 0$,

$$\sup_{|x_k - x_{0,k}| \leq t r_{n,k}} \left| f_0(\boldsymbol{x}) - f_0(\boldsymbol{x}_0) - \sum_{\boldsymbol{l} \in L} \frac{\partial^{\boldsymbol{l}} f_0(\boldsymbol{x}_0)}{\boldsymbol{l}!} (\boldsymbol{x} - \boldsymbol{x}_0)^{\boldsymbol{l}} \right| = o(\omega_n),$$

where
$L = \{\boldsymbol{l} : 0 < \sum_{k=1}^{s} l_k/\beta_k \leq 1 \text{ and } l_k = 0, \text{ for } k = s+1, \dots, d\}$.

Key assumptions

- Local smoothness: let β_k be the order of the first non-zero derivative of f at $\boldsymbol{x}_0$ along the kth coordinate, that is, $\beta_k = \min_{l \geq 1}\{l : \partial_k^l f(\boldsymbol{x}_0) \neq 0\}$.
- Assume $1 \leq \beta_1, \ldots, \beta_s < \infty$, and that $\beta_{s+1} = \ldots = \beta_d = \infty$ for some $0 \leq s \leq d$.
- For a positive sequence $\omega_n \downarrow 0$, set $\boldsymbol{r}_n = (\omega_n^{1/\beta_1}, \ldots, \omega_n^{1/\beta_s}, 1, \ldots, 1)^{\mathrm{T}}$. For any $t > 0$,

$$\sup_{|x_k - x_{0,k}| \leq t r_{n,k}} \left| f_0(\boldsymbol{x}) - f_0(\boldsymbol{x}_0) - \sum_{\boldsymbol{l} \in L} \frac{\partial^{\boldsymbol{l}} f_0(\boldsymbol{x}_0)}{\boldsymbol{l}!} (\boldsymbol{x} - \boldsymbol{x}_0)^{\boldsymbol{l}} \right| = o(\omega_n),$$

  where
  $L = \{\boldsymbol{l} : 0 < \sum_{k=1}^{s} l_k/\beta_k \leq 1 \text{ and } l_k = 0, \text{ for } k = s+1, \ldots, d\}$.
- β_k is odd, $k \leq s$; $\partial^{\boldsymbol{l}} f_0(\boldsymbol{x}_0) = 0$ if $\sum l_k/\beta_k < 1$.

Notations

Let H_1, H_2 be independent centered Gaussian processes on $\mathbb{R}^d_{\geq 0} \times \mathbb{R}^d_{\geq 0}$ with covariance kernel
$K(\boldsymbol{u}, \boldsymbol{v}) = \prod_{k=1}^{s}(u_k \wedge u'_k + v_k \wedge v'_k) D_s(\boldsymbol{u} \wedge \boldsymbol{u}', \boldsymbol{v} \wedge \boldsymbol{v}')$, where

$$D_s(\boldsymbol{u} \wedge \boldsymbol{u}', \boldsymbol{v} \wedge \boldsymbol{v}') = \int_{\substack{x_k \in [(\boldsymbol{x}_0 - \boldsymbol{u})_k, (\boldsymbol{x}_0 + \boldsymbol{v})_k] \cap [0,1] \\ s+1 \leq k \leq d}} g(x_{0,1}, \ldots, x_{0,s}, x_{s+1}, \ldots, x_d) \mathrm{d}x_s$$

with $D_d(\boldsymbol{u}, \boldsymbol{v}) = g(\boldsymbol{x}_0)$, where g is the probability density function of $\boldsymbol{X}$.

Let

$$U(\boldsymbol{u},\boldsymbol{v}) = \frac{\sigma_0 H_1(\boldsymbol{u},\boldsymbol{v})}{\prod_{k=1}^{s}(u_k+v_k)D_s(\boldsymbol{u},\boldsymbol{v})} + \frac{\sigma_0 H_2(\boldsymbol{u},\boldsymbol{v})}{\prod_{k=1}^{s}(u_k+v_k)D_s(\boldsymbol{u},\boldsymbol{v})} + \sum_{\boldsymbol{l}\in L^*} \frac{\partial^{\boldsymbol{l}} f_0(x_0)}{(\boldsymbol{l}+\boldsymbol{1})!} \prod_{k=1}^{s} \frac{v_k^{l_k+1} - (-u_k)^{l_k+1}}{u_k+v_k}.$$

$$Z_* = \sup_{\substack{\boldsymbol{u}\succeq\boldsymbol{0} \\ u_k\le x_{0,k} \\ s+1\le k\le d}} \inf_{\substack{\boldsymbol{v}\succeq\boldsymbol{0} \\ v_k\le 1-x_{0,k} \\ s+1\le k\le d}} U(\boldsymbol{u},\boldsymbol{v}), \quad Z^* = \inf_{\substack{\boldsymbol{v}\succeq\boldsymbol{0} \\ v_k\le 1-x_{0,k} \\ s+1\le k\le d}} \sup_{\substack{\boldsymbol{u}\succeq\boldsymbol{0} \\ u_k\le x_{0,k} \\ s+1\le k\le d}} U(\boldsymbol{u},\boldsymbol{v}).$$

Weak limit

Theorem

Let $\omega_n = n^{-1/(2+\sum_{k=1}^{s}\beta_k^{-1})}$. *Let* $J_k \gg r_{n,k}^{-1}$, *for each* $k = 1, \ldots, d$, *and* $\prod_{k=1}^{d} J_k \ll n\omega_n$. *For any* $z \in \mathbb{R}$, *we have*

$$\Pi(\omega_n^{-1}(f_*(\boldsymbol{x}_0) - f_0(\boldsymbol{x}_0)) \leq z|\mathbb{D}_n) \rightsquigarrow \mathrm{P}(Z_* \leq z|H_1);$$
$$\Pi(\omega_n^{-1}(f^*(\boldsymbol{x}_0) - f_0(\boldsymbol{x}_0)) \leq z|\mathbb{D}_n) \rightsquigarrow \mathrm{P}(Z^* \leq z|H_1);$$
$$\Pi(\omega_n^{-1}(\tilde{f}(\boldsymbol{x}_0) - f_0(\boldsymbol{x}_0)) \leq z|\mathbb{D}_n) \rightsquigarrow \mathrm{P}((Z_* + Z^*)/2 \leq z|H_1).$$

Steps in the proof

- No switch relation (the key technique used in the univariate case), but the dimension doubles, as we consider increments with two free ends.

Steps in the proof

- No switch relation (the key technique used in the univariate case), but the dimension doubles, as we consider increments with two free ends.
- Write the original posterior process as a sum of several terms, one is for posterior variation, one is the discrepancy between the posterior mean and the empirical analog, and another is for sampling variation of the empirical.

Steps in the proof

- No switch relation (the key technique used in the univariate case), but the dimension doubles, as we consider increments with two free ends.
- Write the original posterior process as a sum of several terms, one is for posterior variation, one is the discrepancy between the posterior mean and the empirical analog, and another is for sampling variation of the empirical.
- The first process goes to a Gaussian process conditionally on the data, the second to zero, and the third weakly to an independent Gaussian process.

Steps in the proof

- No switch relation (the key technique used in the univariate case), but the dimension doubles, as we consider increments with two free ends.
- Write the original posterior process as a sum of several terms, one is for posterior variation, one is the discrepancy between the posterior mean and the empirical analog, and another is for sampling variation of the empirical.
- The first process goes to a Gaussian process conditionally on the data, the second to zero, and the third weakly to an independent Gaussian process.
- Key technique: truncate the domain, and transition to the whole in the limit.

Steps in the proof

- No switch relation (the key technique used in the univariate case), but the dimension doubles, as we consider increments with two free ends.
- Write the original posterior process as a sum of several terms, one is for posterior variation, one is the discrepancy between the posterior mean and the empirical analog, and another is for sampling variation of the empirical.
- The first process goes to a Gaussian process conditionally on the data, the second to zero, and the third weakly to an independent Gaussian process.
- Key technique: truncate the domain, and transition to the whole in the limit.
- Tightness verification: Maximal inequality for a martingale with respect to multi-index given a partial order.

Steps in the proof

- No switch relation (the key technique used in the univariate case), but the dimension doubles, as we consider increments with two free ends.
- Write the original posterior process as a sum of several terms, one is for posterior variation, one is the discrepancy between the posterior mean and the empirical analog, and another is for sampling variation of the empirical.
- The first process goes to a Gaussian process conditionally on the data, the second to zero, and the third weakly to an independent Gaussian process.
- Key technique: truncate the domain, and transition to the whole in the limit.
- Tightness verification: Maximal inequality for a martingale with respect to multi-index given a partial order.
- Tightness of the functional — max/min activity within a compact domain with high probability.

Special cases

- If $L = \{\beta_k \boldsymbol{e}_k : 1 \leq k \leq d\}$, then $U =_d \tilde{A}_\beta \tilde{U}$, where

Special cases

- If $L = \{\beta_k \boldsymbol{e}_k : 1 \le k \le d\}$, then $U =_d \tilde{A}_\beta \tilde{U}$, where
  - $\tilde{A}_{\boldsymbol{\beta}} = \left(\frac{\sigma_0^2}{g(\boldsymbol{x}_0)} \prod_{k=1}^d \left(\frac{\partial_k^{\beta_k} f_0(\boldsymbol{x}_0)}{(\beta_k+1)!}\right)^{1/\beta_k}\right)^{1/(2+\sum_{k=1}^d \beta_k^{-1})} > 0,$

Special cases

- If $L = \{\beta_k \boldsymbol{e}_k : 1 \le k \le d\}$, then $U =_d \tilde{A}_\beta \tilde{U}$, where
  - $\tilde{A}_{\boldsymbol{\beta}} = \big(\frac{\sigma_0^2}{g(\boldsymbol{x}_0)} \prod_{k=1}^d \big(\frac{\partial_k^{\beta_k} f_0(\boldsymbol{x}_0)}{(\beta_k+1)!}\big)^{1/\beta_k}\big)^{1/(2+\sum_{k=1}^d \beta_k^{-1})} > 0,$
  - $\tilde{U} := \dfrac{\tilde{H}_1(\boldsymbol{u}, \boldsymbol{v})}{\prod_{k=1}^d (u_k + v_k)} + \dfrac{\tilde{H}_2(\boldsymbol{u}, \boldsymbol{v})}{\prod_{k=1}^d (u_k + v_k)} + \displaystyle\sum_{k=1}^d \dfrac{v_k^{\beta_k+1} - (-u_k)^{\beta_k+1}}{u_k + v_k},$

Special cases

- If $L = \{\beta_k \boldsymbol{e}_k : 1 \leq k \leq d\}$, then $U =_d \tilde{A}_\beta \tilde{U}$, where
  - $\tilde{A}_{\boldsymbol{\beta}} = \left(\frac{\sigma_0^2}{g(\boldsymbol{x}_0)} \prod_{k=1}^d \left(\frac{\partial_k^{\beta_k} f_0(\boldsymbol{x}_0)}{(\beta_k+1)!}\right)^{1/\beta_k}\right)^{1/(2+\sum_{k=1}^d \beta_k^{-1})} > 0,$
  - $\tilde{U} := \dfrac{\tilde{H}_1(\boldsymbol{u}, \boldsymbol{v})}{\prod_{k=1}^d (u_k + v_k)} + \dfrac{\tilde{H}_2(\boldsymbol{u}, \boldsymbol{v})}{\prod_{k=1}^d (u_k + v_k)} + \displaystyle\sum_{k=1}^d \frac{v_k^{\beta_k+1} - (-u_k)^{\beta_k+1}}{u_k + v_k},$
  - $\tilde{H}_1$ and $\tilde{H}_2$ are two independent centered Gaussian processes with covariance kernel $\prod_{k=1}^d (u_k \wedge u_k' + v_k \wedge v_k')$, $(\boldsymbol{u}, \boldsymbol{v}), (\boldsymbol{u}', \boldsymbol{v}') \in \mathbb{R}_{\geq 0}^d \times \mathbb{R}_{\geq 0}^d$.

Special cases

- If $L = \{\beta_k \boldsymbol{e}_k : 1 \leq k \leq d\}$, then $U =_d \tilde{A}_\beta \tilde{U}$, where
  - $\tilde{A}_{\boldsymbol{\beta}} = \Big(\frac{\sigma_0^2}{g(\boldsymbol{x}_0)} \prod_{k=1}^d \big(\frac{\partial_k^{\beta_k} f_0(\boldsymbol{x}_0)}{(\beta_k+1)!}\big)^{1/\beta_k}\Big)^{1/(2+\sum_{k=1}^d \beta_k^{-1})} > 0,$
  - $\tilde{U} := \dfrac{\tilde{H}_1(\boldsymbol{u}, \boldsymbol{v})}{\prod_{k=1}^d (u_k + v_k)} + \dfrac{\tilde{H}_2(\boldsymbol{u}, \boldsymbol{v})}{\prod_{k=1}^d (u_k + v_k)} + \sum_{k=1}^d \dfrac{v_k^{\beta_k+1} - (-u_k)^{\beta_k+1}}{u_k + v_k},$
  - $\tilde{H}_1$ and $\tilde{H}_2$ are two independent centered Gaussian processes with covariance kernel $\prod_{k=1}^d (u_k \wedge u_k' + v_k \wedge v_k')$, $(\boldsymbol{u}, \boldsymbol{v}), (\boldsymbol{u}', \boldsymbol{v}') \in \mathbb{R}_{\geq 0}^d \times \mathbb{R}_{\geq 0}^d$.
- This happens if

Special cases

- If $L = \{\beta_k \boldsymbol{e}_k : 1 \le k \le d\}$, then $U =_d \tilde{A}_\beta \tilde{U}$, where
  - $\tilde{A}_{\boldsymbol{\beta}} = \left(\frac{\sigma_0^2}{g(\boldsymbol{x}_0)} \prod_{k=1}^d \left(\frac{\partial_k^{\beta_k} f_0(\boldsymbol{x}_0)}{(\beta_k+1)!}\right)^{1/\beta_k}\right)^{1/(2+\sum_{k=1}^d \beta_k^{-1})} > 0$,
  - $\tilde{U} := \dfrac{\tilde{H}_1(\boldsymbol{u}, \boldsymbol{v})}{\prod_{k=1}^d (u_k + v_k)} + \dfrac{\tilde{H}_2(\boldsymbol{u}, \boldsymbol{v})}{\prod_{k=1}^d (u_k + v_k)} + \sum_{k=1}^d \dfrac{v_k^{\beta_k+1} - (-u_k)^{\beta_k+1}}{u_k + v_k}$,
  - $\tilde{H}_1$ and $\tilde{H}_2$ are two independent centered Gaussian processes with covariance kernel $\prod_{k=1}^d (u_k \wedge u_k' + v_k \wedge v_k')$, $(\boldsymbol{u}, \boldsymbol{v}), (\boldsymbol{u}', \boldsymbol{v}') \in \mathbb{R}_{\ge 0}^d \times \mathbb{R}_{\ge 0}^d$.
- This happens if
  - $\beta_1, \ldots, \beta_d$ are relative primes;

Special cases

- If $L = \{\beta_k \boldsymbol{e}_k : 1 \le k \le d\}$, then $U =_d \tilde{A}_\beta \tilde{U}$, where
  - $\tilde{A}_{\boldsymbol{\beta}} = \Big(\frac{\sigma_0^2}{g(\boldsymbol{x}_0)} \prod_{k=1}^d \big(\frac{\partial_k^{\beta_k} f_0(\boldsymbol{x}_0)}{(\beta_k+1)!}\big)^{1/\beta_k}\Big)^{1/(2+\sum_{k=1}^d \beta_k^{-1})} > 0$,
  - $\tilde{U} := \dfrac{\tilde{H}_1(\boldsymbol{u}, \boldsymbol{v})}{\prod_{k=1}^d (u_k + v_k)} + \dfrac{\tilde{H}_2(\boldsymbol{u}, \boldsymbol{v})}{\prod_{k=1}^d (u_k + v_k)} + \displaystyle\sum_{k=1}^d \frac{v_k^{\beta_k+1} - (-u_k)^{\beta_k+1}}{u_k + v_k}$,
  - $\tilde{H}_1$ and $\tilde{H}_2$ are two independent centered Gaussian processes with covariance kernel $\prod_{k=1}^d (u_k \wedge u_k' + v_k \wedge v_k')$, $(\boldsymbol{u}, \boldsymbol{v}), (\boldsymbol{u}', \boldsymbol{v}') \in \mathbb{R}_{\ge 0}^d \times \mathbb{R}_{\ge 0}^d$.
- This happens if
  - $\beta_1, \ldots, \beta_d$ are relative primes;
  - $\beta_1 = \cdots = \beta_d = 1$;

Special cases

- If $L = \{\beta_k \boldsymbol{e}_k : 1 \leq k \leq d\}$, then $U =_d \tilde{A}_\beta \tilde{U}$, where
  - $\tilde{A}_{\boldsymbol{\beta}} = \left(\frac{\sigma_0^2}{g(\boldsymbol{x}_0)} \prod_{k=1}^d \left(\frac{\partial_k^{\beta_k} f_0(\boldsymbol{x}_0)}{(\beta_k+1)!}\right)^{1/\beta_k}\right)^{1/(2+\sum_{k=1}^d \beta_k^{-1})} > 0,$
  - $\tilde{U} := \dfrac{\tilde{H}_1(\boldsymbol{u}, \boldsymbol{v})}{\prod_{k=1}^d (u_k + v_k)} + \dfrac{\tilde{H}_2(\boldsymbol{u}, \boldsymbol{v})}{\prod_{k=1}^d (u_k + v_k)} + \sum_{k=1}^d \dfrac{v_k^{\beta_k+1} - (-u_k)^{\beta_k+1}}{u_k + v_k},$
  - $\tilde{H}_1$ and $\tilde{H}_2$ are two independent centered Gaussian processes with covariance kernel $\prod_{k=1}^d (u_k \wedge u_k' + v_k \wedge v_k')$, $(\boldsymbol{u}, \boldsymbol{v}), (\boldsymbol{u}', \boldsymbol{v}') \in \mathbb{R}_{\geq 0}^d \times \mathbb{R}_{\geq 0}^d$.
- This happens if
  - $\beta_1, \ldots, \beta_d$ are relative primes;
  - $\beta_1 = \cdots = \beta_d = 1$;
  - the univariate case considered earlier.

Coverage of a credible interval

- Consider the $(1-\gamma)$-quantile of the $\underline{\iota}$-immersion posterior $Q^{(1)}_{n,\gamma} = \inf\{z : \Pi(f_*(\boldsymbol{x}_0) \leq z | \mathbb{D}_n) \geq 1-\gamma\}$, and the corresponding one-sided credible interval $(-\infty, Q^{(1)}_{n,\gamma}]$.

Coverage of a credible interval

- Consider the $(1-\gamma)$-quantile of the $\underline{\iota}$-immersion posterior $Q^{(1)}_{n,\gamma} = \inf\{z : \Pi(f_*(\boldsymbol{x}_0) \leq z|\mathbb{D}_n) \geq 1-\gamma\}$, and the corresponding one-sided credible interval $(-\infty, Q^{(1)}_{n,\gamma}]$.
- Its asymptotic coverage is $\mathrm{P}(\underbrace{\mathrm{P}(\sup_{\boldsymbol{u}\succeq\mathbf{0}} \inf_{\boldsymbol{v}\succeq\mathbf{0}} \tilde{U}(\boldsymbol{u},\boldsymbol{v}) \leq 0|\tilde{H}_1)}_{Z_B^{(1)}} \leq 1-\gamma)$.

Coverage of a credible interval

- Consider the $(1-\gamma)$-quantile of the $\underline{\iota}$-immersion posterior $Q^{(1)}_{n,\gamma} = \inf\{z : \Pi(f_*(\boldsymbol{x}_0) \le z | \mathbb{D}_n) \ge 1-\gamma\}$, and the corresponding one-sided credible interval $(-\infty, Q^{(1)}_{n,\gamma}]$.
- Its asymptotic coverage is $\mathrm{P}(\underbrace{\mathrm{P}(\sup_{\boldsymbol{u}\succeq\boldsymbol{0}} \inf_{\boldsymbol{v}\succeq\boldsymbol{0}} \tilde{U}(\boldsymbol{u},\boldsymbol{v}) \le 0 | \tilde{H}_1)}_{Z^{(1)}_B} \le 1-\gamma)$.
- Corresponding statement for a two-sided equal-tailed posterior quantile interval follows.

Coverage of a credible interval

- Consider the $(1-\gamma)$-quantile of the $\underline{\iota}$-immersion posterior $Q^{(1)}_{n,\gamma} = \inf\{z : \Pi(f_*(\boldsymbol{x}_0) \leq z | \mathbb{D}_n) \geq 1-\gamma\}$, and the corresponding one-sided credible interval $(-\infty, Q^{(1)}_{n,\gamma}]$.
- Its asymptotic coverage is $\mathrm{P}(\underbrace{\mathrm{P}(\sup_{\boldsymbol{u}\succeq\boldsymbol{0}} \inf_{\boldsymbol{v}\succeq\boldsymbol{0}} \tilde{U}(\boldsymbol{u},\boldsymbol{v}) \leq 0 | \tilde{H}_1)}_{Z^{(1)}_B} \leq 1-\gamma)$.
- Corresponding statement for a two-sided equal-tailed posterior quantile interval follows.
- Similar statements for the other two immersion maps with corresponding changes in the limiting process.

Reverse Cox phenomenon

- We tabulated the limiting distribution through simulation.

Reverse Cox phenomenon

- We tabulated the limiting distribution through simulation.
- Like the Bayes-Chernoff distribution in the univariate case, the limiting distributions are slightly more concentrated than the uniform.

Reverse Cox phenomenon

- We tabulated the limiting distribution through simulation.
- Like the Bayes-Chernoff distribution in the univariate case, the limiting distributions are slightly more concentrated than the uniform.
- As a result, the limiting coverage is slightly higher than the nominal credibility level.

Reverse Cox phenomenon

- We tabulated the limiting distribution through simulation.
- Like the Bayes-Chernoff distribution in the univariate case, the limiting distributions are slightly more concentrated than the uniform.
- As a result, the limiting coverage is slightly higher than the nominal credibility level.
- We ran a simulation experiment with the following multivariate monotone functions: $(x_1 + x_2)^2$; $\sqrt{x_1 + x_2}$; $x_1 x_2$; $e^{x_1+x_2}$; $e^{x_1 x_2}$, independent uniform predictors and normal errors, and sample sizes $n = 200, 500, 1000, 2000$.

Reverse Cox phenomenon

- We tabulated the limiting distribution through simulation.
- Like the Bayes-Chernoff distribution in the univariate case, the limiting distributions are slightly more concentrated than the uniform.
- As a result, the limiting coverage is slightly higher than the nominal credibility level.
- We ran a simulation experiment with the following multivariate monotone functions: $(x_1 + x_2)^2$; $\sqrt{x_1 + x_2}$; x_1x_2; $e^{x_1+x_2}$; $e^{x_1x_2}$, independent uniform predictors and normal errors, and sample sizes $n = 200, 500, 1000, 2000$.
- Overcoverage was noted as predicted by the theory.

Reverse Cox phenomenon

- We tabulated the limiting distribution through simulation.
- Like the Bayes-Chernoff distribution in the univariate case, the limiting distributions are slightly more concentrated than the uniform.
- As a result, the limiting coverage is slightly higher than the nominal credibility level.
- We ran a simulation experiment with the following multivariate monotone functions: $(x_1 + x_2)^2$; $\sqrt{x_1 + x_2}$; $x_1 x_2$; $e^{x_1 + x_2}$; $e^{x_1 x_2}$, independent uniform predictors and normal errors, and sample sizes $n = 200, 500, 1000, 2000$.
- Overcoverage was noted as predicted by the theory.
- Using the numerical tables of the limiting distribution, we can reset the credibility level lower to obtain a targeted limiting coverage.

Reverse Cox phenomenon

- We tabulated the limiting distribution through simulation.
- Like the Bayes-Chernoff distribution in the univariate case, the limiting distributions are slightly more concentrated than the uniform.
- As a result, the limiting coverage is slightly higher than the nominal credibility level.
- We ran a simulation experiment with the following multivariate monotone functions: $(x_1+x_2)^2$; $\sqrt{x_1+x_2}$; x_1x_2; $e^{x_1+x_2}$; $e^{x_1x_2}$, independent uniform predictors and normal errors, and sample sizes $n = 200, 500, 1000, 2000$.
- Overcoverage was noted as predicted by the theory.
- Using the numerical tables of the limiting distribution, we can reset the credibility level lower to obtain a targeted limiting coverage.
- Compared with the Han-Zhang procedure, the corrected intervals are significantly shorter (even the uncorrected ones are shorter), and have better coverage too, in most cases.

Thank you