

Rates and coverage for monotone densities using projection-posterior

MOUMITA CHAKRABORTY* and SUBHASHIS GHOSAL†

*Department of Statistics, North Carolina State University, 2311 Stinson Drive, Raleigh, North Carolina 27695-8203, USA. E-mail: *mchakra@ncsu.edu; †sghosal@ncsu.edu*

We consider Bayesian inference for a monotone density on the unit interval and study the resulting asymptotic properties. We consider a “projection-posterior” approach, where we construct a prior on density functions through random histograms without imposing the monotonicity constraint, but induce a random distribution by projecting a sample from the posterior on the space of monotone functions. The approach allows us to retain posterior conjugacy, allowing explicit expressions extremely useful for studying asymptotic properties. We show that the projection-posterior contracts at the optimal $n^{-1/3}$ -rate. We then construct a consistent test based on the posterior distribution for testing the hypothesis of monotonicity. Finally, we obtain the limiting coverage of a projection-posterior credible interval for the value of the function at an interior point. Interestingly, the limiting coverage turns out to be higher than the nominal credibility level, the opposite of the undercoverage phenomenon observed in a smoothness regime. Moreover, we show that a recalibration method using a lower credibility level gives an intended limiting coverage. We also discuss extensions of the obtained results for densities on the half-line. We conduct a simulation study to demonstrate the accuracy of the asymptotic results in finite samples.

Keywords: Monotone density; contraction rate; Bayesian test for monotonicity; credible interval; coverage

1. Introduction

We consider the problem of making inference about a bounded, monotone density g on a bounded interval, based on independent and identically distributed (i.i.d.) observations. The domain can be taken to be the unit interval, without loss of generality. We shall also discuss extensions of our results to the positive half-line in the end. Shape restrictions like monotonicity or convexity often naturally appear in modeling, especially in the context of inverse problems; see Groeneboom and Jongbloed [29] for details. The maximum likelihood estimator (MLE) of a monotone decreasing density g on the positive half-line exists, and is characterized as the left-continuous derivative of the least concave majorant of the empirical distribution function of the data (Grenander [27]), commonly known as the Grenander estimator. Prakasa Rao [35] obtained the limiting distribution of the Grenander estimator based on n i.i.d. observations at an interior point, centered and scaled by $n^{1/3}$. The asymptotic distribution is given by the distribution of the minimizer Z of a two-sided Brownian motion with a parabolic drift, and is commonly called the Chernoff distribution, named after the work of Chernoff [17]. Groeneboom [28] computed the density of Z , and Groeneboom and Wellner [30] tabulated its quantiles. A confidence interval for the density at a point may be obtained from the limiting distribution, but that approach involves the additional challenging work of estimating the derivative of the function. Estimation of monotone regression, commonly known as isotonic regression, was addressed well in the literature (Barlow and Brunk [5], Barlow et al. [4], Leeuw et al. [20], Brunk [11]). Huang and Zhang [33] and Huang and Wellner [32] respectively obtained estimators of a monotone density and a monotone hazard rate under right-censoring. In shape-restricted inference, confidence sets were constructed by Dümbgen [21], Dümbgen and Johns [22], Cai et al. [14], Schmidt-Hieber et al. [39], Banerjee and Wellner [3] and Banerjee [2]. Testing the hypothesis of monotonicity of a regression function was addressed by Bowman et al. [10], Hall and Heckman [31] Ghosal et al. [24], Gijbels et al. [26], and others.

Bayesian methods for shape-restricted inference and their properties have been studied less extensively in the literature. A prior on a decreasing density may be constructed by using its well-known representation as a scale mixture of the uniform kernel, by putting a Dirichlet process prior on the mixing distribution. Brunner and Lo [13] and Brunner [12] used this technique to construct priors on unimodal densities. Wu and Ghosal [45] showed that the resulting prior assigns positive probabilities to Kullback–Leibler neighborhoods of a decreasing true density, and hence the posterior is weakly consistent by a well-known theorem of Schwartz [40]. Salomond [36] showed that the posterior concentrates at the optimal rate $n^{-1/3}$ up to a logarithmic factor. Salomond [38] proposed a Bayesian approach to testing the hypothesis of monotonicity of regression. Very recently, a “projection-posterior” approach to inference in a monotone regression model has been proposed by Chakraborty and Ghosal [15, 16]. They considered a prior distribution based on random step functions by putting a prior distribution on the step-heights, without imposing the monotonicity constraint on them. This allows us to retain conjugacy and easier expressions for the posterior distribution. To impose the monotonicity constraint, they projected a sample from the posterior distribution on the space of monotone functions, thus inducing a random measure appropriate for inference. This idea of embedding the functions in a larger model where the posterior is easily represented and then using the projection map to comply with the restriction was earlier used by Lin and Dunson [34], who used a Gaussian process prior for the same problem, and by Bhaumik and Ghosal [6–8] for regression models driven by ordinary differential equations. The technique has been recently used by Bhaumik et al. [9] for regression models described by partial differential equations. For monotone regression, the projection map on posterior samples can be easily computed by the Pool–Adjacent–Violators Algorithm (PAVA) (pages 9–15, Section 2.3 of Barlow et al. [4]). Chakraborty and Ghosal [15] showed that the resulting posterior contracts optimally at the rate $n^{-1/3}$ and constructed a universally consistent test for the hypothesis of monotonicity following Salomond’s [38] idea of extending the null region by the amount given by the contraction rate. The asymptotic coverage of a Bayesian credible interval for the value of the regression function at an interior point using quantiles of the projection-posterior distribution was computed by Chakraborty and Ghosal [16].

In the present paper, we propose Bayesian methods for estimation, uncertainty quantification and testing for a bounded, monotone decreasing density g on $[0, 1]$, based on n i.i.d. observations X_1, \dots, X_n from g . We pursue the projection-posterior approach. This is especially useful in studying the coverage of a credible interval for the function value $g(x_0)$ at an interior point x_0 . We assign a random histogram prior on g where the probability contents of the intervals are jointly given a Dirichlet distribution. This is equivalent with a finite random series prior (Shen and Ghosal [42]) using the basis consisting of indicators of equal-length disjoint intervals and a Dirichlet distribution on the vector of coefficients. The resulting unrestricted posterior is conjugate, allowing a simple description and easy generation of samples for projection. In particular, a representation of a Dirichlet distribution in terms of independent gamma variables allows us to use the techniques of empirical processes to study the asymptotic properties of the projection-posterior process through a switch-relation expressing the slope of the least concave majorant with the minimizer of the original function perturbed by a quadratic term. We compute the limiting coverage of a $(1 - \alpha)$ -credible interval for $g(x_0)$. Interestingly, the asymptotic coverage is not $1 - \alpha$, but is higher. This is the opposite of a phenomenon Cox [18] observed about arbitrarily low limiting coverage of Bayesian credible sets in smoothing regimes. The present “reverse-Cox phenomenon” has been recently observed by Chakraborty and Ghosal [16] in the context of monotone regression. Like them, we show that the limit is driven by two independent Brownian motions, and is free of any unknown parameters, thus allowing a recalibration to meet a targeted coverage.

The rest of the paper is organized as follows. Section 2 lists the notations, describes the prior distribution and introduces the notion of a projection-posterior. Section 3 states results on posterior contraction rates, Bayesian testing errors, and coverage of a projection-posterior credible interval for the density

function at a fixed interior point. Simulation results are discussed in Section 4. Proofs of the main results are provided in Section 5. The paper concludes with a discussion on an extension of the proposed methods to a decreasing density on the domain $(0, \infty)$. The appendix section contains the proofs of the auxiliary results.

2. Setup, prior and projection-posterior

We consider n i.i.d. observations

$$X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} g, \quad (2.1)$$

from a distribution G on $(0, 1]$ with density g . We consider the problems of obtaining a posterior contraction rate for Bayesian estimation of g and the coverage of a Bayesian credible interval for $g(x_0)$, the value of the density g at a fixed $x_0 \in (0, 1)$ under the assumption that g is bounded and decreasing, and also the problem of testing the hypothesis that g is decreasing. Let g_0 stand for the true density, and $E_0(\cdot)$ and $\text{Var}_0(\cdot)$ respectively stand for the expectation and variance operators taken under g_0 . We write D_n for the data (X_1, \dots, X_n) . Let a prior distribution on g be denoted by Π and let $\Pi(\cdot | D_n)$ refer to the corresponding posterior distribution.

For $g : (0, 1] \mapsto \mathbb{R}$ and d the \mathbb{L}_1 - or the \mathbb{L}_2 -distance, let the projection of g on \mathcal{F} with respect to d be the function g^* that minimizes $d(g, h)$ over $h \in \mathcal{F}$. Let $d(g, \mathcal{F}) = d(g, g^*)$ denote the value of the minimized distance. The induced posterior distribution of g^* , to be called the projection-posterior distribution, will be viewed as a random measure for inference on g .

We put a prior on g through step functions with J equispaced knots j/J , $j = 1, \dots, J$, and an appropriate joint distribution on the step heights, where J may be chosen deterministically (depending on n) or may be given a further prior. More specifically, let $I_j = ((j-1)/J, j/J]$, $j = 1, \dots, J$, and

$$g = J \sum_{j=1}^J \theta_j \mathbb{1}_{I_j}, \quad (\theta_1, \dots, \theta_J) \sim \text{Dir}(\alpha_1, \dots, \alpha_J), \quad (2.2)$$

where $b_1 J^{-b_2} \leq \alpha_j \leq b_3$ for all $j = 1, \dots, J$, for some positive constants b_1, b_2, b_3 . The parameter J is chosen deterministically as $J = J_n$ suitably depending on n , or is given a prior distribution π . Thus the prior is a finite random series prior with the basis $(\mathbb{1}_{I_j} : 1 \leq j \leq J)$ and Dirichlet prior on the coefficients, as termed by Shen and Ghosal [42]. We shall write $\alpha = \sum_{j=1}^J \alpha_j$ and note that $\alpha = O(J)$ under the assumed condition. The associated sieve on which the prior is supported is thus given by $\mathcal{F}_J = \{g = J \sum_{j=1}^J \theta_j \mathbb{1}_{I_j} : \theta_1 \geq \dots \geq \theta_J\}$. It is immediate that the posterior distribution of g (given J) is again of the same form, with α_j updated to $\alpha_j^* = \alpha_j + N_j$, where $N_j = \sum_{i=1}^n \mathbb{1}\{X_i \in I_j\}$, $j = 1, \dots, J$. Here, to keep the notation simple, we have suppressed the dependence on J , in that, $N_1, \dots, N_J, \theta_1, \dots, \theta_J$, and $\alpha_1, \dots, \alpha_J$ are triangular arrays of numbers, and I_1, \dots, I_J are a triangular array of disjoint intervals partitioning $(0, 1]$. For a given real-valued function h on $(0, 1]$, we denote by h_J the step function $J \sum_{j=1}^J \eta_j \mathbb{1}_{I_j}$, where $\eta_j = \int_{I_j} h(x) dx$. If h is a probability density, so is h_J . In particular, if g_0 denotes the true density, $g_{0J} = J \sum_{j=1}^J (\int_{I_j} g_0) \mathbb{1}_{I_j}$, its projection on the sieve, will serve as a useful approximation to g_0 in the proofs.

Since we do not impose monotonicity on g in the prior, the posterior measure also does not comply with the monotonicity restriction, and needs to be modified. Let Π_n^* be the distribution of the \mathbb{L}_1 -projection g^* of g on \mathcal{F} , that is, for every measurable subset B of \mathcal{F} , $\Pi_n^*(B) = \Pi(g : g^* \in B | D_n)$, and is viewed as a random probability measure quantifying the uncertainty in g after observing the sample.

By arguments similar to those in the proof of Lemma 2.2. of Groeneboom and Jongbloed [29], it follows that the projection of a density $g = J \sum_{j=1}^J \theta_j \mathbb{1}_{I_j}$ is also a piece-wise constant decreasing density $g^* = J \sum_{j=1}^J \theta_j^* \mathbb{1}_{I_j}$, that is, $\theta_1^* \geq \dots \geq \theta_J^*$ and $\sum_{j=1}^J \theta_j^* = 1$. In particular, the projection-posterior charges only decreasing densities, and hence is suitable for inference on a decreasing density. Further, by Theorem 2.1 of Groeneboom and Jongbloed [29], it follows that both \mathbb{L}_1 - and \mathbb{L}_2 -projection of $g = J \sum_{j=1}^J \theta_j \mathbb{1}_{I_j}$ are obtained by minimizing $\sum_{j=1}^J |\theta_j - \theta_j^*|^2$ subject to $\theta_1^* \geq \dots \geq \theta_J^*$. The solution of this constrained minimization problem is given by the slopes (left-derivatives) of the least concave majorant of the graph of the line segments connecting the points

$$\left\{ (0, 0), (1/J, \theta_1/J), (2/J, (\theta_1 + \theta_2)/J), \dots, \left(1, \left(\sum_{j=1}^J \theta_j\right)/J\right) \right\}, \quad (2.3)$$

at the points j/J , $j = 1, \dots, J$; see Section 2.1 of Groeneboom and Jongbloed [29]. In particular, the \mathbb{L}_1 - and \mathbb{L}_2 -projections are the same, and henceforth will be unambiguously referred to as the monotone projection. This can be computed very efficiently by the PAVA for any sample of $(\theta_1, \dots, \theta_J)$ from its conjugate Dirichlet posterior.

We shall use the following notations throughout this paper. For two sequences of real numbers a_n and b_n , $a_n \lesssim b_n$ (equivalently, $b_n \gtrsim a_n$) means that a_n/b_n is bounded, $a_n \ll b_n$ means that $a_n/b_n \rightarrow 0$, $a_n \asymp b_n$ means that they are of the same order, and $a_n \sim b_n$ means that $a_n/b_n \rightarrow 1$. Let $\mathbb{N} = \{1, 2, \dots\}$ be the set of positive integers and let $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. For a random variable X and a sequence of random variables X_n , $X_n \rightsquigarrow X$ means that X_n converges in distribution to X , $X_n \rightarrow_p X$ means that X_n converges to X in probability, $X_n = o_P(Y_n)$ for another sequence Y_n means that $X_n/Y_n \rightarrow_p 0$, $X_n = O_P(Y_n)$ means that X_n/Y_n is bounded with probability arbitrarily close to 1. For random variables X and Y , $X \stackrel{d}{=} Y$ 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 .

Let the space of real-valued, bounded, monotone decreasing functions on $(0, 1]$ be denoted by \mathcal{F}_+ , its \mathbb{L}_1 -closure by $\bar{\mathcal{F}}_+$, and the space of monotone decreasing functions on $(0, 1]$ bounded in absolute value by $K > 0$ be denoted by $\mathcal{F}_+(K)$. A real-valued function f on $(0, 1]$ is called β -Hölder continuous for some $\beta \in (0, 1)$ if there exists some $L > 0$ such that for all $x, y \in (0, 1)$, $|f(x) - f(y)| \leq L|x - y|^\beta$. The set of β -Hölder continuous functions on $(0, 1]$ is denoted by $\mathcal{H}(\beta, L)$. For $T \subset \mathbb{R}$, $\mathbb{L}_\infty(T)$ denotes the space of bounded functions on T .

3. Main results

The following result gives the contraction rate of the projection-posterior distribution with respect to the \mathbb{L}_1 -distance on densities.

Theorem 3.1 (Posterior contraction for monotone density). *Let $g_0 \in \mathcal{F}_+$ and $1 \ll J \ll n$. Then, with $\epsilon_n = \max\{\sqrt{J/n}, J^{-1}\}$, $\mathbb{E}_0 \Pi_n^*(\|g - g_0\|_1 > M_n \epsilon_n) \rightarrow 0$ for every $M_n \rightarrow \infty$. In particular, when $J \asymp n^{1/3}$, the projection-posterior contracts at the optimal rate $\epsilon_n = n^{-1/3}$.*

If g_0 lies between two positive numbers, J has prior probability mass function π and for some $a_1, a_2 > 0$,

$$e^{-a_1 J \log J} \leq \pi(J) \leq e^{-a_2 J \log J}, \quad (3.1)$$

then $\mathbb{E}_0 \Pi_n^(\|g - g_0\|_1 > M_0(n/\log n)^{-1/4}) \rightarrow 0$, for some $M_0 > 0$.*

Remark 3.1. In the above result, the posterior contraction for the projection-posterior is inherited from the unrestricted posterior using the triangle inequality (see the proof of Theorem 3.1). Therefore, the contraction rate is also valid about the unrestricted posterior. It is also not essential that the true density is monotone. The same proof shows that at any true density g_0 , monotone or not, the \mathbb{L}_1 -contraction rate is $\max(\sqrt{J/n}, \|g_0 - g_{0J}\|_1)$. We shall use this fact to construct a Bayesian test for monotonicity.

Putting a prior on J in the second part of the theorem is unnecessary for optimal posterior contraction. In fact, the obtained rate is suboptimal. However, the conclusion is needed to study the asymptotic properties of a Bayesian test with power adapting to smoothness in Theorem 3.3 below.

A natural test for the hypothesis of monotonicity is given by the posterior probability of \mathcal{F}_+ : reject the hypothesis if $\Pi(g \in \mathcal{F}_+ | D_n)$ is smaller than 1/2, say. However, in spite of posterior consistency at a monotone g_0 , the above posterior probability may be low because g_0 may be approximated by non-monotone functions. To compensate for that, the set \mathcal{F}_+ needs to be enlarged to the extent the posterior allows a departure from g_0 . With the choice $J \asymp n^{1/3}$ which corresponds to the best contraction rate $n^{-1/3}$ in the \mathbb{L}_1 -distance, the enlargement needs to be $M_n n^{-1/3}$ for some slowly growing sequence $M_n \rightarrow \infty$. This approach was also pursued by Salomond [37,38] and Chakraborty and Ghosal [15] for monotone regression. We show that the level goes to zero, the power goes to 1 at a fixed alternative, and the power is high against a sufficiently separated smooth alternative even when it approaches the null.

Theorem 3.2 (Test for monotonicity). *Consider a test for monotonicity given by $\phi_n = \mathbb{1}\{\Pi(d(f, \mathcal{F}_+) \leq M_n n^{-1/3} | D_n) < \gamma\}$, where d is the \mathbb{L}_1 -distance, $0 < \gamma < 1$ is any predetermined constant, $J \asymp n^{1/3}$ and $M_n \rightarrow \infty$ is a chosen slowly varying sequence. Then the following assertions hold.*

- (a) (Consistency under H_0): For any fixed $g_0 \in \mathcal{F}_+$, $E_0 \phi_n \rightarrow 0$, and further the convergence is uniform over $\mathcal{F}_+(K)$ for any $K > 0$.
- (b) (Universal Consistency): For any fixed density $g_0 \notin \bar{\mathcal{F}}_+$, $E_0(1 - \phi_n) \rightarrow 0$.
- (c) (High power at converging smooth alternatives): For any $0 < \beta \leq 1$ and $L > 0$, $\sup\{E_0(1 - \phi_n) : f_0 \in \mathcal{H}(\beta, L), d(f_0, \mathcal{F}_+) > \rho_n(\beta)\} \rightarrow 0$, where

$$\rho_n(\beta) = \begin{cases} Cn^{-\beta/3}, & \text{for some } C > 0 \text{ if } \beta < 1, \\ CM_n n^{-1/3}, & \text{for any } C > 1 \text{ if } \beta = 1. \end{cases}$$

The procedure involving the test ϕ_n is computationally simple as it does not involve a prior on J . However, with a deterministic choice of J , the required separation from the null hypothesis does not adapt to the smoothness β of the function class in the alternative. More specifically, an order of separation $n^{-\beta/3}$ (up to a logarithmic factor) is needed, which is larger than the optimal order $n^{-\beta/(1+2\beta)}$ of separation for $\beta < 1$. Adaptation can however be restored by using a prior on J and letting the cut-off value for the discrepancy with \mathcal{F}_+ depend on J .

Theorem 3.3 (Adaptive test for monotonicity). *In a prior for g given by (2.2), let J be also undetermined and be given the prior (3.1). Let a test for monotonicity be given by $\phi_n = \mathbb{1}\{\Pi(d(f, \mathcal{F}_+) \leq M_0 \sqrt{J(\log n)/n} | D_n) < \gamma\}$, where d is the \mathbb{L}_1 -distance, $0 < \gamma < 1$ is a predetermined constant and $M_0 > 0$ is a sufficiently large constant. Then,*

- (a) (Consistency under H_0): for any fixed $g_0 \in \mathcal{F}_+$ lying between two positive numbers, $E_0 \phi_n \rightarrow 0$, and the convergence is uniform over $\mathcal{F}_+(K)$;

- (b) (*Universal Consistency*): for any fixed $g_0 \notin \bar{\mathcal{F}}_+$ lying between two positive numbers, $E_0(1 - \phi_n) \rightarrow 0$;
- (c) (*Adaptive power at converging smooth alternatives*): for $g_0 \notin \mathcal{F}_+$, $g_0 \in \mathcal{H}(\beta, L)$, there exists C depending on β and L only such that

$$\sup\{E_0(1 - \phi_n) : g_0 \in \mathcal{H}(\beta, L), d(g_0, \mathcal{F}_+) > C(n/\log n)^{-\beta/(1+2\beta)}\} \rightarrow 0,$$

provided that g_0 lies between two positive numbers.

Now we turn to the study of the limiting coverage of a posterior credible set. More specifically, we are interested in quantifying the uncertainty in the value of the density function g at a given interior point assuming that g is bounded and globally decreasing, and obtain the limiting coverage of the resulting posterior credible interval. Let $x_0 \in (0, 1)$ be such that $g'(x_0)$ exists and $g'(x_0) < 0$. We consider a projection-posterior credible interval for $g(x_0)$ using posterior quantiles of $g^*(x_0) = J \sum_{j=1}^J \theta_j^* \mathbf{1}\{x_0 \in I_j\}$, where θ_j^* , $j = 1, \dots, J$, are obtained from (2.3). To study the limiting shape of the posterior distribution, following the clue from the classical Bernstein–von Mises theorem, it seems natural to center $g(x_0)$ by a suitable estimator and scale the difference by $n^{1/3}$. A natural candidate is the MLE of g , but it is not structurally similar to the projection-posterior distribution, which is supported on $\mathcal{F}_+ \cap \mathcal{F}_J$. A remedy is to center at the sieve-MLE \hat{g}_n of g obtained by maximizing the likelihood over $g \in \mathcal{F}_+ \cap \mathcal{F}_J$, with a given choice of J . The following result shows that an analog of the Bernstein–von Mises theorem does not hold but it nevertheless gives a useful intermediate result necessary to obtain the limiting coverage in the theorem following it.

Let $a = \sqrt{g_0(x_0)}$, $b = |g'_0(x_0)|/2$ and $C_0 = 2b(a/b)^{2/3}$. For a continuous function w on \mathbb{R} , let $\Delta_w^* = \arg \max\{w(t) - t^2 : t \in \mathbb{R}\}$. Let W_1, W_2 be independent two-sided Brownian motions on \mathbb{R} with $W_1(0) = W_2(0) = 0$, $Z = \Delta_{W_1}^*$ and $Z_B = P(\Delta_{W_1+W_2}^* \leq 0 | W_1)$.

In the first part of the result, we show that the limiting distribution of the sieve-MLE is the classical Chernoff distribution, which is the limiting distribution of the MLE over the whole of \mathcal{F}_+ .

Theorem 3.4 (Point-wise distributional limit). *If $n^{1/3} \ll J \ll n^{2/3}$, then the following assertions hold.*

- (a) *For every $z \in \mathbb{R}$, $P_0(n^{1/3}(\hat{g}_n(x_0) - g_0(x_0)) \leq z) \rightarrow P(C_0 Z \leq z)$.*
- (b) *For every $z \in \mathbb{R}$, $P_0 \times \Pi(n^{1/3}(g^*(x_0) - g_0(x_0)) \leq z | D_n) \rightarrow P(C_0 \Delta_{W_1+W_2}^* \leq z)$.*
- (c) *For any $z \in \mathbb{R}$, the conditional probability $\Pi(n^{1/3}(g^*(x_0) - \hat{g}_n(x_0)) \leq z | D_n)$ does not have a limit in probability.*

Define stochastic processes F_n^* and F^* on \mathbb{R} by $F_n^*(z | D_n) = \Pi(n^{1/3}(g^*(x_0) - g_0(x_0)) \leq z | D_n)$ and $F^*(z | W_1) = P(2b(a/b)^{2/3} \Delta_{W_1+W_2}^* \leq z | W_1)$. For every $n \geq 1$, $\gamma \in [0, 1/2]$, define the $(1 - \gamma)$ -th posterior quantile $Q_{n,\gamma} = \inf\{z \in \mathbb{R} : \Pi(g^*(x_0) \leq z | D_n) \geq 1 - \gamma\}$, the associated two-sided $(1 - \gamma)$ -credible interval $I_{n,\gamma} = [Q_{n,1-\gamma/2}, Q_{n,\gamma/2}]$. We then have the following result. The interesting aspect is that the unknown scaling constant C_0 disappears from the expression for the limiting coverage.

Theorem 3.5 (Coverage of credible interval). *If $n^{1/3} \ll J \ll n^{2/3}$, then*

- (a) *for every $z \in \mathbb{R}$, $F_n^*(z | D_n) \rightsquigarrow F^*(z | W_1)$;*
- (b) *the limiting coverage of $I_{n,\gamma}$ is given by*

$$P_0(g_0(x_0) \in I_{n,\gamma}) \rightarrow P(\gamma/2 \leq Z_B \leq 1 - \gamma/2).$$

For $0 < u < 1$, let $A(u) = P(Z_B \leq u)$ be the distribution function of Z_B , to be called the Bayes–Chernoff distribution. Then it follows from Lemma 3.5 of Chakraborty and Ghosal [16] that $A(1-u) = 1 - A(u)$. Hence the limiting coverage in (b) can be written as $1 - 2A(\gamma/2)$, which only depends on γ , not on any characteristics of g_0 . Thus, even though the limiting coverage is not identical with the nominal credibility $1 - \gamma$, the limiting coverage can be computed for each credibility level. These values were given in Table 1 of Chakraborty and Ghosal [16] numerically using Monte Carlo. This reveals that the limiting coverage is slightly higher than the nominal credibility when it is over 50%, and a targeted coverage $1 - \alpha$ can be obtained by choosing $1 - \gamma = 1 - 2A^{-1}(\alpha/2)$, which can be obtained from Table 2 of Chakraborty and Ghosal [16]. For instance, to obtain 95% asymptotic coverage, the credibility level needs to be set to 93.4%. If viewed as a mechanism to obtain a confidence interval with a targeted coverage, the procedure does not require estimation and plugging-in of any nuisance parameters, which is a big advantage of the proposed method.

4. Simulation

In this section, we conduct a simulation study to investigate the behavior of the proposed Bayesian procedures in finite samples. To keep the discussion concise, we only study the most interesting aspect, namely, the coverage and size of point-wise credible intervals and compare with the corresponding confidence interval based on the sieve-MLE. We consider the density of Beta(1,3) as g_0 and make inference on its value at $x_0 = 0.4$. We take J equal to the greatest integer less than or equal to $n^{1/3} \log n$. For each instance of data, we generate 1000 posterior samples of g , isotonize each g to obtain $g^*(x_0)$. We also compute the sieve-MLE $\hat{g}_n(x_0)$ for every instance of data, using the same value of J .

To obtain a two-sided $100(1 - \alpha)\%$ projection-posterior credible interval for $g(x_0)$, we find the $\alpha/2$ and $(1 - \alpha/2)$ quantiles of the 1000 samples of $g^*(x_0)$. We construct the adjusted projection-posterior credible interval using the $A^{-1}(\alpha/2)$ and $(1 - A^{-1}(1 - \alpha/2))$ quantiles of $g^*(x_0)$ where A^{-1} is found from Table 1 of Chakraborty and Ghosal [16]. The adjusted credible interval thus formed has projection-posterior credibility $A^{-1}(1 - \alpha)$, and its coverage in large samples is expected to be $(1 - \alpha)$. We compute the confidence interval using $\hat{g}_n(x_0)$ based on the quantiles of the Chernoff's distribution.

Table 1. Comparison of obtained coverage and average length of unadjusted and adjusted Bayesian credible intervals and a confidence interval based on the sieve-MLE. In the table, $C_B(\alpha)$, $C_B^*(\alpha)$ and $C_F(\alpha)$ respectively denote the coverages of the $(1 - \alpha)$ -level projection-posterior credible interval, recalibrated projection-posterior credible interval with target coverage $(1 - \alpha)$, and the $(1 - \alpha)$ -level confidence interval based on $\hat{g}_n(x_0)$. The average lengths of the intervals are respectively denoted by $L_B(\alpha)$, $L_B^*(\alpha)$ and $L_F(\alpha)$

$1 - \alpha$	$n = 100$						$n = 500$					
	$C_B(\alpha)$	$L_B(\alpha)$	$C_B^*(\alpha)$	$L_B^*(\alpha)$	$C_F(\alpha)$	$L_F(\alpha)$	$C_B(\alpha)$	$L_B(\alpha)$	$C_B^*(\alpha)$	$L_B^*(\alpha)$	$C_F(\alpha)$	$L_F(\alpha)$
0.99	0.994	0.97	0.986	0.91	0.990	1.33	0.994	0.66	0.990	0.61	0.988	0.79
0.98	0.982	0.89	0.978	0.84	0.984	1.21	0.984	0.61	0.976	0.57	0.982	0.72
0.96	0.974	0.80	0.964	0.75	0.962	1.08	0.976	0.54	0.974	0.50	0.970	0.64
0.94	0.964	0.74	0.952	0.70	0.952	0.99	0.972	0.50	0.954	0.47	0.948	0.59
0.92	0.948	0.69	0.930	0.65	0.930	0.93	0.952	0.47	0.928	0.44	0.918	0.55
0.90	0.930	0.65	0.912	0.62	0.908	0.87	0.928	0.44	0.914	0.41	0.906	0.52
0.88	0.916	0.62	0.896	0.59	0.882	0.83	0.918	0.42	0.902	0.40	0.884	0.49
0.86	0.896	0.59	0.878	0.56	0.866	0.79	0.902	0.40	0.892	0.38	0.862	0.47
0.84	0.878	0.56	0.862	0.54	0.854	0.75	0.892	0.38	0.884	0.36	0.854	0.45
0.82	0.862	0.54	0.840	0.51	0.830	0.72	0.884	0.36	0.866	0.35	0.826	0.43
0.80	0.840	0.52	0.830	0.49	0.812	0.69	0.866	0.35	0.850	0.33	0.812	0.41

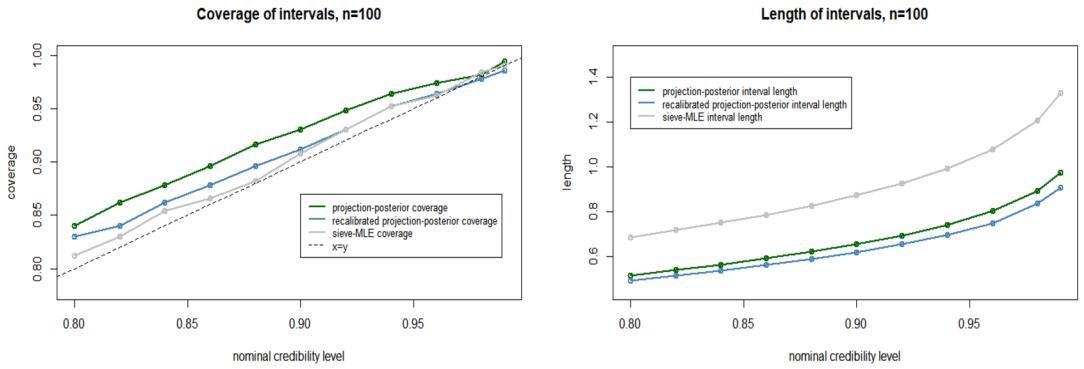


Figure 1. Obtained coverages and lengths of intervals across different nominal credibility levels ($1 - \alpha$), for $n = 100$, averaged over 500 replications.

The constant C_0 that appears in Theorem 3.4 involves $g(x_0)$ and $g'(x_0)$, which are respectively estimated using the density function in R and successive differentiation on a fine grid.

We report the coverages and lengths of the three intervals averaged over 500 replications of data for two values of n in Table 1. For a better understanding, we plot the coverages and lengths in Figures 1–2. We observe that the projection-posterior credible intervals are mildly conservative in that the coverage values are slightly higher than the corresponding credibility, as predicted by Theorem 3.5. The recalibrated intervals are thus slightly shorter, and their coverage values are seen to be closer to the target ($1 - \alpha$). The confidence interval using the asymptotic distribution of the sieve-MLE with the value of the normalizing constant estimated from the data is seen to have adequate coverage, although it is noticeably longer. Thus the Bayesian intervals give more accurate and precise quantification of the uncertainty in the value of the function at the point of interest.

We also explore how the coverages and lengths of the intervals vary with increasing n , and beyond what n the asymptotic regime starts being observed. We display the coverages and lengths averaged over 500 replications for several cases of n , for the target coverage ($1 - \alpha$) equal to 0.95 and 0.98 in Figures 3–4. The chosen sample sizes are $n = 50, 80, 100, 200, 300, 500, 600, 800, 1000, 1200, 1500, 1800, 2000$. As expected, with increasing n , the lengths of the intervals are seen to go down. The

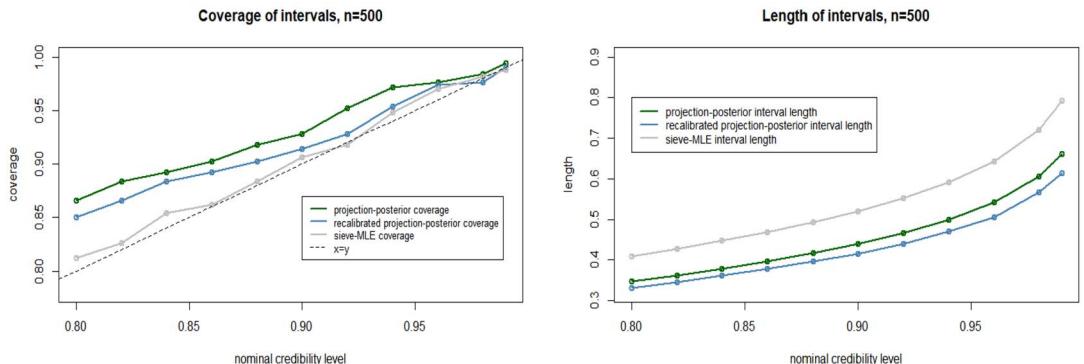


Figure 2. Obtained coverages and lengths of intervals across different nominal credibility levels ($1 - \alpha$), for $n = 500$, averaged over 500 replications.

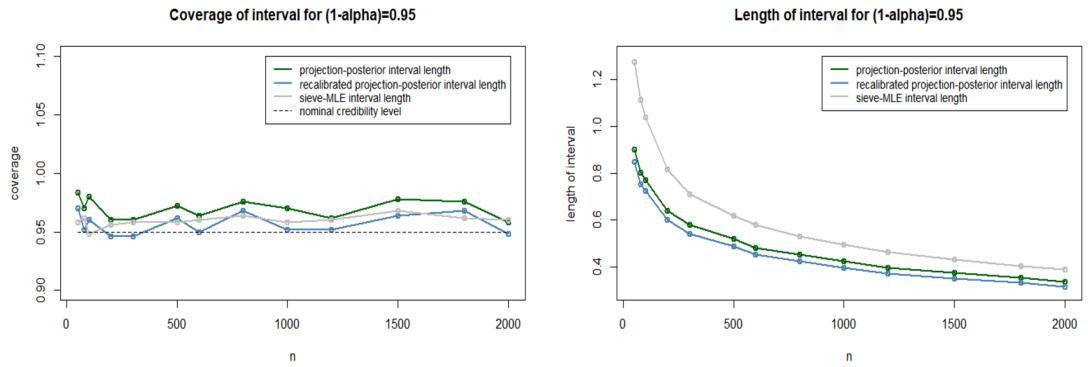


Figure 3. Coverages and lengths of intervals across different sample sizes for $(1 - \alpha) = 0.95$, averaged over 500 replications.

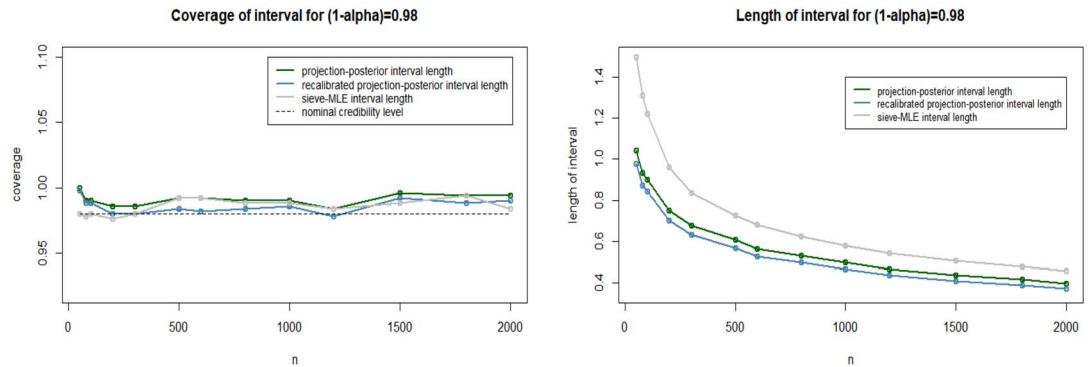


Figure 4. Coverages and lengths of intervals across different sample sizes for $(1 - \alpha) = 0.98$, averaged over 500 replications.

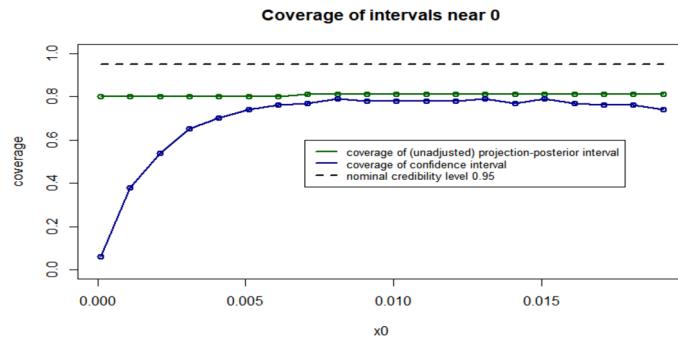


Figure 5. Coverage of (unadjusted) projection-posterior credible interval and that of a confidence interval based on the Grenander's estimator near 0. Samples of size $n = 500$ are generated from an exponential distribution and 100 replications of the data are used.

coverages seem to behave like what we would expect in the asymptotic regime, even for smaller sample sizes like 50 and 80.

It would be interesting to study the coverage of projection-posterior credible intervals when x_0 is close to 0 and g_0 is unbounded. We compare the coverage of our unadjusted Bayesian credible interval (with prior supported on the unbounded domain) to that of a frequentist confidence interval based on the Grenander's estimator and quantiles of the Chernoff distribution, for data generated from the exponential distribution with a rate 1. We use 100 replications of the data, with $n = 500$. The coverages for intervals with level 0.95 are displayed in Figure 5. We observe that although both the methods yield intervals with coverage less than the nominal level 0.95, the Bayesian credible interval has much more coverage than the confidence interval, when x_0 is less than 0.005.

5. Proofs

Proof of Theorem 3.1. If g^* is the monotone projection of a density g and the true density g_0 is monotone, then by the definition of the projection and the triangle inequality, $d(g^*, g_0) \leq d(g^*, g) + d(g, g_0) \leq 2d(g, g_0)$. Thus the contraction rate of the unrestricted posterior is inherited by the projection-posterior, and hence it suffices to obtain the contraction rate of the unrestricted posterior.

For a bounded, monotone density g_0 , with $g_0(0)$ defined to be $g_0(0+)$ without loss of generality, we have that $\int |g_0(x) - g_{0J}(x)|dx$ is

$$\sum_{j=1}^J \int_{I_j} |g_0(x) - J\theta_{0j}|dx \leq \sum_{j=1}^J [g_0((j-1)/J) - g_0(j/J)]|I_j| \leq [g_0(0) - g_0(1)]/J, \quad (5.1)$$

since $J\theta_{0j}$ lies between $g_0(j/J)$ and $g_0((j-1)/J)$ and the length $|I_j|$ of I_j is $1/J$. Thus it suffices to prove that $\int |g(x) - g_{0J}(x)|dx \lesssim \sqrt{J/n}$ with posterior probability exceeding $1 - \delta$ in P_0 -probability for any predetermined $\delta > 0$. Since the \mathbb{L}_2 -distance dominates the \mathbb{L}_1 -distance, in view of Markov's inequality and the standard bias variance decomposition, it suffices to establish that $\sum_{j=1}^J E_0(\hat{\theta}_j - \theta_{0j})^2 \lesssim 1/n$ and $\sum_{j=1}^J E_0 \text{Var}(\theta_j | D_n) \lesssim 1/n$, where $\hat{\theta}_j = (\alpha_j + N_j)/(\alpha_+ + n)$ is the posterior mean of θ_j . The latter follows from $\text{Var}(\theta_j | D_n) = (\alpha_j + N_j)(\alpha_- - \alpha_j + n - N_j)/((\alpha_+ + n)^2(\alpha_+ + n + 1)) \leq (\alpha_j + N_j)/(\alpha_+ + n)^2$. The sum of these is bounded by $1/(\alpha_+ + n) \leq 1/n$. To take care of the squared bias term, we further decompose it into $\text{Var}(\hat{\theta}_j)$ and $(E_0(\hat{\theta}_j) - \theta_{0j})^2$. The former term is $n\theta_{0j}(1 - \theta_{0j})/(\alpha_+ + n)^2$, whose sum over j is bounded by $n/(\alpha_+ + n)^2 \leq 1/n$ since $\sum_{j=1}^J \theta_{0j} = 1$. The latter term is equal to $\sum_{j=1}^J (\alpha_j - \alpha_- \theta_{0j})^2/(\alpha_+ + n)^2 \lesssim J/n^2 \leq 1/n$ because θ_{0j} are bounded by a multiple of J^{-1} and $\alpha_- \lesssim J$ by the assumption that α_j are bounded. This proves the first part.

For the second part, we apply the general theory of posterior contraction with i.i.d. observations with respect to the \mathbb{L}_1 -distance on densities (Ghosal et al. [23] or Ghosal and van der Vaart [25]). If ϵ_n is the targeted posterior contraction rate, we first need to estimate the prior concentration rate and establish that for some $C > 0$,

$$-\log \Pi(g : K(g_0, g) \leq C\epsilon_n^2, V(g_0, g) \leq C\epsilon_n^2) \lesssim n\epsilon_n^2, \quad (5.2)$$

where $K(g_0, g) = \int g_0 \log(g_0/g) dx$ and $V(g_0, g) = \int g_0 \log^2(g_0/g) dx$ stand for the Kullback–Leibler divergences. To this end, we observe from (5.1) that the \mathbb{L}_1 -approximation rate at monotone functions by a step function with J equal intervals is J^{-1} . Hence with $\bar{J}_n \asymp \epsilon_n^{-2}$, we have that $\|g_0 - g_{0\bar{J}_n}\|_1 \leq \epsilon_n^2$. Therefore, their Hellinger distance $d_H(g_0, g_{0\bar{J}_n})$ is bounded by $\|g_0 - g_{0\bar{J}_n}\|_1^{1/2} \leq \epsilon_n$ (cf., Lemma B.1(ii) of Ghosal and van der Vaart [25]). Because g_0 and hence also $g_{0\bar{J}_n}$ are uniformly bounded above and

below, it follows from Lemma B.2 of Ghosal and van der Vaart [25] that $K(g_0, g)$ and $V(g_0, g)$ are also bounded by a constant multiple of ϵ_n^2 . We note that for $J = \bar{J}_n$,

$$K(g_0, g) = K(g_0, g_{0\bar{J}_n}) + \int g_0 \log(g_{0\bar{J}_n}/g) = K(g_0, g_{0\bar{J}_n}) + \sum_{j=1}^{\bar{J}_n} \theta_{0j} \log(\theta_{0j}/\theta_j)$$

and similarly

$$V(g_0, g) \leq 2V(g_0, g_{0\bar{J}_n}) + 2 \int g_0 \log^2(g_{0\bar{J}_n}/g) = 2V(g_0, g_{0\bar{J}_n}) + 2 \sum_{j=1}^{\bar{J}_n} \theta_{0j} \log^2(\theta_{0j}/\theta_j).$$

Therefore, for (5.2), it suffices to lower bound

$$\pi(\bar{J}_n) \Pi\left(\sum_{j=1}^{\bar{J}_n} \theta_{0j} \log(\theta_{0j}/\theta_j) \leq \epsilon_n^2, \sum_{j=1}^{\bar{J}_n} \theta_{0j} \log^2(\theta_{0j}/\theta_j) \leq \epsilon_n^2 | J = \bar{J}_n\right). \quad (5.3)$$

By the assumption that g_0 is upper and lower bounded, we have that $\theta_{0j} \asymp 1/\bar{J}_n$ uniformly in j . Hence if $\sum_{j=1}^{\bar{J}_n} |\theta_j - \theta_{0j}| \leq \epsilon_n^2/\bar{J}_n \ll 1/\bar{J}_n$, it follows that $\theta_j \asymp 1/\bar{J}_n$ uniformly in j . Thus by Lemmas B.1 and B.2 of Ghosal and van der Vaart [25], it follows that

$$\max\left\{\sum_{j=1}^{\bar{J}_n} \theta_{0j} \log(\theta_{0j}/\theta_j), \sum_{j=1}^{\bar{J}_n} \theta_{0j} \log^2(\theta_{0j}/\theta_j)\right\} \lesssim \sum_{j=1}^{\bar{J}_n} |\sqrt{\theta_j} - \sqrt{\theta_{0j}}|^2 \leq \sum_{j=1}^{\bar{J}_n} |\theta_j - \theta_{0j}|.$$

Therefore, the expression in (5.3) is bounded below by

$$\pi(\bar{J}_n) \Pi\left(\sum_{j=1}^{\bar{J}_n} |\theta_j - \theta_{0j}| \leq \epsilon_n^2/\bar{J}_n | J = \bar{J}_n\right) \gtrsim \exp[-a_1 \bar{J}_n \log \bar{J}_n - c_0 \bar{J}_n \log(\bar{J}_n/\epsilon_n^2)]$$

by the estimate given by Lemma G.13 of Ghosal and van der Vaart [25]. Hence, if we choose $\bar{J}_n = \sqrt{n/\log n}$ and $\epsilon_n = \bar{J}_n^{-1/2} = (n/\log n)^{-1/4}$ so that $\log \bar{J}_n \asymp \log(1/\epsilon_n) \asymp \log n$ and $\bar{J}_n \log n \asymp n\epsilon_n^2$, then (5.2) is satisfied.

Choose $\mathcal{S}_n = \cup_{j=1}^{J_n} \mathcal{F}_j$ as the sieve in Theorem 8.9 of Ghosal and van der Vaart [25], where $J_n = L\bar{J}_n$ for a sufficiently large constant $L > 0$. Then

$$\Pi(\mathcal{S}_n^c) \leq \Pi(J > J_n) \leq e^{-a_2 J_n \log J_n} \leq e^{-a n \epsilon_n^2}$$

for any predetermined $a > 0$ if we choose L large enough. Finally, the ϵ_n -covering number $\mathcal{N}(\epsilon_n, \mathcal{S}_n, \|\cdot\|_1)$ of the sieve for the \mathbb{L}_1 -metric is bounded by $\sum_{j=1}^{J_n} (3/\epsilon_n)^j \leq J_n (3/\epsilon_n)^{J_n}$, so clearly the entropy condition $\log \mathcal{N}(\epsilon_n, \mathcal{S}_n, \|\cdot\|_1) \lesssim n\epsilon_n^2$ holds. This shows that the \mathbb{L}_1 -convergence rate is $(n/\log n)^{-1/4}$ and also we have that $\Pi(J > J_n | D_n) \rightarrow_{P_0} 0$. \square

Proof of Theorem 3.2. (a) For $g_0 \in \mathcal{F}_+$, by the definition of the projection,

$$\Pi(\|g - g^*\|_1 > M_n n^{-1/3} | D_n) \leq \Pi(\|g - g_0\| > M_n n^{-1/3} | D_n) \rightarrow_{P_0} 0$$

by Remark 3.1. Further, the convergence is uniform over $g_0 \in \mathcal{F}_+(K)$ for any $K > 0$.

(b) Let $g_0 \notin \bar{\mathcal{F}}_+$ be a fixed density. The martingale convergence theorem gives $\|g_0 - g_{0J}\|_1 \rightarrow 0$ as $J \asymp n^{1/3} \rightarrow \infty$. Also by Remark 3.1, $\Pi(\|g - g_{0J}\|_1 > M_n n^{-1/3} | D_n) \rightarrow_{P_0} 0$. Now by the triangle inequality and the definition of the projection,

$$\begin{aligned} \Pi(d(g, \mathcal{F}_+) \leq M_n n^{-1/3} | D_n) &\leq \Pi(\|g_0 - g\|_1 + M_n n^{-1/3} \geq d(g_0, \mathcal{F}_+) | D_n) \\ &\leq \Pi(\|g - g_{0J}\|_1 \geq d(g_0, \mathcal{F}_+) - \|g_{0J} - g_0\|_1 - M_n n^{-1/3} | D_n), \end{aligned}$$

which goes to 0 in P_0 -probability because $d(g_0, \mathcal{F}_+)$ is fixed and positive.

(c) Let $g_0 \in \mathcal{F}_+^c \cap \mathcal{H}(\beta, L)$ such that $d(g_0, \mathcal{F}_+) \geq \rho_n(\beta)$ and g_{0J} be as in part (b). Then it is well-known (cf. de Boor [19]) that $\|g_0 - g_{0J}\|_1 \leq C(L) J^{-\beta} \asymp n^{-\beta/3}$ for some constant $C(L)$ depending only on L . Therefore, from Remark 3.1, $\Pi(\|g - g_{0J}\|_1 > M'_n n^{-1/3} + C(L) n^{-\beta/3} | D_n) \rightarrow_{P_0} 0$, uniformly for all $g_0 \in \mathcal{H}(\beta, L)$ for any $M'_n \rightarrow \infty$. Using the fact that

$$d(g, \mathcal{F}_+) \geq d(g_0, g^*) - d(g, g_0) \geq d(g_0, \mathcal{F}_+) - d(g, g_0) \geq \rho_n(\beta) - d(g, g_0),$$

the right side will exceed $M_n n^{-1/3}$ with posterior probability tending to 1 in P_0 -probability, provided that $\rho_n(\beta) \geq C n^{-\beta/3}$ for some $C > C(L)$ when $\beta < 1$, while for $\beta = 1$, $\rho_n(\beta) \geq C M_n n^{-1/3}$ for any $C > 1$ ensures that in view of $M_n \rightarrow \infty$. \square

Proof of Theorem 3.3. Let g_0 be a bounded density function that may or may not be decreasing. Let $\theta_{0j} = \int_{I_j} g_0(x) dx$, $j = 1, \dots, J$. We claim that if $\log J_n \asymp \log n$, given any $\epsilon > 0$, for some sufficiently large M_0 ,

$$E_0 \Pi(\|g - g_{0J}\|_1 \geq M_0 \sqrt{J(\log n)/n}, J \leq J_n | D_n) < \epsilon. \quad (5.4)$$

We have $N_j \sim \text{Bin}(n, \theta_{0j})$ and $\theta_{0j} \lesssim 1/J$ by the boundedness of the density, simultaneously for all $j = 1, \dots, J$. If $N \sim \text{Bin}(n, \theta)$, then from Bennett's inequality (cf. Proposition A.6.2 of van der Vaart and Wellner [44]), it easily follows that

$$P(|N/n - \theta| \geq \lambda / \sqrt{n}) \leq 2 \exp[-\lambda^2 / (2\theta)]$$

for any $\lambda > 0$. Hence, upon choosing $\lambda^2 = 6\theta_{0j}$ and using $\theta_{0j} \lesssim 1/J$, we can find $C > 0$ such that for any $j = 1, \dots, J$, $J \leq J_n$, we have that

$$P(|N_j/n - \theta_{0j}| \geq C \sqrt{(\log n)/(nJ)}) \leq 2n^{-3}.$$

Since the cardinality of the set $\{1 \leq j \leq J \leq J_n\} \leq J_n^2 \leq n^2$, it follows that with probability tending to one, simultaneously for all $J \leq J_n$,

$$\max\{|N_j/n - \theta_{0j}| : 1 \leq j \leq J\} \lesssim \sqrt{(\log n)/(nJ)}. \quad (5.5)$$

In particular, this also ensures that

$$\max\{N_j/n : 1 \leq j \leq J\} \lesssim J^{-1} + \sqrt{(\log n)/(nJ)} \lesssim J^{-1} \quad (5.6)$$

provided that $J \log n \lesssim n$. Now, the posterior probability in (5.4) can be written as

$$\sum_{J=1}^{J_n} \Pi(J | D_n) \Pi\left(\sum_{j=1}^J |\theta_j - \theta_{0j}| \geq M_0 \sqrt{J(\log n)/n} | J, D_n\right) \quad (5.7)$$

and that $\Pi(\sum_{j=1}^J |\theta_j - \theta_{0j}| \geq M_0 \sqrt{J(\log n)/n} | J, D_n)$ bounded by

$$\frac{\sqrt{n/J}}{M_0 \sqrt{\log n}} \sum_{j=1}^J \{[\text{Var}(\theta_j | J, D_n)]^{1/2} + |\text{E}(\theta_j | J, D_n) - \theta_{0j}|\}. \quad (5.8)$$

Since the prior parameters are bounded, $\text{Var}(\theta_j | J, D_n) \lesssim N_j/n^2 \lesssim 1/(Jn)$ and

$$|\text{E}(\theta_j | J, D_n) - \theta_{0j}| \leq |N_j/n - \theta_{0j}| + 1/n \lesssim \sqrt{(\log n)/nJ},$$

with P_0 -probability tending to one. Therefore the expression in (5.8) is bounded in P_0 -probability by a constant multiple of M_0^{-1} in view of (5.5). This leads to (5.4).

Note that if $\text{E}_0 \Pi(g : \|g - g_0\|_1 > M_0 \epsilon_n | D_n) \rightarrow 0$ for some $M_0 > 0$, then because g_{0J} is the projection of g_0 on \mathcal{F}_J , we have that

$$\Pi(J : \|g_{0J} - g_0\|_1 > M_0 \epsilon_n | D_n) \leq \Pi(\|g - g_0\|_1 > M_0 \epsilon_n | D_n) \rightarrow 0 \quad (5.9)$$

in P_0 -probability.

(a) If $g_0 \in \mathcal{F}_+$, then by the second part of Theorem 3.1 and Remark 3.1, the \mathbb{L}_1 -contraction rate is $\epsilon_n = (n/\log n)^{-1/4}$ and that $\Pi(J > J_n | D_n) \rightarrow_{P_0} 0$ for $J_n \asymp \epsilon_n^{-2} \asymp \sqrt{n/\log n}$. Using the fact that $g_{0J} \in \mathcal{F}_+$ (and hence g^* is closer to g than g_{0J} for any $g \in \mathcal{F}_J$), we have that for any given $\epsilon > 0$, there exists $M_0 > 0$ such that

$$\text{E}_0 \Pi(\|g - g^*\|_1 > M_0 \sqrt{J(\log n)/n} | D_n) \leq \Pi(\|g - g_{0J}\|_1 > M_0 \sqrt{J(\log n)/n} | D_n) < \epsilon$$

by (5.4); here we have used $\Pi(J > J_n | D_n) \rightarrow_{P_0} 0$ and $\log J_n \lesssim \log n$. Arguing as in the proof of part (a) of Theorem 3.2, the conclusion follows.

(b) Let $g_0 \notin \bar{\mathcal{F}}_+$ be a fixed density bounded away from zero and infinity. By the martingale convergence theorem, $\|g_{0J} - g_0\|_1 \rightarrow 0$ as $J \rightarrow \infty$, so for a given $\epsilon > 0$, we can get J_0 (depending on ϵ but not depending on n) such that $\|g_{0J_0} - g_0\|_1 < \epsilon/2$. Then we have $\Pi(\|g - g_0\|_1 < \epsilon) \geq \Pi(J = J_0) \Pi(J_0 \sum_{j=1}^{J_0} |\theta_j - \theta_{0j}| < \epsilon/2) > 0$. Further, for $J_n \sim cn/\log n$ with a sufficiently small $c > 0$, by the tail-estimate of the prior distribution (3.1), there exists a constant $b > 0$ depending on c such that $\Pi(J > J_n) \leq e^{-bn}$. Considering a sieve $\mathcal{S}_n = \{g = J \sum_{j=1}^J \theta_j \mathbb{1}_{I_j}, J \leq J_n\}$, standard estimates give a bound for its metric entropy a multiple of $J_n \log n \sim cn$, and that $\Pi(\mathcal{S}_n^c) \leq e^{-bn}$. Therefore it follows (see Theorem 6.17 of Ghosal and van der Vaart [25]) that $\text{E}_0 \Pi(J > J_n | D_n) \rightarrow 0$ and the posterior is consistent at g_0 with respect to the \mathbb{L}_1 -metric. Hence it suffices to restrict J to at most J_n .

Observe that for any $g \in \mathcal{F}_J$, the distance to the monotone projection

$$d(g, \mathcal{F}_+) \geq \|g_0 - g_0^*\|_1 - \|g - g_{0J}\|_1 - \|g_{0J} - g_0\|_1. \quad (5.10)$$

Since g_0 does not belong to the topological closure of \mathcal{F}_+ , the first term is fixed and positive. By (5.4), with posterior probability tending to 1 in P_0 -probability, the second term is bounded by $\sqrt{J_n(\log n)/n} \leq c$, which can be taken to as small as we need. By (5.9) and posterior consistency, the third term can also be made arbitrarily small with high posterior probability. Thus $\Pi(d(g, \mathcal{F}_+) \geq M_0 \sqrt{J(\log n)/n} | D_n) \rightarrow_{P_0} 1$, that is, the power tends to 1.

(c) Let $g_0 \in \mathcal{H}(\beta, L)$ be at least $C(n/\log n)^{-\beta/(2\beta+1)}$ away from \mathcal{F}_+ for some sufficiently large $C > 0$ in terms of the \mathbb{L}_1 -metric. Observe that by the Lipschitz continuity of g_0 , with $\theta_{0j} = \int_{I_j} g_0$,

$$\|g_0 - g_{0J}\|_1 = \sum_{j=1}^J \int_{I_j} |g_0(x) - J\theta_{0j}| dx \leq \sup_{x, y \in I_j} |g_0(x) - g_0(y)| \leq LJ^{-\beta}. \quad (5.11)$$

Then standard arguments as in the second part of the proof of Theorem 3.1 with $J^{-\beta}$ replacing $J^{-1/2}$ give bounds for prior concentration and metric entropy, leading to the posterior contraction rate $\epsilon_n = (n/\log n)^{-\beta/(2\beta+1)}$ at g_0 with respect to the \mathbb{L}_1 -distance. Further, it follows that $\Pi(J \leq M'(n/\log n)^{1/(2\beta+1)} | D_n) \rightarrow_{P_0} 1$ for some sufficiently large M' . To complete the proof, we proceed as in part (b) with the following changes. The second term in (5.10) is bounded by a multiple of $(n/\log n)^{-\beta/(2\beta+1)}$ with high posterior probability. By (5.9), the third term is bounded by a multiple of $(n/\log n)^{-\beta/(2\beta+1)}$ with high posterior probability. Therefore, for $C > 0$ large enough, $\|g_0 - g_0^*\|_1$ is larger than any predetermined constant multiple of $(n/\log n)^{-\beta/(2\beta+1)}$, which exceeds $M_0\sqrt{J(\log n)/n}$ with high posterior probability. Thus the power of ϕ_n at g_0 tends to one. \square

In the remaining proofs below, we use the following ‘switch relation’ (Page 56 of Groeneboom and Jongbloed [29]): for a lower semi-continuous function Φ on an interval I with Φ^* its least concave majorant, and Φ^{*l} denoting the left-derivative of Φ^* , for every $t \in I$, $v \in \mathbb{R}$,

$$\{\Phi^{*l}(t) > v\} = \{\arg \min\{\Phi(s) - vs : s \in I\} < t\}, \quad (5.12)$$

where ‘ $\arg \min$ ’ selects the maximum of the minimizers when multiple minimizers exist.

To prove Theorems 3.4 and 3.5, we need to establish a few auxiliary lemmas. The first two lemmas are about the asymptotics for the sieve MLE, showing that the local empirical process, whose maximization leads to the normalized sieve MLE, converges to an appropriate Gaussian process, and its maximizer is tight, respectively. The remaining two results are Bayesian analogs of these two results, replacing the empirical process by the posterior process. The proofs of these lemmas are given in the appendix. Below, W_1 and W_2 will stand for independent two-sided Brownian motions with $W_1(0) = W_2(0) = 0$. We shall use the convention that for $t < 0$, a sum of the form $\sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3}t) J \rceil}$ stands for the corresponding sum $\sum_{j=\lceil (x_0 + n^{-1/3}t) J \rceil + 1}^{\lceil x_0 J \rceil}$.

Lemma 5.1. *If $n^{1/3} \ll J \ll n^{2/3}$, then*

$$\left(n^{2/3} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3}t) J \rceil} \left(\frac{N_j}{n} - \frac{g_0(x_0)}{J} \right) : t \in [-K, K] \right) \rightsquigarrow (aW_1(t) - bt^2 : t \in [-K, K])$$

in $\mathbb{L}_\infty([-K, K])$, for all $K > 0$.

Lemma 5.2. *If $n^{1/3} \ll J \ll n^{2/3}$, then for all $z \in \mathbb{R}$,*

$$\arg \max_{t \in \mathbb{R}} \left\{ n^{2/3} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3}t) J \rceil} \left(\frac{N_j}{n} - \frac{g_0(x_0)}{J} \right) - n^{1/3} z \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3}t) J \rceil} \frac{1}{J} \right\} = O_{P_0}(1).$$

Lemma 5.3. *If $n^{1/3} \ll J \ll n^{2/3}$, then*

$$\begin{aligned} \mathcal{L} \left(n^{2/3} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3}t) J \rceil} \left(\theta_j - \frac{g_0(x_0)}{J} \right) : t \in [-K, K] \middle| D_n \right) \\ \rightsquigarrow \mathcal{L} (aW_1(t) + aW_2(t) - bt^2 : t \in [-K, K] \middle| W_1) \end{aligned} \quad (5.13)$$

in $\mathbb{L}_\infty([-K, K])$, for all $K > 0$.

Lemma 5.4. If $n^{1/3} \ll J \ll n^{2/3}$, then for all $z \in \mathbb{R}$, there exists $K > 0$ such that

$$\Pi \left(\left| \arg \max_{t \in \mathbb{R}} \left\{ \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t) J \rceil} \left(\theta_j - \frac{g_0(x_0)}{J} \right) - n^{-1/3} z \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t) J \rceil} \frac{1}{J} \right\} \right| > K \middle| D_n \right) \rightarrow_{P_0} 0.$$

Proof of Theorem 3.4. Let $U_n(s) = \sum_{j=1}^{\lceil s J \rceil} (N_j/n)$ and $G_n(s) = \sum_{j=1}^{\lceil s J \rceil} J^{-1} = \lceil s J \rceil / J$. Let $\tau_n := n^{1/3} (J^{-1} \lceil x_0 J \rceil - x_0)$. From the switch relation (5.12) and a change of variable $s = x_0 + n^{-1/3} t$, for $z \in \mathbb{R}$,

$$\begin{aligned} & P(\hat{g}_n(x_0) \leq g_0(x_0) + n^{-1/3} z) \\ &= P\left(\arg \max_{s \in [0, 1]} \{U_n(s) - (g_0(x_0) + n^{-1/3} z) G_n(s)\} \geq \lceil x_0 J \rceil / J\right) \\ &= P\left(\arg \max_{s \in [0, 1]} \{n^{2/3} (U_n(s) - U_n(x_0)) - n^{2/3} (g_0(x_0) + n^{-1/3} z) (G_n(s) - G_n(x_0))\} \geq \lceil x_0 J \rceil / J\right) \\ &= P\left(\arg \max_{t \in \mathbb{R}} \{n^{2/3} (U_n(x_0 + n^{-1/3} t) - U_n(x_0)) - n^{2/3} (g_0(x_0) + n^{-1/3} z) (G_n(x_0 + n^{-1/3} t) - G_n(x_0))\} \geq \tau_n\right) \\ &= P\left(\arg \max_{t \in \mathbb{R}} \left\{ n^{2/3} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t) J \rceil} \frac{N_j}{n} - n^{2/3} (g_0(x_0) + n^{-1/3} z) \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t) J \rceil} \frac{1}{J} \right\} \geq \tau_n \right). \end{aligned}$$

Note that

$$n^{1/3} z J^{-1} (\lceil (x_0 + n^{-1/3} t) J \rceil - \lceil x_0 J \rceil - 1) = zt + O(n^{1/3} J^{-1}) \rightarrow zt \quad (5.14)$$

because $J \gg n^{1/3}$. Since $\tau_n \rightarrow 0$, we evaluate

$$P(\hat{g}_n(x_0) \leq g_0(x_0) + n^{-1/3} z) = P\left(\arg \max_{t \in \mathbb{R}} \left\{ n^{2/3} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t) J \rceil} \left(\frac{N_j}{n} - \frac{g_0(x_0)}{J} \right) - zt \right\} \geq 0\right).$$

From Lemma 5.1 and 5.2, the Argmax Theorem and Lemma A.1, we have that the expression converges to

$$\begin{aligned} P\left(\arg \max_{t \in \mathbb{R}} \{a W_1(t) - bt^2 - zt\} \geq 0\right) &= P\left((a/b)^{2/3} \arg \max_{t \in \mathbb{R}} \{W_1(t) - t^2\} + \frac{z}{2b} \geq 0\right) \\ &= P\left(2b(a/b)^{2/3} \arg \max_{t \in \mathbb{R}} \{W_1(t) - t^2\} \geq -z\right) \\ &= P\left(2b(a/b)^{2/3} \arg \max_{t \in \mathbb{R}} \{W_1(t) - t^2\} \leq z\right). \end{aligned}$$

The last step follows from the fact that $\arg \max\{W_1(t) - t^2 : t \in \mathbb{R}\}$ is symmetric about zero. Substituting $C_0 = 2b(a/b)^{2/3}$ in the last expression of the display we get $P(n^{1/3}(\hat{g}_n(x_0) - g_0(x_0)) \leq z) \rightarrow P(C_0 \arg \max\{W_1(t) - t^2 : t \in \mathbb{R}\} \leq z)$ as $n \rightarrow \infty$. This establishes the part (a).

Since g^* is piece-wise constant on each I_j , $g^*(x_0) = \theta_{\lceil x_0 J \rceil}^*$. We first evaluate the expression $P_0 \times \Pi(n^{1/3}(g^*(x_0) - g_0(x_0)) \leq z | D_n) = P_0 \times \Pi(\theta_{\lceil x_0 J \rceil}^* \leq g_0(x_0) + n^{-1/3} z | D_n)$ for $z \in \mathbb{R}$ as $n \rightarrow \infty$.

Let $c_n(\cdot)$ denote the graph of the lines connecting the points

$$\{(0, 0), (J^{-1}, \theta_1), (2J^{-1}, \sum_{k=1}^2 \theta_k), \dots, (1, \sum_{k=1}^J \theta_k)\},$$

along with $c_n(s) = 0$ for $s \leq 0$ and $c_n(s) = \sum_{k=1}^J \theta_k$ for $s \geq 1$. Observe that c agrees with the cumulative distribution function of the density g at the points j/J , $j = 0, 1, \dots, J$. Define a stochastic process $\tilde{U}_n(s) = \sum_{j=1}^{\lceil sJ \rceil} \theta_j$. Since $\theta_{\lceil x_0 J \rceil}^*$ is the left-derivative of the least concave majorant of $c(\cdot)$ at the point $\lceil x_0 J \rceil/J$, by the switch relation (5.12), the fact that $\tilde{U}_n(s) = G_n(s) = 0$ for $s \leq 0$ and that the location of minimum does not change upon adding a constant term or upon multiplication by a positive constant, we have that $g^*(x_0) \leq g_0(x_0) + n^{-1/3}z$ if and only if $\arg \max\{c_n(s) - (g_0(x_0) + n^{-1/3}z)s : s \geq 0\} \geq \lceil x_0 J \rceil/J$. Hence the probability of this event can be written as

$$\begin{aligned} & P_0 \times \Pi \left(\arg \max_{s \in \mathbb{R}} \{c_n(s) - (g_0(x_0) + n^{-1/3}z)s\} \geq \lceil x_0 J \rceil/J | D_n \right) \\ &= P_0 \times \Pi \left(\arg \max_{s \in \mathbb{R}} \{\tilde{U}_n(s) - (g_0(x_0) + n^{-1/3}z)G_n(s)\} \geq J^{-1}\lceil x_0 J \rceil | D_n \right) \\ &= P_0 \times \Pi \left(\arg \max_{s \in \mathbb{R}} \left\{ n^{2/3}(\tilde{U}_n(s) - \tilde{U}_n(x_0)) \right. \right. \\ &\quad \left. \left. - n^{2/3}(g_0(x_0) + n^{-1/3}z)(G_n(s) - G_n(x_0)) \right\} \geq J^{-1}\lceil x_0 J \rceil | D_n \right). \end{aligned} \quad (5.15)$$

The last expression in (5.15) then equals to

$$\begin{aligned} & P_0 \times \Pi \left(\arg \max_{t \in \mathbb{R}} \{n^{2/3}(\tilde{U}_n(x_0 + n^{-1/3}t) - \tilde{U}_n(x_0)) \right. \\ &\quad \left. - n^{2/3}(g_0(x_0) + n^{-1/3}z)(G_n(x_0 + n^{-1/3}t) - G_n(x_0))\} \geq \tau_n | D_n \right) \\ &= P_0 \times \Pi \left(\arg \max_{t \in \mathbb{R}} \left\{ n^{2/3} \sum_{j=\lceil x_0 J \rceil+1}^{\lceil (x_0 + n^{-1/3}t)J \rceil} (\theta_j - J^{-1}g_0(x_0)) \right. \right. \\ &\quad \left. \left. - n^{1/3}zJ^{-1}(\lceil (x_0 + n^{-1/3}t)J \rceil - \lceil x_0 J \rceil - 1) \right\} \geq \tau_n | D_n \right). \end{aligned} \quad (5.16)$$

Using Lemma 5.3 and 5.4, a multivariate version of the Argmax Theorem (see Theorem 3.6.10 of Banerjee [1]), and Lemma A.1, we rewrite (5.16) to obtain

$$\begin{aligned} & P_0 \times \Pi(n^{1/3}(g^*(x_0) - g_0(x_0)) \leq z | D_n) \\ &\rightarrow P \left(\arg \max_{t \in \mathbb{R}} \{aW_1(t) + aW_2(t) - bt^2 - zt\} \geq 0 \right) \\ &= P \left((a/b)^{2/3} \arg \max_{t \in \mathbb{R}} \{W_1(t) + W_2(t) - t^2\} + \frac{z}{2b} \geq 0 \right) \\ &= P \left(2b(a/b)^{2/3} \arg \max_{t \in \mathbb{R}} \{W_1(t) + W_2(t) - t^2\} \leq z \right), \end{aligned}$$

the last step following from the fact that $\arg \max \{W_1(t) + W_2(t) - t^2 : t \in \mathbb{R}\}$ is symmetric about zero. Substituting $C_0 = 2b(a/b)^{2/3}$ above, we get

$$P_0 \times \Pi(n^{1/3}(g^*(x_0) - g_0(x_0)) \leq z | D_n) \rightarrow P(C_0 \arg \max_{t \in \mathbb{R}} \{W_1(t) + W_2(t) - t^2\} \leq z).$$

This completes the proof of part (b).

(c) For every $z_1, z_2, t_1, t_2 \in \mathbb{R}$, define

$$H_{1n}(t_1, z_1) = \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t_1) J \rceil} \left(\frac{N_j}{n} - \frac{g_0(x_0)}{J} \right) - n^{-1/3} z_1 \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t_1) J \rceil} \frac{1}{J},$$

$$H_{2n}(t_2, z_2) = \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t_2) J \rceil} \left(\theta_j - \frac{g_0(x_0)}{J} \right) - n^{-1/3} z_2 \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t_2) J \rceil} \frac{1}{J}.$$

Then from the proofs of parts (a) and (b), and the Multivariate Argmax Theorem, we have $(\arg \max \{H_{1n}(t_1, z_1) : t_1 \in \mathbb{R}\}, \arg \max \{H_{2n}(t_2, z_2) : t_2 \in \mathbb{R}\})$ converges weakly to

$$\left(\arg \max_{t_1 \in \mathbb{R}} \{a W_1(t_1) - b t_1^2 - z_1 t_1\}, \arg \max_{t_2 \in \mathbb{R}} \{a W_1(t_2) + a W_2(t_2) - b t_2^2 - z_2 t_2\} \right). \quad (5.17)$$

Rewriting $n^{1/3}(g^*(x_0) - \hat{g}_n(x_0))$ as $n^{1/3}(g^*(x_0) - g_0(x_0)) - n^{1/3}(\hat{g}_n(x_0) - g_0(x_0))$, and using (5.17) and Lemma A.1, we get that for all $z \in \mathbb{R}$,

$$P_0 \times \Pi(n^{1/3}(g^*(x_0) - \hat{g}_n(x_0)) \leq z | D_n) \rightarrow P(C_0 \Delta_{W_1+W_2}^* - C_0 \Delta_{W_1}^* \leq z).$$

From the proof of part (a) of Theorem 3.5, we also obtain that

$$\Pi(n^{1/3}(g^*(x_0) - \hat{g}_n(x_0)) \leq z | D_n) \rightsquigarrow P(C_0 \Delta_{W_1+W_2}^* - C_0 \Delta_{W_1}^* \leq z | W_1).$$

If $\Pi(n^{1/3}(g^*(x_0) - \hat{g}_n(x_0)) \leq z | D_n)$ converges in probability, then the limit must be the random variable $Q = P(C_0 \Delta_{W_1+W_2}^* - C_0 \Delta_{W_1}^* \leq z | W_1)$. Let $W_n^* = n^{1/3}(g^*(x_0) - \hat{g}_n(x_0))$ and $Q_n = \Pi(n^{1/3}(g^*(x_0) - \hat{g}_n(x_0)) \leq z | D_n)$. If Q_n were to converge in probability, then $\Delta_{W_1+W_2}^* - \Delta_{W_1}^*$ and Q would be independent in view of Lemma 3.1 of Sen et al. [41]. But as both of them depend on W_1 , the convergence in probability cannot happen. \square

Proof of Theorem 3.5. As $n^{1/3}(g^*(x_0) - g_0(x_0))$ is the argmax of the process on the left side of (5.13) in Lemma 5.3 and the argmax itself is conditionally tight by Lemma 5.4, the Argmax theorem applied to the conditional distribution concludes that

$$\begin{aligned} & \Pi(n^{1/3}(g^*(x_0) - g_0(x_0)) \leq z | D_n) \\ & \rightsquigarrow P\left(\arg \max_{t \in \mathbb{R}} \{a W_1(t) + a W_2(t) - b t^2 - z t\} \geq 0 | W_1\right) \\ & \stackrel{d}{=} P\left(C_0 \arg \max_{t \in \mathbb{R}} \{W_1(t) + W_2(t) - t^2\} \leq z | W_1\right). \end{aligned}$$

The last step uses Lemma A.1 and the fact that $t \mapsto -t$ leaves the independent processes W_1 and W_2 distributionally invariant. This completes the proof of part (a).

Now to prove Part (b). By the definition of posterior quantile $Q_{n,\gamma}$, we have that $g_0(x_0) \leq Q_{n,\gamma}$ if and only if $\Pi(g^*(x_0) \leq g_0(x_0) | D_n) \leq 1 - \gamma$. Therefore,

$$\begin{aligned} \mathbb{P}_0(g_0(x_0) \leq Q_{n,\gamma}) &= \mathbb{P}_0\left(\Pi(n^{1/3}(g^*(x_0) - g_0(x_0)) \leq 0 | D_n) \leq 1 - \gamma\right) \\ &\rightarrow \mathbb{P}(F^*(0 | W_1) \leq 1 - \gamma), \end{aligned} \quad (5.18)$$

from Part (a) with $z = 0$. Observing that $F^*(0 | W_1) = \mathbb{P}(C_0 \Delta_{W_1 + W_2}^* \leq 0 | W_1) = \mathbb{P}(\Delta_{W_1 + W_2}^* \leq 0 | W_1) = Z_B$, the right hand side of (5.18) reduces to $\mathbb{P}(Z_B \leq 1 - \gamma)$. \square

6. Extension to unbounded domain

In this section, we seek analogous results on contraction rates, testing, and coverage of credible intervals for a bounded, decreasing density g on the half-line $(0, \infty)$. We modify the random histogram prior on an increasing, but still a finite, subinterval of the positive half-line, whose length is controlled by another parameter. More specifically, let $I_{jk} = (k - 1 + (j - 1)/J, k - 1 + j/J]$, $j = 1, \dots, J$, $k = 1, \dots, K$. Note that $\bigcup_{j=1}^J I_{jk} = (k - 1, k]$, that is, like the unit interval in the previous sections, the interval $(k - 1, k]$ is also split into J equal pieces having length $|I_{jk}| = 1/J$ for all $j = 1, \dots, J$, $k = 1, \dots, K$. Then analogously, we put a prior Π_{JK} for g by

$$g = J \sum_{j=1}^J \sum_{k=1}^K \theta_{jk} \mathbb{1}_{I_{jk}}, \quad (6.1)$$

$$(\theta_{jk} : 1 \leq j \leq J, 1 \leq k \leq K) \sim \text{Dir}(\alpha_{jk} : 1 \leq j \leq J, 1 \leq k \leq K),$$

where α_{jk} , $j = 1, \dots, J$, $k = 1, \dots, K$, are bounded by a constant $b > 0$. The parameters J and K can be appropriately chosen depending on n , or may be given prior distributions. Assuming an exponential-type tail condition, which allows choosing K relatively low, the following result obtains a contraction rate using deterministic J and K .

Since under this prior construction with a given K , some observations call be larger than K , creating a conflict with the proposed model, we need to clarify what is meant by the posterior distribution in this case. The natural consideration is to ignore such observations, as they contain no information about θ_{jk} , $j = 1, \dots, J$, $k = 1, \dots, K$. Then the number of admissible observations $X_i \leq K$ is random, but as $K \rightarrow \infty$, the proportion of such observation $G_0(K) \rightarrow 1$, and hence asymptotically almost all observations are admissible.

Theorem 6.1. *Let g_0 be a bounded, decreasing density on $(0, \infty)$. Then using the prior Π_{JK} defined by (6.1) with $J, K \rightarrow \infty$ with n , the \mathbb{L}_1 -posterior contraction rate at g_0 is $\max\{\sqrt{JK/n}, J^{-1}, 1 - G_0(K)\}$. In particular, if $1 - G_0(x) \lesssim e^{-ax^r}$ for some $a, r > 0$, upon choosing $K \sim (\log n/(3a))^{1/r}$ and $J \asymp n^{1/3}(\log n)^{-1/(3r)}$, the contraction rate reduces to the nearly optimal rate $n^{-1/3}(\log n)^{1/(3r)}$.*

Proof. Define $g_0(0) = g_0(0+)$ and

$$g_{0,JK} = \sum_{j=1}^J \sum_{k=1}^K \frac{\int_{I_{jk}} g_0(u) du}{|I_{jk}|} \mathbb{1}_{\{x \in I_{jk}\}} = \sum_{j=1}^J \sum_{k=1}^K J \theta_{0,jk} \mathbb{1}_{\{x \in I_{jk}\}},$$

where $\theta_{0,jk} = \int_{I_{jk}} g_0(x) dx$. Then $g_{0,JK}$ is a decreasing sub-probability density on $(0, \infty)$. We now show that $g_{0,JK}$ approximates g_0 within $J^{-1} + 1 - G_0(K)$ in \mathbb{L}_1 . This follows since for every j, k and $x \in I_{jk}$, $g_0(k - 1 + j/J) \leq g_{0,JK}(x) \leq g_0(k - 1 + (j - 1)/J)$, leading to

$$\begin{aligned} & \int |g_0(x) - g_{0,JK}(x)| dx \\ &= \sum_{j=1}^J \sum_{k=1}^K \int_{I_{jk}} |g_0(x) - g_{0,JK}(x)| dx + \int_K^\infty g_0(x) dx \\ &\leq \sum_{j=1}^J \sum_{k=1}^K \{g_0(k - 1 + (j - 1)/J) - g_{0,JK}(k - 1 + j/J)\} \frac{1}{J} + 1 - G_0(K) \\ &\leq \frac{1}{J} (g_0(0) - g_0(K)) + 1 - G_0(K) \end{aligned}$$

using the decreasing property of g_0 and the telescoping nature of the sum. To finish the proof, we show that $E_0 \Pi(\|g - g_{0,JK}\|_1 > M_n \sqrt{JK/n} + J^{-1} | D_n) \rightarrow 0$ for any $M_n \rightarrow \infty$.

Let $N_{jk} = \sum_{i=1}^n \mathbb{1}\{X_i \in I_{jk}\}$, $j = 1, \dots, J$, $k = 1, \dots, K$, and $N_+ = \sum_{j=1}^J \sum_{k=1}^K N_{jk}$. Observe that $1 - N_+/n = O_P(1 - G_0(K))$.

Now $\|g - g_{0,JK}\|_1 = \sum_{j=1}^J \sum_{k=1}^K |\theta_{jk} - \theta_{0,jk}|$, so it is enough to show that

$$E_0 \Pi \left(\sum_{j=1}^J \sum_{k=1}^K |\theta_{jk} - \theta_{0,jk}| > M_n \sqrt{JK/n} | D_n \right) \rightarrow 0. \quad (6.2)$$

This can be established exactly as in proof of the first part of Theorem 3.1, after observing that the effective sample size, that is, the number of observations falling in $(0, K]$ is $N_+ \sim n$ in P_0 -probability, and the observations falling outside do not alter the posterior distribution. The extra factor K appears in the bound because now there are JK intervals, instead of just J previously.

If we assume the bound $g_0(x) \lesssim e^{-ax^r}$, we have the tail estimate $\int_K^\infty g_0(x) dx \lesssim \int_K^\infty e^{-ax^r} dx = r^{-1} \int_{K^r}^\infty e^{-ay} y^{-1+1/r} dy \lesssim e^{-aK^r}$. Choosing $J \asymp n^{1/3}(\log n)^{-1/(3r)}$ and $K \sim (\log n/(3a))^{1/r}$, the rate $n^{-1/3}(\log n)^{1/(3r)}$ is immediately obtained. \square

Based on the above convergence result, a Bayesian test as in Theorem 3.2 with $J \asymp n^{1/3}(\log n)^{-1/(3r)}$ and $K \sim (\log n/(3a))^{1/r}$ under the tail condition $g_0(x) \lesssim e^{-ax^r}$ is immediate, with an enlargement of the null hypothesis by the amount $M_n n^{-1/3}(\log n)^{1/(3r)}$ in \mathbb{L}_1 for any $M_n \rightarrow \infty$. It is not clear whether an adaptive test analogous to that in Theorem 3.3 has the stated asymptotic properties, since the proof with random J and K depends on the conclusion in the second part of Theorem 3.1. This needs a lower bound for the true density to control the Kullback–Leibler divergence, which is not possible on an unbounded domain.

Finally, the coverage result for the credible interval for $g(x_0)$ holds, because this is a local result not affected by the tail part. More precisely, in the proof of the result, we only need to deal with locations within a constant multiple of $n^{-1/3}$ of the point of interest x_0 to study the convergence of the processes. It is easy to see that this involves only N_{jk} with k bounded in n , and hence these are uniformly of the order n/J . The rest of the arguments apply.

Appendix

The following result is an argmax analog of Lemma A.3 of Chakraborty and Ghosal [16]. Its proof is analogous and is omitted.

Lemma A.1. *Let W_1, W_2 be independent two-sided standard Brownian motions starting at zero. Then for $a, b > 0$ and $c \in \mathbb{R}$,*

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

$$(b) \arg \max \{aW_1(t) - bt^2 + ct : t \in \mathbb{R}\} \stackrel{d}{=} (a/b)^{2/3} \arg \max \{W_1(t) - t^2 : g \in \mathbb{R}\} - \frac{c}{2b}.$$

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

$$\begin{aligned} & (\arg \max \{aW_1(t_1) - bt_1^2 + c_1 t_1 : t_1 \in \mathbb{R}\}, \\ & \quad \arg \max \{aW_1(t_2) + aW_2(t_2) - bt_2^2 + c_2 t_2 : t_2 \in \mathbb{R}\}) \\ & \stackrel{d}{=} ((a/b)^{2/3} \arg \max \{W_1(t_1) - t_1^2 : t_1 \in \mathbb{R}\} - \frac{c_1}{2b}, \\ & \quad (a/b)^{2/3} \arg \max \{W_1(h_2) + W_2(h_2) - h_2^2 : h_2 \in \mathbb{R}\} - \frac{c_2}{2b}). \end{aligned}$$

Proof of Lemma 5.1. We consider the part $t \geq 0$ in the proof; the proof for the part $t \leq 0$ follows similarly. Let $f_{n,t}(x) = n^{1/6}(\mathbb{1}\{x \leq \lceil(x_0 + n^{-1/3}t)J\rceil/J\} - \mathbb{1}\{x \leq \lceil x_0 J \rceil/J\})$. Note that

$$\begin{aligned} & n^{2/3} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil(x_0 + n^{-1/3}t)J\rceil} \left(\frac{N_j}{n} - \frac{g_0(x_0)}{J} \right) \\ & = \mathbb{G}_n(f_{n,t}) + \sqrt{n} \mathbb{E}_0 f_{n,t} - n^{2/3} \frac{g_0(x_0)}{J} (\lceil(x_0 + n^{-1/3}t)J\rceil - \lceil x_0 J \rceil), \end{aligned}$$

where $\{\mathbb{G}_n(f_{n,t}) : t \in [0, K]\}$ is the empirical process corresponding to the class $\mathcal{F}_n := \{f_{n,t} : t \in [0, K]\}$.

We first show that

$$\sqrt{n} \mathbb{E}_0 f_{n,t} - n^{2/3} g_0(x_0) (\lceil(x_0 + n^{-1/3}t)J\rceil - \lceil x_0 J \rceil) / J \rightarrow -bt^2. \quad (\text{A.1})$$

Using Taylor's expansion $g_0(u) - g_0(x_0) = (g'(x_0) + o(1))(u - x)$ of g around x_0 , we write the expression above as

$$n^{2/3} \int_{\lceil x_0 J \rceil / J}^{\lceil(x_0 + n^{-1/3}t)J\rceil / J} (g'(x_0) + o(1))(u - x) du = \frac{1}{2} g'_0(x_0) t^2 + o(1),$$

establishing the assertion (A.1), as $b = -g'_0(x_0)/2$.

The class \mathcal{F}_n is easily seen to be a VC-class. Hence by Example 2.11.24 of van der Vaart and Wellner [44], the entropy condition in Theorem 2.11.22 of van der Vaart and Wellner [44] holds. We check the remaining conditions of that theorem to conclude that $\{\mathbb{G}_n(f_{n,t}) : t \in [0, K]\}$ converges to a tight centered Gaussian process on $\mathbb{L}_\infty([0, K])$.

Define $F_n(x) = n^{1/6}(\mathbb{1}\{x \leq \lceil(x_0 + n^{-1/3}K)J\rceil/J\} - \mathbb{1}\{x \leq \lceil x_0 J \rceil/J\})$. Then F_n is an envelope for \mathcal{F}_n , that is, $|f_{n,t}| \leq F_n$ for every $t \in [0, K]$. Using the continuity of g in a neighborhood of x_0 , we have that

$$\mathbb{E}_0 F_n^2 = n^{1/3} \int_{(\lceil x_0 J \rceil)/J}^{\lceil(x_0 + n^{-1/3}K)J\rceil/J} [g_0(x_0) + o(1)] du = K g_0(x_0) + o(1),$$

which is bounded. Next we note that $\mathbb{E}_0 F_n^2 \mathbb{1}\{F_n > \eta\sqrt{n}\} \rightarrow 0$ for every $\eta > 0$, which is immediate because for any x , the indicator vanishes for all sufficiently large n .

Next, fix any sequence $\delta_n \downarrow 0$. Then by the boundedness of g_0 , for $0 \leq t \leq s$,

$$\begin{aligned} \mathbb{E}_0(f_{n,s} - f_{n,t})^2 &= n^{1/3} \int_{(\lceil(x_0 + n^{-1/3}t)J\rceil)/J}^{\lceil(x_0 + n^{-1/3}s)J\rceil/J} g_0(u) du \\ &\lesssim \frac{n^{1/3}}{J} (\lceil(x_0 + n^{-1/3}s)J\rceil - \lceil(x_0 + n^{-1/3}t)J\rceil) \\ &\leq |s - t| + o(n^{1/3}/J), \end{aligned}$$

which goes to zero uniformly on $\{(s, t) : 0 \leq t \leq s, |s - t| < \delta_n\}$.

We shall now find the limit of the covariance $\mathbb{E}_0(f_{n,s} f_{n,t}) - \mathbb{E}_0(f_{n,s})\mathbb{E}_0(f_{n,t})$ for all $0 \leq t \leq s$. By the orthogonality of $f_{n,s} - f_{n,t}$ and $f_{n,t}$, we have that

$$\mathbb{E}_0(f_{n,s} f_{n,t}) = \mathbb{E}_0(f_{n,t}^2) = n^{1/3} \int_{\lceil x_0 J \rceil/J}^{\lceil(x_0 + n^{-1/3}t)J\rceil/J} g_0(u) du \rightarrow g_0(x_0)t.$$

Further, $\mathbb{E}_0 f_{n,t} = n^{1/6}[g_0(x_0) + o(1)][n^{-1/3}(t + o(1))] \rightarrow 0$. Therefore, for any s, t , we have $\text{Cov}(f_{n,s}, f_{n,t}) \rightarrow \min(s, t)g_0(x_0)$ as $n \rightarrow \infty$. Hence all assumptions of Theorem 2.11.22 of van der Vaart and Wellner [44] are satisfied for the limit process given by $aW_1(t) - bt^2$, completing the proof of the lemma. \square

Proof of Lemma 5.2. Fix z and denote the maximizer in the statement by \hat{h}_n . As in the proof of Lemma 5.1, we restrict attention to $t \geq 0$. The tightness can be similarly proved for the argmax with t restricted to the left of zero, and the final result by combining the two conclusions.

For $r \geq 0$, define a stochastic process

$$K_n(r) = \mathbb{P}_n \mathbb{1}\{(\lceil x_0 J \rceil/J, \lceil(x_0 + r)J\rceil/J] - (g_0(x_0) + zn^{-1/3})(\lceil(x_0 + r)J\rceil - \lceil x_0 J \rceil)/J\}$$

and a function $K(r) = G_0(x_0 + r) - G_0(x_0) - g_0(x_0)r$ on the positive half-line. Note that $\hat{h}_n = n^{1/3} \arg \max\{K_n(r) : r \in \mathbb{R}\}$. We apply Theorem 3.2.5 of van der Vaart and Wellner [44] on K_n and K to show that $\arg \max\{K_n(r) : r \in \mathbb{R}\} = O_P(n^{-1/3})$. We have that $K_n(0) = K(0) = 0$ and $\mathbb{E}_0 K_n(r) \rightarrow K(r)$ for all r . Also, by a Taylor series expansion $g_0(s) - g_0(x_0) = (g_0'(x_0) + o(1))(s - x_0)$, we have that

$$K(r) = g_0(x_0)r + \frac{1}{2}(g_0'(x_0) + o(1))r^2 - g_0(x_0)r \lesssim -r^2. \quad (\text{A.2})$$

Next, for sufficiently small δ , we bound $\mathbb{E}_0^* \sup\{\sqrt{n}|K_n(r) - K(r)| : |r| < \delta\}$. We can write $\sqrt{n}(K_n(r) - K(r))$ as the difference of $\mathbb{G}_n \mathbb{1}\{(\lceil x_0 J \rceil/J, \lceil(x_0 + r)J\rceil/J]\}$ and

$$[\sqrt{n}(G_0(x_0 + r) - G_0(x_0) - g_0(x_0)r) + zn^{1/6}](\lceil(x_0 + r)J\rceil - \lceil x_0 J \rceil)/J. \quad (\text{A.3})$$

Clearly, the class of indicators $\mathcal{M}_\delta := \{\mathbb{1}\{(\lceil x_0 J \rceil / J, \lceil (x_0 + r) J \rceil / J) : 0 \leq r < \delta\} : 0 \leq r < \delta\}$ has envelope $M_\delta = \mathbb{1}\{(x_0, x_0 + \delta)\} \leq 1$ and forms a VC-class. From the discussion on page 291 of van der Vaart and Wellner [44], we have that

$$\mathbb{E}_0^* \sup\{|\mathbb{G}_n(f)| : f \in \mathcal{M}_\delta\} \lesssim (\mathbb{E}_0^* M_\delta^2)^{1/2} \sup_Q \int_0^\eta \sqrt{\log \mathcal{N}(\epsilon \|M_\delta\|_{2,Q}, \mathcal{M}_\delta, \|\cdot\|_{2,Q})} d\epsilon.$$

Observe that $\mathbb{E}_0^* M_\delta^2 \lesssim \delta$, while the uniform entropy integral is finite by Theorem 2.6.9 of van der Vaart and Wellner [44]. Thus the display above is bounded by $\sqrt{\delta}$.

The second factor in (A.3) is bounded by $r < \delta$. The first term in the first factor in (A.3) is bounded in absolute value by a multiple of $\sqrt{n}\delta^2$ in view of (A.2), while the second term is bounded by a multiple of $n^{1/6}$. Therefore, the expression in (A.3) is bounded by $\sqrt{n}\delta^3 + n^{1/6}\delta$. Piecing these bounds together, we have that

$$\mathbb{E}_0 \sup\{|K_n(r) - K(r)| : 0 \leq r < \delta\} \lesssim \sqrt{\delta} + \sqrt{n}\delta^3 + n^{1/6}\delta.$$

Hence by Theorem 3.2.5 of van der Vaart and Wellner [44] with $\phi_n(\delta) = \sqrt{\delta} + \sqrt{n}\delta^3 + n^{1/6}\delta$, the rate $n^{-1/3}$ is obtained by solving the rate equation $\sqrt{\delta} + \sqrt{n}\delta^3 + n^{1/6}\delta \leq \sqrt{n}\delta^2$.

Finally, the condition $K_n(\hat{h}_n) \geq K_n(0) \geq K_n(0) - O_{P_0}(r_n^{-2})$ holds by the definition of the maximizer. \square

For the remaining proofs, the dependence of $(\theta_1, \dots, \theta_J)$ makes it harder to deal with the expressions directly. To bypass the dependence in $(\theta_1, \dots, \theta_J)$, we represent θ_j as $V_j / (\sum_{l=1}^J V_l)$, where $V_j \sim \text{Gamma}(\alpha_j + N_j, 1)$ independently. Introduce the processes

$$\begin{aligned} A_n(V, D_n, t) &= n^{2/3} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3}t) J \rceil} (V_j - \mathbb{E}(V_j)) / (\sum_{l=1}^J V_l), \\ \bar{A}_n(D_n, t) &= n^{2/3} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3}t) J \rceil} \{ \mathbb{E}(V_j) / (\sum_{l=1}^J V_l) - N_j / n \}, \\ B_n(D_n, t) &= n^{2/3} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3}t) J \rceil} (N_j / n - J^{-1} g_0(x_0)), \end{aligned}$$

where $V = (V_1, \dots, V_J)$. Observe that

$$A_n(V, D_n, t) + \bar{A}_n(D_n, t) + B_n(D_n, t) = n^{2/3} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3}t) J \rceil} (\theta_j - J^{-1} g_0(x_0)). \quad (\text{A.4})$$

Proof of Lemma 5.3. Note that $\sum_{l=1}^J V_l | D_n \sim \text{Gamma}(\alpha. + n, 1)$, and so we have $\mathbb{E}(\sum_{l=1}^J V_l | D_n) = \alpha. + n$, and $\text{Var}(\sum_{l=1}^J V_l | D_n) = \alpha. + n$. By the assumptions on the prior, $\max\{\alpha_j : 1 \leq j \leq J\} = O(1)$, and therefore

$$n^{-1} \sum_{l=1}^J V_l = 1 + O(J/n) + O_{P_0}(n^{-1/2}) \rightarrow_{P_0} 1. \quad (\text{A.5})$$

We shall show that $A_n(\theta, D_n, \cdot) \rightsquigarrow aW_2$ in $\mathbb{L}_\infty([-K, K])$ and $\sup\{|\bar{A}_n(D_n, t)| : t \in [-K, K]\} \rightarrow 0$ in probability for all $K > 0$. To prove the first assertion, define $Y_n(V, D_n, t) = n^{-1/3} \times \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3}t)J \rceil} (V_j - \mathbb{E}(V_j))$. In view of (A.5), the assertion reduces to showing that for all $K > 0$,

$$\mathcal{L}(Y_n(V, D_n, t) : t \in [-K, K] | D_n) \rightsquigarrow \mathcal{L}(aW_2(t) : t \in [-K, K] | W_1) \quad (\text{A.6})$$

in $\mathbb{L}_\infty[-K, K]$.

We first verify the finite-dimensional convergence to Gaussian limits. Since $N_j \sim \text{Bin}(n; G_0(I_j))$ for $1 \leq j \leq J$, and g_0 is continuous and positive in a neighborhood of x_0 , with probability tending to one, $\sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3}t)J \rceil} N_j \sim g_0(x_0)tn^{2/3}$. Recall that the fourth central moment of $\text{Gamma}(\alpha, 1)$ is given by $3\alpha^2 + 9\alpha \lesssim \alpha^2$ for $\alpha \geq 1$. As $V_j \sim \text{Gamma}(\alpha_j + N_j, 1)$ and $\min\{N_j : 1 \leq j \leq J\} \rightarrow \infty$, we obtain that

$$\mathbb{E}[|V_j - \mathbb{E}(V_j)|^2 | D_n] = \alpha_j + N_j, \quad \mathbb{E}[|V_j - \mathbb{E}(V_j)|^4 | D_n] \lesssim (\alpha_j + N_j)^2. \quad (\text{A.7})$$

Therefore, the sum of the variances is

$$\sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3}t)J \rceil} \text{Var}[n^{-1/3}(V_j - \mathbb{E}(V_j)) | D_n] = n^{-2/3} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3}t)J \rceil} (\alpha_j + N_j) \rightarrow g_0(x_0)t \quad (\text{A.8})$$

in P_0 -probability.

Also by (A.7) and (5.6), for any $|t| \leq K$, $\sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3}t)J \rceil} \mathbb{E}[|n^{-1/3}(V_j - \mathbb{E}(V_j))|^4 | D_n]$ is bounded in probability by a constant multiple of

$$n^{-4/3} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3}t)J \rceil} (\alpha_j + N_j)^2 \lesssim n^{-4/3} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3}t)J \rceil} N_j^2 \lesssim n^{-4/3} n^{-1/3} K J (n^2/J^2),$$

which is at most of the order $n^{1/3}/J \rightarrow 0$. This verifies Lyapunov's condition with the fourth moment, and hence the Lindeberg condition, for the central limit theorem to hold, giving that $Y_n(V, D_n, t) \rightsquigarrow \sqrt{g_0(x_0)t} N(0, 1)$, conditionally on D_n , in P_0 -probability. The joint convergence can be verified by evaluating the limit of the covariance between $Y_n(V, D_n, t_1)$ and $Y_n(V, D_n, t_2)$ for $t_1 < t_2$ as

$$\begin{aligned} & \left[\mathbb{E} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3}t_1)J \rceil} (V_j - \mathbb{E}(V_j)) \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3}t_2)J \rceil} (V_j - \mathbb{E}(V_j)) | D_n \right] \\ &= \left[\sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3}t_1)J \rceil} \text{Var}(V_j) | D_n \right] \rightarrow g_0(x_0)t_1 \end{aligned}$$

in P_0 -probability.

Next, we show the tightness of $Y_n(V, D_n, \cdot)$ in $\mathbb{L}_\infty([-K, K])$. It suffices to restrict t to $[0, K]$. Similarly, we can prove tightness with t restricted to $[-K, 0]$, and conclude tightness in $\mathbb{L}_\infty([-K, K])$ after combining. The verification of tightness in $\mathbb{L}_\infty([0, K])$ involves the verification of asymptotic stochastic equicontinuity as in Section 2.1.2 of van der Vaart and Wellner [44], or more specifically, as in Theorem 18.14 of van der Vaart [43]. To derive the required bound, we apply Theorem 2.2.4 of van der Vaart and Wellner [44] with the \mathbb{L}_4 -norm (i.e., $\psi(x) = x^4$ in the theorem). It suffices to show that

$$\mathbb{E}[|Y_n(V, D_n, s) - Y_n(V, D_n, t)|^4 | D_n] \lesssim |s - t|^2, \quad (\text{A.9})$$

because then, with $d(s, t) = \sqrt{|s - t|}$ and $\eta = \delta^{2/3}$, conditionally on D_n , we obtain

$$\left\| \sup_{d(s,t) \leq \delta} |Y_n(V, D_n, s) - Y_n(V, D_n, t)| \right\|_4 \lesssim \int_0^\eta (1/\epsilon^2)^{-1/4} d\epsilon + \delta(\eta^4)^{-1/4} \lesssim \delta^{1/3},$$

as the ϵ -packing number with respect to d is of the order ϵ^{-2} .

To show (A.9), let $0 \leq t \leq s$ and write

$$Y_n(V, D_n, s) - Y_n(V, D_n, t) = n^{-1/3} \sum_{j=\lceil(x_0+n^{-1/3}t)J\rceil+1}^{\lceil(x_0+n^{-1/3}s)J\rceil} (V_j - \mathbb{E}(V_j))$$

as the sum of independent centered random variables. Hence the left hand side of (A.9) is given by

$$\begin{aligned} & n^{-4/3} \left\{ \sum_{j=\lceil(x_0+n^{-1/3}t)J\rceil+1}^{\lceil(x_0+n^{-1/3}s)J\rceil} \mathbb{E}[|V_j - \mathbb{E}(V_j)|^4 | D_n] \right. \\ & \quad \left. + \sum_{\lceil(x_0+n^{-1/3}t)J\rceil+1 \leq j \neq j' \leq \lceil(x_0+n^{-1/3}s)J\rceil} \mathbb{E}[|V_j - \mathbb{E}(V_j)|^2 | D_n] \mathbb{E}[|V_{j'} - \mathbb{E}(V_{j'})|^2 | D_n] \right\}. \end{aligned}$$

Using (A.7) and (5.6), we obtain the bound a constant multiple of

$$n^{-4/3} (n^{-1/3} |s - t| J \max\{N_j : 1 \leq j \leq J\})^2 \lesssim |s - t|^2$$

with probability tending to one, completing the verification of (A.9).

Next, we write $\bar{A}_n(D_n, t)$ as

$$\begin{aligned} & \left| n^{2/3} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil(x_0+n^{-1/3}t)J\rceil} \left(\frac{\alpha_j + N_j}{\sum_{l=1}^J V_l} - \frac{N_j}{\sum_{l=1}^J V_l} + \frac{N_j}{\sum_{l=1}^J V_l} - \frac{N_j}{n} \right) \right| \\ & \leq \frac{n^{2/3}}{\sum_{l=1}^J V_l} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil(x_0+n^{-1/3}t)J\rceil} \max_{1 \leq l \leq J} \alpha_l + n^{2/3} \left| \frac{n}{\sum_{l=1}^J V_l} - 1 \right| \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil(x_0+n^{-1/3}t)J\rceil} \frac{N_j}{n}. \end{aligned} \tag{A.10}$$

By the assumptions on the prior and (A.5), the first term of (A.10) is bounded in probability by a constant multiple of $n^{-2/3} J K$, while the second term is bounded in probability by a constant multiple of $n^{2/3} (J/n) n^{-1/3} J (1/J) = J n^{-2/3} \rightarrow 0$. Therefore, $\bar{A}_n(D_n, \cdot) \rightarrow_{P_0} 0$ in $\mathbb{L}_\infty([-K, K])$.

The weak convergence of $B_n(D_n, \cdot)$ to $a W_1$ in $\mathbb{L}_\infty([-K, K])$ has been established in Lemma 5.1. Combining all three assertions, we get the result. \square

Proof of Lemma 5.4. The main idea of the proof is similar to that of Lemma 5.2 using Theorem 3.2.5 of van der Vaart and Wellner [44] applied to the posterior process. We need to establish the tightness of the conditional distribution of $n^{1/3} h_n$ given D_n in probability, where

$$h_n = \arg \max_{r \in \mathbb{R}} \left\{ \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil(x_0+r)J\rceil} \left(\theta_j - \frac{g_0(x_0)}{J} \right) - n^{-1/3} z(\lceil(x_0+r)J\rceil - \lceil x_0 J \rceil)/J \right\}.$$

Observe that this is equivalent to proving the tightness of $n^{1/3}h_n$ with respect to the joint distribution of $(\theta_1, \dots, \theta_J)$ and the observations X_1, \dots, X_n respectively following the posterior distribution and the sampling distribution.

For $r \geq 0$, let $M_n(r)$ and $M(r)$ be defined as

$$M_n(r) = \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + r) J \rceil} \left(\theta_j - \frac{g_0(x_0)}{J} \right) - z n^{-1/3} (\lceil (x_0 + r) J \rceil - \lceil x_0 J \rceil) / J$$

and $M(r) = G_0(x_0 + r) - G_0(x_0) - g_0(x_0)r$, the same as $K(r)$ in the proof of Lemma 5.2. We restrict to $r \geq 0$, and observe that $M(0) = 0$. The condition $M(r) - M(0) \lesssim -r^2$ has been verified within the proof of Lemma 5.2. We verify other conditions of Theorem 3.2.5 of van der Vaart and Wellner [44] to show that $h_n = O_P(n^{-1/3})$ with respect to the joint probability. As in the proof of Theorem 3.5, we use the gamma representation in (A.4) for the posterior distribution of $(\theta_1, \dots, \theta_J)$. Then $\sqrt{n}(M_n(r) - M(r))$ can be rewritten as the sum of

$$\begin{aligned} H_{1n}(r) &= \sqrt{n} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + r) J \rceil} (V_j - \mathbb{E}(V_j)) / (\sum_{l=1}^J V_l); \\ H_{2n}(r) &= \sqrt{n} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + r) J \rceil} \{(N_j + \alpha_j) / (\sum_{l=1}^J V_l) - N_j / n\}; \\ H_{3n}(r) &= \sqrt{n} \mathbb{P}_n [\mathbb{1}\{(\lceil x_0 J \rceil / J, \lceil (x_0 + r) J \rceil / J) - \mathbb{1}\{(x_0, x_0 + r)\}\} - z n^{1/6} (\lceil (x_0 + r) J \rceil - \lceil x_0 J \rceil) / J]. \end{aligned}$$

We estimate the maximal size of all these processes over $0 \leq r < \delta$ in the joint probability.

To estimate $\sup\{|H_{1n}(r)| : 0 \leq r < \delta\}$, we first bound

$$\mathbb{P}\left(\sup\left\{|\sqrt{n} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + r) J \rceil} (V_j - \mathbb{E}(V_j))| : 0 \leq r < \delta\right\} \geq b | D_n\right) \leq n \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + r) J \rceil} (N_j + \alpha_j) / b^2$$

by the Kolmogorov–Doob maximal inequality. Hence, in view of (5.6), $\alpha_j = O(1)$ uniformly for all $1 \leq j \leq J$ and (A.5), it follows that $\sup\{|H_{1n}(r)| : 0 \leq r < \delta\} = O_P(\sqrt{\delta})$. For H_{2n} , we have, using arguments similar to (A.10), $\sup\{|H_{2n}(r)| : 0 \leq r < \delta\} = O_P(J n^{-1/2} \delta + \delta + n^{-3/2} J^2 \delta)$, which reduces to $O_P(n^{1/6} \delta)$ because of $n^{1/3} \ll J \ll n^{2/3}$. The last term $H_{3n}(r)$ was $K_n(r)$ in the proof of Lemma 5.2 and was shown to be $O_P(\sqrt{\delta} + \sqrt{n} \delta^3 + n^{1/6} \delta)$.

Hence by applying Theorem 3.2.5 of van der Vaart and Wellner [44] with $\phi_n(\delta) = \sqrt{\delta} + \sqrt{n} \delta^3 + n^{1/6} \delta$, the rate $n^{-1/3}$ is obtained by solving the rate equation $\sqrt{\delta} + \sqrt{n} \delta^3 + n^{1/6} \delta \leq \sqrt{n} \delta^2$.

Finally, the condition $M_n(h_n) \geq M_n(0) \geq M_n(0) - O_{P_0}(r_n^{-2})$ holds by the definition of the maximizer. \square

Funding

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

References

- [1] Banerjee, M. (2000). *Likelihood Ratio Inference in Regular and Non-regular Problems*. Ann Arbor, MI: ProQuest LLC. Thesis (Ph.D.)–University of Washington. [MR2701606](#)
- [2] Banerjee, M. (2007). Likelihood based inference for monotone response models. *Ann. Statist.* **35** 931–956. [MR2341693](#) <https://doi.org/10.1214/009053606000001578>
- [3] 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>

- [4] 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 Series in Probability and Mathematical Statistics. London–Sydney: Wiley. [MR0326887](#)
- [5] Barlow, R.E. and Brunk, H.D. (1972). The isotonic regression problem and its dual. *J. Amer. Statist. Assoc.* **67** 140–147. [MR0314205](#)
- [6] 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>
- [7] 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>
- [8] Bhaumik, P. and Ghosal, S. (2017). Bayesian inference for higher-order ordinary differential equation models. *J. Multivariate Anal.* **157** 103–114. [MR3641739](#) <https://doi.org/10.1016/j.jmva.2017.03.003>
- [9] Bhaumik, P., Shi, W. and Ghosal, S. (2021). Bayesian generalized regression in partial differential equation models. *Bernoulli*. To appear.
- [10] Bowman, A.W., Jones, M.C. and Gijbels, I. (1998). Testing monotonicity of regression. *J. Comput. Graph. Statist.* **7** 489–500.
- [11] Brunk, H.D. (1970). Estimation of isotonic regression. In *Nonparametric Techniques in Statistical Inference (Proc. Sympos., Indiana Univ., Bloomington, Ind., 1969)* 177–197. London: Cambridge Univ. Press. [MR0277070](#)
- [12] Brunner, L.J. (1992). Bayesian nonparametric methods for data from a unimodal density. *Statist. Probab. Lett.* **14** 195–199. [MR1173617](#) [https://doi.org/10.1016/0167-7152\(92\)90021-V](https://doi.org/10.1016/0167-7152(92)90021-V)
- [13] Brunner, L.J. and Lo, A.Y. (1989). Bayes methods for a symmetric unimodal density and its mode. *Ann. Statist.* **17** 1550–1566. [MR1026299](#) <https://doi.org/10.1214/aos/1176347381>
- [14] 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>
- [15] Chakraborty, M. and Ghosal, S. (2021). Convergence rates for Bayesian estimation and testing in monotone regression. *Electron. J. Stat.* **15** 3478–3503. [MR4280172](#) <https://doi.org/10.1214/21-ejs1861>
- [16] Chakraborty, M. and Ghosal, S. (2021). Coverage of credible intervals in nonparametric monotone regression. *Ann. Statist.* **49** 1011–1028. [MR4255117](#) <https://doi.org/10.1214/20-aos1989>
- [17] Chernoff, H. (1964). Estimation of the mode. *Ann. Inst. Statist. Math.* **16** 31–41. [MR0172382](#) <https://doi.org/10.1007/BF02868560>
- [18] 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>
- [19] de Boor, C. (2001). *A Practical Guide to Splines*, revised ed. *Applied Mathematical Sciences* **27**. New York: Springer. [MR1900298](#)
- [20] 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**. <https://doi.org/10.18637/jss.v032.i05>
- [21] Dümbgen, L. (2003). Optimal confidence bands for shape-restricted curves. *Bernoulli* **9** 423–449. [MR1997491](#) <https://doi.org/10.3150/bj/1065444812>
- [22] 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>
- [23] Ghosal, S., Ghosh, J.K. and van der Vaart, A.W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28** 500–531. [MR1790007](#) <https://doi.org/10.1214/aos/1016218228>
- [24] 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>
- [25] Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. Cambridge Series in Statistical and Probabilistic Mathematics **44**. Cambridge: Cambridge Univ. Press. [MR3587782](#) <https://doi.org/10.1017/9781139029834>
- [26] 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>
- [27] Grenander, U. (1957). On the theory of mortality measurement. II. *Skand. Aktuarietidskr.* **39** 125–153. [MR0093415](#) <https://doi.org/10.1080/03461238.1956.10414944>
- [28] 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>

- [29] Groeneboom, P. and Jongbloed, G. (2014). *Nonparametric Estimation Under Shape Constraints: Estimators, Algorithms and Asymptotics*. Cambridge Series in Statistical and Probabilistic Mathematics **38**. New York: Cambridge Univ. Press. [MR3445293](#) <https://doi.org/10.1017/CBO9781139020893>
- [30] 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>
- [31] 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>
- [32] 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](#)
- [33] 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>
- [34] 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>
- [35] Prakasa Rao, B.L.S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31** 23–36. [MR0267677](#)
- [36] Salomond, J.-B. (2014). Concentration rate and consistency of the posterior distribution for selected priors under monotonicity constraints. *Electron. J. Stat.* **8** 1380–1404. [MR3263126](#) <https://doi.org/10.1214/14-EJS929>
- [37] Salomond, J.-B. (2014). Adaptive Bayes test for monotonicity. In *The Contribution of Young Researchers to Bayesian Statistics*. Springer Proc. Math. Stat. **63** 29–33. Cham: Springer. [MR3133254](#) https://doi.org/10.1007/978-3-319-02084-6_7
- [38] 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>
- [39] 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>
- [40] Schwartz, L. (1965). On Bayes procedures. *Z. Wahrsch. Verw. Gebiete* **4** 10–26. [MR0184378](#) <https://doi.org/10.1007/BF00535479>
- [41] Sen, B., Banerjee, M. and Woodroffe, M. (2010). Inconsistency of bootstrap: The Grenander estimator. *Ann. Statist.* **38** 1953–1977. [MR2676880](#) <https://doi.org/10.1214/09-AOS777>
- [42] Shen, W. and Ghosal, S. (2015). Adaptive Bayesian procedures using random series priors. *Scand. J. Stat.* **42** 1194–1213. [MR3426318](#) <https://doi.org/10.1111/sjos.12159>
- [43] van der Vaart, A.W. (2000). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics **3**. Cambridge: Cambridge Univ. Press. [MR1652247](#) <https://doi.org/10.1017/CBO9780511802256>
- [44] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Process with Applications to Statistics*. New York: Springer-Verlag.
- [45] Wu, Y. and Ghosal, S. (2008). Kullback–Leibler property of kernel mixture priors in Bayesian density estimation. *Electron. J. Stat.* **2** 298–331. [MR2399197](#) <https://doi.org/10.1214/07-EJS130>

Received August 2020 and revised January 2021