The Annals of Statistics
2021, Vol. 49, No. 2, 1011–1028
https://doi.org/10.1214/20-AOS1989

COVERAGE OF CREDIBLE INTERVALS IN NONPARAMETRIC MONOTONE REGRESSION

BY MOUMITA CHAKRABORTY* AND SUBHASHIS GHOSAL†

Department of Statistics, North Carolina State University, *mchakra@ncsu.edu; †sghosal@ncsu.edu*

For nonparametric univariate regression under a monotonicity constraint on the regression function f, we study the coverage of a Bayesian credible interval for $f(x_0)$, where x_0 is an interior point. Analysis of the posterior becomes a lot more tractable by considering a "projection-posterior" distribution based on a finite random series of step functions with normal basis coefficients as a prior for f. A sample f from the resulting conjugate posterior distribution is projected on the space of monotone increasing functions to obtain a monotone function f^* closest to f, inducing the "projection-posterior." We use projection-posterior samples to obtain credible intervals for $f(x_0)$. We obtain the asymptotic coverage of the credible interval thus constructed and observe that it is free of nuisance parameters involving the true function. We observe a very interesting phenomenon that the coverage is typically higher than the nominal credibility level, the opposite of a phenomenon observed by Cox (*Ann. Statist.* **21** (1993) 903–923) in the Gaussian sequence model. We further show that a recalibration gives the right asymptotic coverage by starting from a lower credibility level that can be explicitly calculated.

1. Introduction. We consider a nonparametric regression model for a response variable Y with respect to a predictor variable $X \in [0, 1]$ given by $Y = f(X) + \varepsilon$, where f is a monotone increasing function on $[0, 1]$ and ε is a mean-zero random error with finite variance σ^2. Inference on f under monotonicity or some other shape restriction may arise naturally in many fields such as epidemiology, climate change, reliability and biomedical studies. Occasionally, an understanding of the physical phenomenon suggests such a shape restriction. The restriction needs to be incorporated to draw meaningful conclusions. Moreover, the shape information often allows inference on the underlying function without requiring global smoothness. Regression under monotonicity restriction is commonly known as isotonic regression. A graphical representation of isotonic regression in terms of the greatest convex minorant (GCM) of a cumulative sum diagram can be found in Barlow and Brunk [7]. The Pool-Adjacent-Violators Algorithm (PAVA) gives successive approximation to the GCM, and is the most commonly used algorithm to obtain the isotonic regression estimator (see Ayer et al. [2], pp. 9–15, Section 2.3 of Barlow et al. [6] or Leeuw et al. [24]). If the working model for the error distribution is Gaussian, the nonparametric maximum likelihood estimator (MLE) under the monotonicity restriction is given by the isotonic regression estimator. In this paper, we consider a Bayesian approach to isotonic regression, quantify the uncertainty in the value of f at a point through a credible interval and study its coverage.

Properties of estimators under a shape restriction were studied by various authors. The nonparametric MLE of a monotone decreasing density was characterized by Grenander [30]. Rao [47] showed that Grenander's estimator based on n independent observations, centered and scaled by $n^{1/3}$, converges to the Chernoff distribution (Chernoff [22]). Groeneboom and

Received March 2020; revised June 2020.
MSC2020 subject classifications. Primary 62G08, 62F15, 62G05, 62G10, 62G15, 62G20.
Key words and phrases. Nonparametric regression, monotonicity, credible set, coverage, Chernoff's distribution, projection-posterior.

Wellner [33] derived the pointwise asymptotic distribution of the MLE of the distribution function in current status time-to-event data. Brunk [15] derived the asymptotic distribution of the isotonic regression estimator at an interior point. Similar results can be found in Huang and Zhang [38] and Huang and Wellner [37], respectively, in the context of a monotone density and a monotone hazard rate with right-censored data. In all these scenarios, the Chernoff distribution is obtained as the limit. It can be described as the probability distribution of $Z = \arg\min\{W(t) + t^2 : t \in \mathbb{R}\}$, where W is a two-sided standard Brownian motion on the real line with $W(0) = 0$. The density of Z was obtained by Groeneboom [31] and its quantiles were computed in Groeneboom and Wellner [34].

Constructing optimally sized confidence regions for a function or its value at a given point is a challenging problem, primarily because of a bias issue—under the optimal smoothing, the order of the bias matches with an estimator's variability, rendering a shift in the limit distribution causing undercoverage. Such a problem does not happen in the parametric setting as the order of the bias is smaller than that of the variability. Adapting a confidence region to smoothness is further complicated by the fact that maintaining coverage at all functions of different smoothness and adapting the size to the smoothness is not possible (Li [42], Low [44]). In a multiresolution normal sequence model, Cai and Low [16] obtained a lower bound for the radius of a confidence region for the mean vector at a resolution level and of the whole infinite-dimensional vector. They also constructed an adaptive confidence ball adapting to the Besov smoothness with the center at an adaptive estimator. Robins and van der Vaart [49] constructed a confidence region for a parameter in a subset of a Hilbert space by estimating the risk of an adaptive estimator, and also obtained a lower bound for the size. Gine and Nickl [29] constructed a uniform confidence band for a density under a self-similarity-type condition (Picard and Tribouley [46]). Hoffmann and Nickl [36] obtained the necessary and sufficient condition under which an adaptive uniform confidence band is possible. Patchkowski and Rohde [45] obtained locally adaptive confidence regions under weaker restrictions.

In shape-restricted inference, confidence sets were constructed by several authors. In a white noise model, Dümbgen [25] obtained confidence bands for monotone or convex signals adapting to the smoothness of order $0 < \beta \le 1$ and $1 \le \beta \le 2$, respectively. Cai et al. [17] obtained a lower bound to the expected length of a confidence interval of the function value at a point in terms of the local modulus of continuity in the function class and constructed confidence intervals attaining the bounds in monotone and convex regression settings. Dümbgen and Johns [26] and Schmidt-Hieber et al. [51] constructed confidence regions respectively for median regression and deconvolution problems. The existence of a Chernoff limit distribution provides a method for constructing a pointwise confidence interval. However, the limit depends on the derivative of the true function, estimation of which is a harder problem. Therefore, methods of uncertainty quantification that avoid the estimation of nuisance parameters are more desirable. A nuisance parameter-free method to obtain confidence intervals was provided by Banerjee and Wellner [5] in current status models, and by Banerjee [4] in monotone response models. Their method is based on the asymptotic distribution of the likelihood ratio statistic for testing $H_0 : f(x_0) = \theta_0$ against $H_1 : f(x_0) \ne \theta_0$, and inverting the test to build confidence intervals for $f(x_0)$ based on the quantiles of the limiting distribution.

Testing the hypothesis of monotonicity of a regression function was addressed by Bowman et al. [14], Hall and Heckman [35], Ghosal et al. [27], Gijbels et al. [28], and others, using frequentist methods. Armstrong [1] constructed a test for a sign or shape restriction in nonparametric heteroschedastic regression under adaptive optimal separation. A Bayesian test for testing monotonicity of regression was developed by Salomond [50].

The Bayesian paradigm offers a conceptually simpler way of uncertainty quantification through posterior sampling, which is usually easier to implement with the help of modern

computers and advanced posterior computational techniques. Credible sets are often formed using posterior quantiles, giving a range of likely values of the parameter. In parametric models, credible sets have asymptotically correct frequentist coverages because of the Bernstein–von Mises (BvM) theorem, which asserts that Bayesian and frequentist measures of uncertainty agree in large samples. In nonparametric models, however, a BvM theorem may not hold, and credible regions may fail to have adequate asymptotic coverage. For a Gaussian sequence model, Cox [23] showed that Bayesian credible sets can have arbitrarily low asymptotic coverage under the optimal smoothing. The works of Leahu [41] and Knapik et al. [39] further clarified the reason behind this "Cox phenomenon" is the bias problem. Positive coverage results for optimal sized credible sets were obtained by Knapik et al. [39] in the Gaussian sequence model by undersmoothing the prior, and by Yoo and Ghosal [57] for smooth multivariate nonparametric regression by inflating an $\mathbb{L}_\infty$-norm credible region by an appropriate constant. In both approaches, the coverage tends to one. A different approach through BvM theorems in negative Sobolev spaces, which gives exact limiting coverage, was developed by Castillo and Nickl [18, 19], obtaining optimal $\mathbb{L}_2$- and $\mathbb{L}_\infty$-sized credible regions, respectively, in the Gaussian sequence model. Positive coverage results for adaptive Bayesian credible sets were obtained in the Gaussian sequence model by Szabó et al. [54] under a "polished tail" condition, and by Belitser [8] for an empirical Bayes procedure under a more general "excessive bias restriction" condition. Both papers use inflation of a credible region and show that the coverage tends to one while maintaining the optimal order for the size of the region. An adaptive size credible set was obtained by Ray [48] by extending the approach of weak Bernstein–von Mises theorem of Castillo and Nickl [18] in the adaptive setting using spike-and-slab priors, under a stronger "self-similarity" condition, but with the exact asymptotic coverage. Sniekers and van der Vaart [53] obtained adaptive credible sets for nonparametric regression under an analog of the polished tail condition in the regression setting. Following the approach of Belitser [8], Belitser and Nurushev [11] and Belitser and Ghosal [9], respectively, obtained adaptive credible regions for sparse normal sequence models and sparse regression under the excessive bias restriction condition. Castillo and Szabó [20] considered a different empirical Bayes approach, provided coverage results for adaptive credible sets, and showed the necessity of the excessive bias restriction condition. Adaptive credible regions with adequate coverage in a general framework of projection structures encompassing the above-discussed models and more under an excessive bias restriction condition were recently obtained in Belitser and Nurushev [10].

The goal of the present paper is to develop an easy-to-use Bayesian method for constructing credible intervals for the function value $f(x_0)$ at an interior point x_0 of a monotone regression function f, and obtain the asymptotic frequentist coverage of the resulting interval. Although incorporating the monotonicity constraint into the prior appears a natural way to comply with the constraint, it is also very challenging to analyze the asymptotic properties of the resulting posterior. For this reason, we use an unconstrained conjugate prior on f, and project posterior samples on the space of monotone functions using the PAVA to obtain an induced posterior distribution, to be called the "projection-posterior." The idea of embedding the parameter space in a larger space where the posterior is easier to compute and analyze, and then applying the projection to obtain the induced posterior distribution for inference, was earlier also used by Lin and Dunson [43] and Bhaumik and Ghosal [12, 13], respectively, for monotone regression and regression models driven by ordinary differential equations. In this paper, the projection-posterior will be used to obtain credible intervals for monotone regression with an asserted asymptotic frequentist coverage.

Our findings suggest that the coverage of credible regions behaves very differently from the situation when models are indexed by smooth functions. We find that the limiting coverage of a projection-posterior credible interval for $f(x_0)$ is given by a functional of two independent

two-sided standard Brownian motions W_1 and W_2, provided that the true f is differentiable at x_0 with a positive derivative. For a one-sided $(1-\gamma)$-level credible interval, the limit is $A(1-\gamma) := \mathrm{P}(Z_B \le 1-\gamma)$, where $Z_B = \mathrm{P}(\arg\min\{W_1(t) + W_2(t) + t^2 : t \in \mathbb{R}\} \le 0 | W_1)$. It is interesting to note that the limit is free of the true regression function and depends only on γ. Although an analytic evaluation of the function A seems difficult at present, a numerical evaluation through simulations is possible. Numerical evaluations based on Monte Carlo show that $A(1-\gamma)$ is larger than the nominal coverage level $(1-\gamma)$, at least for all values where this function has been evaluated. This is the opposite of the Cox phenomenon for smoothness regimes. Moreover, the evaluation of the function A allows recalibration of the nominal credibility using a back-calculation. A targeted coverage $(1-\alpha)$ may be obtained by starting with a lower credibility level $(1-\gamma)$ such that $A(1-\gamma) = 1-\alpha$. A similar conclusion remains valid for two-sided credible intervals. For instance, a two-sided 95%-credible interval approximately yields 96.5% coverage in large samples, while 95% asymptotic coverage is obtained from a 93.2%-credible interval. These conclusions are also supported by our simulation experiments.

An important aspect of our approach to uncertainty quantification is that credible intervals are obtained directly through (projection) posterior sampling, without requiring us to estimate any normalizing constant, as in Banerjee [3]. The asymptotic coverage is free of constants involving the true f or its derivative. We provide a comparison with an analog of his method.

The rest of the paper is organized as follows. Section 2 introduces some notation, assumptions and the prior on f and σ. Results on coverage of credible intervals are presented in Section 3. Computation procedures for the cut-offs of the credible intervals are discussed in Section 4. A simulation experiment, comparing coverage and size of unadjusted and adjusted credible intervals with those of the confidence interval obtained by inverting the likelihood ratio, is also given in this section. Proofs of the results are provided in Section 5. Some auxiliary results are given in the Appendix, and the rest in the Supplementary Material [21].

2. Notation, assumptions and the prior. In this section, we describe the notation and assumptions used in the paper. For two sequences of real numbers a_n and b_n, $a_n \lesssim b_n$ or $a_n = O(b_n)$ means that a_n/b_n is bounded, $a_n \ll b_n$ or $a_n = o(b_n)$ means that $a_n/b_n \to 0$. For a sequence of random variables Z_n with distribution P, $Z_n = O_P(a_n)$ means that $P(|Z_n| > C_n a_n) \to 0$ for every $C_n \to \infty$. Let $\boldsymbol{I}_m$ stand for the $m \times m$ identity matrix.

We say that $\boldsymbol{Z} \sim \mathrm{N}_J(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ if $\boldsymbol{Z}$ has a J-dimensional normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The probability distribution of a random element Z will be denoted by $\mathcal{L}(Z)$. Let the space of real-valued monotone increasing functions on $[0,1]$ be denoted by $\mathcal{F}$, and the space of monotone increasing functions on $[0,1]$ bounded in absolute value by $K > 0$ be denoted by $\mathcal{F}(K)$. For $T \subset \mathbb{R}$, $\mathbb{L}_\infty(T)$ denotes the set of all uniformly bounded real functions on T. The indicator function of a set U is denoted by $\mathbb{1}_U$. For $f : [0,1] \mapsto \mathbb{R}$ and d a distance on real-valued functions, let the projection of f on $\mathcal{F}$ be the function f^* that minimizes $d(f,h)$ over $h \in \mathcal{F}$.

For a random variable Y and a sequence of random variables X_n, $X_n \rightsquigarrow Y$ means that X_n converges in distribution to Y, $X_n \to_P Y$ means that X_n converges to Y in P-probability. For random variables X and Y, $X \overset{d}{=} Y$ means that X and Y have the same distribution. The ϵ-covering number of a set A with respect to a metric d, denoted by $\mathcal{N}(\epsilon, A, d)$, is the minimum number of balls of radius ϵ needed to cover A.

We make the following assumption on the data generating process P_0.

ASSUMPTION D. The predictor variables $X_1, \ldots, X_n$ are independent and identically distributed (i.i.d.) with a probability measure G having a positive and continuous density g on $[0,1]$, the response variables are $Y_i = f_0(X_i) + \varepsilon_i$, and the random errors $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. sub-Gaussian with mean 0 and variance σ_0^2.

Let $0 < a_1 < a_2$ be such that $a_1 \le g(x) \le a_2$ for all $x \in [0,1]$. With a slight abuse of notation, we let G denote the corresponding distribution function as well. Let $\mathrm{E}_0(\cdot)$ and $\mathrm{Var}_0(\cdot)$ be the expectation and variance operators taken under the true distribution P_0. We write $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$, $\boldsymbol{X} = (X_1, \ldots, X_n)$, $D_n = (\boldsymbol{Y}, \boldsymbol{X})$, $\boldsymbol{F}_0 = (f_0(X_1), \ldots, f_0(X_n))$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)$. Let $\mathbb{P}_n$ denote the empirical measure.

We represent f as a piecewise constant function on $[0,1]$ with $J = J_n$ the number of pieces, and the regression model as $\boldsymbol{Y} = \boldsymbol{B\theta} + \boldsymbol{\varepsilon}$. For a deterministic J, let $I_j = ((j-1)/J, j/J]$ be the jth interval and its count $N_j = \sum_{i=1}^n \mathbb{1}\{X_i \in I_j\}$, $1 \le j \le J$. Let $f = \sum_{j=1}^J \theta_j \mathbb{1}_{I_j}$. We use the prior $\boldsymbol{\theta} \sim \mathrm{N}_J(\boldsymbol{\zeta}, \sigma^2 \boldsymbol{\Lambda})$ where $\|\boldsymbol{\zeta}\|_\infty$ is bounded, and $\boldsymbol{\Lambda}$ a $J \times J$ diagonal matrix with diagonal entries $\lambda_1^2, \ldots, \lambda_J^2$, with $B_1 < \lambda_j < B_2$ for some $B_1, B_2 > 0$. Given σ, θ_j are a posteriori independently distributed as $\mathrm{N}((N_j \bar{Y}_j + \zeta_j/\lambda_j^2)/(N_j + 1/\lambda_j^2), \sigma^2/(N_j + 1/\lambda_j^2))$.

The error variance σ^2 may be estimated by maximizing the marginal likelihood of σ. Observe that $(\boldsymbol{Y}|\sigma) \sim \mathrm{N}_n(\boldsymbol{B\zeta}, \sigma^2(\boldsymbol{B\Lambda B}^{\mathrm{T}} + \boldsymbol{I}_n))$. Therefore, the MLE is $\hat{\sigma}_n^2 = n^{-1}(\boldsymbol{Y} - \boldsymbol{B\zeta})^{\mathrm{T}}(\boldsymbol{B\Lambda B}^{\mathrm{T}} + \boldsymbol{I}_n)^{-1}(\boldsymbol{Y} - \boldsymbol{B\zeta})$. The plug-in posterior distribution of f is obtained by substituting $\hat{\sigma}_n$ for σ. A fully Bayes alternative is to endow σ^2 with an inverse-gamma prior $\mathrm{IG}(\beta_1, \beta_2)$ with parameters (β_1, β_2) for some $\beta_1 > 2$ and $\beta_2 > 0$.

3. Uncertainty quantification in pointwise estimation. Let $x_0 \in (0,1)$ be such that $f'(x_0)$ exists and $f'(x_0) > 0$. Without loss of generality, we assume that $X_1 \le \cdots \le X_n$. The isotonic regression estimator of f is obtained by minimizing the sum of squares $\sum_{i=1}^n (Y_i - f(X_i))^2$ subject to the constraint that f is nondecreasing on $[0,1]$. The resulting estimator of f is a nondecreasing step function, constant on the pieces $(X_{i-1}, X_i]$, with X_0 defined to be zero. The estimated value of $f(x_0)$ is given by the left derivative of the greatest convex minorant of the graph of the line segments connecting $\{(0,0), (1/n, Y_1/n), (2/n, (Y_1+Y_2)/n), \ldots, (1, (\sum_{i=1}^n Y_i)/n)\}$, at the point $i(x_0)/n$, where $i(x_0)$ is the integer such that $X_{i-1} < x_0 \le X_i$.

Now suppose that f is a monotone increasing piecewise constant function on $[0,1]$. For $J > 1$, let $f = \sum_{j=1}^J q_j \mathbb{1}_{I_j}$, where $(q_1, \ldots, q_J) \in \mathbb{R}^J$ with $q_1 \le \cdots \le q_J$. Consider a sieve $\Theta_J = \{f = \sum_{j=1}^J q_j \mathbb{1}_{I_j} : q_1 \le \cdots \le q_J\}$, where $I_j = ((j-1)/J, j/J]$, $1 \le j \le J$. We define the sieve-maximum likelihood estimator $\hat{f}_n$ of f, as the element of Θ_J that maximizes the likelihood function using the working model hypothesis $\varepsilon_i \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, \sigma^2)$. This amounts to minimizing the error sum of squares $\sum_{i=1}^n (Y_i - f(X_i))^2$, in $f \in \Theta_J$. Now $\sum_{i=1}^n (Y_i - f(X_i))^2 = \sum_{j=1}^J \sum_{i:X_i \in I_j} (Y_i - q_j)^2$, which can be decomposed as $\sum_{j=1}^J \sum_{i:X_i \in I_j} (Y_i - \bar{Y}_j)^2 + \sum_{j=1}^J N_j (\bar{Y}_j - q_j)^2$, where $\bar{Y}_j = N_j^{-1} \sum_{i:X_i \in I_j} Y_i$. The first term here is free of $q_1, \ldots, q_J$, so it suffices to minimize the weighted sum of squares $\sum_{j=1}^J N_j (\bar{Y}_j - q_j)^2$ subject to the ordering constraint $q_1 \le \cdots \le q_J$. Using the algorithm for weighted isotonic regression, we get that the optimal value of q_j is the left derivative at $\sum_{k=1}^j N_k/n$ of the greatest convex minorant of the graph of the lines connecting $\{(0,0), (N_1/n, N_1\bar{Y}_1/n), \ldots, (\sum_{j=1}^J N_j/n, \sum_{j=1}^J N_j \bar{Y}_j/n)\}$.

The following result tells us that for a certain range for J, the sieve-MLE at x_0 has the same rate of convergence and asymptotic distribution as those of the isotonic regression estimator at x_0.

THEOREM 3.1. *Let $n^{1/3} \ll J \ll n^{2/3}$ and Assumption D hold. Then the sieve-MLE $\hat{f}_n$ satisfies $\mathrm{P}_0(n^{1/3}(\hat{f}_n(x_0) - f_0(x_0)) \le z) \to \mathrm{P}(C_0 Z \le z)$ for every $z \in \mathbb{R}$, where $Z = \arg\min\{W_1(t) + t^2 : t \in \mathbb{R}\}$, and $C_0 = 2b(a/b)^{2/3}$ with $a = \sqrt{\sigma_0^2/g(x_0)}$ and $b = f_0'(x_0)/2$, and W_1 is a two-sided Brownian motion on $\mathbb{R}$ with $W_1(0) = 0$.*

The asymptotic distribution of the sieve-MLE under the condition $n^{1/3} \ll J \ll n^{2/3}$ is the same as that of the MLE, that is, the Chernoff distribution, and hence both have the same level of accuracy. However, the joint treatment of the estimator and the posterior distribution is easier with the latter, ostensibly due to their structural similarity.

Now we study the projection-posterior of f with a deterministic $J = J_n$, depending on n. We have that $f = \sum_{j=1}^{J} \theta_j \mathbb{1}_{I_j}$, with the posterior of $\theta_1, \ldots, \theta_J$ being independent Gaussian because of conjugacy. As the posterior samples of f may not be monotone, we impose monotonicity by projecting each f on $\mathcal{F}$ using the PAVA. In particular, we find the f^* in the closest monotone function to f in that it minimizes $\sum_{i=1}^{n}(f(X_i) - f^*(X_i))^2$ subject to $f^* \in \mathcal{F}$. Expanding the sum of squares, we get

$$\sum_{j=1}^{J} \sum_{X_i \in I_j} \big(f(X_i) - f^*(X_i)\big)^2 = \sum_{j=1}^{J} \sum_{X_i \in I_j} \big(\theta_j - f^*(X_i)\big)^2. \tag{3.1}$$

Now note that for every j, $\sum_{X_i \in I_j}(\theta_j - f^*(X_i))^2$ is minimized when $f^*(X_i)$ values are the same for all i with $X_i \in I_j$. Thus f^* has to be constant on each I_j, and for any $f = \sum_{j=1}^{J} \theta_j \mathbb{1}_{I_j}$, the minimization of (3.1) is equivalent to that of $\sum_{j=1}^{J} N_j(\theta_j - \theta_j^*)^2$ subject to $\theta_1^* \le \cdots \le \theta_J^*$, yielding a monotone function $f^* = \sum_{j=1}^{J} \theta_j^* \mathbb{1}_{I_j}$ closest to f. From the theory of weighted isotonic regression (cf. Lemma 2.1 of Groeneboom and Jongbloed [32]), the value of θ_j^* is the left derivative at the point $\sum_{k=1}^{j} N_k/n$ of the greatest convex minorant of the graph of the line segments connecting

$$\left\{(0,0), (N_1/n, N_1\theta_1/n), \ldots, \left(\sum_{k=1}^{J} N_k/n, \sum_{k=1}^{J} N_k\theta_k/n\right)\right\}. \tag{3.2}$$

Our method of uncertainty quantification for the parameter $f(x_0)$ is based on using the quantiles of the pointwise projection-posterior $f^*(x_0)$. For every f generated from the unrestricted posterior, we project f on $\mathcal{F}$ to obtain a monotone function f^* as described above, and evaluate $f^*(x_0)$. We then form a $(1-\gamma)$-level projection-posterior credible interval using the $\gamma/2$ and $(1-\gamma/2)$th quantiles of $f^*(x_0)$. The projection-posterior credibility of this interval is therefore $(1-\gamma)$.

While the asymptotic distribution of the MLE or the sieve-MLE of $f(x_0)$ allows the construction of confidence intervals for $f(x_0)$ to meet a coverage target, the unknown parameters in the limit distributions still need to be estimated. This inconvenience and its excessive reliance on asymptotics make a Bayesian route more appealing, which immediately provides a credible interval based on easily doable posterior sampling. A natural question that arises here is whether an analog of the Bernstein–von Mises theorem holds, reconciling the frequentist and Bayesian measures of uncertainty quantification. However, as the following result shows, unlike in the Bernstein–von Mises theorem for a regular parametric model, the centered and scaled posterior distribution of $f(x_0)$ does not even converge in probability to a limit, conditioned on the data.

PROPOSITION 3.2. *If $n^{1/3} \ll J \ll n^{2/3}$ and Assumption D holds, then $\Pi(n^{1/3}(f^*(x_0) - \hat{f}_n(x_0)) \le z | D_n)$ does not converge in probability for any $z \in \mathbb{R}$.*

Note that we scale the projection-posterior $f^*(x_0)$ around $\hat{f}_n(x_0)$ by $n^{1/3}$ because the sieve-MLE $\hat{f}_n(x_0)$ converges to $f_0(x_0)$ at the rate $n^{-1/3}$ by Theorem 3.1. The conclusion of Proposition 3.2 resonates with the results on the inconsistency of the bootstrap in the Grenander estimator, as shown by Kosorok [40] and Sen et al. [52]. It implies that the conditional distribution of $n^{1/3}(f^*(x_0) - \hat{f}_n(x_0))$ given the data cannot be approximated by a deterministic probability measure. This is confirmed by our simulation study on the behavior of

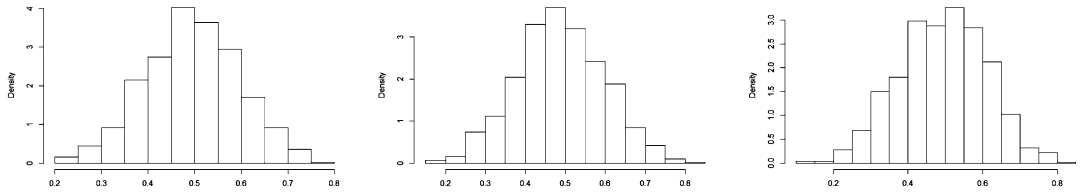


FIG. 1. *Plot demonstrating that* $\Pi(n^{1/3}(f^*(x_0) - \hat{f}_n(x_0)) \le 0|D_n)$ *does not have a limit in probability, using three instances of the data.*

$\Delta_n := \Pi(n^{1/3}(f^*(x_0) - \hat{f}_n(x_0)) \le 0|D_n)$. For the sample size 2000, we obtained 1000 posterior samples of Δ_n. Figure 1 shows the histograms of Δ_n for three different sets of simulated data. It is evident that the posterior probability is not concentrating at any fixed number, and hence the conditional distribution of $n^{1/3}(f^*(x_0) - \hat{f}_n(x_0))$ for an observed sample is not convergent.

Nevertheless, $\Pi(n^{1/3}(f^*(x_0) - f_0(x_0)) \le z|D_n)$ has a weak limit.

THEOREM 3.3. *Let* $n^{1/3} \ll J \ll n^{2/3}$ *and Assumption* D *hold. Let* W_1, W_2 *be independent two-sided Brownian motions on* $\mathbb{R}$ *where* $W_1(0) = W_2(0) = 0$, *and* $C_0 = 2b(a/b)^{2/3}$ *with* $a = \sqrt{\sigma_0^2/g(x_0)}$ *and* $b = f_0'(x_0)/2$. *Then* $\Pi(n^{1/3}(f^*(x_0) - f_0(x_0)) \le z|D_n) \rightsquigarrow \mathrm{P}(C_0 \arg\min\{W_1(t) + W_2(t) + t^2 : t \in \mathbb{R}\} \le z|W_1)$ *for every* $z \in \mathbb{R}$.

The weak limit has a key role in obtaining the limiting coverage of the credible interval. For $n \ge 1$, $\gamma \in [0, 1]$, let $I_{n,\gamma} = [Q_{n,1-\gamma/2}, Q_{n,\gamma/2}]$, where $Q_{n,\gamma} = \inf\{z \in \mathbb{R} : \Pi(f^*(x_0) \le z|D_n) \ge 1 - \gamma\}$ is the $(1-\gamma)$-quantile of the projection-posterior distribution of $f(x_0)$. Then $f_0(x_0) \le Q_{n,\gamma}$ if and only if $\Pi(f^*(x_0) \le f_0(x_0)|D_n) \le 1 - \gamma$, and hence by Theorem 3.3,

$$\begin{aligned}\mathrm{P}_0\big(f_0(x_0) \le Q_{n,\gamma}\big) &= \mathrm{P}_0\big(\Pi\big(n^{1/3}\big(f^*(x_0) - f_0(x_0)\big) \le 0|D_n\big) \le 1 - \gamma\big) \\ &\to \mathrm{P}\big(\mathrm{P}\big(\arg\min_{t\in\mathbb{R}}\big\{W_1(t) + W_2(t) + t^2\big\} \le 0|W_1\big) \le 1 - \gamma\big),\end{aligned}$$

which can be written as $\mathrm{P}(Z_B \le 1 - \gamma)$, where

$$Z_B = \mathrm{P}\big(\arg\min\big\{W_1(t) + W_2(t) + t^2 : t \in \mathbb{R}\big\} \le 0|W_1\big). \tag{3.3}$$

Note that the unknown constant C_0 disappears from the limit. By a similar argument, the limiting coverage of a two-sided credible interval $I_{n,\gamma}$ is obtained in the following result. In the Bayesian context, the role of the distribution of Z_B is similar to that of the Chernoff distribution for the MLE, and hence we shall call it the Bayes–Chernoff distribution. The notable difference is that the former is a distribution on the whole real line, while the latter is a distribution on the unit interval, but both are symmetric, respectively, around 0 and $1/2$ (see Lemma 3.5 below). This leads to the following main conclusion of this paper.

THEOREM 3.4. $\mathrm{P}_0(f_0(x_0) \in I_{n,\gamma}) \to \mathrm{P}(\gamma/2 \le Z_B \le 1 - \gamma/2)$.

Thus the limiting coverage of a $(1-\gamma)$-credible interval is not $(1-\gamma)$, but it can be calculated from the completely known distribution of Z_B. Consider a function $A(u) = \mathrm{P}(Z_B \le u)$, $u \in [0, 1]$. Observe that, A is continuous, strictly increasing and onto $[0, 1]$. This follows by considering the functions $(t - a)^2$ for different values of a in place of $W_1(t)$ to conclude that all values of Z_B are possible. Since all continuous functions are in the support of a Brownian motion, A is strictly increasing on $[0, 1]$. Thus the inverse function is well-defined and is also increasing. Moreover, the following result shows that A has a symmetry property.

TABLE 1
Table of the values of the function A

u	0.700	0.750	0.800	0.850	0.900	0.910	0.920	0.930	0.940
$A(u)$	0.724	0.779	0.830	0.878	0.923	0.931	0.940	0.948	0.956
u	0.950	0.960	0.965	0.970	0.975	0.980	0.985	0.990	0.995
$A(u)$	0.964	0.970	0.974	0.979	0.982	0.986	0.990	0.993	0.997

LEMMA 3.5. *The random variable $Z_B \in [0,1]$ is symmetrically distributed about* $1/2$, *and hence $A(1-u) = 1 - A(u)$, for all $u \in [0,1]$, $A^{-1}(1-v) = 1 - A^{-1}(v)$ for all $v \in [0,1]$, and that $A(1/2) = 1/2 = A^{-1}(1/2)$.*

A recalibration technique will allow us to obtain any desired asymptotic coverage $(1-\alpha)$ by starting from a credibility level $(1-\gamma)$ related to, but not the same as, $(1-\alpha)$. The limiting coverage of a one-sided credible interval $(-\infty, Q_{n,\gamma}]$ is $A(1-\gamma)$. By choosing γ such that $A(1-\gamma) = 1-\alpha$, that is, $1-\gamma = A^{-1}(1-\alpha)$, where $A^{-1} : [0,1] \to [0,1]$ is the inverse function, we can achieve the desired coverage target. For a two-sided credible interval, Theorem 3.4 and Lemma 3.5 give us that the asymptotic coverage of $I_{n,\gamma} = [Q_{n,1-\gamma/2}, Q_{n,\gamma/2}]$ is equal to $A(1-\gamma/2) - A(\gamma/2) = 2A(1-\gamma/2) - 1$. Hence a desired coverage $(1-\alpha)$ can be obtained by setting $2A(1-\gamma/2) - 1 = 1-\alpha$, or $1-\gamma/2 = A^{-1}(1-\alpha/2) = 1 - A^{-1}(\alpha/2)$, that is, $\gamma = 2A^{-1}(\alpha/2)$ and $1-\gamma = 1 - 2A^{-1}(\alpha/2)$.

COROLLARY 3.6. *For any* $0 < \alpha < 1$, $\mathrm{P}_0(f_0(x_0) \in I_{n,2A^{-1}(\alpha/2)}) \to 1-\alpha$.

The back-calculation of the required credibility level $1-\gamma = 1 - 2A^{-1}(\alpha/2)$ to achieve a coverage level $1-\alpha$ can be obtained from a table of A^{-1}. In the next section, we present tables of the functions A and A^{-1}. It is also observed from numerical calculations based on Monte Carlo that $A(u) \geq u$ for all values of $u \geq 1/2$ where the computation has been performed. Hence a credible interval has asymptotic coverage more than its nominal credibility, leading to the reverse Cox phenomenon mentioned in the Introduction. The undercoverage phenomenon does not happen here because by the choice $J \gg n^{1/3}$, the order of the bias J^{-1} is smaller than the rate of variability $n^{-1/3}$. In this context, the parameter J controls the complexity and the approximation error, but primarily the regularization is provided by the isotonization step, which is a global procedure, rather than the smoothing operation regulated by J. This is clear from the fact that a wide range of values of J leads to the same rate in Theorems 3.1 and 3.3. In other words, unlike in smoothing problems, the initial undersmoothing needed to reduce the bias does not lead to an inferior convergence rate. It may be noted that the asymptotic distribution of the sieve-MLE is also centered, and the weak nature of the convergence in Theorem 3.3 results in more variation in the posterior than that in the estimator. This leads to longer posterior credible intervals and higher limiting coverage than its credibility.

4. Numerical study. In this section, we numerically obtain the distribution function and the quantile functions of the Bayes–Chernoff distribution and study the coverage of unadjusted and adjusted Bayesian credible intervals through a simulation study.

4.1. *Calculation of A and A^{-1}.* Analytic evaluations of the functions A and A^{-1} appear too difficult at this stage, due the to lack of the necessary probabilistic tools. However, as the functions are probabilistic characteristics of standard Brownian motions, not involving any unknown parameters, Monte Carlo simulations can compute these approximately.

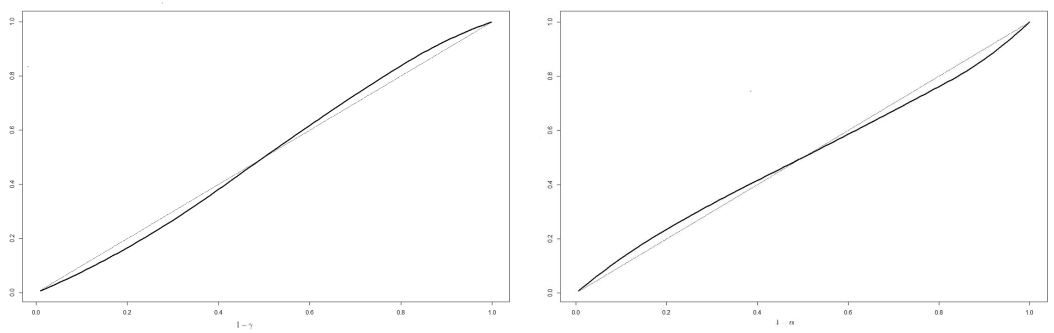
TABLE 2
Table of the values of the function A^{-1}

v	0.700	0.750	0.800	0.850	0.900	0.910	0.920	0.930	0.940	0.950	0.960
$A^{-1}(v)$	0.677	0.723	0.771	0.820	0.874	0.885	0.897	0.909	0.922	0.934	0.946
v	0.965	0.970	0.975	0.980	0.985	0.990	0.995	0.996	0.997	0.998	0.999
$A^{-1}(v)$	0.952	0.960	0.966	0.973	0.980	0.986	0.993	0.995	0.996	0.997	0.999

A discrete approximation to the standard two-sided Brownian motions on $\mathbb{R}$ is obtained by generating $2m$ independent standard normal variables $\{Z_j, Z'_j\}$, $j = 1, \ldots, m$, and computing $W(t) = m^{-1/2}\{\mathbb{1}(t \geq 0)\sum_{j=1}^{\lceil mt \rceil} Z_j + \mathbb{1}(t < 0)\sum_{j=1}^{\lceil -mt \rceil} Z'_j\}$, for $t \in [-2, 2]$ and $m = 10^4$. For each simulation step, we generate an independent copy W_1 of W. For each such instance, we generate an independent copy W_2 of W, form the processes $W_1(t) + W_2(t) + t^2$ and compute its argmin. For each W_1, we evaluate these 5000 times independently and compute the proportion of times $\arg\min\{W_1(t) + W_2(t) + t^2\}$ is less than zero to obtain an approximate sample of size 1 from Z_B. We then repeat this $M = 30{,}000$ times. In Table 1, we tabulate the Monte Carlo estimates of the probabilities $A(u) = \mathrm{P}(Z_B \leq u)$ up to three decimal values for u on a fine grid over [0, 1]. We evaluate the values of A^{-1} by numerically inverting the table of A. We provide the estimated values of $A^{-1}(v)$ for some selected values of v in Table 2.

Thus, to obtain 95%-confidence, the nominal level of credibility should be chosen to be $1 - \gamma = 2A^{-1}(0.975) - 1 = 2 \times 0.966 - 1 = 0.932$. The nature of the functional relationship between the credibility and the coverage is clearly demonstrated by the two plots in Figure 2.

4.2. *Simulation*. We now study coverage of credible intervals for the value of the regression function at an interior point through simulations. We generate data from a regression model with monotone regression function $f_0(x) = x^2 + x/5$, $x \in [0, 1]$, and normal error with standard deviation 0.1 and the predictor variable X sampled from the uniform distribution on [0, 1]. We study the coverage and size of the proposed Bayesian credible interval for $f_0(x_0)$ at $x_0 = 0.5$ in finite sample sizes. We consider a prior supported on piecewise constant functions with number of intervals J the integer closest to $n^{1/3}\log n$, where n stands for the sample size. The prior on the step heights θ_j is chosen to be independent normal with mean 0 and variance $1000\hat{\sigma}_n^2$, where $\hat{\sigma}_n^2$ is the maximum marginal likelihood estimate of σ^2. We vary the sample size n over four different values 500, 1000, 1500 and 2000. For each n, we consider 1000 replications of the simulated data. For each set of data, we generate 1000 posterior samples of $\boldsymbol{\theta}$ and obtain the corresponding projection f^*. Using its $\alpha/2$ and $1 - \alpha/2$



(a) Obtained coverage for a given credibility.

(b) Required credibility for a targeted coverage.

FIG. 2. *Plots comparing coverage and credibility of Bayesian credible intervals. The dotted line denotes* $x = y$.

TABLE 3
Comparison of average length and size of unadjusted and adjusted Bayesian credible intervals and confidence interval based on inverting the likelihood ratio statistics

$1-\alpha$	$C_B(\alpha)$	$L_B(\alpha)$	$C_B^*(\alpha)$	$L_B^*(\alpha)$	$C_F(\alpha)$	$L_F(\alpha)$	$C_B(\alpha)$	$L_B(\alpha)$	$C_B^*(\alpha)$	$L_B^*(\alpha)$	$C_F(\alpha)$	$L_F(\alpha)$
			$n=500$						$n=1000$			
0.99	0.994	0.48	0.983	0.43	0.987	0.56	0.996	0.39	0.991	0.35	0.991	0.41
0.95	0.958	0.38	0.935	0.35	0.951	0.42	0.967	0.30	0.951	0.28	0.949	0.32
0.90	0.911	0.32	0.893	0.30	0.902	0.38	0.929	0.26	0.900	0.24	0.897	0.27
			$n=1500$						$n=2000$			
0.99	0.994	0.34	0.984	0.31	0.989	0.35	0.996	0.31	0.988	0.28	0.993	0.33
0.95	0.967	0.27	0.949	0.25	0.955	0.29	0.968	0.25	0.939	0.23	0.952	0.28
0.90	0.914	0.23	0.894	0.21	0.902	0.24	0.914	0.21	0.895	0.19	0.904	0.25

quantiles, we construct the $100(1-\alpha)\%$ credible interval for $1-\alpha = 0.99$, 0.95 and 0.90. We also obtain the corresponding recalibrated credible intervals using the $A^{-1}(\alpha)/2$ and $[1-A^{-1}(\alpha)]/2$ quantiles of $f^*(x_0)$. We then compare the proposed Bayesian procedures with the confidence intervals obtained by inverting the acceptance region of a likelihood ratio test in Banerjee [4]. The cut-off values are obtained from the limiting distribution of the likelihood ratio statistics in Banerjee [4] and are given in Table 3.3 of Banerjee [3]. For all three types of intervals, we check whether $f_0(x_0)$ is contained in there, and estimate the coverage by Monte Carlo. In Table 3, $C_B(\alpha)$, $C_B^*(\alpha)$ and $C_F(\alpha)$, respectively, denote the coverage of $(Q_{n,\alpha/2}, Q_{n,1-\alpha/2})$, $(Q_{n,A^{-1}(\alpha)/2}, Q_{n,[1-A^{-1}(\alpha)]/2})$ and that of $(1-\alpha)$-confidence interval of Banerjee [4], and $L_B(\alpha)$, $L_B^*(\alpha)$ and $L_F(\alpha)$ denote their respective lengths.

5. Proofs. As Theorem 3.1 and Proposition 3.2 are results of independent interest and are not of the primary of this paper, we present their proofs in the Supplementary Material [21]. In this section, we present the proof of Theorem 3.3 and auxiliary results needed in the proof. We frequently use the "switch relation": for a lower semicontinuous function Φ on an interval I with Φ^* its GCM, and Φ^{*l} denoting the left derivative of Φ^*, for every $t \in I$, $v \in \mathbb{R}$,

$$\{\Phi^{*l}(t) > v\} = \{\arg\min\{\Phi(s) - vs : s \in I\} < t\}, \tag{5.1}$$

where "arg min" selects the maximum of the minimizers when multiple minimizers exist; see page 56 of Groeneboom and Jongbloed [32] for the details regarding the switch relations.

Throughout the proof section, we use the notation $\mathcal{J}_{n,t} = \{\lceil x_0 J\rceil + 1, \ldots, \lceil (x_0 + n^{-1/3}t)J\rceil\}$. If $t < 0$, we interpret a sum over $\mathcal{J}_{n,t}$ as that over $\{\lceil (x_0 + n^{-1/3}t)J\rceil + 1, \ldots, \lceil x_0 J\rceil\}$ but with a negative sign.

PROOF OF THEOREM 3.3. Since f^* is piecewise constant on each I_j, $f^*(x_0) = \theta^*_{\lceil x_0 J\rceil}$. Let $s_j = \sum_{k=1}^j N_k/n$, $\omega_j = \sum_{k=1}^j (N_k/n)\theta_k$ for $1 \le j \le J$, and $s_0 = 0$, $\omega_0 = 0$. Let $\mathcal{G}$ denote the graph obtained by joining the points $\{(s_0, \omega_0), (s_1, \omega_1), \ldots, (s_J, \omega_J)\}$, and define $\mathcal{G}(s) = 0$ for $s \le 0$ and $\mathcal{G}(s) = \omega_J$ for $s \ge 1$. Then $f^*(\lceil x_0 J\rceil/J) = \theta^*_{\lceil x_0 J\rceil}$ is the left derivative of the GCM of $\mathcal{G}$ at $s_{\lceil x_0 J\rceil}$. For $z \in \mathbb{R}$, by the switch relation (5.1),

$$\{n^{1/3}(f^*(x_0) - f_0(x_0)) \le z\} = \{\theta^*_{\lceil x_0 J\rceil} \le f_0(x_0) + n^{-1/3}z\} \tag{5.2}$$

As $\mathcal{G}(s) - (f_0(x_0) + n^{-1/3}z)s$ is piecewise linear in $(s_{j-1}, s_j]$, it is minimized at one of the points in $\mathcal{S} = \{s_0, s_1, \ldots, s_J\}$. Therefore, the event in (5.2) is

$$\begin{aligned}&\{\arg\min\{\mathcal{G}(s) - (f_0(x_0) + n^{-1/3}z)s : s \in [0,1]\} \ge s_{\lceil x_0 J\rceil}\}\\ &\quad = \{\arg\min\{\omega_j - (f_0(x_0) + n^{-1/3}z)s_j : j \in \{0, 1, \ldots, J\}\} \ge \lceil x_0 J\rceil\}.\end{aligned}$$

Now define two stochastic processes V_n and G_n on $\mathbb{R}$, such that $V_n(s) = G_n(s) = 0$ for $s \le 0$, and $V_n(t) = \omega_{\lceil tJ \rceil}$, $G_n(t) = s_{\lceil tJ \rceil}$, $0 \le t \le 1$. Also let $V_n(t) = \omega_J$ and $G_n(t) = 1$ for $t \ge 1$. Thus V_n and G_n are random step functions with jumps at the points j/J, $1 \le j \le J$. Therefore,

$$\begin{aligned}
&\{\arg\min\{\omega_j - (f_0(x_0) + n^{-1/3}z)s_j : 0 \le j \le J\} \ge \lceil x_0 J \rceil\} \\
&\quad = \{\arg\min\{V_n(j/J) - (f_0(x_0) + n^{-1/3}z)G_n(j/J) : 0 \le j \le J\} \ge \lceil x_0 J \rceil\} \\
&\quad = \{\arg\min\{V_n(s) - (f_0(x_0) + n^{-1/3}z)G_n(s) : s \in [0,1]\} \ge \lceil x_0 J \rceil / J\} \\
&\quad = \{\arg\min\{V_n(s) - (f_0(x_0) + n^{-1/3}z)G_n(s) : s \in \mathbb{R}\} \ge \lceil x_0 J \rceil / J\}.
\end{aligned}$$

The last step follows from the fact that V_n and G_n are constants on $s \le 0$ and $s \ge 1$. As the location of the minimum does not change upon addition of a constant or multiplication by a positive constant,

$$\begin{aligned}
&\{\arg\min\{V_n(s) - (f_0(x_0) + n^{-1/3}z)G_n(s) : s \in \mathbb{R}\} \ge \lceil x_0 J \rceil / J\} \\
&\quad = \left\{\arg\min\left\{\frac{n^{2/3}}{g(x_0)}(V_n(s) - V_n(x_0))\right.\right. \\
&\qquad \left.\left. - \frac{n^{2/3}}{g(x_0)}(f_0(x_0) + n^{-1/3}z)(G_n(s) - G_n(x_0)) : s \in \mathbb{R}\right\} \ge \frac{\lceil x_0 J \rceil}{J}\right\}.
\end{aligned} \tag{5.3}$$

We make a change of variable $s = x_0 + n^{-1/3}t$. Then the right-hand side of (5.3) can then be written as

$$\begin{aligned}
&\left\{\arg\min_{t \in \mathbb{R}}\left\{\frac{n^{2/3}}{g(x_0)} \sum_{j \in \mathcal{J}_{n,t}} \frac{N_j}{n}\theta_j - \frac{n^{2/3}}{g(x_0)}(f_0(x_0) + n^{-1/3}z) \sum_{j \in \mathcal{J}_{n,t}} \frac{N_j}{n}\right\}\right. \\
&\qquad \left. \ge n^{1/3}\left(\frac{\lceil x_0 J \rceil}{J} - x_0\right)\right\}.
\end{aligned} \tag{5.4}$$

Therefore, $\Pi(n^{1/3}(f^*(x_0) - f_0(x_0)) \le z | D_n)$ is equal to

$$\begin{aligned}
&\Pi\left(\arg\min_{t \in \mathbb{R}}\left\{\frac{n^{2/3}}{g(x_0)} \sum_{j \in \mathcal{J}_{n,t}} \frac{N_j}{n}(\theta_j - f_0(x_0))\right.\right. \\
&\qquad \left.\left. - \frac{n^{1/3}z}{g(x_0)} \sum_{j \in \mathcal{J}_{n,t}} \frac{N_j}{n}\right\} \ge n^{1/3}\left(\frac{\lceil x_0 J \rceil}{J} - x_0\right) | D_n\right).
\end{aligned} \tag{5.5}$$

Because $N_j \sim \mathrm{Bin}(n; G(I_j))$, and G has bounded and positive density, the expectation of $(n^{1/3}z/g(x_0)) \sum_{j \in \mathcal{J}_{n,t}} N_j/n$ is given by

$$\frac{n^{1/3}z}{g(x_0)}\left(G\left(\frac{\lceil (x_0 + n^{-1/3}t)J \rceil}{J}\right) - G\left(\frac{\lceil x_0 J \rceil}{J}\right)\right) = z(1 + o(1))(t + O(n^{1/3}J^{-1})),$$

which converges to zt uniformly in $t \in [-K, K]$ as $n^{1/3} \ll J \ll n^{2/3}$. By a similar calculation, its variance is of the order $n^{-1/3}z^2 \to 0$. Therefore, $(n^{1/3}z/g(x_0)) \sum_{j \in \mathcal{J}_{n,t}} N_j/n \to_{P_0} zt$. Also $|\lceil x_0 J \rceil / J - x_0| = O(J^{-1})$ and $J \gg n^{1/3}$, so $n^{1/3}|\lceil x_0 J \rceil / J - x_0| \to 0$. Hence by Lemmas 5.1 and 5.2 below, the Argmax theorem applied conditionally on the data, and part (a) of Lemma A.3, we rewrite (5.5) as

$$\begin{aligned}
&\Pi(n^{1/3}(f^*(x_0) - f_0(x_0)) \le z | D_n) \\
&\quad \rightsquigarrow \mathrm{P}(\arg\min\{aW_1(t) + aW_2(t) + bt^2 - zt : t \in \mathbb{R}\} \ge 0 | W_1)
\end{aligned}$$

$$\stackrel{d}{=} \mathrm{P}\big((a/b)^{2/3} \arg\min\{W_1(t) + W_2(t) + t^2 : t \in \mathbb{R}\} + z/(2b) \geq 0 | W_1\big)$$

$$\stackrel{d}{=} \mathrm{P}\big(2b(a/b)^{2/3} \arg\min\{W_1(t) + W_2(t) + t^2 : t \in \mathbb{R}\} \leq z | W_1\big).$$

The last step follows by using the transformation $t \mapsto -t$ and the fact that $\arg\min\{W_1(t) + W_2(t) + t^2 : t \in \mathbb{R}\}$ is symmetric about zero. $\square$

LEMMA 5.1. *Let $a = \sqrt{\sigma_0^2/g(x_0)}$, $b = f_0'(x_0)/2$, and W_1, W_2 be independent two-sided Brownian motions on $\mathbb{R}$ starting at zero and $n^{1/3} \ll J \ll n^{2/3}$. Then $\mathcal{L}((n^{2/3}/g(x_0)) \times \sum_{j \in \mathcal{J}_{n,t}} N_j(\theta_j - f_0(x_0))/n : t \in [-K, K] | D_n)$ converges weakly to $\mathcal{L}(aW_1(t) + aW_2(t) + bt^2 : t \in [-K, K] | W_1)$ as random probability measures on the space $\mathbb{L}_\infty([-K, K])$, for any $K > 0$.*

PROOF. We consider a sequence $\sigma_n \in \mathcal{U}_n$, where $\mathcal{U}_n$ is a shrinking neighborhood of σ_0 with $\Pi(\sigma \in \mathcal{U}_n | D_n) \to 1$. The existence of such a sequence is guaranteed by Lemma A.1. We can condition on $\sigma = \sigma_n$. Write $\tilde{D}_n = \{D_n, \sigma_n\}$ and

$$\frac{n^{2/3}}{g(x_0)} \sum_{j \in \mathcal{J}_{n,t}} \frac{N_j}{n}\big(\theta_j - f_0(x_0)\big) = A_n(t; \theta, \tilde{D}_n) + A_n'(t; \tilde{D}_n) + B_n(t; \tilde{D}_n),$$

where the processes A_n, A_n' and B_n are defined by

$$A_n(t; \theta, \tilde{D}_n) = \big(n^{2/3}/g(x_0)\big) \sum_{j \in \mathcal{J}_{n,t}} (N_j/n)\big(\theta_j - \mathrm{E}(\theta_j | \tilde{D}_n)\big),$$

$$A_n'(t; \tilde{D}_n) = \big(n^{2/3}/g(x_0)\big) \sum_{j \in \mathcal{J}_{n,t}} (N_j/n)\big(\mathrm{E}(\theta_j | \tilde{D}_n) - \bar{Y}_j\big),$$

$$B_n(t; \tilde{D}_n) = \big(n^{2/3}/g(x_0)\big) \sum_{j \in \mathcal{J}_{n,t}} (N_j/n)\big(\bar{Y}_j - f_0(x_0)\big).$$

We claim that $\mathcal{L}(A_n(t; \theta, \tilde{D}_n) : t \in [-K, K] | \tilde{D}_n) \rightsquigarrow (aW_2(t) : t \in [-K, K])$ in P_0-probability in $\mathbb{L}_\infty([-K, K])$ for all $K > 0$. As $((\theta_j - \mathrm{E}(\theta_j | \tilde{D}_n)) | \tilde{D}_n) \sim \mathrm{N}(0, \sigma_n^2/(N_j + \lambda_j^{-2}))$ and $\sup\{|\lambda_j| : 1 \leq j \leq J\}$ is bounded by the assumptions on the prior, for any $t \in [-K, K]$, we have that $\mathrm{Var}[A_n(t; \theta, \tilde{D}_n) | \tilde{D}_n]$ is equal to

$$\frac{n^{4/3}}{g^2(x_0)} \sum_{j \in \mathcal{J}_{n,t}} \left(\frac{N_j}{n}\right)^2 \frac{\sigma_n^2}{N_j + \lambda_j^{-2}} = \frac{n^{-2/3}\sigma_n^2}{g^2(x_0)} \sum_{j \in \mathcal{J}_{n,t}} N_j + O\left(\frac{J}{n}\right).$$

As $N_j \sim \mathrm{Bin}(n; G(I_j))$, the expectation of the first term is

$$\frac{n^{1/3}\sigma_0^2}{g^2(x_0)} \int_{\lceil x_0 J \rceil / J}^{\lceil (x_0 + n^{-1/3}t) J \rceil / J} g(u)\, du = \frac{n^{1/3}\sigma_0^2}{g^2(x_0)}\big(g(x_0) + o(1)\big)n^{-1/3}t \to \frac{\sigma_0^2 t}{g(x_0)},$$

uniformly in $t \in [-K, K]$, and its variance

$$\frac{n^{-4/3}\sigma_n^4}{g^4(x_0)} \mathrm{Var}\Big(\sum_{j \in \mathcal{J}_{n,t}} N_j\Big) \lesssim n^{-1/3}\big(g(x_0) + o(1)\big)n^{-1/3}t \to 0,$$

leading to $\mathrm{Var}[A_n(t; \theta, \tilde{D}_n) | \tilde{D}_n] \to_{P_0} \sigma_0^2 t/g(x_0)$ uniformly in $t \in [-K, K]$.

Similarly, $\mathrm{Cov}(A_n(t; \theta, \tilde{D}_n), A_n(s; \theta, \tilde{D}_n) | \tilde{D}_n) \to_{P_0} \sigma_0^2 \min(s, t)/g(x_0) \mathbb{1}\{ts > 0\}$ uniformly in $t, s \in [-K, K]$. Thus the finite-dimensional distributions of $\mathcal{L}(A_n(\cdot; \theta, \tilde{D}_n) | \tilde{D}_n)$ converge to those of aW_1 in probability.

We now show that $\mathcal{L}(A_n(\cdot;\theta,\tilde{D}_n): t\in[0,K]|\tilde{D}_n)$ is tight on $\mathbb{L}_\infty[0,K]$ in probability. The calculations are similar for $[-K,0]$. We shall use the characterization of tightness given by Theorem 18.14 of van der Vaart [55]: for every $\epsilon>0$ and $\eta>0$, there exists a partition $\{T_1,\dots,T_k\}$ of $[0,K]$ with k depending only on ϵ and η such that

$$\mathrm{P}\big(\sup\{|A_n(s,\theta,\tilde{D}_n)-A_n(t,\theta,\tilde{D}_n)|: s,t\in T_l,\ 1\le l\le k\}>\epsilon|\tilde{D}_n\big)<\eta$$

with P_0-probability tending to 1, as $n\to\infty$. To verify this, let $\delta>0$, to be chosen later depending only on ϵ and η, and consider a partition of $[0,K]$ into $T_l=(s_{l-1},s_l]$ with equal lengths at most δ, and $k\le 2/\delta$. Then it suffices to verify that $\mathrm{P}(\sup\{|A_n(s,\theta,\tilde{D}_n)-A_n(t,\theta,\tilde{D}_n)|: s,t\in T_l\}>\epsilon|\tilde{D}_n)<\eta\delta/2$ for all l. Let $s_{l-1}\le t\le s\le s_l$. Then $A_n(s,\theta,\tilde{D}_n)-A_n(t,\theta,\tilde{D}_n)=S_{m_s}-S_{m_t}$, where $m_s=\lceil(x_0+n^{-1/3}s)J\rceil$, $m_t=\lceil(x_0+n^{-1/3}t)J\rceil$, and $S_m=(n^{2/3}/g(x_0))\sum_{j\le m}(N_j/n)(\theta_j-\mathrm{E}(\theta_j|\tilde{D}_n))$, $m=1,2,\dots,J$, are partial sums of independent centered normal variables. To verify the criterion for tightness, it then suffices to verify that

$$\mathrm{P}\left(\max_{m_{s_{l-1}}<m\le m_{s_l}}\left\{\frac{n^{2/3}}{g(x_0))}\left|\sum_{j=m_{s_{l-1}}+1}^{m}\frac{N_j}{n}\big(\theta_j-\mathrm{E}(\theta_j|\tilde{D}_n)\big)\right|\right\}>\epsilon\middle|\tilde{D}_n\right)<\frac{\eta\delta}{4}$$

for all l. By Doob's maximal inequality applied to the submartingale given by the fourth power of the partial sums, the left hand side is bounded by a constant multiple of $\epsilon^{-4}n^{8/3}\sum_{j=m_{s_{l-1}}+1}^{m_{s_l}}\mathrm{E}(|(N_j/n)(\theta_j-\mathrm{E}(\theta_j|\tilde{D}_n))|^4|\tilde{D}_n)$. The terms inside the sum are independent centered normal with variance of the order $(N_j/n)^2(1/N_j)=N_j/n^2\lesssim 1/(Jn)$ uniformly with P_0-probability tending to 1, and there are at most $(\delta n^{-1/3}J+1)$ terms. From this and the fact that the fourth central moment of a normal variable is 3 times the square of its variance, it easily follows that the expression in the last display is bounded by a constant multiple of $n^{8/3}(Jn)^{-2}(\delta n^{-1/3}J)^2=\delta^2$. Thus, choosing δ a sufficiently small multiple of $\eta\epsilon^4$, the tightness criterion is verified.

We now show that $A_n'(\cdot;\tilde{D}_n)\to_{P_0}0$ uniformly in $\mathbb{L}_\infty([-K,K])$. We write $\bar{Y}_j$ as $N_j^{-1}\sum_{i:X_i\in I_j}f_0(X_i)+N_j^{-1}\sum_{i:X_i\in I_j}\varepsilon_i$. Let $M_n=\{\min(N_1,\dots,N_J)\ge a_1n/(2J)\}$. From Lemma A.2, $P_0(M_n)\to 1$. If $\xi_1,\dots,\xi_N$ are i.i.d. centered sub-Gaussian variables, then $\mathrm{P}(|N^{-1}\sum_{i=1}^N\xi_i|>T)\le 2\exp(-C'NT^2)$ for some constant $C'>0$. Choosing $N=N_j$, $\xi_1,\dots,\xi_N$ to be $(\varepsilon_i: X_i\in I_j)$ and using the fact that $X_1,\dots,X_n$ are independent of $\varepsilon_1,\dots,\varepsilon_n$, we obtain

$$P_0\left(\max_{1\le j\le J}\left|N_j^{-1}\sum_{i:X_i\in I_j}\varepsilon_i\right|>T\right)\le\sum_{j=1}^{J}P_0\left(\left|N_j^{-1}\sum_{i:X_i\in I_j}\varepsilon_i\right|>T|M_n\right)+P_0(M_n^c),$$

which is bounded by $2Je^{-na_1C'T^2/(2J)}+o(1)\to 0$, because $n/J\gg J^{1/3}\gg\log J$ as $J\to\infty$. Hence using the monotonicity of f_0, we have

$$\max_{1\le j\le J}|\bar{Y}_j|\le\max_{1\le j\le J}\left|N_j^{-1}\sum_{i:X_i\in I_j}f_0(X_i)\right|+\max_{1\le j\le J}\left|N_j^{-1}\sum_{i:X_i\in I_j}\varepsilon_i\right|=O_{P_0}(1),$$

because the first term is bounded by $\max(|f_0(0)|,|f_0(1)|)$. Then, as the prior means of $\theta_1,\dots,\theta_J$ are bounded and prior variances lie in a compact subinterval of $(0,\infty)$,

$$\begin{aligned}|A_n'(t;\tilde{D}_n)|&=\left|\frac{n^{2/3}}{g(x_0)}\sum_{j\in\mathcal{J}_{n,t}}\frac{N_j}{n}\left(\frac{(\zeta_j-\bar{Y}_j)/\lambda_j^2}{N_j+1/\lambda_j^2}\right)\right|\\&\lesssim\left|\frac{n^{2/3}}{g(x_0)}\max_{1\le j\le J}\left|\zeta_j-\bar{Y}_j\right|\sum_{j\in\mathcal{J}_{n,t}}\frac{N_j}{n}O\big(N_j^{-1}\big)\right|,\end{aligned}\tag{5.6}$$

which is $n^{-2/3}tJO_{P_0}(1)$. As $J \ll n^{2/3}$, we obtain $\sup\{A_n'(t;\tilde{D}_n) : t \in [-K,K]\} \to_{P_0} 0$.

Finally, Lemma S.1 of the Supplementary Material [21] establishes that $(B_n(t;\tilde{D}_n) : t \in [-K,K]) \rightsquigarrow (aW_1(t) + bt^2 : t \in [-K,K])$ in $\mathbb{L}_\infty([-K,K])$, for all $K > 0$. □

LEMMA 5.2. *Let $n^{1/3} \ll J \ll n^{2/3}$. Then $\Pi(g_n^* \notin [-K,K]|D_n) \to_{P_0} 0$ for some $K > 0$, where $g_n^* = \arg\min\{\sum_{j\in\mathcal{J}_{n,t}} \frac{N_j}{n}(\theta_j - f_0(x_0)) - \frac{z}{n^{1/3}}\sum_{j\in\mathcal{J}_{n,t}} \frac{N_j}{n} : t \in \mathbb{R}\}$.*

PROOF. We use the change of variable $r = n^{-1/3}t$, and define

$$h_n^* = \arg\min_{r\in\mathbb{R}}\left\{\sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+r)J\rceil} \frac{N_j}{n}\big(\theta_j - f_0(x_0)\big) - n^{-1/3}z\sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+r)J\rceil} \frac{N_j}{n}\right\}.$$

Therefore, $g_n^* = n^{1/3}h_n^*$. We only prove the tightness of the argmin for r restricted to $[0,\infty)$; the proof for the case $r < 0$ follows similarly, and the intended result by combining. As in the proof of Lemma 5.1, we can condition $\sigma = \sigma_n$, where $\sigma_n \to \sigma_0$. For $r \geq 0$, let $M_n(r)$ and $M(r)$ be defined as

$$\begin{aligned} M_n(r) &= \sum_{j\in\mathcal{J}_r} \frac{N_j}{n}\big(\theta_j - f_0(x_0)\big) \\ &\quad + zn^{-1/3}\mathbb{P}_n\big[\mathbb{1}\{X \leq \lceil (x_0+r)J\rceil/J\} - \mathbb{1}\{X \leq \lceil x_0 J\rceil/J\}\big], \\ M(r) &= \mathrm{E}_0\big[\big(Y - f_0(x_0)\big)\big(\mathbb{1}\{X \leq x_0 + r\} - \mathbb{1}\{X \leq x_0\}\big)\big], \end{aligned}$$

where $\mathcal{J}_r = \{\lceil x_0 J\rceil + 1, \ldots, \lceil (x_0+r)J\rceil\}$, and we have suppressed the dependence of $M_n(r)$ on $\theta_1,\ldots,\theta_J$. We apply Theorem 3.2.5 of van der Vaart and Wellner [56], conditionally on $\tilde{D}_n = \{D_n, \sigma_n\}$, to the process $r \mapsto -M_n(r)$ and the deterministic function $r \mapsto -M(r)$ on the domain $[0,\infty)$, to establish the tightness of the conditional distribution of g_n^* given $\tilde{D}_n$. Note that $-M_n(0) = -M(0) = 0$ and the condition $-M(r) + M(0) \lesssim -r^2$ is verified within the proof of Theorem S.2 of the Supplementary Material [21]. We need to construct functions ϕ_n such that

$$\mathrm{E}_0\sqrt{n}\Big[\mathrm{E}^* \sup_{|r|<\delta} \big|M_n(r) - M(r)\big|\big|\tilde{D}_n\Big] \lesssim \phi_n(\delta), \tag{5.7}$$

for all sufficiently small δ, where $\phi_n(\delta)/\delta^\alpha$ is decreasing in δ for some $\alpha \in (0,2)$. We can write $\sqrt{n}(M_n(r) - M(r))$ as $\sum_{l=1}^3 H_{ln}(r)$, where

$$\begin{aligned} H_{1n}(r) &= \sqrt{n}\textstyle\sum_{j\in\mathcal{J}_r}(N_j/n)\big(\theta_j - \mathrm{E}(\theta_j|\tilde{D}_n)\big), \\ H_{2n}(r) &= \sqrt{n}\textstyle\sum_{j\in\mathcal{J}_r}(N_j/n)\big(\mathrm{E}(\theta_j|\tilde{D}_n) - \bar{Y}_j\big), \\ H_{3n}(r) &= \sqrt{n}\textstyle\sum_{j\in\mathcal{J}_r}(N_j/n)\big(\bar{Y}_j - f_0(x_0)\big) - n^{1/6}z\sum_{j\in\mathcal{J}}(N_j/n) \\ &\qquad -\sqrt{n}\,\mathrm{E}_0\big[\big(Y - f_0(x_0)\big)\big(\mathbb{1}\{X \in (x_0, x_0+r]\}\big)\big]. \end{aligned} \tag{5.8}$$

To bound the variation of H_{1n}, we proceed as in the proof of Lemma 5.1. We view the maximum of the piecewise constant process $r \mapsto H_{1n}(r)$ as the maximum of a partial-sum sequence of centered Gaussian random variables, and apply Doob's submartingale maximal inequality on the square of the partial sum sequence. Thus it suffices to bound the variance of $H_{1n}(\delta)$ and then extract its square root. As the number of terms is of the order δJ and the variance of each summand in $H_{1n}(\delta)$ is bounded by a multiple of $n(N_j/n)^2(1/N_j) \lesssim 1/J$ in P_0-probability simultaneously for all j, by it follows that $\mathrm{E}[|H_{1n}(\delta)|^2|\tilde{D}_n] \leq \delta$, and hence $\mathrm{E}^*[\sup\{|H_{1n}(r)| : |r| < \delta\}|\tilde{D}_n] \lesssim \delta^{1/2}$, in P_0-probability.

For H_{2n}, using arguments similar to (5.6) with multiplication by $n^{1/2}$ instead of $n^{2/3}$ and $n^{-1/3}t$ replaced by r and K by δ, we obtain the estimate $\sqrt{n}\mathrm{E}_0 \sup_{|r|<\delta}\sum_{j\in\mathcal{J}_r}\frac{N_j}{n} \times$

$|\frac{(\zeta_j-\bar{Y}_j)/\lambda_j^2}{N_j+1/\lambda_j^2}| \lesssim n^{-1/2}J\delta$, for sufficiently small δ. As $J \ll n^{2/3}$, we have $n^{-1/2}J\delta \lesssim n^{1/6}\delta$, giving the estimate $\mathrm{E}_0^*[\sup\{|H_{2n}(r)| : |r| < \delta] \lesssim n^{1/6}\delta$. From (S4), we have that $\mathrm{E}_0[\sup\{|H_{3n}(r)| : |r| < \delta] \lesssim \delta^{1/2} + n^{1/6}\delta + n^{-1/6}$. Combining these estimates, for sufficiently small δ, (5.7) holds for $\phi_n(\delta) = C[\delta^{1/2} + n^{1/6}\delta + n^{-1/6}]$ for some $C > 0$. The condition $r_n^2\phi_n(r_n^{-1}) \le \sqrt{n}$ is satisfied for $r_n = n^{1/3}$. The condition that the (conditional) distribution of the $\arg\min M_n(r)$ concentrates at the $\arg\min M(r) = 0$ (in P_0-probability) holds by Lemma S.3. Hence by Theorem 3.2.5 of van der Vaart and Wellner [56] applied conditionally on the data D_n, the posterior distribution of $n^{1/3}h_n^*$ is tight in P_0-probability. □

PROOF OF LEMMA 3.5. Let $W_1'(t) = W_1(-t)$ and $W_2'(t) = W_2(-t)$. Using the transformation $t \mapsto -t$ and the fact that $W_1 \overset{d}{=} W_1'$ and $W_2' \overset{d}{=} W_2$, we have

$$\begin{aligned}
&\mathrm{P}\big(\arg\min\{W_1(t) + W_2(t) + t^2 : t \in \mathbb{R}\} \ge 0 | W_1\big) \\
&\quad = \mathrm{P}\big(\arg\min\{W_1(-t) + W_2(-t) + (-t)^2 : t \in \mathbb{R}\} \le 0 | W_1\big) \\
&\quad \overset{d}{=} \mathrm{P}\big(\arg\min\{W_1'(t) + W_2(t) + t^2 : t \in \mathbb{R}\} \le 0 | W_1'\big).
\end{aligned}$$

Therefore, $\mathrm{P}(\Delta^*_{W_1,W_2} \ge 0 | W_1) \overset{d}{=} \mathrm{P}(\Delta^*_{W_1,W_2} \le 0 | W_1)$, where $\Delta^*_{W_1,W_2} = \arg\min\{W_1(t) + W_2(t) + t^2 : t \in \mathbb{R}\}$, and $\mathrm{P}(\Delta^*_{W_1,W_2} \ge 0 | W_1)$ is symmetric about $1/2$. □

APPENDIX

To establish the asymptotic coverage of pointwise credible intervals, we have used the result that the maximum marginal likelihood estimator for σ^2 in the plug-in Bayes approach or the marginal posterior distribution of σ^2 in the fully Bayes approach, are consistent uniformly for $f_0 \in \mathcal{F}(K)$. The following lemma states this result.

LEMMA A.1. *Let $f_0 \in \mathcal{F}(K)$ for some $K > 0$. If $1 \ll J \ll n$ and Assumption D holds, then*:

(a) *the maximum marginal likelihood estimator $\hat{\sigma}_n^2$ converges in probability to σ_0^2 at the rate* $\max\{n^{-1/2}, n^{-1}J, J^{-1}\}$.

(b) *If $\sigma^2 \sim \mathrm{IG}(\beta_1, \beta_2)$ with $\beta_1 > 2$, $\beta_2 > 0$, then the marginal posterior distribution of σ^2 contracts at σ_0^2 at the rate* $\max\{n^{-1/2}, n^{-1}J, J^{-1}\}$.

The proof follows from similar (actually simpler, because of the univariate setting) calculations using posterior conjugacy as in the proof of Proposition 4.1 of Yoo and Ghosal [57].

LEMMA A.2. *Under Assumption D, $P_0(A_n) \to 1$, where*

$$A_n = \{a_1 n/(2J) \le \min(N_1, \ldots, N_J) \le \max(N_1, \ldots, N_J) \le 2a_2 n/J\}.$$

In particular, $N_1, \ldots, N_J$ are simultaneously of the order n/J in probability.

PROOF. Recall that $N_j \sim \mathrm{Bin}(n; G(I_j))$ and $a_1/J \le G(I_j) \le a_2/J$ for every $1 \le j \le J$. The large deviation probability $\mathrm{P}(N_j \ge 2a_2 n/J)$ is bounded by $e^{-2a_2 nt/J}\mathrm{E}(e^{tN_j})$, which optimized over $t > 0$ leads to the bound

$$\exp\left[-n\left(\frac{2a_2}{2J}\log\left(\frac{2a_2}{JG(I_j)}\right) + \left(1 - \frac{2a_2}{J}\right)\log\left(\frac{1-2a_2/J}{1-G(I_j)}\right)\right)\right] \le 2e^{-Cn/J}$$

for some constant $C > 0$. Similarly, $\mathrm{P}(N_j \le a_1 n/(2J))$ is also bounded by $2e^{-Cn/J}$ for some $C > 0$. Combining, and because $n/J \gg J^{1/3} \gg \log J$, we get the desired result.

Alternatively, we can apply Bennet's inequality (cf. Proposition A.6.2 of van der Vaart and Wellner [56]) with $\lambda = c\sqrt{n}/J$ for some $c > 0$. □

LEMMA A.3. *Let W_1, W_2 be independent two-sided standard Brownian motions starting at zero. Then for $a, b > 0$ and $c \in \mathbb{R}$:*

(a) $\mathcal{L}(\arg\min\{aW_1(t) + aW_2(t) + bt^2 + ct : t \in \mathbb{R}\}|W_1)$
$\stackrel{d}{=} \mathcal{L}((a/b)^{2/3} \arg\min\{W_1(t) + W_2(t) + t^2 : t \in \mathbb{R}\} - c/(2b)|W_1)$;

(b) *For $c_1, c_2 \in \mathbb{R}$,*

$$\begin{aligned}&\big(\arg\min\{aW_1(t_1) + bt_1^2 + c_1t_1 : t_1 \in \mathbb{R}\},\\ &\qquad \arg\min\{aW_1(t_2) + aW_2(t_2) + bt_2^2 + c_2t_2 : t_2 \in \mathbb{R}\}\big)\\ &\stackrel{d}{=} \big((a/b)^{2/3} \arg\min\{W_1(t_1) + t_1^2 : t_1 \in \mathbb{R}\} - c_1/(2b),\\ &\qquad (a/b)^{2/3} \arg\min\{W_1(t_2) + W_2(t_2) + t_2^2 : t_2 \in \mathbb{R}\} - c_2/(2b)\big).\end{aligned}$$

Lemma A.3 follows from Problem 5 of Chapter 3.2 of van der Vaart and Wellner [56], using the transformation $t \mapsto \psi(t) = (a/b)^{2/3}t - c/(2b)$.

Acknowledgments. The second author's research is partially supported by NSF Grant DMS-1916419.

SUPPLEMENTARY MATERIAL

Supplement to "Coverage of credible intervals in nonparametric monotone regression" (DOI: 10.1214/20-AOS1989SUPP; .pdf). The proofs of Theorem 3.1 and Proposition 3.2 are provided in supplementary material [21].

REFERENCES

[1] ARMSTRONG, T. (2015). Adaptive testing on a regression function at a point. *Ann. Statist.* **43** 2086–2101. MR3375877 https://doi.org/10.1214/15-AOS1342

[2] AYER, M., BRUNK, H. D., EWING, G. M., REID, W. T. and SILVERMAN, E. (1955). An empirical distribution function for sampling with incomplete information. *Ann. Math. Stat.* **26** 641–647. MR0073895 https://doi.org/10.1214/aoms/1177728423

[3] BANERJEE, M. (2000). Likelihood ratio inference in regular and non-regular problems. Ph. D. Dissertation—University of Washington. MR2701606

[4] BANERJEE, M. (2007). Likelihood based inference for monotone response models. *Ann. Statist.* **35** 931–956. MR2341693 https://doi.org/10.1214/009053606000001578

[5] BANERJEE, M. and WELLNER, J. A. (2001). Likelihood ratio tests for monotone functions. *Ann. Statist.* **29** 1699–1731. MR1891743 https://doi.org/10.1214/aos/1015345959

[6] BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference Under Order Restrictions. The Theory and Application of Isotonic Regression*. Wiley, London-Sydney. MR0326887

[7] BARLOW, R. E. and BRUNK, H. D. (1972). The isotonic regression problem and its dual. *J. Amer. Statist. Assoc.* **67** 140–147. MR0314205

[8] BELITSER, E. (2017). On coverage and local radial rates of credible sets. *Ann. Statist.* **45** 1124–1151. MR3662450 https://doi.org/10.1214/16-AOS1477

[9] BELITSER, E. and GHOSAL, S. (2020). Empirical Bayes oracle uncertainty quantification for regression. *Ann. Statist.* To appear.

[10] BELITSER, E. and NURUSHEV, N. (2019). Robust inference for general framework of projection structures. Preprint. Available at arXiv:1904.01003.

[11] BELITSER, E. and NURUSHEV, N. (2020). Needles and straw in a haystack: Robust confidence for possibly sparse sequences. *Bernoulli* **26** 191–225. MR4036032 https://doi.org/10.3150/19-BEJ1122
[12] BHAUMIK, P. and GHOSAL, S. (2015). Bayesian two-step estimation in differential equation models. *Electron. J. Stat.* **9** 3124–3154. MR3453972 https://doi.org/10.1214/15-EJS1099
[13] BHAUMIK, P. and GHOSAL, S. (2017). Efficient Bayesian estimation and uncertainty quantification in ordinary differential equation models. *Bernoulli* **23** 3537–3570. MR3654815 https://doi.org/10.3150/16-BEJ856
[14] BOWMAN, A. W., JONES, M. C. and GIJBELS, I. (1998). Testing monotonicity of regression. *J. Comput. Graph. Statist.* **7** 489–500.
[15] BRUNK, H. D. (1970). Estimation of isotonic regression. In *Nonparametric Techniques in Statistical Inference* (*Proc. Sympos., Indiana Univ., Bloomington, Ind.*, 1969) 177–197. Cambridge Univ. Press, London. MR0277070
[16] CAI, T. T. and LOW, M. G. (2006). Adaptive confidence balls. *Ann. Statist.* **34** 202–228. MR2275240 https://doi.org/10.1214/009053606000000146
[17] CAI, T. T., LOW, M. G. and XIA, Y. (2013). Adaptive confidence intervals for regression functions under shape constraints. *Ann. Statist.* **41** 722–750. MR3099119 https://doi.org/10.1214/12-AOS1068
[18] CASTILLO, I. and NICKL, R. (2013). Nonparametric Bernstein–von Mises theorems in Gaussian white noise. *Ann. Statist.* **41** 1999–2028. MR3127856 https://doi.org/10.1214/13-AOS1133
[19] CASTILLO, I. and NICKL, R. (2014). On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures. *Ann. Statist.* **42** 1941–1969. MR3262473 https://doi.org/10.1214/14-AOS1246
[20] CASTILLO, I. and SZABÓ, B. (2020). Spike and slab empirical Bayes sparse credible sets. *Bernoulli* **26** 127–158. MR4036030 https://doi.org/10.3150/19-BEJ1119
[21] CHAKRABORTY, M. and GHOSAL, S. (2021). Supplement to "Coverage of credible intervals in nonparametric monotone regression." https://doi.org/10.1214/20-AOS1989SUPP
[22] CHERNOFF, H. (1964). Estimation of the mode. *Ann. Inst. Statist. Math.* **16** 31–41. MR0172382 https://doi.org/10.1007/BF02868560
[23] COX, D. D. (1993). An analysis of Bayesian inference for nonparametric regression. *Ann. Statist.* **21** 903–923. MR1232525 https://doi.org/10.1214/aos/1176349157
[24] DE LEEUW, J., KURT, H. and MAIR, P. (2009). Isotone optimization in R: Pool-adjacent-violators algorithm (PAVA) and active set methods. *J. Stat. Softw.* **32**.
[25] DÜMBGEN, L. (2003). Optimal confidence bands for shape-restricted curves. *Bernoulli* **9** 423–449. MR1997491 https://doi.org/10.3150/bj/1065444812
[26] DÜMBGEN, L. and JOHNS, R. B. (2004). Confidence bands for isotonic median curves using sign tests. *J. Comput. Graph. Statist.* **13** 519–533. MR2063998 https://doi.org/10.1198/1061860043506
[27] GHOSAL, S., SEN, A. and VAN DER VAART, A. W. (2000). Testing monotonicity of regression. *Ann. Statist.* **28** 1054–1082. MR1810919 https://doi.org/10.1214/aos/1015956707
[28] GIJBELS, I., HALL, P., JONES, M. C. and KOCH, I. (2000). Tests for monotonicity of a regression mean with guaranteed level. *Biometrika* **87** 663–673. MR1789816 https://doi.org/10.1093/biomet/87.3.663
[29] GINÉ, E. and NICKL, R. (2010). Confidence bands in density estimation. *Ann. Statist.* **38** 1122–1170. MR2604707 https://doi.org/10.1214/09-AOS738
[30] GRENANDER, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuarietidskr.* **39** 125–153. MR0093415 https://doi.org/10.1080/03461238.1956.10414944
[31] GROENEBOOM, P. (1989). Brownian motion with a parabolic drift and Airy functions. *Probab. Theory Related Fields* **81** 79–109. MR0981568 https://doi.org/10.1007/BF00343738
[32] GROENEBOOM, P. and JONGBLOED, G. (2014). *Nonparametric Estimation Under Shape Constraints*: *Estimators, Algorithms and Asymptotics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **38**. Cambridge Univ. Press, New York. MR3445293 https://doi.org/10.1017/CBO9781139020893
[33] GROENEBOOM, P. and WELLNER, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. *DMV Seminar* **19**. Birkhäuser, Basel. MR1180321 https://doi.org/10.1007/978-3-0348-8621-5
[34] GROENEBOOM, P. and WELLNER, J. A. (2001). Computing Chernoff's distribution. *J. Comput. Graph. Statist.* **10** 388–400. MR1939706 https://doi.org/10.1198/10618600152627997
[35] HALL, P. and HECKMAN, N. E. (2000). Testing for monotonicity of a regression mean by calibrating for linear functions. *Ann. Statist.* **28** 20–39. MR1762902 https://doi.org/10.1214/aos/1016120363
[36] HOFFMANN, M. and NICKL, R. (2011). On adaptive inference and confidence bands. *Ann. Statist.* **39** 2383–2409. MR2906872 https://doi.org/10.1214/11-AOS903
[37] HUANG, J. and WELLNER, J. A. (1995). Estimation of a monotone density or monotone hazard under random censoring. *Scand. J. Stat.* **22** 3–33. MR1334065
[38] HUANG, Y. and ZHANG, C.-H. (1994). Estimating a monotone density from censored observations. *Ann. Statist.* **22** 1256–1274. MR1311975 https://doi.org/10.1214/aos/1176325628

[39] KNAPIK, B. T., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2011). Bayesian inverse problems with Gaussian priors. *Ann. Statist.* **39** 2626–2657. MR2906881 https://doi.org/10.1214/11-AOS920
[40] KOSOROK, M. R. (2008). Bootstrapping in Grenander estimator. In *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen. Inst. Math. Stat. (IMS) Collect.* **1** 282–292. IMS, Beachwood, OH. MR2462212 https://doi.org/10.1214/193940307000000202
[41] LEAHU, H. (2011). On the Bernstein–von Mises phenomenon in the Gaussian white noise model. *Electron. J. Stat.* **5** 373–404. MR2802048 https://doi.org/10.1214/11-EJS611
[42] LI, K.-C. (1989). Honest confidence regions for nonparametric regression. *Ann. Statist.* **17** 1001–1008. MR1015135 https://doi.org/10.1214/aos/1176347253
[43] LIN, L. and DUNSON, D. B. (2014). Bayesian monotone regression using Gaussian process projection. *Biometrika* **101** 303–317. MR3215349 https://doi.org/10.1093/biomet/ast063
[44] LOW, M. G. (1997). On nonparametric confidence intervals. *Ann. Statist.* **25** 2547–2554. MR1604412 https://doi.org/10.1214/aos/1030741084
[45] PATSCHKOWSKI, T. and ROHDE, A. (2019). Locally adaptive confidence bands. *Ann. Statist.* **47** 349–381. MR3909936 https://doi.org/10.1214/18-AOS1690
[46] PICARD, D. and TRIBOULEY, K. (2000). Adaptive confidence interval for pointwise curve estimation. *Ann. Statist.* **28** 298–335. MR1762913 https://doi.org/10.1214/aos/1016120374
[47] PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31** 23–36. MR0267677
[48] RAY, K. (2017). Adaptive Bernstein–von Mises theorems in Gaussian white noise. *Ann. Statist.* **45** 2511–2536. MR3737900 https://doi.org/10.1214/16-AOS1533
[49] ROBINS, J. and VAN DER VAART, A. (2006). Adaptive nonparametric confidence sets. *Ann. Statist.* **34** 229–253. MR2275241 https://doi.org/10.1214/009053605000000877
[50] SALOMOND, J.-B. (2018). Testing un-separated hypotheses by estimating a distance. *Bayesian Anal.* **13** 461–484. MR3780431 https://doi.org/10.1214/17-BA1059
[51] SCHMIDT-HIEBER, J., MUNK, A. and DÜMBGEN, L. (2013). Multiscale methods for shape constraints in deconvolution: Confidence statements for qualitative features. *Ann. Statist.* **41** 1299–1328. MR3113812 https://doi.org/10.1214/13-AOS1089
[52] SEN, B., BANERJEE, M. and WOODROOFE, M. (2010). Inconsistency of bootstrap: The Grenander estimator. *Ann. Statist.* **38** 1953–1977. MR2676880 https://doi.org/10.1214/09-AOS777
[53] SNIEKERS, S. and VAN DER VAART, A. (2015). Credible sets in the fixed design model with Brownian motion prior. *J. Statist. Plann. Inference* **166** 78–86. MR3390135 https://doi.org/10.1016/j.jspi.2014.07.008
[54] SZABÓ, B., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2015). Frequentist coverage of adaptive nonparametric Bayesian credible sets. *Ann. Statist.* **43** 1391–1428. MR3357861 https://doi.org/10.1214/14-AOS1270
[55] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.
[56] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Process with Applications to Statistics*. Springer, New York.
[57] YOO, W. W. and GHOSAL, S. (2016). Supremum norm posterior contraction and credible sets for nonparametric multivariate regression. *Ann. Statist.* **44** 1069–1102. MR3485954 https://doi.org/10.1214/15-AOS1398