ABSTRACT

CHAKRABORTY, MOUMITA. Bayesian Inference Under Shape Constraints. (Under the direction of Subhashis Ghosal).

In the context of nonparametric univariate monotone regression with unknown error variance σ^2, we study global posterior contraction rates and the point-wise coverage of a Bayesian credible interval. Using a finite random series of piece-wise constant functions with normal basis coefficients as a prior for a regression function f, we obtain a conjugate posterior and project each sample from the unrestricted posterior on the shape-constrained space of functions, inducing a distribution to be called the “projection-posterior". For global contraction rates, point estimation and credible sets, a lot more tractability is obtained by considering this projection-posterior distribution. We show that the projection-posterior contracts at the minimax rate $n^{-1/3}$ with respect to the L_1-distance, and also with respect to the empirical L_p-distances up to a logarithmic factor. For the limiting coverage of a credible interval of f evaluated at a fixed point x_0, we observe a very interesting phenomenon that the coverage may be higher than the nominal credibility level, the opposite of a phenomenon observed by Cox (1993) in the context of smooth nonparametric signal estimation. We then show that a re-calibration technique can give the right coverage. By employing the distance of the unrestricted posterior from its projection on the monotone class, we develop an asymptotically optimal Bayesian test for the hypothesis of monotonicity. Using the idea of enlarging the region near the null hypothesis of f being in the monotone class, as proposed by Salomond (2018), we construct a universally consistent Bayesian test for monotonicity with optimal power properties.

For a monotone decreasing probability density on the positive real line, we study global posterior contraction rates and the point-wise coverage of a Bayesian credible interval. Using a random histogram with Dirichlet basis coefficients as a prior for a density g, we obtain a conjugate posterior and use the projection-based approach to project each sample from the unrestricted posterior on the space of monotone decreasing densities. We show that the projection-posterior contracts at the minimax rate $n^{-1/3}$ with respect to the L_1-distance for g on a bounded domain, and the minimax rate up to a logarithmic factor for g on an unbounded domain. For g evaluated at a fixed point, we obtain a projection-posterior credible interval with the asymptotic coverage more than the credibility level, a phenomenon similar to that observed in the case of monotone regression.

We use the projection-posterior for a regression function to estimate the regression

quantile of a monotone increasing function. We obtain a credible set for the regression quantile that has the appropriate asymptotic coverage. Next, we improve the posterior contraction rate by sampling in two stages. We obtain a sampling interval from the posterior of the regression quantile and generate the second-stage observations with the predictors belonging to that sample. By approximating the regression function on this interval as a linear function, we use a Gaussian prior on the coefficients and analyze the second-stage posterior distribution of the regression quantile and find that the second stage posterior contracts at the parametric rate $n^{-1/2}$.

Bayesian Inference Under Shape Constraints

by
Moumita Chakraborty

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Statistics

Raleigh, North Carolina
2019

APPROVED BY:

Soumendra Lahiri	Ryan Martin
Sujit Ghosh	Subhashis Ghosal Chair of Advisory Committee

BIOGRAPHY

The author was born in Allahabad, Uttar Pradesh, India on 24th November, 1991. After finishing high school from DAV Model School, Kharagpur, in May 2009, she joined the Bachelor of Statistics program at Indian Statistical Institute (ISI), Kolkata. In June 2012, she enrolled in the same institute to pursue a master's degree in statistics. After graduating from ISI Kolkata, she joined the statistics department at North Carolina State University in the Fall of 2014, as a doctoral student in statistics.

ACKNOWLEDGEMENTS

I would like to express my deepest thanks to my advisor Dr. Subhashis Ghosal for being the most patient, caring and helpful advisor one could think of. Without his support, this work would not have been possible. His in-depth knowledge of his subject and immense patience in guiding his students through challenging problems have been the sole factors in helping me build the foundations of my learning, and working my way out through some of the most challenging phases of my academic life as a PhD student. I am extremely grateful to have him as my advisor and mentor. I would also like to thank my committee members Dr. Soumendranath Lahiri, Dr. Sujit Ghosh and Dr. Ryan Martin for being in my doctoral committee and providing me helpful advice from time to time. Their insightful suggestions on different aspects of my academic career have been of great help throughout these years. I want to thank Dr. Zheng Li for serving as the graduate representative in my committee. I would also like to thank two staff members from our department – Alison McCoy and Lanakila Alexander. Thanks to the friendly atmosphere that they have created by their caring nature, this department has been a second home to me. I would like to thank my fellow graduate students and friends from ISI who have all enriched my experience of staying here for the past five years. Lastly, I owe all my accomplishments to my parents. Their unconditional love is my biggest source of strength.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER

1

INTRODUCTION

Statistical models often involve unknown functions which are described qualitatively in terms of the shape of the function. Shape constraints arise naturally in several contexts such as climate science, ecology, genomics, survival analysis, finance and so on. The most commonly encountered cases of shape restrictions are monotonicity, convexity (or concavity) and unimodality. For example, the aging process of hardware equipments often indicates a monotone increasing hazard rate. In situations where a particular shape restriction is known to be present, one needs an estimator of the function complying with the restriction for an appropriate interpretation of the results. In many situations, estimation of a shape-restricted function is possible without assuming smoothness on it. They use the qualitative assumption of shape only, eliminating the need for the extra smoothness assumptions that may sometimes impose unnecessary restrictions on the model. Several studies involving nonparametric statistical methods for estimation and inference under shape restrictions were carried out previously, with wide applicability to real world problems and theoretical justifications. In this dissertation, we explore a nonparamteric Bayesian approach to some common shape-restricted problems. In this chapter, we shall provide a background of the statistical tools used for such problems, some relevant discussion on Bayesian nonparametrics, and lay out the key ideas and novelties pertaining to the works in the subsequent

chapters. We start with describing some real life examples involving shape constraints.

1.1 Some motivating examples

1.1.1 Global warming

The global warming dataset provide by Jones et al. (2016) contains the mean annual temperature anomalies in Celsius, from 1850 to 2010. Temperature anomaly denotes the departure from a long-term average. For this dataset, the long-term average is the mean of the temperatures from 1961 to 1990. The data were collected from different meteorological stations both on the land and the sea and corrected for non-climatic errors. The data are tabulated in Table 1.1. Figure 1.1 provides a scatterplot of the temperatures with respect to the years, along with a nonparametric constrained estimator of the underlying function. The plot depicts an increasing trend of temperature with time.

Under the assumption that the function relating the mean annual temperatures to time is monotone increasing, estimation of the function calls for the known problem of isotonic regression. The isotonic regression provides an estimate of the underlying regression function by minimizing the error sum of squares under the monotonicity constraint.

1.1.2 Warming-up of Lake Mendota

The historical data considered by Barlow and Brunk (1972) on the number of winter days Lake Mendota in Wisconsin freezes per year, shows a decreasing trend over time. The data have been obtained from the R package "isotone". Similar to the data on annual temperatures, the nonparamteric estimator in this case is also given by the isotonic regression of the number of freezing days on time. The constraint on the regression function is that of monotone decreasing. Table 1.2 gives the data, and Figure 1.2 gives its scatterplot along with the isotonic estimator and a smoothed version of the estimator.

Several other situations call for shape-restricted estimation. The Canadian income data (Ruppert et al. 2003) suggests a concave relation between the logarithm of income and the age of the earner; see Figure 1.3. Also, the aging process of certain hardware equipments show a monotone hazard rate. Estimation and inference in such problems should take into account the underlying shape constraint for accurate interpretations of the results.

Table 1.1: Temperatures adjusted for the mean of 1961 – 1990 temperatures, based on the global warming dataset.

−0.38	−0.22	−0.22	−0.27	−0.25	−0.27	−0.35	−0.46	−0.47	−0.29	−0.35
−0.41	−0.52	−0.28	−0.49	−0.27	−0.24	−0.31	−0.23	−0.26	−0.28	−0.33
−0.23	−0.30	−0.37	−0.40	−0.38	−0.08	0.04	−0.23	−0.23	−0.21	−0.21
−0.29	−0.41	−0.39	−0.37	−0.42	−0.31	−0.17	−0.42	−0.34	−0.46	−0.47
−0.41	−0.39	−0.19	−0.21	−0.42	−0.29	−0.20	−0.26	−0.41	−0.48	−0.53
−0.38	−0.29	−0.47	−0.52	−0.53	−0.49	−0.55	−0.44	−0.43	−0.25	−0.15
−0.39	−0.47	−0.34	−0.28	−0.25	−0.19	−0.30	−0.27	−0.29	−0.21	−0.11
−0.21	−0.21	−0.35	−0.14	−0.08	−0.13	−0.27	−0.13	−0.17	−0.14	−0.02
0.00	−0.05	0.02	0.02	-0.02	0.00	0.15	0.03	−0.07	−0.04	−0.04
−0.07	−0.17	−0.05	0.03	0.10	−0.13	−0.19	−0.26	−0.00	0.05	0.02
−0.05	0.04	0.02	0.05	−0.22	−0.14	−0.07	−0.08	−0.11	0.03	−0.03
−0.19	−0.07	0.06	−0.21	−0.15	−0.24	0.04	−0.06	0.06	0.09	0.14
0.01	0.19	−0.01	−0.03	0.05	0.19	0.20	0.12	0.30	0.25	0.10
0.14	0.21	0.32	0.18	0.39	0.54	0.31	0.29	0.44	0.50	0.51
0.45	0.54	0.50	0.49	0.40	0.51	0.56	0.42	0.47	0.50	0.57
0.75										

The order is in increasing years from left to right and (next) row-wise.

1.2 Shape-restricted estimation

Nonparametric estimation of the underlying function in a major portion of shape-restricted statistical models are based on the concepts of the greatest convex minorant (GCM) and the least concave majorant (LCM) of a cumulative sum diagram of a certain set of points corresponding to the context of the problem. For instance, in monotone regression, the constrained estimator of the regression function, obtained by minimizing the error sum of squares subject to the monotonicity constraint, is the slope (left derivative) of the GCM of

$$\left\{(0,0),(1/n,Y_1/n),(2/n,(Y_1+Y_2)/n),\ldots,\left(1,\sum_{i=1}^{n}Y_i/n\right)\right\}.$$

Similar formulae exist for shape-constrained estimation in current status and monotone density scenarios. The algorithm for obtaining the constrained estimator using this approach has been described by De Leeuw et al. (2009) using the Pool-Adjacent-Violators Algorithm (PAVA). The implementation of PAVA has been carried out through the R package "isotone".

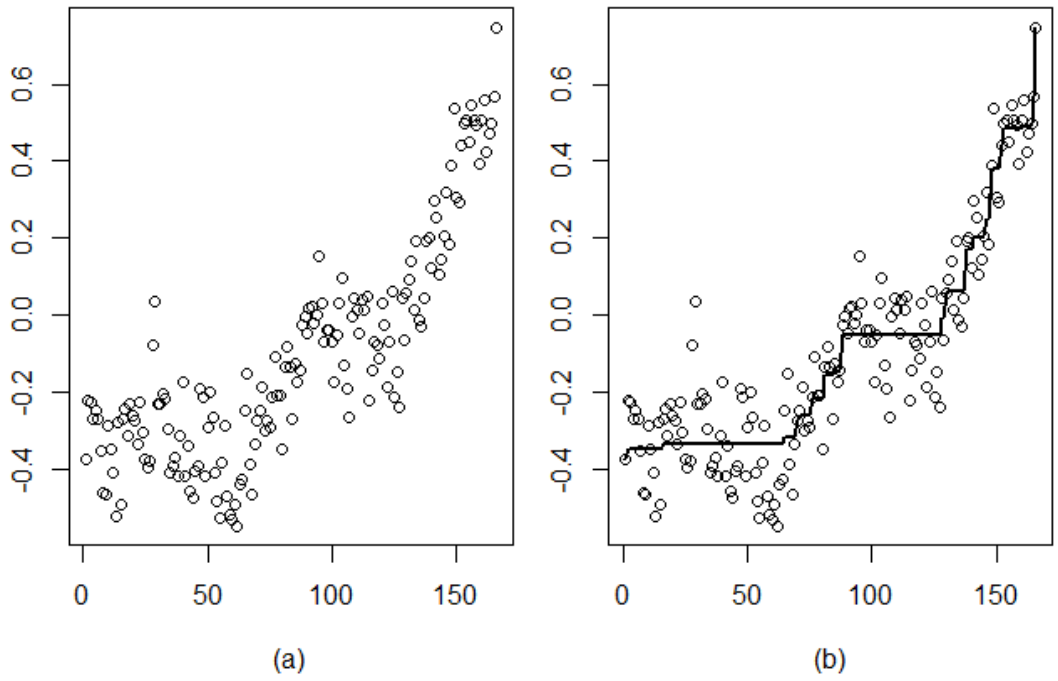

Figure 1.1: (a) Plot of mean annual temperatures with time, based on the global warming data. (b) The plot in (a) with the isotonic regression estimator.

1.2.1 Asymptotic properties of the constrained estimators

Durot (2002) established that when the true regression function f_0 is monotone increasing and differentiable, the isotonic regression estimator $\hat{f}$ converges to f_0 at the rate $n^{-1/3}$ under the L_1-metric. In other words,

$$\int_0^1 |\hat{f}(t) - f_0(t)| dt = O(n^{-1/3}).$$

For $x_0 \in (0,1)$ such that $f'(x_0)$ exists and $f'(x_0) > 0$, Brunk (1970) established that $\hat{f}(x_0)$ is a consistent estimator of $f(x_0)$ and

$$(\hat{f}(x_0) - f_0(x_0)) = O_P(n^{-1/3}).$$

For x_0 an interior point of $[0,1]$ with $f_0'(x_0) > 0$, Brunk (1970) established the asymptotic distribution of the constrained point estimator $\hat{f}(x_0)$. For W_1 a two-sided standard Brownian motion on $\mathbb{R}$ and $Z_1 = \arg\min\{W_1(t) + t^2 : t \in \mathbb{R}\}$, and $C = 2b\,(a/b)^{2/3}$ with $a = \sqrt{\sigma_0^2 / g(x_0)}$

Table 1.2: The number of days Lake Mendota was frozen every year, starting from 1855

118	151	119	94	111	129	133	119	131	117	125	128	108	126	138
100	142	141	136	123	91	132	62	115	93	160	77	127	120	121
131	131	137	87	74	113	96	117	99	104	119	123	101	129	112
118	103	85	140	111	115	106	106	118	96	103	113	104	83	115
111	116	116	83	116	85	96	127	103	103	126	102	108	96	104
98	60	107	103	98	107	123	111	102	102	127	85	114	101	89
99	108	108	92	113	118	112	93	72	83	109	92	109	125	101
98	113	112	92	114	65	88	87	101	92	93	95	90	94	95
90	103	121	117	104	86	100	56	107	92	113	85	91	93	98
88	91	108	95	74	107	96	49	70	52	108	56	83	77	97
94	67	122	104	88	112	69	101	121	105	62	81	93	97	

The order is in increasing years from left to right and (next) row-wise.

and $b = f_0'(x_0)/2$, the asymptotic distribution of $\hat{f}(x_0)$ is given by

$$n^{1/3}(\hat{f}(x_0) - f_0(x_0)) \rightsquigarrow C Z_1.$$

A similar result was established in the context for decreasing densities by Prakasa Rao (1969). The distribution of Z_1 is known as the standard Chernoff distribution. It is symmetric about zero, with tails tending to zero faster than the standard normal distribution. Its quantiles have been tabulated in Groeneboom and Wellner (2001).

1.2.2 Shape-restricted estimation as a 'projection'

The shape-restricted estimator of the parameter of interest in all such cases as described above is obtained by minimizing a certain 'loss function' under the constraint of shape on the parameter. In regression, for instance, the loss function is the error sum of squares, while for monotone density estimation, the loss function is the negative log-likelihood. The PAVA provides an algorithm for constrained estimation under a wide range of convex loss functions; see Section 3.1 of De Leeuw et al. (2009). The constrained estimator of a function f can therefore be viewed as a 'projection' onto the class of the shape-restricted functions, with respect to the 'distance' induced by the corresponding loss function.

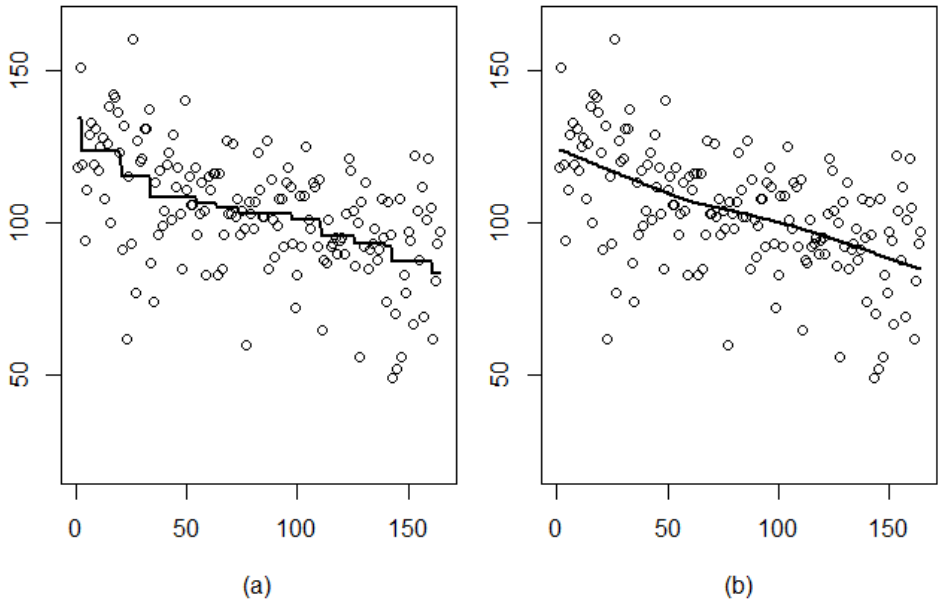

Figure 1.2: Scatterplot and the isotonic regression estimate for the warming trend of Lake Mendota. (a) Raw estimate. (b) A smoothed version of the raw estimate.

1.3 Some background on Bayesian nonparametrics

The application of Bayesian principles on nonparametric estimation and inference problems has gained huge popularity in recent times. Bayesian nonparametrics consists of two main steps: specifying a prior distribution on the infinite-dimensional parameter space, and combining the prior with the likelihood using the Bayes formula to obtain the posterior distribution. The development of Markov Chain Monte Carlo (MCMC) sampling methodologies has popularized the use of Bayesian nonparametrics by providing fast and efficient ways of sampling from the posterior distribution. This subject has enormous applications in engineering, machine learning, genetics, climate science, neuroscience and many more.

There are, however some challenges associated with choosing an appropriate prior when dealing with nonparametric models. Some natural choices of priors may result in inconsistent posteriors that do not concentrate near the true parameter asymptotically. Posterior consistency is not guaranteed simply by the inclusion of the true parameter value in the topological support of the prior. A list of default priors have been introduced in the literature to deal with these issues. These priors are analogous to noninformative priors in the parametric context in the sense that they have very low information content in

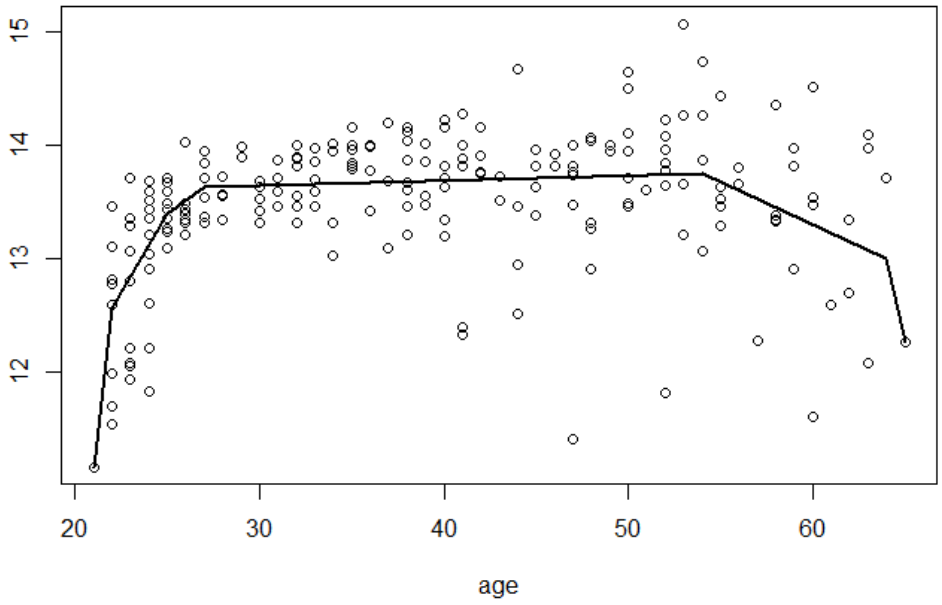

Figure 1.3: Plot of the Canadian income data with the convex regression estimate.

them as compared with the data. The most common methods of prior construction in nonparametric models are

1. finite random series with priors on the coefficients,
2. mixture models,
3. Gaussian processes.

The basis functions used in a finite random series prior can consist of polynomials, trigonometric polynomials, B-splines, wavelets and so on. The number of terms used in the basis expansion is used for controlling the degree of smoothness. The number of terms may be chosen deterministically, guided by asymptotic calculations, or may be chosen using some model selection criterion such as the Akaike Information Criterion (AIC), or may be given a further prior and mixed over different values. Kernel mixture models are popularly used for density estimation; see examples such as Dirichlet mixtures (Ferguson 1983; Lo 1984). Gaussian processes are used to construct priors on different object such as density, regression and classification functions.

1.3.1 Posterior consistency and contraction rates

Posterior consistency is the characteristic of the posterior distribution concentrating near the point mass of the true parameter value when the sample size grows large. In finite dimensional parametric models, posterior consistency holds under mild assumptions, but in nonparametric models, more involved conditions are required to be satisfied for posterior consistency to hold. Schwartz (1965) introduced the concept of Kullback-Leibler neighborhoods and exponentially consistent tests to prove posterior consistency in dominated statistical models. Ghosal et al. (1999) and Barron et al. (1999) provided extensions of Schwartz's results.

Let Θ be a parameter space equipped with the Borel-sigma field $\mathscr{B}$. Let $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ denote the observed data. Let d be a metric defined on Θ. Let $P_\theta^{(n)}$ denote the joint distribution of $\boldsymbol{X}$ indexed by θ. Let $\boldsymbol{X} \sim P_\theta^{(n)}$ and let $\theta_0 \in \Theta$ be the true value of the parameter. Let $P_{\theta_0}^{(n)}$ denote the corresponding true joint distribution. Let $\Pi(\cdot|\boldsymbol{X})$ denote the posterior corresponding to the prior Π on the Borel sets of Θ. The posterior distribution $\Pi(\cdot|\boldsymbol{X})$ is said to be consistent at θ_0 if for every $\epsilon > 0$,

$$\Pi(\theta : d(\theta, \theta_0) > \epsilon | \boldsymbol{X}) \rightarrow^{P_{\theta_0}^{(n)}} 0, \tag{1.1}$$

as $n \rightarrow \infty$. Posterior consistency tells us that as the sample size grows large, the posterior distribution converges to the point mass at θ_0. It also tells us that the posterior distributions obtained using different priors will agree in the sense that their distances in the weak topology will tend to zero, with sample size going to infinity.

The rate at which the posterior distribution converges to the point mass of θ_0 is formalized in what is known as the contraction rate. A sequence ϵ_n of real numbers with $\epsilon_n \rightarrow 0$ is a contraction rate at θ_0 with respect to the metric d if

$$\Pi(\theta : d(\theta, \theta_0) > M_n \epsilon_n \Big| \boldsymbol{X}) \rightarrow^{P_{\theta_0}^{(n)}} 0, \tag{1.2}$$

for every $M_n \rightarrow \infty$. A natural problem is then to obtain the fastest possible rate, i.e. the smallest ϵ_n satisfying (1.2). In many cases, the optimal rate coincides with the minimax estimation rate in the frequentist sense, possibly up to a logarithmic factor. In this dissertation, our parameter space Θ will be the space of regression or density functions, with d the global or empirical L_p-distance between two functions. If the posterior probability in (1.2) has an analytical form through conjugacy, then it can be decomposed using Markov's inequality and contraction rates can be obtained without much difficulty. However, in situations

where an analytic representation is not available, more elaborate steps are needed. The general theory of posterior contraction developed by Ghosal et al. (2000a) gives a set of conditions on the parameter space and the prior distribution, under which (1.2) holds true. The foremost condition is the existence of test functions ϕ_n with exponentially small error for testing

$$H_0 : P = P_{\theta_0}^{(n)}, \;\; H_1 : P = P_{\theta}^{(n)} \text{ for some } \theta : d(\theta, \theta_1) > \epsilon,$$

for some $\epsilon > 0$, $\theta_1 \in \Theta$ and $\boldsymbol{X} \sim P$ for some joint distribution P. The null hypothesis is rejected when $\phi_n = 1$. Such tests are usually constructed by first considering a series of hypothesis tests where the alternative is a neighborhood separated from the null parameter, and then combining these tests to obtain ϕ_n. Combining the tests over the parameter values when Θ is non-compact is more complicated since the number of balls covering the alternative parameter space may be infinite. For this reason, a condition on the size of an appropriate sieve $\Theta_n \subset \Theta$ is imposed so that the tests can be combined over the parameter values belonging to the sieve, giving the desired ϕ_n. This condition controlling the size of Θ_n is given by a condition on the entropy (logarithm of the number of small balls required to cover Θ_n) of the sieve. Another key condition requires that the prior gives sufficient mass to Kullback-Leibler neighborhoods of the true parameter θ_0. Finally, the prior mass of the complement of the sieve Θ_n needs to be small, so that Θ can be substituted by Θ_n in the testing and entropy conditions. An elaborate theory with technical details on the conditions, depending on the dependence structure of $\boldsymbol{X}$ has been provided in Ghosal and van der Vaart (2017).

1.3.2 Credible sets and coverage

For a parameter space Θ, $\theta_0 \in \Theta$ the true parameter value and $\gamma \in [0,1]$, a subset $C(\boldsymbol{X})$ of Θ is a $1-\gamma$ credible set if the posterior mass of $C(\boldsymbol{X})$ is $1-\gamma$. In other words,

$$\Pi(\theta : \theta \in C(\boldsymbol{X}) \,|\boldsymbol{X}) = 1-\gamma. \tag{1.3}$$

The posterior probability of θ falling in $C(\boldsymbol{X})$ is therefore $1-\gamma$, and the spread of the posterior distribution can be used as a margin of error in quantifying the uncertainty in estimating θ. In the finite dimensional parametric setting, some standard methods for obtaining credible sets have been developed. For example, the smallest $C(\boldsymbol{X})$ satisfying (1.3) gives the highest posterior probability density (HPD) credible set. These methods in many parametric models yield credible intervals with the right asymptotic coverage, i.e. as

$n \to \infty$,

$$P_{\theta_0}^{(n)}(\theta_0 \in C(\boldsymbol{X})) \to 1-\gamma, \tag{1.4}$$

in view of the Bernstein-von Mises theorem. Thus in finite-dimensional models, the frequentist and Bayesian methods of uncertainty quantification nearly agree for large sample sizes. The credible intervals therefore asymptotically behave as confidence intervals and give the right coverage. In infinite-dimensional parameter spaces, however, the Bernstein-von Mises theorem may not hold, and it has been shown in various instances that credible sets fail to yield the right asymptotic coverage. Under the Gaussian white noise model, Cox (1993) showed that the asymptotic coverage of Bayesian credible sets at any level may be arbitrarily low. The problem of inadequate coverage was further emphasized by Freedman (1999). It may be noted that both Cox and Freedman drew the true function from the prior itself, and formulated their conclusions by showing the inadequate coverage in the almost sure sense with respect to the prior. The work of Knapik et al. (2011) provided a more detailed picture of this problem by covering inverse problems. They showed that asymptotically zero coverage occurs by an oversmoothing prior over the truth, while by an undersmoothing prior, conservative credible sets can be obtained with the asymptotic coverages equal to 1. This technique was used in the problem of estimating the initial condition of the heat equation in Knapik et al. (2013).

Leahu (2011) and Castillo and Nickl (2013) showed that if the parameter space is extended beyond the ℓ_2-space of sequences, then the credible sets can yield adequate frequentist coverage. Castillo and Nickl (2013) further established parametric posterior convergence rates under negative Sobolev norms, which are weaker than the original ℓ_2-norms. They established a Bernestein-von Mises theorem under this weaker norm and under an appropriate condition on the prior. Based on such a weak Bernstein-von Mises theorem, they constructed credible sets under the weaker norm that gave the right asymptotic coverage for white noise models. Positive coverage results have also been obtained in Sniekers and van der Vaart (2015b) by undersmoothing the prior, and Yoo and Ghosal (2016) using inflated uniform balls. Adaptive credible sets, where the radius adapts to the unknown regularity of θ_0, have been constructed and shown to have adequate frequentist coverage by Sniekers and van der Vaart (2015a), Szabó et al. (2015), Serra and Krivobokova (2017), Ray (2017) and Belitser (2017) in different settings. More relevant work can be found in Belitser and Ghosal (2017), Belitser and Nurushev (2019), and Belitser and Nurushev (2015).

1.4 Bayesian nonparametrics in shape-restricted problems

In spite of the spectacular development of Bayesian theory and methods for function estimation under smoothness regimes, shape-constrained analysis has received much less attention. Bayesian methods are however well suited for shape-restricted problems as the additional qualitative information on shape can be naturally incorporated in the prior. In some situations where the shape constraint on the function leads to a particular representation of the function, Bayesian methods have been used by exploiting that representation, resulting in a posterior distribution complying with the shape restriction because of that representation. For instance, Williamson (1956) showed that a monotone non-increasing density on $[0,\infty)$ can be represented as

$$g(x) = \int_0^\infty \mathbb{1}_{[0,z]}(z) z^{-1} dP(z), \tag{1.5}$$

where P is a probability distribution on $[0,\infty)$ known as the mixture distribution. Using this mixture representation, Salomond (2014) established nearly minimax contraction rates for a decreasing density by using a Dirichlet mixture of uniform distributions as a prior on g. Further, in the same context, Martin (2017) obtained nearly minimax contraction rates using an empirical prior. Contraction rates have been obtained in a few other occasions such as the current status censoring model with a Dirichlet process prior as an illustration of the general theory in Ghosal et al. (2000a) and Salomond (2018) within testing for monotonicity of a regression function. A more general Bayesian framework for analyzing contraction rates in different shape-restricted problems is therefore lacking. Also, point-wise posterior contraction rates and asymptotic representation of the posterior distribution of the function at a specific point, like the Bernstein-von Mises theorem in parametric models have not been addressed in the literature. In this dissertation, we shall focus on some of the fundamental problems encountered while using nonparametric Bayesian methods in shape-restricted problems.

1.5 Introducing the 'projection-posterior'

Let Θ be a parameter space equipped with a distance metric d. Let $\Theta^* \subset \Theta$ be the targeted parameter space embedded in Θ. Suppose that the true parameter generating the data belongs to Θ^*. For a Bayesian analysis on the constrained parameter space, we would need a

prior on Θ^*. Because of the structure of Θ^*, however, the resulting posterior distribution may have a complicated form that may be quite difficult to analyze for frequentist properties. As an alternative approach, consider a prior on Θ, not necessarily concentrated on Θ. The flexibility of a conjugate prior on Θ would result in a posterior distribution easier to analyze and sample from. However, this posterior may not be concentrated on Θ^* where the true parameter is known to lie. To rectify this issue, let us define the 'projection' of $\theta \in \Theta$ with respect to the distance d as the element in Θ^* closest to θ in terms of d. Formally stated, for $\theta \in \Theta$, the projection θ^* is defined as

$$\theta^* = \underset{\eta \in \Theta^*}{\arg\min}\, d(\theta, \eta). \tag{1.6}$$

Note that if $\theta \in \Theta^*$, then by (1.6), $\theta^* = \theta$. Under the mapping $\theta \mapsto \theta^*$, the induced distribution of the posterior of θ is what we would denote by the 'projection-posterior'. In other words, let Π be a prior distribution on Θ, and let $\Pi(\cdot|\boldsymbol{X})$ denote the posterior distribution. With θ^* as defined in (1.6), define the projection-posterior Π^* on Θ^* as

$$\Pi_n^*(\theta \in B) = \Pi(\theta : \theta^* \in B|\boldsymbol{X}), \tag{1.7}$$

for every measurable $B \subset \Theta^*$. As an example, consider the differential equation model by Bhaumik and Ghosal (2015). We have a regression model $\boldsymbol{Y} = \boldsymbol{f}_\theta(t) + \boldsymbol{\varepsilon}$, $\theta \in \Theta \subset \mathbb{R}^p$. The function $\boldsymbol{f}$ is assumed to satisfy a constraint given by the system of ordinary differential equations

$$\frac{d\boldsymbol{f}_\theta(t)}{dt} = \boldsymbol{F}(t, \boldsymbol{f}_\theta(t), \theta),\ t \in [0,1]. \tag{1.8}$$

The function $\boldsymbol{F}$ is a known appropriately smooth function. The parameter θ controls the regression function. As the system of the differential equations may not have a closed form expression, the resulting posterior is hard to compute and analyze. To solve this problem, the regression model is first embedded in a nonparametric model, and a prior is put using a tensor-product of B-splines on the regression function $\boldsymbol{f}$. In the second step, for a sampled value of the regression function from its posterior, the corresponding value of the parameter is obtained by minimizing the distance between the derivative of the $\boldsymbol{f}$ obtained in the first step, and the derivative suggested by the system of differential equations. The procedure of minimizing the distance between the derivatives can be viewed as projecting the first-stage posterior on the constrained space of functions satisfying the system of differential

equations.

Lin and Dunson (2014) proposed a method of Bayesian monotone curve and surface estimation using Gaussian process projections. Samples from the Gaussian posterior were projected on the smooth monotone class of functions. They established posterior consistency and obtained contraction rates in terms of the smoothness of the true function. A similar technique was used by Neelon and Dunson (2004).

1.6 Contraction rates under shape constraints

A natural Bayesian approach to a problem where the underlying function has some shape restriction, is to use a prior on the function complying with the shape constraint, and obtain the corresponding posterior satisfying the shape restriction. The main problem with this approach is, however, that the resulting posterior has a complicated form which makes it very challenging to analyze the distribution for contraction rates. On the other hand, an unrestricted prior will result in a conjugate posterior that may not be supported on the shape-restricted class of functions, yielding a posterior not complying with the shape constraint. A possible way to rectify this problem, the one which we will explore in this dissertation, is using an unrestricted prior on the function space, and analyzing a 'projection-posterior' introduced in the previous section for consistency and contraction rates, instead of the unrestricted posterior. The distance used to introduce the projection-posterior in our context could be the global or empirical L_p-metrics on function spaces, with $1 \leq p < \infty$. In this case, the projection of a step function generated from the unrestricted posterior can be computed using algorithms similar to the PAVA (see Section 3.1 of De Leeuw et al. (2009)). Let θ^* denote the projection of θ with respect to a distance d. Then by the property of projection and the triangle inequality, we have that

$$d(\theta^*, \theta_0) \leq d(\theta^*, \theta) + d(\theta, \theta_0) \leq d(\theta_0, \theta) + d(\theta, \theta_0) = 2d(\theta, \theta_0), \tag{1.9}$$

where θ_0 is the true function lying in the constrained parameter space. Using (1.9), we write (1.2) as

$$\Pi(\theta : d(\theta^*, \theta_0) > M_n \epsilon_n | \boldsymbol{X}) \leq \Pi(\theta : d(\theta, \theta_0) > M_n \epsilon_n / 2 | \boldsymbol{X}). \tag{1.10}$$

From (1.10), we obtain that in situations where the projection is computable by appropriate algorithms and the true function lies in the constrained space, the projection-posterior

inherits the rate of the unrestricted posterior. The unrestricted posterior, in most cases, is easier to analyse for rates, because of the simpler structure resulting from conjugacy.

1.7 Uncertainty quantification in point-wise estimation under shape constraints

With the concept of projection-posterior as a tool for Bayesian estimation of a shape-constrained function, it seems worth exploring if a point estimator resulting from the induced posterior can be used in quantifying uncertainty in estimating the value of the function at a fixed point. Credible sets from the projection-posterior are easy to obtain from posterior sampling and projection using the PAVA, and do not need any bootstrapping or asymptotic distribution. In this dissertation, we obtain a credible set for the value of the shape-restricted function at a fixed point using the projection-posterior in the context of monotone regression and density function. Using concepts from the empirical processes theory, and results on the greatest convex minorant, we establish that the asymptotic coverage of our credible interval has a quantifiable form, and that a re-calibration method will give us a credible interval with the right asymptotic coverage.

Two very interesting implications seem to result from our study. First, we observe that the coverage may be higher than the nominal credibility level, the opposite of the phenomenon observed by Cox (1993) in the context of smooth nonparametric signal estimation. The second, more crucial finding is that both the method of constructing the credible interval and its asymptotic coverage do not involve the constants involving the true function. This is in contrast with the uncertainty quantification using confidence intervals in the frequentist approach, where calculating the margin of error requires estimating the derivative of the function at the fixed point. Our construction of the credible interval thus eliminates the uncertainties resulting from estimating any nuisance parameters.

1.8 Bayesian testing for monotonicity

A traditional Bayesian approach for testing the hypothesis $H_0 : \theta \in \mathscr{F}$ versus $H_1 : \theta \notin \mathscr{F}$, where $\mathscr{F}$ is a class of shape-restricted functions, would be employing the Bayes factor and rejecting H_0 whenever θ falls outside $\mathscr{F}$. This approach, however, has a major limitation as pointed out by Salomond (2018) in that the objects near the boundary of $\mathscr{F}$ are incorrectly classified as not belonging to $\mathscr{F}$, with high posterior probability, when the function has flat

parts. Salomond proposed the idea of 'enlarging' the null hypothesis by a multiple of the contraction rate corresponding to $\theta \in \mathscr{F}$. In other words, he showed that for an appropriate τ_n and a distance d, testing for

$$H_0 : d(\theta, \mathscr{F}) \leq \tau_n \quad \text{versus} \quad H_1 : d(\theta, \mathscr{F}) > \tau_n$$

leads to consistency under H_0 and high power against α-smooth alternatives with an optimal separation rate. Adapting Salomond's idea, we propose a Bayes test for monotonicity of a regression or density function by employing the distance of f from its projection f^* on the constrained space. Our test is shown to be computationally simpler than the test proposed by Salomond and universally consistent. For any fixed $f_0 \in \mathscr{F}$, the Type I error goes to zero, and for any fixed $f_0 \notin \mathscr{F}$, the Type II error goes to zero. Also, our test has high power against smooth alternatives sufficiently separated from $\mathscr{F}_0$.

1.9 Bayesian two-stage monotone regression quantile estimation

Estimating a regression quantile of a monotone increasing regression function can be viewed as an inverse regression problem. In this dissertation, we use the projection-posterior for a regression function to estimate a monotone increasing function. We obtain a credible set for a regression quantile that has the appropriate asymptotic coverage. We then obtain a sampling interval from the posterior of the regression quantile and generate the second-stage observations with the predictors belonging to that sample. By representing the regression function on this interval as a linear function, we use a Gaussian prior on the coefficients and analyze the second-stage posterior distribution of the regression quantile and find that the second stage estimator obtains the parametric $n^{-1/2}$-rate of contraction.

1.10 Notations

In this section, we describe the notations and assumptions used in this dissertation. For two sequences of real numbers a_n and b_n, $a_n \lesssim b_n$ or $a_n = O(b_n)$ means that a_n/b_n is bounded, $a_n \ll b_n$ or $a_n = o(b_n)$ means that $a_n/b_n \to 0$, and $a_n \asymp b_n$ means that $a_n = O(b_n)$ and $b_n = O(a_n)$. For a sequence of random variables Z_n, $Z_n = O_P(a_n)$ means that $\mathrm{P}(|Z_n| > C\,a_n) \to 0$ for some $C > 0$. Let $\mathbb{N} = \{1, 2, \ldots\}$ be the set of positive integers and let $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$.

We say $\boldsymbol{Z} \sim \mathrm{N}_J(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ if $\boldsymbol{Z}$ has a J-dimensional normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Let the space of real-valued monotone increasing functions on $[0,1]$ be denoted by $\mathscr{F}$, and the space of monotone increasing functions on $[0,1]$ bounded in absolute value by $K > 0$ be denoted by $\mathscr{F}(K)$. A real-valued function f on $[0,1]$ is called α-Hölder continuous for some $\alpha \in (0,1)$ if there exists some $L > 0$ such that for all $x, y \in [0,1]$, $|f(x) - f(y)| \leq |x - y|^\alpha$. The set of α-Hölder continuous functions on $[0,1]$ is denoted by $\mathscr{H}(\alpha, L)$. For $T \subset \mathbb{R}$, $L^\infty(T)$ denotes the set of all uniformly bounded real functions on T.

For $1 \leq p < \infty$ and $\boldsymbol{x} \in \mathbb{R}^k$, let $\|\boldsymbol{x}\|_p = (\sum_{l=1}^k |x_l|^p)^{1/p}$ and $\|\boldsymbol{x}\|_\infty = \max_{1 \leq l \leq k} |x_l|$. For G a probability measure on $[0,1]$ and $f : [0,1] \mapsto \mathbb{R}$ satisfying $\int_0^1 |f|^p \, dG < \infty$, let $\|f\|_{p,G} = (\int_0^1 |f|^p \, dG)^{1/p}$. The indicator function of a set U is denoted by $\mathbb{1}_U(\cdot)$.

Let $\mathrm{E}_0(\cdot)$ and $\mathrm{Var}_0(\cdot)$ be the expectation and variance operators taken under the true distribution P_0. We write $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$, $\boldsymbol{X} = (X_1, \ldots, X_n)$, $D_n = (\boldsymbol{Y}, \boldsymbol{X})$, $\boldsymbol{F}_0 = (f_0(X_1), \ldots, f_0(X_n))$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)$. Let $\mathbb{P}_n$ denote the expectation with respect to the empirical measure. Let $p^{(n)}_{f,\sigma}(\boldsymbol{Y}|\boldsymbol{X})$ be the conditional probability density of $\boldsymbol{Y}$ given $\boldsymbol{X}$ under the working assumption of normality of errors with variance σ^2 in (2.0.1). For an $m \times m$ matrix $\boldsymbol{A}_1$, let $\lambda_{\min}(\boldsymbol{A}_1)$ and $\lambda_{\max}(\boldsymbol{A}_1)$ be the smallest and largest eigenvalues respectively. For another $m \times m$ matrix $\boldsymbol{A}_2$, the relation $\boldsymbol{A}_1 \leq \boldsymbol{A}_2$ means that $\boldsymbol{A}_2 - \boldsymbol{A}_1$ is non-negative definite. Let $\boldsymbol{I}_m$ stand for the $m \times m$ identity matrix.

For a random variable Y and a sequence of random variables X_n, $X_n \rightsquigarrow Y$ means that X_n converges in distribution to Y, $X_n \to^P Y$ means that X_n converges to Y in P-probability. For random variables X and Y, $X \stackrel{d}{=} Y$ 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 $\mathscr{N}(\epsilon, A, d)$, is the minimum number of balls of radius ϵ needed to cover A.

For $f : [0,1] \mapsto \mathbb{R}$ and d a distance metric on real valued functions, let the projection of f on $\mathscr{F}$ with respect to a distance d be the function f^* that minimizes $d(f, h)$ over $h \in \mathscr{F}$. Also, let $d(f, \mathscr{F})$ denote $d(f, f^*)$. Let Π^* denote the induced posterior distribution of f^*.

1.11 Chapter organization

In this dissertation, we follow a Bayesian approach to shape-constrained problems by using a finite random step function prior on the function parameterizing the model. The conjugate posterior is projected on the space of monotone functions using the algorithm for isotonic regression. We establish contraction rates for the induced distribution of the projection under different metrics. In regression, when $p = 1$, the error due to approximation using

piece-wise constant functions having J knots is of the order J^{-1} and this rate can be achieved when the knots are equidistant on $[0,1]$. For $p>1$ however, equidistant knots may only give the approximation rate $J^{-1/p}$, but the rate J^{-1} of approximation can be restored by varying the locations of the knots. For the Bayesian method, this will therefore need assigning a prior on the location of the knots, in addition to the prior on the coefficients of the step function. We develop contraction rates under both these settings and achieve the minimax rate $n^{-1/3}$ for $p=1$ and up to a logarithmic factor for $p>1$, under the empirical L_p-distance. For a point-wise confidence interval, we derive the asymptotic distribution of the sieve-maximum likelihood estimator under monotonicity. The asymptotic distribution of the point-wise projection-posterior is then developed, and a credible interval is obtained, whose asymptotic coverage can be analytically characterized. We then use a re-calibration method to obtain a credible interval with the right asymptotic coverage. Similar procedures are performed for monotone density estimation as well. In summary, we demonstrate the properties of Bayesian nonparametric methods through the following examples of shape-restricted problems.

1.11.1 Rates, coverage and tests in Bayesian monotone regression

Using a finite random step function prior on the regression function f with equispaced knots on $[0,1]$ and Gaussian priors on the coefficients, we analyze the projection-posterior to obtain contraction rates in terms of the global and empirical L_1-distances. Using a finite random step function prior on f with an additional prior on the location of the knots, we obtain nearly minimax contraction rates in terms of the empirical L_p-distance. We construct a point-wise projection-posterior credible interval and show that it gives the right asymptotic coverage. We finally develop a universally consistent Bayes test for monotonicity and show that it has optimal power properties. The asymptotic distribution of a constrained sieve-maximum likelihood estimator has been established.

1.11.2 Rates, coverage and tests in Bayesian monotone density estimation

In the setup of a monotone non-increasing density on the positive real line, we use a finite random histogram prior on the density function g with a Dirichlet prior on the coefficients and analyze the projection-posterior for contraction rates in terms of the L_1-distance, to obtain the minimax rate $n^{-1/3}$. We construct a point-wise projection-posterior

credible interval and show that it gives the right asymptotic coverage. We finally develop a universally consistent Bayes test for monotonicity of the density and show that it has optimal power properties. The asymptotic distribution of a constrained sieve-maximum likelihood estimator is also established.

1.11.3 Bayesian two-stage regression quantile estimation with accelerated contraction rate

Using concepts from the projection-posterior credible set for the point-wise regression estimator, we establish a $n^{-1/3}$ contraction rate for and obtain a credible set that has the right asymptotic coverage. We then construct a sampling interval from the first-stage posterior and show that its credibility and coverage tend to 1 asymptotically. Using this sampling interval, we sample the second stage observations with the predictors from this interval. Representing f as a linear function on the interval, we use a Gaussian prior on the coefficients to obtain a posterior distribution of the regression quantile using the second-stage data. We show that the second-stage posterior achieves the parametric $n^{-1/2}$-rate of contraction.

CHAPTER

2

RATES, COVERAGE AND TESTS FOR BAYESIAN MONOTONE REGRESSION

Consider the nonparametric regression model for a response variable Y with respect to a predictor variable $X \in [0,1]$ given by

$$Y = f(X) + \varepsilon, \tag{2.0.1}$$

where f is a monotone increasing function on $[0,1]$ and ε is a mean-zero random error with finite variance σ^2. Let us observe n replications $(Y_1, X_1), \ldots, (Y_n, X_n)$, where the design points $X_1, \ldots, X_n$ are either deterministic or are assumed to have been randomly sampled from a distribution G having a density g which is positive on $[0,1]$. The error is assumed to be distributed independently of the predictors.

A graphical representation of the isotonic regression problem was provided by Barlow and Brunk (1972), in terms of the greatest convex minorant (GCM) of a cumulative sum diagram. The Pool-Adjacent-Violators Algorithm (PAVA) algebraically describes a method of successive approximation to the GCM, and is the most commonly used algorithm for isotonic regression (see Ayer et al. (1955), pages 9–15 and Section 2.3 of Barlow et al. (1972),

and De Leeuw et al. (2009)). The constrained point estimator was shown to converge at a rate $n^{-1/3}$, and its asymptotic distribution was evaluated by Brunk (1970). Durot (2002) studied the global convergence rates of the constrained maximum likelihood estimator of f and established that when the true regression function f_0 is monotone increasing and differentiable, the isotonic regression estimator $\hat{f}$ converges to f_0 at the rate $n^{-1/3}$ under the L_1-metric.

A Bayesian approach to such problems would involve approximating f with a random series function, putting a prior on the coefficients and analyzing the behavior of the posterior distribution for contraction rates and uncertainty quantification. Posterior contraction rates in the nonparametric regression setup have been well studied in the literature, but Bayeisan inference under the monotonicity constraint has received relatively less attention. Shively et al. (2009) followed a Bayesian approach by using a mixture of constrained normal distributions as a prior for the spline coefficients. Bayesian nonparametric methods have been developed for other shape-constrained problems, such as monotone density and current status. Salomond (2014) established the nearly minimax rate $n^{-1/3}$ for a decreasing density using a prior of a mixture of uniform densities.

Testing for monotonicity of a regression function has been studied following various approaches both in frequentist and Bayesian literature; see Akakpo et al. (2014), Hall and Heckman (2000), Baraud et al. (2005), Ghosal et al. (2000b) and Bowman et al. (1998). Salomond (2018) showed that when a monotone f is near the boundary, the traditional Bayes factor approach would often fail to identify it correctly, a problem that can be rectified by enlarging the null hypothesis so that monotone functions near the boundary are correctly classified with high posterior probability.

The rest of the chapter is organized as follows. Section 2.1 states some assumptions and describes the priors. Results on the contraction rates of f and those on the point-wise credible intervals are presented in Sections 2.2 and 2.3 respectively. The Bayesian test for monotonicity is developed in Section 2.4. Simulation results related to point-wise credible intervals and the testing procedures are discussed in Section 3.4. Proofs of the results are provided in Section 2.6 and Section 2.7.

2.1 Assumptions and Priors

For $X_i \in [0,1]$, $i = 1,\ldots,n$, and $I_1,\ldots,I_J$ a partition of $[0,1]$ into J disjoint intervals, let $\boldsymbol{B} = (\!(b_{ij})\!)$ with $b_{ij} = 1$ when X_i belongs to I_j and $b_{ij} = 0$ otherwise. Let G_n denote the

empirical distribution of the design points. Let G be a measure on $[0,1]$ with continuous and positive density g satisfying

$$a_1 < g < a_2,$$

for some $0 < a_1 < a_2 < \infty$. If the design points are fixed, we assume that G_n satisfies

$$\sup_{x\in[0,1]} \Big|G_n([0,x]) - G([0,x])\Big| = o(n^{-1/3}). \tag{2.1.1}$$

For random $\boldsymbol{X}$, we assume that

$$X_i \overset{\text{i.i.d.}}{\sim} G. \tag{2.1.2}$$

With a slight abuse of notation, we shall use G_n and G to denote the corresponding distribution functions as well. We have the following the assumption on the true distribution P_0.

Assumption A. *Under the true distribution P_0, $Y_i = f_0(X_i) + \varepsilon_i$, such that ε_i are i.i.d. sub-Gaussian with mean 0 and variance σ_0^2 for $i = 1,\dots,n$.*

We construct priors on f by representing f as a piece-wise constant function on $[0,1]$ with J the number of pieces. The knots of f may be equally-spaced on $[0,1]$ or may be chosen randomly from the design points $X_1,\dots,X_n$. We represent the model in (2.0.1) as $\boldsymbol{Y} = \boldsymbol{B}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$. We use the following types of priors on f.

1. **Type** 1: For a deterministic J, let $I_j = ((j-1)/J, j/J]$, $1 \le j \le J$. We represent f as $f(\cdot) = \sum_{j=1}^J \theta_j \mathbb{1}_{I_j}(\cdot)$, and use the prior $\boldsymbol{\theta} \sim \mathrm{N}_J(\boldsymbol{\zeta}, \sigma^2\boldsymbol{\Lambda})$ where $\|\boldsymbol{\zeta}\|_\infty$ is bounded, and $\boldsymbol{\Lambda}$ a $J \times J$ diagonal matrix with diagonal entries $\lambda_1^2,\dots,\lambda_J^2$, with $B_1 < \lambda_j < B_2$ for some $B_1, B_2 > 0$.

2. **Type 2**: Using a deterministic J, let $I_j = [\xi_{j-1}, \xi_j)$, $1 \le j \le J$, with $\xi_0 = 0$. We write $f(\cdot) = \sum_{j=1}^J \theta_j \mathbb{1}_{I_j}(\cdot)$, with $\boldsymbol{\xi} = (\xi_1,\dots,\xi_J)$ and $\boldsymbol{\theta}$ having the joint prior

$$\pi(\boldsymbol{\xi}) = \binom{n}{J-1}^{-1} \mathbb{1}\big[\xi_j \in \boldsymbol{X}, \xi_j < \xi_{j+1},\ 1 \le j \le J-1,\ \xi_J = 1\big],$$
$$\boldsymbol{\theta} \sim \mathrm{N}_J(\boldsymbol{\zeta}, \sigma^2\boldsymbol{\Lambda}).$$

   The assumptions on $\boldsymbol{\zeta}$ and $\boldsymbol{\Lambda}$ in Prior 1 carry to Prior 2.

For the Prior 1, the error variance σ^2 may be estimated by maximizing the marginal likelihood of σ or by endowing σ^2 with an inverse-gamma prior with parameters $(\beta_1/2, \beta_2/2)$, where $\beta_1 > 4, \beta_2 > 0$. Observe that $\boldsymbol{Y}|\sigma \sim \mathrm{N}_n\big(\boldsymbol{B\zeta}, \sigma^2(\boldsymbol{B\Lambda B}^T + \boldsymbol{I}_n)\big)$. Maximizing the corresponding marginal log-likelihood with respect to σ^2, we obtain

$$\widehat{\sigma}_n^2 = n^{-1}(\boldsymbol{Y} - \boldsymbol{B\zeta})^T(\boldsymbol{B\Lambda B}^T + \boldsymbol{I}_n)^{-1}(\boldsymbol{Y} - \boldsymbol{B\zeta}). \tag{2.1.3}$$

The plug-in posterior distribution of f is obtained by substituting $\widehat{\sigma}_n$ for σ. In Prior 2, for a pre-fixed large constant S_0, we use the uniform distribution on $(0, S_0)$ as a prior for σ.

2.2 Posterior contraction rates under monotonicity

In order to establish posterior contraction rates for f with unknown σ, we first show that the maximum marginal likelihood estimator for σ^2 in the plug-in Bayes approach or the marginal posterior distribution of σ^2 in the fully Bayes approach, are consistent uniformly for $f_0 \in \mathscr{F}(K)$.

Proposition 2.2.1. *For $f_0 \in \mathscr{F}(K)$, and the prior on f of Type 1, if $J \to \infty$ and $J \ll n$, then under Assumption A on the errors,*

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

(b) If $\sigma^2 \sim \mathrm{IG}(\beta_1/2, \beta_2/2)$ with $\beta_1 > 4, \beta_2 > 0$, then the marginal posterior distribution of σ^2 contracts at the rate $\max\{n^{-1/2}, n^{-1}J\}$.

We impose monotonicity on f by projecting f on $\mathscr{F}$ using the algorithm for isotonic regression. Let Π_n^* be the distribution induced by the posterior of f defined as

$$\Pi_n^*(f \in B) = \Pi(f : f^* \in B | D_n),$$

for every B measurable with respect to Pi. The induced distribution Π_n^*, or the projection-posterior, is easier to handle theoretically than a posterior distribution obtained by a prior complying with the monotonicity constraint. For a monotone f_0 and f^* denoting the projection of f onto $\mathscr{F}$ with respect to a distance d, we have $d(f^*, f) \leq d(f_0, f)$ by the property of projection. From this, and using the triangle inequality, we have that

$$d(f^*, f_0) \leq d(f^*, f) + d(f, f_0) \leq d(f_0, f) + d(f, f_0) = 2d(f, f_0). \tag{2.2.1}$$

The projection f^* corresponding to a distance d therefore inherits the contraction rate of f at a monotone f_0 with respect to d. For $p \geq 1$, the L_p-projection of a step function is computable, by algorithms similar to the PAVA (see Section 3.1 of De Leeuw et al. (2009)).

2.2.1 Contraction rates under the L_1-metric

In the following theorem, we shall derive contraction rates for the projection-posterior when f_0 is monotone increasing, in terms of the L_1-metric.

Theorem 2.2.2. *Let $f_0 \in \mathscr{F}(K)$ and let the prior on f be of Type 1, with $J \to \infty$ and $J \ll n$. Let σ^2 be estimated using the plug-in Bayes approach or endowed with the inverse-gamma prior using a fully Bayes approach. Then under Assumption A on the errors, for $\epsilon_n = \max\{J^{-1}, (J/n)^{1/2}\}$ and every $M_n \to \infty$, the following assertions hold.*

(a) For fixed design points satisfying (2.1.1)*,*

$$\mathrm{E}_0 \, \Pi_n^* \left(\|f - f_0\|_{1,G_n} > M_n \epsilon_n \right) \to 0.$$

(b) For random design points satisfying (2.1.2)*,*

$$\mathrm{E}_0 \, \Pi_n^* \left(\|f - f_0\|_{1,G} > M_n \epsilon_n \right) \to 0.$$

In particular, when $J \asymp n^{1/3}$, the projection-posterior contracts at the minimax rate $\epsilon_n = n^{-1/3}$ in either case.

2.2.2 Contraction rates under the empirical L_p-metric

When the metric under consideration is L_p with $p > 1$, the approximating step function with equal intervals without additional smoothness may have only $J^{-1/p}$ approximation rate. A prior on the knots is therefore required for obtaining the optimal contraction rate. For the case of L_p-metric, we shall assume that the true error variance σ_0^2 is restricted to the interval $(0, S_0^2)$ for some known S_0 large enough. We shall consider deterministic predictors, and derive a posterior contraction rate based on the empirical L_p-metric.

The following result gives a contraction rate for the projection-posterior under the empirical L_p-distance when the prior on f constitutes of a prior on the knots and independent normal prior on the coefficients.

Theorem 2.2.3. *Let $f_0 \in \mathscr{F}(K)$ and the prior on f be of Type 2, with $J \asymp n^{1/3}(\log n)^{t_1}$ for some $t_1 > 0$. Let the prior on σ be the uniform distribution on $(0, S_0)$ for a pre-fixed S_0. Let f^* be the projection of f on $\mathscr{F}$ with respect to the empirical L_p-metric. Then under the additional assumption that the errors ε_i, $i = 1, \ldots, n$, are i.i.d. normal with mean zero and known variance σ_0^2, the posterior distribution of f^* contracts at the rate $\epsilon_n = n^{-1/3}(\log n)^{t_2}$, for $t_2 \geq (1 + t_1)/2$.*

For the empirical L_2-distance, Theorem 2.2.3 also holds true under Assumption A on the error distribution. As the true regression function is correctly specified, the modified neighborhood $\tilde{B}(p_{f^*}, \epsilon_n; f_0)$ in Theorem 8.40 of Ghosal and van der Vaart (2017) reduces to an L_2-ball around f_0, and the assumptions of the theorem follow in an exactly similar manner as derived in the proof of Theorem 2.2.3.

The assumption of σ_0 being restricted to a bounded domain $(0, S_0)$ is fairly mild, and allows one to construct the exponentially consistent tests required in the general theory of posterior contraction, using arguments similar to Lemma 8.27 of Ghosal and van der Vaart (2017). Nearly optimal contraction rates may be possible to obtain by relaxing this assumption, if we use the arguments from the proof of Theorem 4.1 of Ning and Ghosal (2018) to construct the tests, with a prior on σ having sufficiently quickly decreasing tails.

It would be interesting to obtain nearly optimal contraction rates for the continuous L_p-metrics, but we do not know about an appropriate prior that would yield the desired result. In the situation of continuous metrics, the weak approximation with equal intervals produces only a sub-optimal rate $n^{-1/(p+2)}$, while with a prior on the knots, the completely arbitrary nature of the knots for arbitrary monotone f_0 does not allow a good prior concentration needed for the optimal contraction rate.

2.3 Uncertainty quantification in point-wise estimation

Let the data be observed from the model $Y = f(X) + \varepsilon$, where f is a monotone increasing function from $[0, 1]$ to $\mathbb{R}$, bounded in absolute value by some $K > 0$, and the error ε under the working model is Gaussian with mean zero and variance σ^2. Let $x_0 \in (0, 1)$ be such that $f_0'(x_0)$ exists and $f_0'(x_0) > 0$. The isotonic regression estimator of f under the constraint that f is monotone increasing, is obtained by minimizing the sum of squares $\sum_{i=1}^n (Y_i - f(X_i))^2$ subject to f is non-decreasing on $[0, 1]$. Without loss of generality, we assume that $X_1 \leq \cdots \leq X_n$. From the algorithm for isotonic regression as described in Section 2.1 of Groeneboom and Jongbloed (2014), the estimate of $f(x_0)$ under the monotonicity constraint is the value

of the slope (left-derivative) of the greatest convex minorant of the graph of

$$\left\{(0,0),\left(\frac{1}{n},\frac{Y_1}{n}\right),\left(\frac{2}{n},\frac{Y_1+Y_2}{n}\right),\ldots,\left(1,\frac{\sum_{i=1}^n Y_i}{n}\right)\right\},$$

at the point x_0.

Now let f be a monotone increasing piece-wise constant function on $[0,1]$. For $J>1$, let $f=\sum_{j=1}^J q_j \mathbb{1}_{I_j}$, where $(q_1,\ldots,q_J)\in\mathbb{R}^J$ with $q_1\le\cdots\le q_J$. Consider a sieve Θ_J defined as

$$\Theta_J=\left\{f=\sum_{j=1}^J q_j\mathbb{1}_{I_j}: q_1\le\cdots\le q_J,\quad I_j=\left(\frac{j-1}{J},\frac{j}{J}\right],\quad 1\le j\le J\right\}.$$

The sieve-maximum likelihood estimator of f, denoted by $\hat{f}_n$, under the monotonicity constraint on f, is obtained by minimizing the error sum of squares $\sum_{i=1}^n(Y_i-\hat{f}_n(X_i))^2$ as

$$\sum_{j=1}^J\sum_{i:X_i\in I_j}(Y_i-q_j)^2=\sum_{j=1}^J\sum_{i:X_i\in I_j}(Y_i-\bar{Y}_j)^2+\sum_{j=1}^J N_j(\bar{Y}_j-q_j)^2.$$

As the first term is free of q, it is sufficient to minimize the second term $\sum_{j=1}^J N_j(\bar{Y}_j-q_j)^2$ subject to $q_1\le\cdots\le q_J$. This is equivalent to minimizing the weighted sum of squares $\sum_{j=1}^J(N_j/n)(\bar{Y}_j-q_j)^2$ subject to the ordering constraint $q_1\le\cdots\le q_J$. From the algorithm for weighted isotonic regression, we get that the optimal value of q_j is the left-derivative of the greatest convex minorant of the graph of

$$\left\{(0,0),\left(\frac{N_1}{n},\frac{N_1}{n}\bar{Y}_1\right),\ldots,\left(\sum_{k=1}^J\frac{N_k}{n},\sum_{k=1}^J\frac{N_k}{n}\bar{Y}_k\right)\right\},\tag{2.3.1}$$

at the point $\sum_{k=1}^j N_k/n$.

We consider the projection-posterior credible interval for $f(x_0)$ for a Type 1 prior. The posterior samples of f need not be monotone, so we impose monotonicity by projecting each sample onto $\mathscr{F}$ using the isotonization algorithm, to obtain the projection $f^*=\sum_{j=1}^J\theta_j^*\mathbb{1}_{I_j}$ where θ_j^* equals the left-derivative of the greatest convex minorant of the graph of

$$\left\{(0,0),\left(\frac{N_1}{n},\frac{N_1}{n}\theta_1\right),\ldots,\left(\sum_{k=1}^J\frac{N_k}{n},\sum_{k=1}^J\frac{N_k}{n}\theta_k\right)\right\},\tag{2.3.2}$$

at the point $\sum_{k=1}^{j} N_k/n$.

Define the unconditional probability $(\mathrm{P}_0 \times \Pi_n)$ as

$$(\mathrm{P}_0 \times \Pi_n)\big(f : f \in B\big) = \mathrm{E}_0\big(\Pi(f : f \in B|D_n)\big),$$

for every B measurable with respect to Π. The following result gives the unconditional asymptotic distributions of $n^{1/3}(f^*(x_0) - f_0(x_0))$ and $n^{1/3}(\hat{f}_n(x_0) - f_0(x_0))$.

Theorem 2.3.1. *Let the prior on f be of Type 1 with $n^{1/3} \ll J \ll n^{2/3}$. Let W_1, W_2 be independent two-sided Brownian motions on $\mathbb{R}$ with $W_1(0) = W_2(0) = 0$, $Z_2 = \operatorname{argmin}\{W_1(t) + W_2(t) + t^2 : t \in \mathbb{R}\}$, $Z_1 = \operatorname{argmin}\{W_1(t) + t^2 : t \in \mathbb{R}\}$, and $C = 2b\,(a/b)^{2/3}$ with $a = \sqrt{\sigma_0^2/g(x_0)}$ and $b = f_0'(x_0)/2$. Let $\boldsymbol{X}$ be random, satisfying* (2.1.2)*. Then under Assumption A on the errors, the following assertions hold.*

(a) For every $z \in \mathbb{R}$, $\mathrm{P}_0(n^{1/3}(\hat{f}_n(x_0) - f_0(x_0)) \le z) \to \mathrm{P}(CZ_1 \le z)$.

(b) For every $z \in \mathbb{R}$, $\big(\mathrm{P}_0 \times \Pi_n\big)(n^{1/3}(f^*(x_0) - f_0(x_0)) \le z) \to \mathrm{P}(CZ_2 \le z)$.

(c) The conditional process $z \mapsto \Pi(n^{1/3}(f^*(x_0) - \hat{f}_n(x_0)) \le z \,\big|\, D_n)$ does not have a limit in probability.

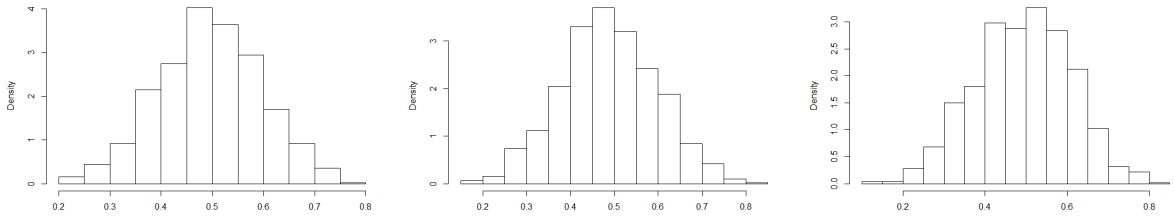

Figure 2.1: Plot demonstrating that $\Pi(n^{1/3}(f^*(x_0) - \hat{f}_n(x_0)) \le 0 \,\big|\, D_n)$ does not have a limit in probability, using three instances of the data.

It follows from part (c) of Theorem 2.3.1 that for $\Delta_n^* = n^{1/3}(f^*(x_0) - \hat{f}_n(x_0))$, and any $z \in \mathbb{R}$, $\Pi(\Delta_n^* \le z \mid D_n)$ cannot converge to a constant (depending on z). This phenomenon is similar to the problem of the inconsistency of the bootstrap estimator as pointed out by Kosorok (2008) and Sen et al. (2010), in the context of monotone density estimation. We illustrate this by drawing samples of $\Pi(\Delta_n^* \le z \mid D_n)$ for 1000 instances of the data,

with $n = 2000$ and $f_0(x) = x^2 + x/5$. Figure 2.1 shows that the limit of $\Pi(\Delta_n^* \le 0 \mid D_n)$ is undoubtedly non-degenerate. It is therefore not possible to construct a credible interval for $f(x_0)$ by an approach along the lines of the classical Bernstein-von Mises theorem.

We shall now evaluate the weak limit of $\Pi(n^{1/3}(f^*(x_0) - f_0(x_0)) \le z \mid D_n)$ and use it to quantify the limiting coverage of credible sets using projected posterior quantiles. As defined in Theorem 2.3.1, let W_1, W_2 be independent two-sided standard Brownian motions starting at zero, $a = \sqrt{\sigma_0^2/g(x_0)}$, and $b = f_0'(x_0)/2$. Define stochastic processes F_n^* and F^* on $\mathbb{R}$ as

$$\begin{aligned} F_n^*(z|D_n) &= \Pi\big(n^{1/3}(f^*(x_0) - f_0(x_0)) \le z \,|D_n\big), \\ F_{a,b}^*(z|W_1) &= \mathrm{P}\left(2b(a/b)^{2/3} \operatorname*{arg\,min}_{t\in\mathbb{R}} V(t) \le z \,|W_1\right), \end{aligned} \tag{2.3.3}$$

where $V(t) = W_1(t) + W_2(t) + t^2$. For every $n \ge 1$, $\gamma \in [0,1]$, define

$$\begin{aligned} &Q_{n,\gamma} = \inf\big\{z \in \mathbb{R} : \Pi(f^*(x_0) \le z|D_n) \ge 1-\gamma\big\}, \\ &I_{n,\gamma} = [Q_{n,1-\frac{\gamma}{2}}, Q_{n,\frac{\gamma}{2}}], \quad \Delta_{W_1,W_2}^* = \operatorname*{arg\,min}_{t\in\mathbb{R}} \big\{W_1(t) + W_2(t) + t^2\big\}. \end{aligned} \tag{2.3.4}$$

We then have the following result.

Theorem 2.3.2. *Let $F_n^*(\cdot)$, $F_{a,b}^*(\cdot)$, $Q_{n,\gamma}$, $I_{n,\gamma}$ and Δ_{W_1,W_2}^* be as defined in* (2.3.3)–(4.2.1)*, with $a = \sqrt{\sigma_0^2/g(x_0)}$ and $b = f_0'(x_0)/2$. Then for $\boldsymbol{X}$ satisfying* (2.1.2) *and under Assumption A on the errors,*

(a) for every $z \in \mathbb{R}$, $F_n^*(z|D_n) \rightsquigarrow F_{a,b}^*(z|W_1)$;

(b) the distribution of $F_{a,b}^*(0|W_1)$ is symmetric about $1/2$;

(c) the limiting coverage of $I_{n,\gamma}$ is characterized as follows:

$$\mathrm{P}_0\big(f_0(x_0) \in I_{n,\gamma}\big) \to \mathrm{P}\left(\frac{\gamma}{2} \le \mathrm{P}(\Delta_{W_1,W_2}^* \ge 0 \,\big|\, W_1) \le 1 - \frac{\gamma}{2}\right) = \mathrm{P}\left(\frac{\gamma}{2} \le F_{a,b}^*(0|W_1) \le 1 - \frac{\gamma}{2}\right).$$

For every $u \in [0,1]$, define

$$A(u) = \mathrm{P}(\mathrm{P}(\Delta_{W_1,W_2}^* \ge 0 \,\big|\, W_1) \le u). \tag{2.3.5}$$

Also, for every $v \in [0,1]$, define $A^{-1}(v)$ as the solution to $A(u) = v$. For a one-sided credible interval of the form $(-\infty, Q_{n,\gamma}]$ for $\gamma \in [0,1]$, the credibilty is $(1-\gamma)$, and the obtained asymptotic coverage is $A(1-\gamma)$. Now, if we want a one-sided credible interval with the

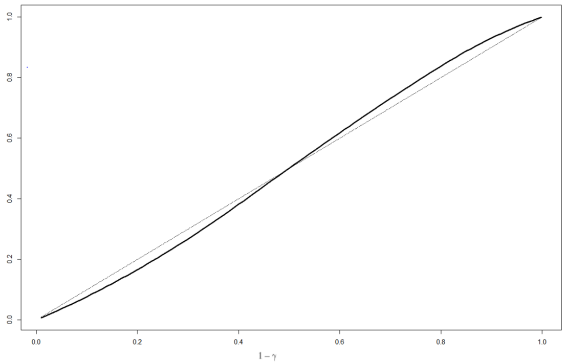

Figure 2.2: Obtained coverage $A(1-\gamma)$ versus the chosen credibilty $(1-\gamma)$. The dotted line denotes $x = y$.

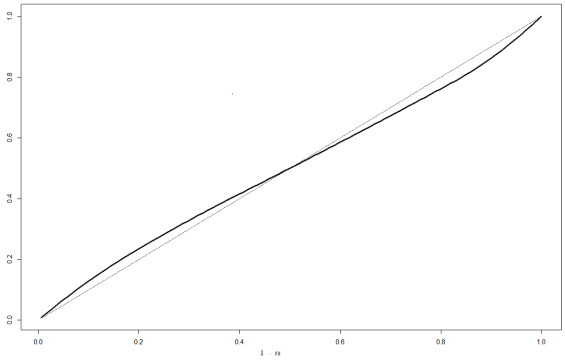

Figure 2.3: Required credibilty $A^{-1}(1-\gamma)$ versus target coverage $(1-\gamma)$. The dotted line denotes $x = y$.

Table 2.1: Comparison of (i) the limiting coverage $A(1-\gamma)$ versus the credibilty $(1-\gamma)$ (ii) the required credibiltyt $A^{-1}(1-\gamma)$ versus the target coverage $(1-\gamma)$ of a one-sided credible interval.

$1-\gamma$	0.500	0.550	0.600	0.650	0.700	0.750	0.800	0.850	0.900	0.910	0.920	0.930	0.940
$A(1-\gamma)$	0.494	0.553	0.612	0.669	0.724	0.779	0.830	0.878	0.923	0.931	0.940	0.948	0.956
$1-\gamma$	0.950	0.960	0.965	0.970	0.975	0.980	0.985	0.990	0.995	0.996	0.997	0.998	0.999
$A(1-\gamma)$	0.964	0.970	0.974	0.979	0.982	0.986	0.990	0.993	0.997	0.997	0.998	0.999	0.999

$1-\gamma$	0.500	0.550	0.600	0.650	0.700	0.750	0.800	0.850	0.900	0.910	0.920	0.930	0.940
$A^{-1}(1-\gamma)$	0.505	0.548	0.590	0.633	0.677	0.723	0.771	0.820	0.874	0.885	0.897	0.909	0.922
$1-\gamma$	0.950	0.960	0.965	0.970	0.975	0.980	0.985	0.990	0.995	0.996	0.997	0.998	0.999
$A^{-1}(1-\gamma)$	0.934	0.946	0.952	0.960	0.966	0.973	0.980	0.986	0.993	0.995	0.996	0.997	0.999

target coverage $(1-\gamma)$, then the credibilty of this interval would need to be $A^{-1}(1-\gamma)$. The corresponding credible interval would be given by $(-\infty, Q_{n,1-A^{-1}(1-\gamma)}]$.

For a two-sided credible interval, part (c) of Theorem 2.3.2 gives us that the asymptotic coverage of $[Q_{n,1-\gamma/2}, Q_{n,\gamma/2}]$ is equal to

$$\begin{aligned}
&\mathrm{P}\left(\frac{\gamma}{2} \leq \mathrm{P}(\Delta^*_{W_1,W_2} \geq 0 \Big| W_1) \leq 1-\frac{\gamma}{2}\right) \\
&\qquad = \mathrm{P}\left(\mathrm{P}(\Delta^*_{W_1,W_2} \geq 0 \Big| W_1) \leq 1-\frac{\gamma}{2}\right) - \mathrm{P}\left(\mathrm{P}(\Delta^*_{W_1,W_2} \geq 0 \Big| W_1) \leq \frac{\gamma}{2}\right) \\
&\qquad = 1-\mathrm{P}\left(\mathrm{P}(\Delta^*_{W_1,W_2} \geq 0 \Big| W_1) > 1-\frac{\gamma}{2}\right) - \mathrm{P}\left(\mathrm{P}(\Delta^*_{W_1,W_2} \geq 0 \Big| W_1) \leq \frac{\gamma}{2}\right).
\end{aligned}$$

By part (b) of Theorem 2.3.2, $\mathrm{P}(\Delta^*_{W_1,W_2} \geq 0 \Big| W_1)$ is symmetric about $1/2$, and therefore the asymptotic coverage of $[Q_{n,1-\gamma/2}, Q_{n,\gamma/2}]$ is equal to

$$1-2\mathrm{P}\left(\mathrm{P}(\Delta^*_{W_1,W_2} \geq 0 \Big| W_1) \leq \frac{\gamma}{2}\right) = 1-2A(\gamma/2).$$

Therefore the obtained coverage of a two-sided $(1-\gamma)$-credible interval is asymptotically equal to $1-2A(\gamma/2)$, for $\gamma \in [0,1]$. For obtaining a two-sided credible interval with asymptotic coverage $(1-\gamma)$, we equate the obtained coverage to $(1-\gamma)$ and solve the equation

$$1-2A(\alpha/2) = 1-\gamma.$$

Solving the above equation in α, we obtain that $\alpha = 2A^{-1}(\gamma/2)$. Therefore the required credibilty of a two sided credible interval with coverage $(1-\gamma)$ would be $1-2A^{-1}(\gamma/2)$. This credible interval would be given by $[Q_{n,1-A^{-1}(\gamma/2)}, Q_{n,A^{-1}(\gamma/2)}]$. The following corollary to Theorem 2.3.2 shows that the asymptotic coverage of this re-calibrated credible interval is

equal to the pre-specified $(1-\gamma)$.

Corollary 2.3.3. *For $\gamma \in [0,1]$,*

$$\begin{aligned} \mathrm{P}_0\left(Q_{n,1-A^{-1}(\frac{\gamma}{2})} \le f_0(x_0) \le Q_{n,A^{-1}(\frac{\gamma}{2})}\right) &\to \mathrm{P}\left(A^{-1}\left(\frac{\gamma}{2}\right) \le \mathrm{P}(\Delta^*_{W_1,W_2} \ge 0 \Big| W_1) \le 1-A^{-1}\left(\frac{\gamma}{2}\right)\right) \\ &= 1-2A\left(A^{-1}\left(\frac{\gamma}{2}\right)\right) \\ &= 1-\gamma. \end{aligned}$$

By simulating Brownian motions, we evaluate $A(1-\gamma)$ and $A^{-1}(1-\gamma)$ over a fine grid of values of γ on $[0,1]$; see Figure 2.3 and Table 2.1 for the comparisons. An interesting phenomenon we observe here is that $A(1-\gamma)$ is bigger than $1-\gamma$ when $1-\gamma$ is greater than 0.5. Because of this, the credible interval $I_{n,\gamma}$ has limiting coverage more than the nominal level $1-\gamma$, for all reasonably small values of γ. Note that $\mathrm{P}(\Delta^*_{W_1,W_2} \ge 0 \Big| W_1) = 0.5$ by part (b) of Theorem 2.3.2, because $\Delta^*_{W_1,W_2}$ is symmetrically distributed about zero.

2.4 Bayesian testing for monotonicity of f

We employ the distance of f from its projection f^* on $\mathscr{F}$ in testing whether f is monotone increasing. Let d denote the global L_1-metric. Instead of looking at the proportion of times f is monotone increasing, we consider testing for an enlarged null hypothesis $H_0 : d(f,\mathscr{F}) \le \tau$ against $H_1 : d(f,\mathscr{F}) > \tau$. For a $\gamma > 0$ and a sequence of positive real numbers τ_n, let ϕ_n be defined as

$$\phi_n = \begin{cases} 1, & \Pi(d(f,\mathscr{F}) \le \tau_n | D_n) < \gamma, \\ 0, & \text{otherwise.} \end{cases} \tag{2.4.1}$$

The following lemma establishes that for appropriate choices of τ_n and J, the test ϕ_n satisfies certain consistency properties.

Theorem 2.4.1. *Let $\tau_n = M_0 n^{-1/3}$ for some predefined constant M_0. Let the prior on f be of Type 1 with $J \asymp n^{1/3}$, and let ϕ_n be the test described in* (2.4.1)*. Let σ^2 be estimated using* (4.1.2) *or $\sigma^2 \sim \mathrm{IG}(\beta_1/2, \beta_2/2)$ with $\beta_1 > 4, \beta_2 > 0$. Then the following assertions hold.*

(a) (Consistency under H_0): For any fixed $f_0 \in \mathscr{F}$, $\mathrm{E}_0\phi_n \to 0$.

(b) (Universal Consistency): For any fixed f_0 integrable on $[0,1]$ and $f_0 \notin \mathscr{F}$, $\mathrm{E}_0(1-\phi_n) \to 0$.

(c) (High power with sub-optimal separation rate): For $f_0 \notin \mathscr{F}$, $f_0 \in \mathscr{H}(\alpha, L)$, and $\rho_n(\alpha) = n^{-\alpha/3}(\log n)^T$ for any $T > 0$,

$$\sup\{\mathrm{E}_0(1-\phi_n) : d(f_0, \mathscr{F}) > \rho_n(\alpha)\} \to 0.$$

The procedure involving the test ϕ_n is computationally simple as it does not involve a prior on J. The algorithm for median isotonic regression (see Robertson and Wright (1973) and De Leeuw et al. (2009)) allows us to compute $d(f, \mathscr{F})$. However, with a deterministic choice of J, the posterior contraction is not adaptive on classes of functions with different smoothness α, and therefore an order of separation $n^{-\alpha/3}$ (up to a logarithmic factor), which is larger than the optimal order $n^{-\alpha/(1+2\alpha)}$, is required to make the power approach one. Adaptation can however be restored by using a prior on J and letting τ_n depend on J as in Salomond (2018). The following prior will be employed for the test.

Type 3: We partition the unit interval into J intervals of equal lengths, the prior on J satisfying for sufficiently large j,

$$\exp(-b_1 j(\log j)^{t_1}) \leq \Pi(J = j) \leq \Pi(J \geq j) \leq \exp(-b_2 j(\log j)^{t_2}), \tag{2.4.2}$$

for some $b_1, b_2 > 0$ and $t_1 \geq t_2 \geq 0$. The prior on $\boldsymbol{\theta}$ given J is the same as in Type 1. Now let $\tau_{n,J}$ be a multiple of $J((\log n)/n)^{1/2}$, and let ψ_n be the test ϕ_n, with τ_n replaced with $\tau_{n,J}$ and d the empirical L_1-metric. The following lemma shows that ψ_n has high power under the optimal separation rate.

Lemma 2.4.2. *Let $\tau_{n,J} = M_0(J(\log n)/n)^{1/2}$ for some $M_0 > 0$. Let the prior on f be of Type 3, and $\sigma^2 \sim \mathrm{IG}(\beta_1/2, \beta_2/2)$ with $\beta_1 > 4, \beta_2 > 0$. Then the following assertions hold for the test ψ_n.*

(a) For any fixed $f_0 \in \mathscr{F}$, $\mathrm{E}_0\phi_n \to 0$.

(b) For $f_0 \notin \mathscr{F}$, $f_0 \in \mathscr{H}(\alpha, L)$, and $\rho_n(\alpha)$ equal to $n^{-\alpha/(1+2\alpha)}(\log n)^{t'}$ where $t' > 2t_1 + 4t_2/3$, with t_1, t_2 as defined in (2.4.2),

$$\sup\{\mathrm{E}_0(1-\phi_n) : d(f_0, \mathscr{F}) > \rho_n(\alpha)\} \to 0.$$

As a summary of our results in this section, the consistency properties of our tests ϕ_n and ψ_n compare with those of Salomond's test through the following findings.

1. Consistency under $f_0 \in \mathscr{F}$ holds in all the tests.

2. We have proved the universal consistency property for ϕ_n, whereas this property has not been established for Salomond's test. Using our approach to proving the consistency, we found that because of the choice of L_∞-distance in Salomond's test for obtaining $d(f, \mathscr{F})$, consistency is not guaranteed when f_0 is a fixed non-monotone function.

3. The test ϕ_n is computationally simpler than Salomond's test as it does not involve a prior on J. The trade-off is the sub-optimal rate of separation, which is rectified by using a prior on J as in ψ_n.

2.5 Simulation

2.5.1 Performance of the projection-posterior

We first demonstrate the validity of our projection-posterior approach for estimating f. We show that the projection-posterior is a much more reliable estimator and has a lower spread, than the corresponding posterior distribution resulting from a truncated normal prior on f satisfying monotonicity. We consider the following three cases of the true functions:

1. $h_1(x) = x^2 + x/5$,
2. $h_2(x) = \frac{e^{4(x-0.5)}}{1+e^{4(x-0.5)}}$,
3. $h_3(x) = (-4.5 + 10x)\mathbb{1}\{x \in [0.45, 0.55]\}$.

For each function, we use two instances of the data. We also compare the two types of posteriors on the Global Warming data, and observe that the projection-posterior fits the data better and has lower spread.

2.5.2 Point-wise uncertainty quantification

We study the behavior of point-wise credible intervals and tests for monotonicity in finite samples. We first compare finite sample performance of Bayesian credible intervals with confidence intervals in terms of their coverages of $f_0(x_0)$, using $f_0(x) = x^2 + x/5$ for $x \in [0,1]$, $x_0 = 0.5$, G the uniform distribution on $[0,1]$, and $\sigma_0 = 1$. We take J to be the greatest integer less than or equal to $n^{1/3} \log n$. We vary sample size n across four different values. For each n, we consider 1000 instances of the data D_n. For each instance, we generate 1000

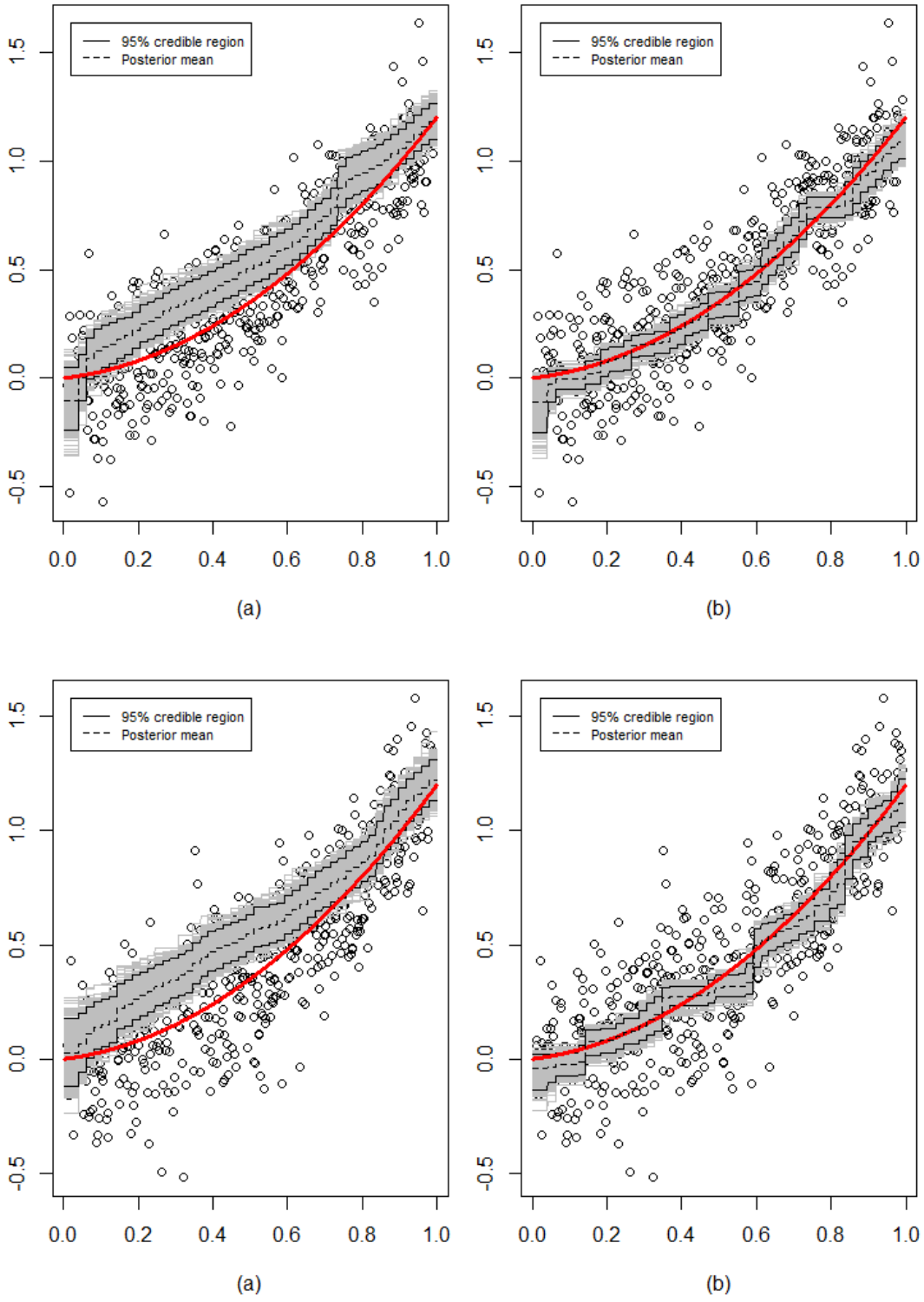

Figure 2.4: (a) Samples from a truncated normal posterior. (b) Samples from a projection-posterior. The red line denotes the true function. $f_0 = h_1$.

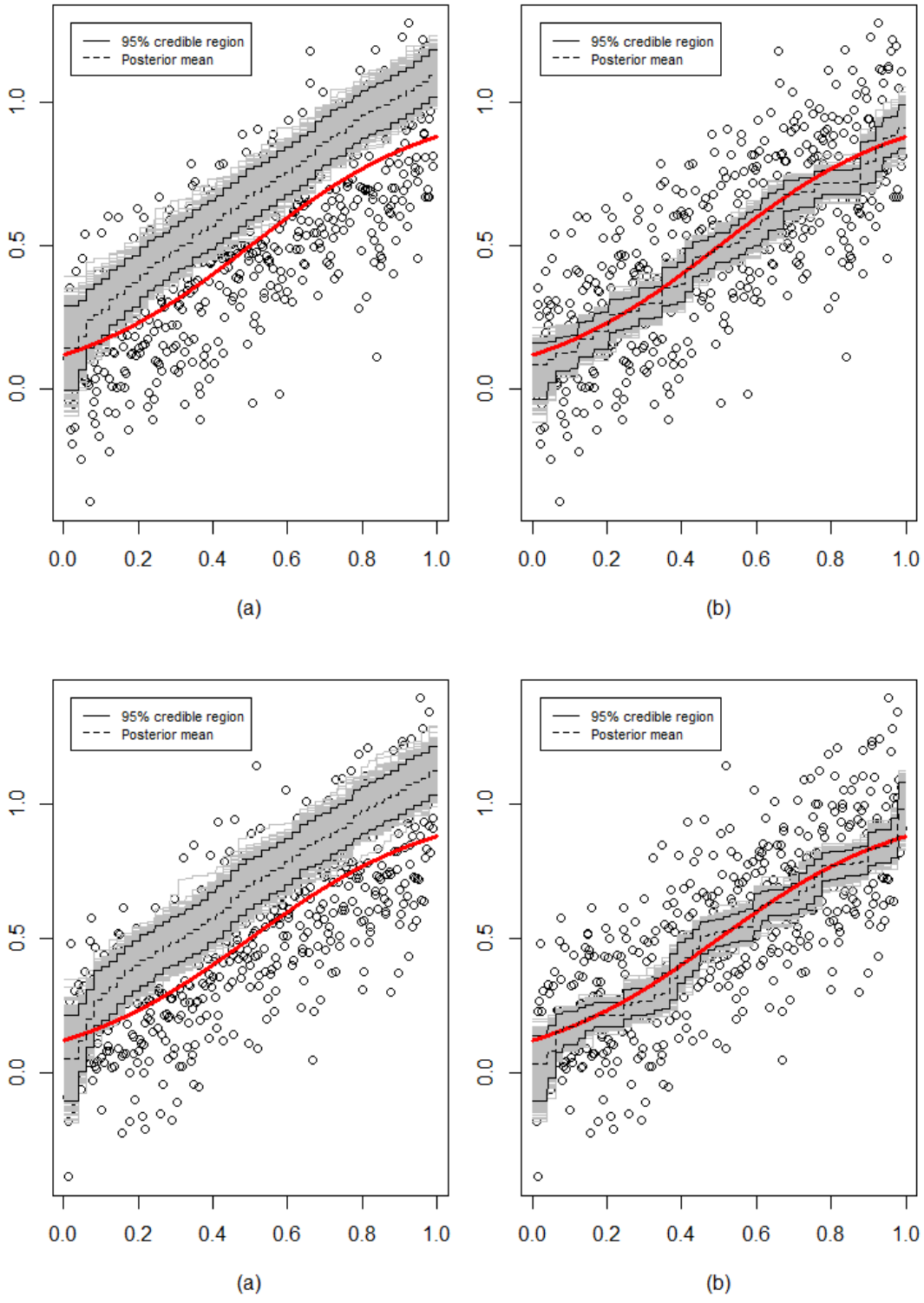

Figure 2.5: (a) Samples from a truncated normal posterior. (b) Samples from a projection-posterior. The red line denotes the true function. $f_0 = h_2$.

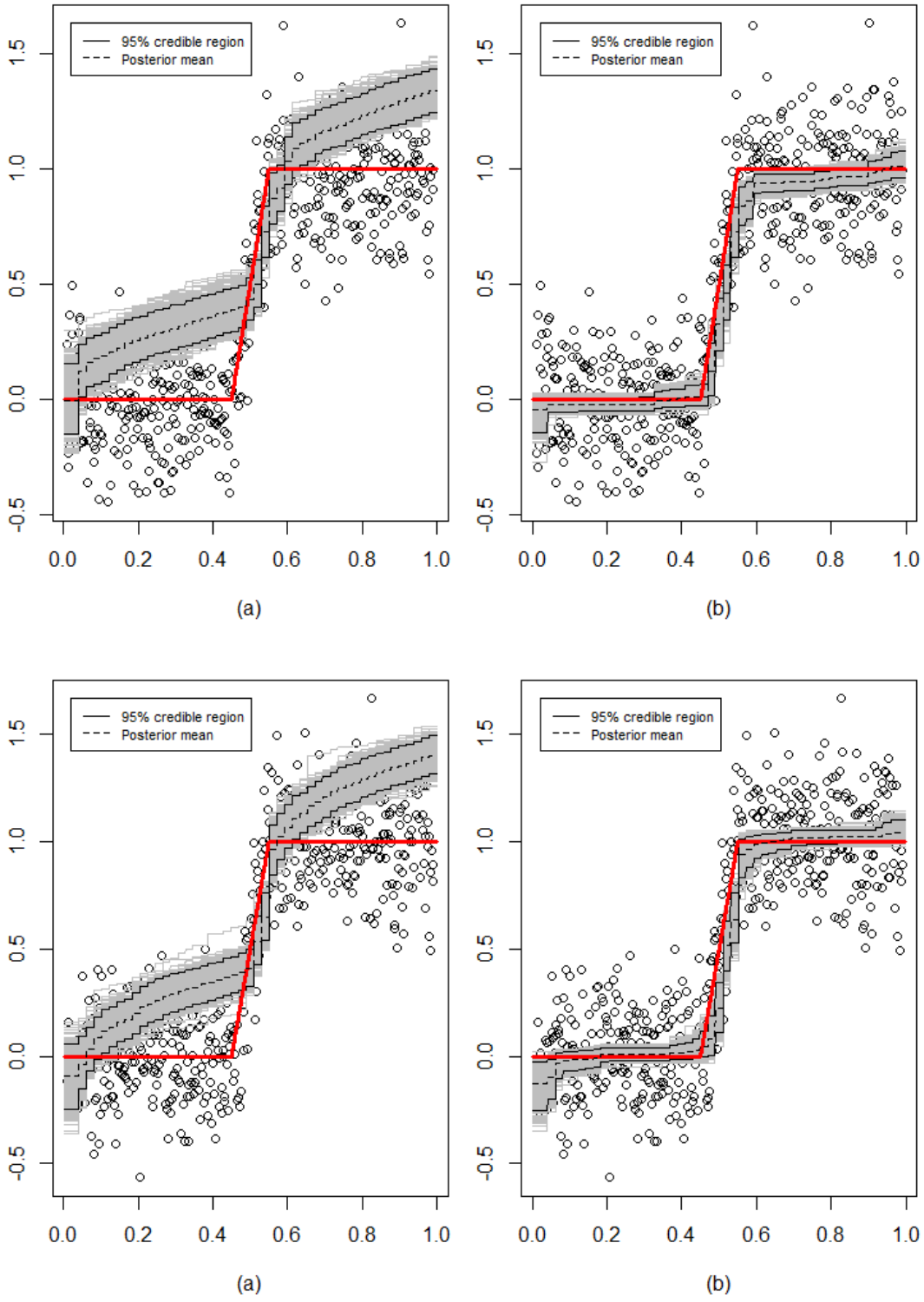

Figure 2.6: (a) Samples from a truncated normal posterior. (b) Samples from a projection-posterior. The red line denotes the true function. $f_0 = h_3$.

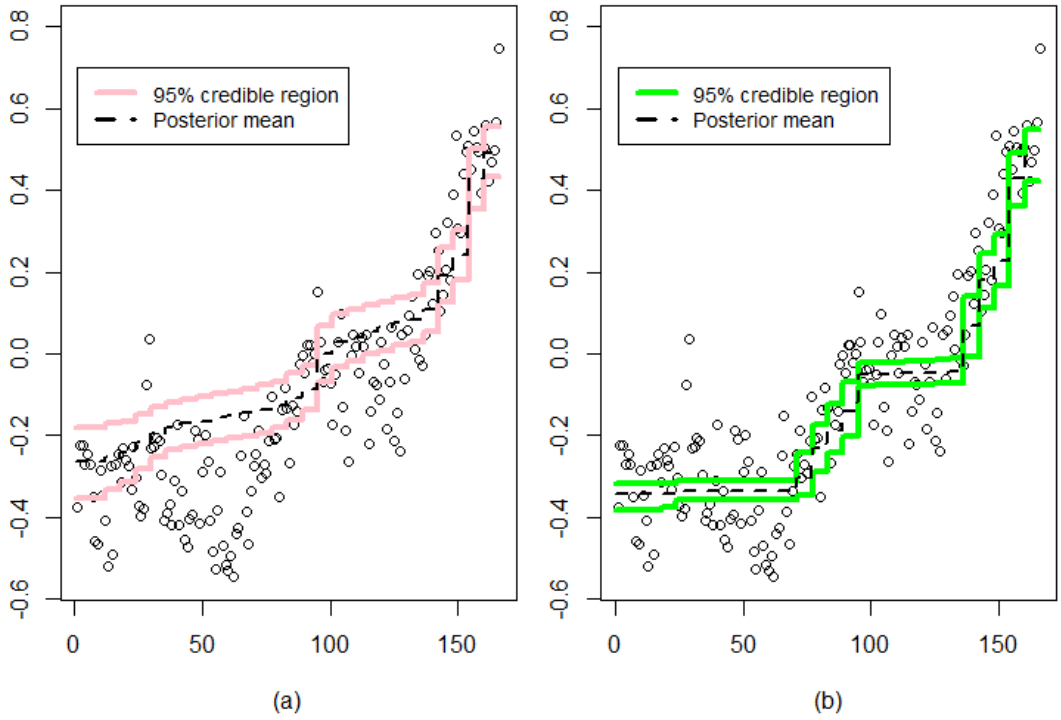

Figure 2.7: Global Warming dataset: 95% credible regions. (a) Credible region from a truncated normal posterior. (b) Credible region from a projection-posterior.

posterior samples of $\boldsymbol{\theta}$ and obtain the corresponding projection f^* from $\boldsymbol{\theta}$ by the algorithm for isotonic regression. We also compute $\hat{f}_n(x_0)$ for every replication. We then obtain the recalibrated $100(1-\alpha)\%$ credible interval using the $[A^{-1}(1-\alpha)]/2$ and $1-[A^{-1}(1-\alpha)]/2$ quantiles of the projected posterior (see Table 2.1 for the values of $A^{-1}(1-\alpha)$), and we check if it contains $f_0(x_0)$. The coverage is determined by the proportion of times the credible interval contains $f_0(x_0)$. We also compute the $100(1-\alpha)\%$ confidence intervals from $\hat{f}_n(x_0)$ based on the quantiles of Chernoff's distribution and the estimated value of the constant C that appears in Theorem 2.3.1. We also obtain the $100(1-\alpha)\%$ credible interval using the $\alpha/2$ and $1-\alpha/2$ projection posterior quantiles. In the results, $C_B(\alpha)$, $C_B^*(\alpha)$ and $C_F(\alpha)$ respectively denote the coverage of $(Q_{n,1-\alpha/2}, Q_{n,\alpha/2})$, $(Q_{n,1-A^{-1}(\alpha/2)}, Q_{n,A^{-1}(\alpha/2)})$, and that of level $(1-\alpha)$ confidence interval, and $L_B(\alpha)$, $L_B^*(\alpha)$ and $L_F(\alpha)$ respectively denote their lengths.

We observe that the in most cases, the obtained coverage is higher than the nominal credibility level, but the average length of the interval is smaller than that of the confidence interval as obtained from the sieve-MLE. The higher coverage of the credible intervals happens as a result of the reverse-Cox phenomenon, as explained in Theorem 2.3.2. However, the coverage of the credible interval obtained from the corrected credibility levels is seen to have fallen slightly below the nominal credibility in a few cases. This can be explained by the error in tabulating the values of $A(\gamma)$ (and hence $A^{-1}(\gamma)$) versus γ that might have happened because of simulating Brownian motions. A more precise calculation of the function $A^{-1}(\gamma)$ is expected to give a credible interval with coverage closer to the desired coverage.

2.5.3 Simulations for testing monotonicity

Following examples from Ghosal et al. (2000b) and Salomond (2018), we consider the following functions on $[0,1]$ for testing monotonicity of f:

$$
\begin{aligned}
f_1(x) &= 0,\\
f_2(x) &= 0.5x^2,\\
f_3(x) &= x(1-x),\\
f_4(x) &= 10(x-0.5)^3 - \exp\left[-100(x-0.25)^2\right]\mathbb{1}\{x<0.5\}\\
&\qquad + 0.1(x-0.5) - \exp\left[-100(x-0.25)^2\right],\\
f_5(x) &= 0.5\cos(6\pi x),\\
f_6(x) &= 1 + x - 0.45e^{-50(x-0.5)^2}.
\end{aligned}
$$

Table 2.2: Comparison of obtained coverage and average length of Bayesian credible intervals for $f(x_0)$, with that obtained from a confidence interval.

$1-\alpha$	n=500						n=1000					
	$C_B(\alpha)$	$L_B(\alpha)$	$C_B^*(\alpha)$	$L_B^*(\alpha)$	$C_F(\alpha)$	$L_F(\alpha)$	$C_B(\alpha)$	$L_B(\alpha)$	$C_B^*(\alpha)$	$L_B^*(\alpha)$	$C_F(\alpha)$	$L_F(\alpha)$
0.99	0.994	0.48	0.983	0.43	0.992	0.54	0.996	0.39	0.991	0.35	0.986	0.43
0.98	0.985	0.44	0.969	0.40	0.985	0.49	0.991	0.36	0.981	0.32	0.979	0.39
0.95	0.958	0.38	0.935	0.35	0.957	0.42	0.967	0.30	0.951	0.28	0.952	0.34
0.92	0.929	0.34	0.913	0.32	0.928	0.38	0.939	0.27	0.929	0.26	0.918	0.30
0.90	0.911	0.32	0.893	0.30	0.907	0.36	0.929	0.26	0.900	0.24	0.891	0.28
0.86	0.876	0.29	0.853	0.26	0.835	0.31	0.883	0.23	0.851	0.21	0.826	0.24
$1-\alpha$	n=1500						n=2000					
	$C_B(\alpha)$	$L_B(\alpha)$	$C_B^*(\alpha)$	$L_B^*(\alpha)$	$C_F(\alpha)$	$L_F(\alpha)$	$C_B(\alpha)$	$L_B(\alpha)$	$C_B^*(\alpha)$	$L_B^*(\alpha)$	$C_F(\alpha)$	$L_F(\alpha)$
0.99	0.994	0.34	0.984	0.31	0.989	0.38	0.996	0.31	0.988	0.28	0.993	0.34
0.98	0.984	0.31	0.979	0.28	0.984	0.34	0.988	0.29	0.976	0.26	0.984	0.31
0.95	0.967	0.27	0.949	0.25	0.953	0.29	0.968	0.25	0.939	0.23	0.951	0.27
0.92	0.939	0.24	0.916	0.23	0.921	0.26	0.932	0.22	0.917	0.21	0.911	0.24
0.90	0.914	0.23	0.894	0.21	0.905	0.25	0.914	0.21	0.895	0.19	0.889	0.23
0.86	0.881	0.2	0.860	0.19	0.842	0.21	0.879	0.19	0.850	0.17	0.828	0.19

The functions f_1 and f_2 are monotone increasing functions, and the rest are non-monotone, with f_6 having a slight departure from monotonicity near 0.3. For different sample sizes n, 1000 replications of the data are generated from f_i, $1 \le i \le 6$, with $\sigma_0 = 0.1$, and σ^2 is estimated by the marginal maximum likelihood estimator. We draw 1000 posterior samples for each instance, and reject the test if the proportion of times $d(f_0, \mathscr{F}) > \tau_n$ is at least 1/2. The results have been tabulated in Tables 2.3–2.5. We observe that the performance of our test is comparable to those of the existing tests.

Table 2.3: Proportion of times the test is rejected: f_1 and f_2.

$f_0 = f_1$	Salomond	ϕ_n
$n = 500$	0.02	0.04
$n = 1000$	0.02	0.01
$f_0 = f_2$	Ghosal et.al.	ϕ_n
$n = 500$	0	0
$n = 1000$	0	0

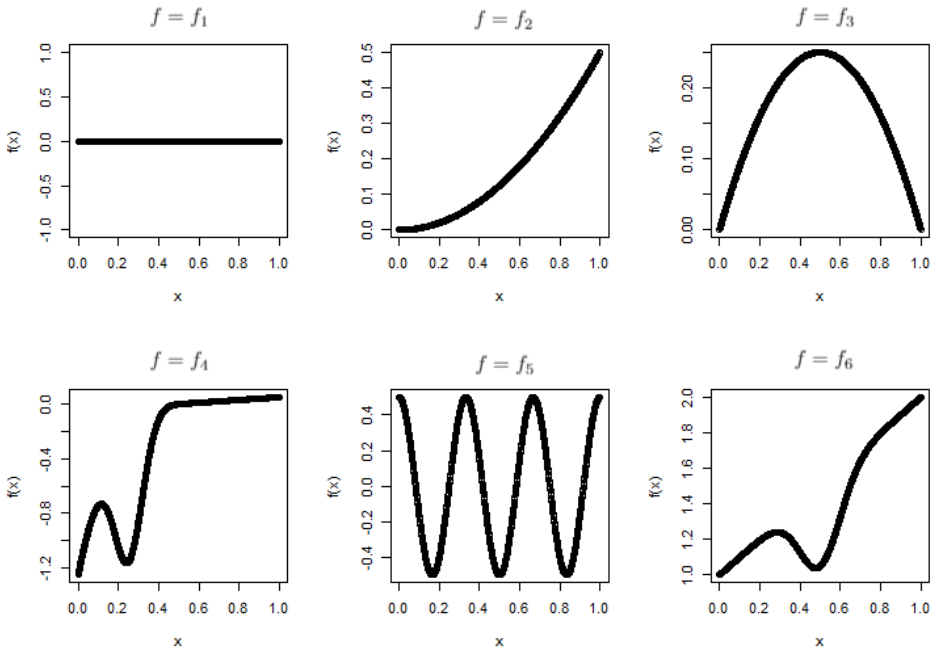

Figure 2.8: Plots of the functions used for testing monotonicity.

Table 2.4: Proportion of times the test is rejected: f_3 and f_4.

$f_0 = f_3$	Ghosal et.al.	ϕ_n
$n = 200$	0.96	1
$n = 500$	1	1
$f_0 = f_4$	Ghosal et.al.	ϕ_n
$n = 200$	0.76	0.43
$n = 500$	1	0.26

2.6 Proofs

The proof of Proposition 2.2.1 uses arguments similar to those used in the proof of Proposition 4.1 of Yoo and Ghosal (2016), and has been provided in Section 2.7.

Proof of Theorem 2.2.2. Let $\mathscr{U}_n$ be a shrinking neighborhood of σ_0 with $\Pi(\sigma \in \mathscr{U}_n \big| D_n) \to 1$. We prove both the parts for the plug-in Bayes case; the fully Bayes case can be dealt with similarly.

(a) For any $j = 1, \ldots, J$, assuming the design points satisfy (2.1.1), we have

$$\max_{1 \le j \le J} \left\{ G_n(j/J) - G_n((j-1)/J) \right\} = \max_{1 \le j \le J} \left\{ G(((j-1)/J, j/J]) \right\} + o(J^{-1})$$

Table 2.5: Proportion of times the test is rejected: f_5 and f_6.

$f_0 = f_5$	Salomond	ϕ_n
$n = 500$	1	1
$n = 1000$	1	1
$f_0 = f_6$	Salomond	ϕ_n
$n = 500$	1	0.93
$n = 1000$	1	1

$$= J^{-1} + o(J^{-1}).$$

Therefore,

$$J^{-1} + o(J^{-1}) \leq \min_{1\leq j\leq J}(N_j/n) \leq \max_{1\leq j\leq J}(N_j/n) \leq J^{-1} + o(J^{-1}). \tag{2.6.1}$$

Now let $f_{0J} = \sum_{j=1}^{J} \theta_{0j} \mathbb{1}_{I_j}$ with $\theta_{0j} = \sum_{i:X_i\in I_j} f_0(X_i)/N_j$ for all $1 \leq j \leq J$. From the monotonicity of f_0 and (2.6.1) we write $\|f_{0J} - f_0\|_{1,G_n}$ as

$$\begin{aligned} n^{-1}\sum_{j=1}^{J}\sum_{i:X_i\in I_j} |\theta_{0j} - f_0(X_i)| &\leq n^{-1}\sum_{j=1}^{J}\sum_{i:X_i\in I_j} |f_0(j/J) - f_0((j-1)/J)| \\ &\leq \max_{1\leq j\leq J}(N_j/n)\sum_{j=1}^{J}(f_0(j/J) - f_0((j-1)/J)) \\ &\lesssim J^{-1}(f_0(1) - f_0(0)). \end{aligned} \tag{2.6.2}$$

Therefore $\|f_{0J} - f_0\|_{1,G_n} \lesssim J^{-1}$, and $\epsilon_n = \max\{J^{-1}, \sqrt{J/n}\} \geq \|f_{0J} - f_0\|_{1,G_n}$. Let $M_n \to \infty$. Using (2.2.1), (2.6.1)–(2.6.2), and Markov's inequality, we bound $\Pi(\|f^* - f_0\|_{1,G_n} > M_n\epsilon_n \big| D_n)$ by

$$\begin{aligned} \Pi(\|f - f_0\|_{1,G_n} > M_n\epsilon_n/2 \big| D_n) &\leq \Pi(\|f - f_{0J}\|_{1,G_n} > M_n\epsilon_n/2 \big| D_n) \\ &= \Pi\left(\sum_{j=1}^{J}(N_j/n)|\theta_j - \theta_{0j}| > M_n\epsilon_n/2 \big| D_n\right) \\ &\leq \frac{2}{M_n\epsilon_n}\sum_{j=1}^{J}(N_j/n)\mathrm{E}(|\theta_j - \theta_{0j}| \big| D_n) \\ &\lesssim M_n^{-1}\epsilon_n^{-1}J^{-1}\sum_{j=1}^{J}\mathrm{E}(|\theta_j - \theta_{0j}| \big| D_n). \end{aligned} \tag{2.6.3}$$

For any $1 \le j \le J$, we bound $\mathrm{E}(|\theta_j - \theta_{0j}| \big| D_n)$ above as $[\mathrm{Var}(\theta_j \big| D_n)]^{1/2} + |\mathrm{E}(\theta_j \big| D_n) - \theta_{0j}|$. Using the fact that $\sigma = O_p(1)$ for $\sigma \in \mathscr{U}_n$, the assumptions on the prior, and (2.6.1), we bound $\mathrm{E}_0[\mathrm{Var}(\theta_j \big| D_n)]^{1/2}$ as

$$\sup_{\sigma \in \mathscr{U}_n} \frac{\sigma}{[N_j + \lambda_j^{-2}]^{1/2}} \lesssim \{\min_{1 \le j \le J} N_j\}^{-1/2} \lesssim J^{1/2} n^{-1/2} + o(J^{1/2} n^{-1/2}). \tag{2.6.4}$$

The last expression goes to zero as $(J/n) \to 0$. We bound $\mathrm{E}_0 |\mathrm{E}(\theta_j \big| D_n) - \theta_{0j}|$ as

$$\begin{aligned}
\mathrm{E}_0 \left| \frac{N_j \bar{Y}_j + \zeta_j / \lambda_j^2}{N_j + 1/\lambda_j^2} - \frac{\sum_{i: X_i \in I_j} f_0(X_i)}{N_j} \right| &\lesssim \mathrm{E}_0 \left| \frac{N_j \bar{Y}_j - \sum_{i: X_i \in I_j} f_0(X_i) + \zeta_j / \lambda_j^2}{N_j} \right| \\
&= \mathrm{E}_0 \left| \frac{\sum_{i: X_i \in I_j} (Y_i - f_0(X_i)) + \zeta_j / \lambda_j^2}{N_j} \right| \\
&\lesssim N_j^{-1} \mathrm{E}_0 \Big| \sum_{i: X_i \in I_j} \varepsilon_i \Big| + N_j^{-1} \\
&\le N_j^{-1} \left[\mathrm{Var}_0 \left(\sum_{i: X_i \in I_j} \varepsilon_i \right) \right]^{1/2} + N_j^{-1}.
\end{aligned} \tag{2.6.5}$$

The last expression is bounded by a constant multiple of $N_j^{-1/2}$ as the number of X_i's in I_j equals N_j. This, using (2.6.1), is $J^{1/2} n^{-1/2} + o(J^{1/2} n^{-1/2})$. Substituting the bounds obtained in (2.6.4)–(2.6.5) in (2.6.3), we obtain that $\mathrm{E}_0 \Pi^*(\|f - f_0\|_{1, G_n} > M_n \epsilon_n \big| D_n) \lesssim M_n^{-1}$, and the proof follows.

(b) For C a positive constant, define the event

$$A_n = \left\{ C n/(2J) \le \min_{1 \le j \le J} N_j \le \max_{1 \le j \le J} N_j \le 3C n/(2J) \right\}. \tag{2.6.6}$$

As $N_j \sim \mathrm{Bin}(n; G(I_j))$ for every $1 \le j \le J$, the large deviation probability $\mathrm{P}(N_j \ge 3C n/(2J))$ is bounded by

$$\exp\left[-n \left(\frac{3C}{2J} \log\left(\frac{3C}{2JG(I_j)} \right) + \left(1 - \frac{3C}{2J} \right) \log\left(\frac{1 - 3C/(2J)}{1 - G(I_j)} \right) \right) \right].$$

Using (2.1.2) and $1 - G(I_j) \le 1$, we obtain that for some $C_1 > 0$ and sufficiently large n, the last expression is bounded by $\exp(-C_1 n/J)$. Therefore, for $1 \le j \le J$, $\mathrm{P}\left(N_j \ge 3C n/(2J)\right) \lesssim$

$\exp(-C_1 n/J)$. Similarly, $\mathrm{P}(N_j \le C n/(2J)) \le \exp(-C_1 n/J)$. Therefore for some $C' > 0$,

$$\mathrm{P}(A_n^c) \le 2J \exp(-C_1 n/J) \lesssim \exp(-C' n/J) \to 0, \tag{2.6.7}$$

as $(n/J) \to \infty$. The rest of the proof is similar to part (a), with $\theta_{0j} = \int_{I_j} f_0 dG/G(I_j)$, $1 \le j \le J$ substituted in (2.6.2), and bounding $\mathrm{E}_0 \Pi(\|f - f_0\|_{1,G} > M_n \epsilon_n \big| D_n)$ by

$$\begin{aligned} &\mathrm{E}_0\big[\Pi(\|f - f_0\|_{1,G} > M_n \epsilon_n \big| D_n)\big| A_n\big]\mathrm{P}(A_n) + P(A_n^c) \\ &\qquad \le \mathrm{E}_0\big[\Pi(\|f - f_0\|_{1,G} > M_n \epsilon_n \big| D_n)\big| A_n\big] + o(1). \end{aligned} \tag{2.6.8}$$

The first term of (2.6.8) is bounded above by a constant multiple of M_n^{-1} by arguments similar to those used in (2.6.3)–(2.6.5), completing the proof. □

Proof of Theorem 2.2.3. We shall use general theory of posterior contraction for establishing the result. We shall verify the assumptions of Theorem B.3.1 for our choice of the sieve $\mathscr{P}_{n1}$ and ϵ_n. The construction of exponentially consistent tests follows by adapting the arguments used in the proof of Lemma 8.27 of Ghosal and van der Vaart (2017) to the L_p-distance setting. By direct calculations, the Kullback-Leibler divergence and the square Kullback-Leibler variation between $p^{(n)}_{f_0,\sigma_0}$ and $p^{(n)}_{f,\sigma_0}$ are respectively equal to

$$\begin{aligned} K(p^{(n)}_{f_0,\sigma_0}; p^{(n)}_{f,\sigma}) &= n\|f - f_0\|^2_{2,G_n}/(2\sigma_0^2) + \big[\log(\sigma_0^2/\sigma^2) + (\sigma^2/\sigma_0^2) - 1\big]/2, \\ V_{2,0}(p^{(n)}_{f_0,\sigma_0}; p^{(n)}_{f,\sigma}) &= n(\sigma/\sigma_0)^2\|f - f_0\|^2_{2,G_n}/\sigma_0^2 + (\sigma^2 - \sigma_0^2)^2/(2\sigma_0^4). \end{aligned}$$

Therefore for any $\epsilon \in (0,1)$, there exists $C > 0$ such that

$$\begin{aligned} B_{n,0}((f_0,\sigma_0),\epsilon) &:= \{(f,\sigma) : K(p^{(n)}_{f_0,\sigma_0}, p^{(n)}_{f,\sigma}) \le n\epsilon^2\} \\ &\supset \{(f,\sigma) : \|f - f_0\|^2_{2,G_n} \le C^2\epsilon^2, |\sigma^2 - \sigma_0^2|^2 \le C^2\epsilon^2\}. \end{aligned}$$

By Theorem A.1.1, there exists f_{0J} with J satisfying $J = \lceil c_1 \epsilon^{-1} \rceil$ for some $c_1 > 0$ and $f_{0J}(\cdot) = \sum_{j=1}^J \theta_{0j} \mathbb{1}_{[\xi_{0,j-1}, \xi_{0,j})}(\cdot)$ with $\{\xi_{0,1}, \ldots, \xi_{0,J-1}\} \subset \{X_1, \ldots, X_n\}$ and $\|f_{0J} - f_0\|_{2,G_n} \lesssim \epsilon$. For $\epsilon = \epsilon_n$, there exists $J = J_n$ such that $J \lesssim \epsilon_n^{-1}$ and $\|f_{0J} - f_0\|_{2,G_n} \le C\epsilon_n/2$. Therefore,

$$B_{n,0}((f_0,\sigma_0),\epsilon_n) \supset \{(f,\sigma) : \|f - f_{0J}\|_{2,G_n} \le C\epsilon_n/2, |\sigma^2 - \sigma_0^2| \le C\epsilon_n\}.$$

For some $\bar{C} > 0$, $\Pi(B_{n,0}((f_0, \sigma_0), \epsilon_n))$ is bounded below by

$$
\begin{aligned}
&\Pi(\|f - f_{0J}\|_{2,G_n} \leq C\epsilon_n/2, |\sigma^2 - \sigma_0^2| \leq C\epsilon_n) \\
&\qquad = \Pi\left(n^{-1} \sum_{i=1}^{n} (f(X_i) - f_{0J}(X_i))^2 \leq C^2\epsilon_n^2/4, |\sigma^2 - \sigma_0^2| \leq C\epsilon_n \right) \\
&\qquad \geq \Pi\left(\sum_{i=1}^{n} (f(X_i) - f_{0J}(X_i))^2 \leq C^2 n\epsilon_n^2/4 \Big| \boldsymbol{\xi} = \boldsymbol{\xi}_0, |\sigma^2 - \sigma_0^2| \leq C\epsilon_n \right) \\
&\qquad\qquad \times \Pi(\boldsymbol{\xi} = \boldsymbol{\xi}_0) \times \Pi(|\sigma^2 - \sigma_0^2| \leq C\epsilon_n) \\
&\qquad \asymp \Pi\left(\sum_{j=1}^{J} N_j(\theta_j - \theta_{0j})^2 \leq n\epsilon_n^2/2 \right) \binom{n}{J-1}^{-1} \epsilon_n \\
&\qquad \geq \Pi\left(\bigcap_{j=1}^{J} \{|\theta_j - \theta_{0,j}| \leq \epsilon_n/\sqrt{2}\} \right) \frac{(J-1)!}{\prod_{k=2}^{J}(n-J+k)} \epsilon_n \\
&\qquad \gtrsim (\bar{C}\epsilon_n)^{J+1}\ n^{-(J-1)} \\
&\qquad \gtrsim \exp\big(-J[\log n + \log(1/\epsilon_n)]\big).
\end{aligned}
$$

For $J \asymp n^{1/3}(\log n)^{t_1}$ and $\epsilon_n \asymp n^{-1/3}(\log n)^{t_2}$ with $1 + t_1 \leq 2t_2$, the last expression is greater than or equal to

$$\exp\big[-(4/3)n^{1/3}(\log n)^{t_1+1}\big] \geq \exp\big[-(4/3)n\epsilon_n^2\big].$$

Next for $K_n > 0$, we define the set

$$\mathscr{P}_{n1} = \left\{ f : f(\cdot) = \sum_{j=1}^{J} \theta_j \mathbb{1}_{[\xi_{j-1}, \xi_j)}(\cdot),\ \xi_1, \ldots, \xi_{J-1} \in \boldsymbol{X}, \|\theta_j\|_\infty \leq K_n \right\}.$$

For $K_n = n^{1/6}(\log n)^{t_2+1}$, ϵ_n as defined in the theorem, and some constant $\bar{C} > 0$, we bound $\mathscr{N}(\xi\epsilon, \mathscr{P}_{n1}, \|\cdot\|_{p,G_n})$ above by

$$\exp\big[J_n \log(K_n/\xi\epsilon_n)\big] \leq \exp\big[\bar{C}n^{1/3}(\log n)^{1+t_1}\big] \leq \exp(n\epsilon_n^2).$$

Finally, note that for our choices of ϵ_n, J and $K_n = n^{1/6}(\log n)^{t_2+1}$, the prior probability of the complement of the sieve $\mathscr{P}_{n1}$ is bounded as

$$J\exp(-K_n^2/2) \leq \exp(-K_n^2/2 + \log J) \leq \exp(-(\log n)^2\ n^{1/3}(\log n)^{2t_2}).$$

Part (iii) of Theorem B.3.1 therefore holds, and the result follows. □

Proof of Theorem 2.3.1. We shall prove parts (b) and (c) here. The proof of part (a) is provided in the Appendix A as the proof of Theorem A.2.1.

(b) Since f^* is piece-wise constant on each I_j, $f^*(x_0) = \theta^*_{\lceil x_0 J \rceil}$. Let $z \in \mathbb{R}$. We shall evaluate $\big(\mathrm{P}_0 \times \Pi_n\big)(n^{1/3}(f^*(x_0) - f_0(x_0)) \le z)$ as $n \to \infty$. Note that $\big(\mathrm{P}_0 \times \Pi_n\big)(n^{1/3}(f^*(x_0) - f_0(x_0)) \le z) = \big(\mathrm{P}_0 \times \Pi_n\big)(\theta^*_{\lceil x_0 J \rceil} \le f_0(x_0) + n^{-1/3} z)$. Let $c(\cdot)$ denote the graph of

$$\left\{(0,0), \left(\frac{N_1}{n}, \frac{N_1}{n}\theta_1\right), \left(\sum_{k=1}^{2}\frac{N_k}{n}, \sum_{k=1}^{2}\frac{N_k}{n}\theta_k\right), \ldots, \left(\sum_{k=1}^{J}\frac{N_k}{n}, \sum_{k=1}^{J}\frac{N_k}{n}\theta_k\right)\right\},$$

with $c(s) = 0$ for $s \le 0$ and $c(s) = \sum_{k=1}^{J}(N_k/n)\theta_k$ for $s \ge 1$. Let V_n, G_n be stochastic processes denoting $V_n(s) = \sum_{j=1}^{\lceil sJ \rceil}(N_j/n)\theta_j$ and $G_n(s) = \sum_{j=1}^{\lceil sJ \rceil} N_j/n$. Now since $\theta^*_{\lceil x_0 J \rceil}$ is the left-derivative of the greatest convex minorant of $c(\cdot)$ at the point $\sum_{k=1}^{\lceil x_0 J \rceil} N_k/n$, by the switch relation (B.2.1), we have that $f^*(x_0) \le f_0(x_0) + n^{-1/3} z$ if and only if $\operatorname{argmin}\{c(s) - (f_0(x_0) + n^{-1/3}z)s : s \ge 0\} \ge \sum_{k=1}^{\lceil x_0 J \rceil} N_k/n$. Since the function inside the argmin is piece-wise linear in s, the minimum occurs on the set $\{0, N_1/n, (N_1+N_2)/n, \ldots, 1\}$, and so $\operatorname{argmin}\big\{c(s) - (f_0(x_0) + n^{-1/3}z)s : s \in \mathbb{R}\big\}$ is equal to $\operatorname{argmin}\big\{V_n(s) - (f_0(x_0) + n^{-1/3}z)G_n(s) : s \in \mathbb{R}\big\}$. Therefore,

$$\begin{aligned}
&\big(\mathrm{P}_0 \times \Pi_n\big)\big(f^*(x_0) \le f_0(x_0) + n^{-1/3} z\big) \\
&\quad = \big(\mathrm{P}_0 \times \Pi_n\big)\left(\operatorname*{argmin}_{s \in \mathbb{R}}\big\{V_n(s) - (f_0(x_0) + n^{-1/3}z)G_n(s)\big\} \ge \sum_{k=1}^{\lceil x_0 J \rceil}\frac{N_k}{n}\right). \qquad (2.6.9)
\end{aligned}$$

Using the fact that $V_n(s) = G_n(s) = 0$ for $s \le 0$, and that the location of minimum does not change upon adding a constant term or upon multiplication by a positive constant, (2.6.9) can be written as

$$\begin{aligned}
&\big(\mathrm{P}_0 \times \Pi_n\big)\left(\operatorname*{argmin}_{s \in \mathbb{R}}\left\{\frac{n^{2/3}}{g(x_0)}(V_n(s) - V_n(x_0))\right.\right. \\
&\qquad \left.\left. - \frac{n^{2/3}}{g(x_0)}(f_0(x_0) + n^{-1/3}z)(G_n(s) - G_n(x_0))\right\} \ge \sum_{k=1}^{\lceil x_0 J \rceil}\frac{N_k}{n}\right). \qquad (2.6.10)
\end{aligned}$$

We use a change of variable $s = x_0 + n^{-1/3}t$. For notational convenience, we shall use the notation $\sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3}t) J \rceil}(N_j/n)\theta_j$ to also denote the expression $\sum_{j=\lceil (x_0 + n^{-1/3}t) J \rceil + 1}^{\lceil x_0 J \rceil}(N_j/n)\theta_j$ when $t < 0$. A similar rule shall be applied for $\sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3}t) J \rceil}(N_j/n)$. For $\tau_n = n^{1/3}(\sum_{k=1}^{\lceil x_0 J \rceil}(N_k/n) - x_0)$, (2.6.10) can then be written as

$$\big(\mathrm{P}_0 \times \Pi_n\big)\left(\operatorname*{argmin}_{t \in \mathbb{R}}\left\{\frac{n^{2/3}}{g(x_0)}\big(V_n(x_0 + n^{-1/3}t) - V_n(x_0)\big)\right.\right.$$

$$
\left.
-\frac{n^{2/3}}{g(x_0)}\big(f_0(x_0)+n^{-1/3}z\big)\big(G_n(x_0+n^{-1/3}t)-G_n(x_0)\big)\right\}\geq\tau_n\Bigg)
$$

$$
=\big(\mathrm{P}_0\times\Pi_n\big)\Bigg(\underset{t\in\mathbb{R}}{\arg\min}\left\{\frac{n^{2/3}}{g(x_0)}\sum_{j=\lceil x_0J\rceil+1}^{\lceil(x_0+n^{-1/3}t)J\rceil}\frac{N_j}{n}\theta_j\right.
$$

$$
\left.
-\frac{n^{2/3}}{g(x_0)}\big(f_0(x_0)+n^{-1/3}z\big)\sum_{j=\lceil x_0J\rceil+1}^{\lceil(x_0+n^{-1/3}t)J\rceil}\frac{N_j}{n}\right\}\geq\tau_n\Bigg). \tag{2.6.11}
$$

Now note that $\sum_{k=1}^{\lceil x_0J\rceil}N_k$ is distributed as Binomial$(n;G(\lceil x_0J\rceil/J))$. Using (2.1.2) on G, we get that

$$
G(\lceil x_0J\rceil/J)-x_0=\lceil x_0J\rceil/J-x_0=x_0+O(J^{-1})-x_0=O(J^{-1}).
$$

Therefore $\mathrm{E}\,\tau_n=n^{1/3}[G(\lceil x_0J\rceil/J)-x_0]$, which is of the order $n^{1/3}J^{-1}$ and $\mathrm{Var}\,\tau_n\lesssim n^{-1/3}$. Therefore $\tau_n\to^{P_0}0$. Similarly,

$$
\mathrm{E}\left[\frac{n^{1/3}z}{g(x_0)}\sum_{j=\lceil x_0J\rceil+1}^{\lceil(x_0+n^{-1/3}t)J\rceil}\frac{N_j}{n}\right]
$$

$$
=\frac{n^{1/3}z}{g(x_0)}\left(G\left(\frac{\lceil(x_0+n^{-1/3}t)J\rceil}{J}\right)-G\left(\frac{\lceil x_0J\rceil+1}{J}\right)\right).
$$

The last expression is equal to $n^{1/3}z(tn^{-1/3}+O(J^{-1}))$, which is $zt+o(1)$. The variance is of the order $n^{-2/3}z^2J$, which goes to zero. We have therefore obtained that

$$
\frac{n^{1/3}z}{g(x_0)}\sum_{j=\lceil x_0J\rceil+1}^{\lceil(x_0+n^{-1/3}t)J\rceil}\frac{N_j}{n}\to^{P_0}zt,\qquad\tau_n\to^{P_0}0. \tag{2.6.12}
$$

Using (2.6.12), Lemma 2.6.1, 2.6.2, the argmin version of Theorem B.1.6, and part (a) of Lemma A.2.7, we rewrite (2.6.11) to obtain

$$
\begin{aligned}
&\big(\mathrm{P}_0\times\Pi_n\big)(n^{1/3}(f^*(x_0)-f_0(x_0))\leq z)\\
&\quad\to\mathrm{P}\left(\underset{t\in\mathbb{R}}{\arg\min}\big\{aW_1(t)+aW_2(t)+bt^2-zt\big\}\geq0\right)\\
&\quad=\mathrm{P}\left((a/b)^{2/3}\underset{t\in\mathbb{R}}{\arg\min}\big\{W_1(t)+W_2(t)+t^2\big\}+z/(2b)\geq0\right)\\
&\quad=\mathrm{P}\left(2b(a/b)^{2/3}\underset{t\in\mathbb{R}}{\arg\min}\big\{W_1(t)+W_2(t)+t^2\big\}\geq-z\right)
\end{aligned}
$$

$$= \mathrm{P}\left(2b(a/b)^{2/3} \operatorname*{arg\,min}_{t\in\mathbb{R}} \left\{ W_1(t) + W_2(t) + t^2 \right\} \le z \right).$$

The last step follows from the fact that $\operatorname{arg\,min}\left\{ W_1(t) + W_2(t) + t^2 : t \in \mathbb{R} \right\}$ is symmetric about zero. Substituting $C = 2b(a/b)^{2/3}$ in the last expression of the display we get that

$$\left(\mathrm{P}_0 \times \Pi_n\right)\left(n^{1/3}(f^*(x_0) - f_0(x_0)) \le z\right) \to \mathrm{P}\left(C \operatorname*{arg\,min}_{t\in\mathbb{R}} \left\{ W_1(t) + W_2(t) + t^2 \right\} \le z \right).$$

This completes the proof of part (b).

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

$$H_{1n}(t_1, z_1) = \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t_1) J \rceil} \frac{N_j}{n} (\bar{Y}_j - f_0(x_0)) - n^{-1/3} z_1 \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t_1) J \rceil} \frac{N_j}{n},$$

$$H_{2n}(t_2, z_2) = \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t_2) J \rceil} \frac{N_j}{n} (\theta_j - f_0(x_0)) - n^{-1/3} z_2 \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t_2) J \rceil} \frac{N_j}{n}.$$

Then from the proofs of part (a) and (b), and Theorem B.1.7, we obtain that for every $z_1, z_2 \in \mathbb{R}$,

$$\begin{aligned} &\left(\operatorname{arg\,min}\{ H_{1n}(t_1, z_1) : t_1 \in \mathbb{R} \}, \operatorname{arg\,min}\{ H_{2n}(t_2, z_2) : t_2 \in \mathbb{R} \} \right) \\ &\quad \rightsquigarrow \left(\operatorname{arg\,min}\left\{ a\, W_1(t_1) + b\, t_1^2 - z_1 t_1 : t_1 \in \mathbb{R} \right\}, \right. \\ &\qquad\qquad \left. \operatorname{arg\,min}\left\{ a\, W_1(t_2) + a\, W_2(t_2) + b\, t_2^2 - z_2 t_2 : t_2 \in \mathbb{R} \right\} \right). \end{aligned} \tag{2.6.13}$$

Using (2.6.13), part (c) of Lemma A.2.7, and rewriting $n^{1/3}(f^*(x_0) - \hat{f}_n(x_0))$ as $n^{1/3}[f^*(x_0) - f_0(x_0)] - n^{1/3}[\hat{f}_n(x_0) - f_0(x_0)]$, we get that for all $z \in \mathbb{R}$,

$$\left(\mathrm{P}_0 \times \Pi_n\right)\left(n^{1/3}(f^*(x_0) - \hat{f}_n(x_0)) \le z\right) \to \mathrm{P}\left(CZ_2 - CZ_1 \le z\right).$$

From the proof of part (a) of Theorem 2.3.2, we also obtain that

$$\Pi\left(n^{1/3}(f^*(x_0) - \hat{f}_n(x_0)) \le z | D_n\right) \rightsquigarrow \mathrm{P}\left(CZ_2 - CZ_1 \le z | W_1\right).$$

If $\Pi(n^{1/3}(f^*(x_0) - \hat{f}_n(x_0)) \le z | D_n)$ converges in probability, then the limit must be the random variable $Q = \mathrm{P}(CZ_2 - CZ_1 \le z | W_1)$. Using Theorem B.4.2 for $W_n = n^{1/3}(f^*(x_0) - \hat{f}_n(x_0))$ and $W_n^* = \Pi(n^{1/3}(f^*(x_0) - \hat{f}_n(x_0)) \le z | D_n)$, we obtain that if W_n^* converges in probability, then $(Z_2 - Z_1)$ and Q are independent. This is clearly not true, since both the random variables

are dependent on W_1. This completes the proof of part (c). □

Lemma 2.6.1. *Let $a = \sqrt{\sigma_0^2/g(x_0)}$, $b = f_0'(x_0)/2$, and W_1, W_2 be independent two-sided Brownian motions on $\mathbb{R}$ starting at zero. Then for the prior of Type 1 with $n^{1/3} \ll J \ll n^{2/3}$, for all $K > 0$,*

$$\left\{ \frac{n^{2/3}}{g(x_0)} \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} \frac{N_j}{n}(\theta_j - f_0(x_0)), \quad t \in [-K,K] \right\}$$

converges weakly in $L^\infty([-K,K])$ to $\{a W_1(t) + a W_2(t) + b\,t^2 : t \in [-K,K]\}$.

Proof of Lemma 2.6.1. We consider a sequence $\sigma_n \in \mathscr{U}_n$, where $\mathscr{U}_n$ is a shrinking neighborhood of σ_0 with $\Pi(\sigma \in \mathscr{U}_n \big| D_n) \to 1$. Define

$$\begin{aligned} A_n(\theta, \sigma_n, D_n, t) &= \frac{n^{2/3}}{g(x_0)} \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} \frac{N_j}{n}(\theta_j - \mathrm{E}(\theta_j|D_n)), \\ \bar{A}_n(D_n, t) &= \frac{n^{2/3}}{g(x_0)} \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} \frac{N_j}{n}(\mathrm{E}(\theta_j|D_n) - \bar{Y}_j), \\ B_n(D_n, t) &= \frac{n^{2/3}}{g(x_0)} \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} \frac{N_j}{n}(\bar{Y}_j - f_0(x_0)). \end{aligned} \tag{2.6.14}$$

Therefore,

$$\frac{n^{2/3}}{g(x_0)} \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} \frac{N_j}{n}(\theta_j - f_0(x_0)) = A_n(\theta, \sigma_n, D_n, t) + \bar{A}_n(D_n, t) + B_n(D_n, t).$$

We shall show that for all $K > 0$, the processes $\{A_n(\theta, \sigma_n, D_n, t) : t \in [-K,K]\}$ and $\{\bar{A}_n(D_n, t) : t \in [-K,K]\}$ converge weakly in $L^\infty([-K,K])$ to $\{a W_2(t) : t \in [-K,K]\}$ and 0 respectively. The proof of the weak convergence of $\{B_n(D_n, t) : t \in [-K,K]\}$ to $\{a W_1(t) + b\,t^2 : t \in [-K,K]\}$ has been provided in Section 2.7 within the proof of Theorem 2.3.1 part (a).

Let $t \in [-K,K]$. Then $(\theta_j - \mathrm{E}(\theta_j|D_n))\big|(D_n, \sigma_n) \sim \mathrm{N}(0, \sigma_n^2/(N_j + \lambda_j^{-2}))$. Using the fact that $\lambda_j, 1 \le j \le J$ are uniformly bounded, the conditional variance of $A_n(\theta, \sigma_n, D_n, t)$ given (D_n, σ_n) equals to

$$\frac{n^{4/3}}{g^2(x_0)} \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} \left(\frac{N_j}{n}\right)^2 \frac{\sigma_n^2}{N_j + \lambda_j^{-2}} = \frac{n^{-2/3}\sigma_n^2}{g^2(x_0)} \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} N_j + O\left(\frac{J}{n}\right).$$

Since $\sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} N_j \sim \mathrm{Bin}(n, \int_{(\lceil x_0 J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}t)J\rceil/J} g(u)du)$, the expectation of the first term in the last display is

$$\begin{aligned}\frac{n^{-2/3}\sigma_n^2}{g^2(x_0)} \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} N_j &= \frac{n^{1/3}\sigma_n^2}{g^2(x_0)} \int_{(\lceil x_0 J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}t)J\rceil/J} g(u)du \\ &= \frac{n^{1/3}\sigma_n^2}{g^2(x_0)} \left[g(x_0) n^{-1/3} t + O(J^{-1}) \right] \\ &= \frac{\sigma_0^2 t}{g(x_0)} + o_P(1),\end{aligned}$$

and its variance is equal to

$$\frac{n^{-4/3}\sigma_n^4}{g^4(x_0)} \mathrm{Var}\left(\sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} N_j \right) \lesssim n^{-1/3} \left[g(x_0) n^{-1/3} t\, J + O(J^{-1}) \right] \to 0.$$

Therefore $\mathrm{Var}[A_n(\theta, \sigma_n, D_n, t)] \to^{P_0} \sigma_0^2 t / g(x_0)$. Using exactly the same arguments, it follows that $\mathrm{Cov}(A_n(\theta, \sigma_n, D_n, t), A_n(\theta, \sigma_n, D_n, s))$ asymptotically equals to zero if s and t are of the opposite signs, and $\sigma_0^2(s \wedge t)/g(x_0)$, if s and t have the same signs. Now note that, given (D_n, σ_n), $A_n(\theta, \sigma_n, D_n, \cdot)$ is a Gaussian process with mean zero and the asymptotic covariance thus obtained. If the conditional distribution of $A_n(\theta, \sigma_n, D_n, \cdot)$ given (D_n, σ_n) can be shown tight, then the conditional distribution of $A_n(\theta, \sigma_n, D_n, \cdot)$ given (D_n, σ_n) shall be asymptotically equal to the distribution of $a W_2$, with $a^2 = \sigma_0^2 / g(x_0)$. Since the limit distribution is free of (D_n, σ_n), this also proves that $\{A_n(\theta, \sigma_n, D_n, t) : t \in [-K, K]\}$ unconditionally converges weakly to $\{a W_2 : t \in [-K, K]\}$ in $L^\infty([-K.K])$ for all $K > 0$.

To show tightness, we define two semimetrics ρ_n and ρ for t_1 and t_2 of the same sign as

$$\begin{aligned}\rho_n(t_1, t_2) &= \left(\mathrm{E}\left[(A_n(\theta, \sigma_n, D_n, t_1) - A_n(\theta, \sigma_n, D_n, t_2))^2 \middle| D_n, \sigma_n \right] \right)^{1/2}, \\ \rho(t_1, t_2) &= \left(\sigma_0^2 \, |t_1 - t_2| / g(x_0) \right)^{1/2}.\end{aligned}$$

Then by calculations similar to that for evaluating the limit of variance of $A_n(\theta, \sigma_n, D_n, t)$, it follows that

$$\sup_{t_1, t_2 \in [-K, K]} \left| \mathrm{E}\left[(A_n(\theta, \sigma_n, D_n, t_1) - A_n(\theta, \sigma_n, D_n, t_2))^2 \middle| D_n, \sigma_n \right] - \sigma_0^2 \, |t_1 - t_2| / g(x_0) \right| \to^{P_0}.$$

On a set of P_0-probability 1, for all large n, $\rho_n(t_1, t_2) \le 2\rho(t_1, t_2)$, and therefore for $\delta > 0$ and

all large n,

$$\sup_{\rho(t_1,t_2)<\delta} |(A_n(\theta,\sigma_n,D_n,t_1)-A_n(\theta,\sigma_n,D_n,t_2)| \\ \leq \sup_{\rho_n(t_1,t_2)<\delta/2} |(A_n(\theta,\sigma_n,D_n,t_1)-A_n(\theta,\sigma_n,D_n,t_2)|.$$

As $A_n(\theta,\sigma_n,D_n,\cdot)$ given (D_n,σ_n) is a Gaussian process with mean zero, by Lemma B.1.3, we have for $\delta'=\sigma_0^2\delta/g(x_0)$ and large n,

$$\begin{aligned}
&\mathrm{E}\left[\sup_{|t_1-t_2|<\delta} |A_n(\theta,\sigma_n,D_n,t_1)-A_n(\theta,\sigma_n,D_n,t_2)| \,\middle|\, D_n,\sigma_n\right] \\
&\quad= \mathrm{E}\left[\sup_{\rho(t_1,t_2)<\sqrt{\delta'}} |A_n(\theta,\sigma_n,D_n,t_1)-A_n(\theta,\sigma_n,D_n,t_2)| \,\middle|\, D_n,\sigma_n\right] \\
&\quad\leq \mathrm{E}\left[\sup_{\rho_n(t_1,t_2)<\delta'/2} |A_n(\theta,\sigma_n,D_n,t_1)-A_n(\theta,\sigma_n,D_n,t_2)| \,\middle|\, D_n,\sigma_n\right] \\
&\quad\lesssim \int_0^{\sqrt{\delta'}/2} \sqrt{\log \mathscr{N}(\epsilon,[0,K],\rho_n)}\, d\epsilon \\
&\quad\lesssim \int_0^{\sqrt{\delta'}/2} \sqrt{\log \mathscr{N}(\epsilon,[0,1],\|\cdot\|_2)}\, d\epsilon \\
&\quad\lesssim \int_0^{\sqrt{\delta'}} (1/\epsilon)^{1/4}\sqrt{\log(1/\epsilon)}\, d\epsilon,
\end{aligned}$$

the last step being a consequence of Lemma B.1.4. Making the change of variable $u=\log(1/\epsilon)$ and using $(u/2)<e^{u/2}$ for $u\geq 0$, we get

$$\int_0^{\sqrt{\delta'}} (1/\epsilon)^{1/4}\sqrt{\log(1/\epsilon)}\, d\epsilon = \int_{-\log(\sqrt{\delta'})}^{\infty} e^{u/4}u^{1/2}e^{-u}du \lesssim \int_{-\log(\delta')/2}^{\infty} e^{-u/2}du,$$

which is bounded by a constant multiple of $\delta^{1/4}$. Therefore as $\delta\to 0$,

$$\limsup_{n\to\infty} \mathrm{E}\left[\sup_{|t_1-t_2|<\delta} |A_n(\theta,\sigma_n,D_n,t_1)-A_n(\theta,\sigma_n,D_n,t_2)| \,\middle|\, D_n,\sigma_n\right] \to^{P_0} 0.$$

We now show that $\bar{A}_n(D_n,\cdot)\to^{P_0} 0$ in $L^\infty([-K,K])$. We write

$$\bar{Y}_j = N_j^{-1}\sum_{i:X_i\in I_j} f_0(X_i) + N_j^{-1}\sum_{i:X_i\in I_j}\varepsilon_i.$$

From (2.6.7), we have that for some $C > 0$, $\mathrm{P}(M_n) \to 1$, where $M_n = \{\min_{1\le j\le J} N_j \ge C\,n/J\}$. From Assumption A, there exists $C' > 0$ such that on M_n, for every $\lambda > 0$,

$$\mathrm{E}_0\left[\exp(\lambda N_j^{-1}\sum_{i:X_i\in I_j}\varepsilon_i)\right] \le \mathrm{E}\left[\exp(\lambda^2\sigma_0^2/2N_j)\right] \le \exp(C'\lambda^2 J/2n).$$

Therefore, for some $C'' > 0$ and any $T > 0$,

$$\begin{aligned}
&\mathrm{P}_0\left(\max_{1\le j\le J}\left|N_j^{-1}\sum_{i:X_i\in I_j}\varepsilon_i\right| > T\right)\\
&\qquad\le \mathrm{P}_0\left(\max_{1\le j\le J}\left|N_j^{-1}\sum_{i:X_i\in I_j}\varepsilon_i\right| > T\,\middle|\,M_n\right) + \mathrm{P}_0(M_n^c)\\
&\qquad\le \sum_{j=1}^{J}\mathrm{P}_0\left(\left|N_j^{-1}\sum_{i:X_i\in I_j}\varepsilon_i\right| > T\,\middle|\,M_n\right) + o(1)\\
&\qquad\le 2Je^{-nC''T^2/J} + o(1).
\end{aligned}$$

The last expression goes to zero as $(J/n) \to 0$. We therefore have that

$$\max_{1\le j\le J}\left|N_j^{-1}\sum_{i:X_i\in I_j}\varepsilon_i\right| \to^{P_0} 0. \tag{2.6.15}$$

By the monotonicity of f_0 and (2.6.15), we have

$$\begin{aligned}
\max_{1\le j\le J}\left|\bar{Y}_j\right| &\le \max_{1\le j\le J}\left|N_j^{-1}\sum_{i:X_i\in I_j} f_0(X_i)\right| + \max_{1\le j\le J}\left|N_j^{-1}\sum_{i:X_i\in I_j}\varepsilon_i\right|\\
&\le \max_{1\le j\le J}\left|N_j^{-1}N_j f_0(1)\right| + o_{P_0}(1),
\end{aligned} \tag{2.6.16}$$

which is $O_{P_0}(1)$. Using (2.6.16) and the assumptions on the prior, we have

$$\begin{aligned}
|\bar{A}_n(D_n,t)| &= \left|\frac{n^{2/3}}{g(x_0)}\sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil}\frac{N_j}{n}\left(\frac{(\zeta_j-\bar{Y}_j)/\lambda_j^2}{N_j+1/\lambda_j^2}\right)\right|\\
&\lesssim \left|\frac{n^{2/3}}{g(x_0)}\max_{1\le j\le J}|\zeta_j-\bar{Y}_j|\sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil}\frac{N_j}{n}O(N_j^{-1})\right|\\
&\le n^{2/3}O_{P_0}(1)n^{-1/3}KJn^{-1}O(1),
\end{aligned} \tag{2.6.17}$$

which is $n^{-2/3} t\, J\, O_{P_0}(1)$. Therefore, $\sup\{|\bar{A}_n(D_n, t)| : t \in [-K, K]\}$ goes to zero in probability, since $J \ll n^{2/3}$. □

Lemma 2.6.2. *For the prior of Type* 1 *with* $n^{1/3} \ll J \ll n^{2/3}$,

$$\operatorname*{arg\,min}_{t\in\mathbb{R}} \left\{ \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} \frac{N_j}{n}(\theta_j - f_0(x_0)) - \frac{z}{n^{1/3}} \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} \frac{N_j}{n} \right\} = O_{P_0}(1).$$

Proof of Lemma 2.6.2. Using the change of variable $r = n^{-1/3} t$, we establish the tightness of

$$\hat{h}_n = \operatorname*{arg\,min}_{r\in\mathbb{R}} \left\{ \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+r)J\rceil} \frac{N_j}{n}(\theta_j - f_0(x_0)) - n^{-1/3} z \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+r)J\rceil} \frac{N_j}{n} \right\} - x_0.$$

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

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

Then $\hat{h}_n = \arg\min\{M_n(r) : r \in \mathbb{R}\} - x_0$. We use Theorem B.1.5 with $\mathbb{M}_n = -M_n$, $\mathbb{M} = -M$, $\Theta = (-\infty, 0]$, $\hat{\theta}_n = \arg\min_{r\in\mathbb{R}} M_n(r)$, $\theta_0 = 0$ and d the Euclidean distance on $\mathbb{R}$. With these notations, we have $-M_n(0) = -M(0) = 0$ and $\hat{\theta}_n$ maximizes $-M_n$. The condition $\mathbb{M}(\theta) - \mathbb{M}(\theta_0) \lesssim -d^2(\theta, \theta_0)$ has been verified within the proof of Theorem 2.3.1, part (a). We need to construct functions ϕ_n such that

$$\sqrt{n}\, \mathrm{E}^* \sup_{|r|<\delta} \big|(\mathbb{M}_n - \mathbb{M})(\theta) - (\mathbb{M}_n - \mathbb{M})(\theta_0)\big| \lesssim \phi_n(\delta), \tag{2.6.18}$$

for sufficiently small δ, and $\phi_n(\delta)/\delta^\alpha$ is decreasing in δ for some $\alpha \in (0, 2)$. We write the left-hand side of (2.6.18) as $\sqrt{n}\, \mathrm{E}_0^* \sup\{|\sum_{l=1}^5 H_{ln}(r)| : |r| < \delta\}$, where $H_{ln}, 1 \leq l \leq 5$ are defined as

$$\begin{aligned} H_{1n}(r) &= \sqrt{n} \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+r)J\rceil} \frac{N_j}{n}(\theta_j - \mathrm{E}(\theta_j | D_n, \sigma_n)), \\ H_{2n}(r) &= \sqrt{n} \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+r)J\rceil} \frac{N_j}{n}(\mathrm{E}(\theta_j | D_n, \sigma_n) - \bar{Y}_j), \end{aligned}$$

$$
\begin{aligned}
H_{3n}(r) &= \sqrt{n}\mathbb{P}_n\big(Y - f_0(x_0)\big)[\mathbb{1}\{X \in ([x_0J]/J, [(x_0+r)J]/J]\} \\
&\qquad\qquad -\mathbb{1}\{X \in (x_0, x_0+r]\}], \\
H_{4n}(r) &= \sqrt{n}\big[(\mathbb{P}_n - \mathrm{E}_0)(Y - f_0(x_0))\mathbb{1}\{X \in (x_0, x_0+r]\}\big], \qquad (2.6.19) \\
H_{5n}(r) &= n^{1/6} z \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+r)J\rceil} \frac{N_j}{n}.
\end{aligned}
$$

As $r \mapsto \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+r)J\rceil}(N_j/n)(\theta_j - \mathrm{E}(\theta_j|D_n))$ is a zero mean Gaussian process, we have by arguments similar to those for proving Lemma 2.6.1, with

$$
\rho_n(r_1, r_2) = [\mathrm{E}(H_{1n}(r_1) - H_{1n}(r_2))^2 | (D_n, \sigma_n)]^{1/2},
$$

and $\rho(r_1, r_2) = |r_1 - r_2|^{1/2}$, for sufficiently small δ,

$$
\sqrt{n}\mathrm{E}_0^* \sup_{|r|<\delta} \left| \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+r)J\rceil} \frac{N_j}{n}(\theta_j - \mathrm{E}(\theta_j|D_n)) \right| \lesssim \int_0^{\sqrt{\delta'}/2} \epsilon^{-1/4}\sqrt{\log(1/\epsilon)}\, d\epsilon,
$$

which is bounded by a constant multiple of $\delta^{1/4}$. For H_{2n}, we have, using arguments similar to (2.6.17), for sufficiently small δ,

$$
\sqrt{n}\mathrm{E}_0^* \sup_{|r|<\delta} \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+r)J\rceil} \frac{N_j}{n} \left| \frac{(\zeta_j - \bar{Y}_j)/\lambda_j^2}{N_j + 1/\lambda_j^2} \right| \lesssim n^{-1/2} J\delta,
$$

which is bounded by a constant multiple of $n^{1/6}\delta$, as $J \ll n^{2/3}$. Using the last two displays and the proof of Lemma A.2.3 as provided in Appendix A, we have that for sufficiently small δ, (2.6.18) holds for $\phi_n(\delta) = C[\delta^{1/4} + \delta^{1/2} + n^{1/6}\delta + n^{-1/3}]$ for some $C > 0$. The rest of the proof follows similarly to that of Lemma A.2.3, with $r_n = n^{1/3}$ and the condition of uniform consistency of $f^*(x)$ in a neighborhood of x_0 satisfied by the consistency of the maximum likelihood estimator near x_0 as shown by Brunk (1970), and the fact that

$$
\Pi\left(\sup_{r\in[0,1]} \sum_{j=1}^{\lceil rJ\rceil} \frac{N_j}{n} \left|\theta_j - \bar{Y}_j\right| > \epsilon \Big| D_n \right) \to^{P_0} 0.
$$

The difference in the slope of the greatest convex minorants of (2.3.2) and (A.2.1) therefore goes to zero in P_0-probability, uniformly in a neighborhood of x_0, and the uniform consistency of $f^*(x)$ in a neighborhood of x_0 follows. Therefore $n^{1/3}\hat{h}_n$ is $O_{P_0}(1)$. □

Proof of Theorem 2.3.2. (a) Let $z \in \mathbb{R}$. Let $A_n(\theta, \sigma_n, D_n, t)$, $\bar{A}_n(D_n, t)$, and $B_n(D_n, t)$ be as defined in (2.6.14). By using similar reasoning as in the proof of part (b) of Theorem 2.3.1, we can rewrite $F_n^*(z|D_n)$ as

$$\begin{aligned}
&\Pi\left(n^{1/3}(f^*(x_0) - f_0(x_0)) \le z \middle| D_n\right) \\
&\quad = \Pi\Bigg(\operatorname*{arg\,min}_{t\in\mathbb{R}} \Big\{A_n(\theta, \sigma_n, D_n, t) + \bar{A}_n(D_n, t) + B_n(D_n, t) \\
&\qquad\qquad - n^{1/3} z \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0 + n^{-1/3}t)J\rceil} \frac{N_j}{n}\Big\} \ge \tau_n \middle| D_n\Bigg).
\end{aligned}$$

In the proof of Lemma 2.6.1 it has been shown that $\sup\{|\bar{A}_n(D_n, t)| : t \in [-K, K]\} \to^{P_0} 0$, and the conditional distribution of $\{A_n(\theta, \sigma_n, D_n, t) : t \in [-K, K]\}$ given (D_n, σ_n) converges weakly to $\{a W_2(t) : t \in [-K, K]\}$ in $L^\infty([-K, K])$. Also, from the proof of part (a) of Theorem 2.3.1, $\{B_n(D_n, t) : t \in [-K, K]\}$ converges weakly to $\{a W_1(t) : t \in [-K, K]\}$ in $L^\infty([-K, K])$. As W_1 and W_2 are independent, the posterior distribution of $A_n(\theta, \sigma_n, D_n, \cdot) + B_n(D_n, \cdot)$ given (D_n, σ_n) converges weakly to the conditional distribution of $a(W_1 + W_2)$ given W_1. We therefore obtain that for every $T \in \mathbb{R}$ and $K > 0$,

$$\begin{aligned}
&\left\{\Pi(A_n(\theta, \sigma_n, D_n, t) + \bar{A}_n(D_n, t) + B_n(D_n, t) - z\,t \le T \middle| D_n), \quad t \in [-K, K]\right\} \\
&\qquad \rightsquigarrow \left\{\mathrm{P}\left(a W_1(t) + a W_2(t) + b\,t^2 - z\,t \ \le T \middle| W_1\right), \quad t \in [-K, K]\right\}.
\end{aligned}$$

Adapting the arguments used in the proof of Lemma 2.6.2 to the posterior distribution of $\hat{h}_n$ given D_n, we have that the conditional distribution of $\operatorname{arg\,min}\{A_n(\theta, \sigma_n, D_n, t) + \bar{A}_n(D_n, t) + B_n(D_n, t) - z\,t\}$ given (D_n, σ_n) is tight with respect to P_0. By (2.6.12), the argmin version of Theorem B.1.6, and part (a) of Lemma A.2.7 for the conditional distribution, we get that

$$\begin{aligned}
&\Pi\left(n^{1/3}(f^*(x_0) - f_0(x_0)) \le z \middle| D_n\right) \\
&\quad \rightsquigarrow \mathrm{P}\left(\operatorname*{arg\,min}_{t\in\mathbb{R}} \left\{a W_1(t) + a W_2(t) + b\,t^2 - z\,t\right\} \ge 0 \middle| W_1\right) \\
&\quad = \mathrm{P}\left((a/b)^{2/3} \operatorname*{arg\,min}_{t\in\mathbb{R}} \left\{W_1(t) + W_2(t) + t^2\right\} + z/(2b) \ge 0 \middle| W_1\right) \\
&\quad = \mathrm{P}\left(C \operatorname*{arg\,min}_{t\in\mathbb{R}} \left\{W_1(t) + W_2(t) + t^2\right\} \le z \middle| W_1\right).
\end{aligned}$$

(b) Let $W_1'(t) = W_1(-t)$. Using the transformation $t \mapsto -t$ and the fact that $W_1 \stackrel{d}{=} W_1'$, we

have that

$$\begin{aligned}
&\mathrm{P}\left(\operatorname*{arg\,min}_{t\in\mathbb{R}}\{W_1(t)+W_2(t)+t^2\}\geq 0\Big|W_1\right)\\
&\qquad=\mathrm{P}\left(\operatorname*{arg\,min}_{t\in\mathbb{R}}\{W_1(-t)+W_2(-t)+(-t)^2\}\leq 0\Big|W_1\right)\\
&\qquad=\mathrm{P}\left(\operatorname*{arg\,min}_{t\in\mathbb{R}}\{W_1'(t)+W_2(t)+t^2\}\leq 0\Big|W_1'\right).
\end{aligned}$$

Therefore $\mathrm{P}(\Delta^*_{W_1,W_2}\geq 0\Big|W_1)\stackrel{d}{=}\mathrm{P}(\Delta^*_{W_1,W_2}\leq 0\Big|W_1)$, and $\mathrm{P}(\Delta^*_{W_1,W_2}\geq 0\Big|W_1)$ is symmetric about $1/2$.

(c) By the definition of posterior quantile $Q_{n,\gamma}$, we have that $f_0(x_0)\leq Q_{n,\gamma}$ if and only if $\Pi(f^*(x_0)\leq f_0(x_0)|D_n)\leq 1-\gamma$. Therefore,

$$\begin{aligned}
\mathrm{P}_0\left(f_0(x_0)\leq Q_{n,\gamma}\right)&=\mathrm{P}_0\left(\Pi(f^*(x_0)\leq f_0(x_0)|D_n)\leq 1-\gamma\right)\\
&=\mathrm{P}_0\left(\Pi(n^{1/3}(f^*(x_0)-f_0(x_0))\leq 0|D_n)\leq 1-\gamma\right)\\
&=\mathrm{P}_0\left(F_n^*(0\Big|D_n)\leq 1-\gamma\right)\\
&\to\mathrm{P}\left(F^*_{a,b}(0\Big|W_1)\leq 1-\gamma\right),
\end{aligned}$$

where the last step follows from part (a), substituting $z=0$. Now rewriting $F^*_{a,b}(0\Big|W_1)$ in terms of $\Delta^*_{W_1,W_2}$, we have that $\mathrm{P}\left(F^*_{a,b}(0\Big|W_1)\leq 1-\gamma\right)$ is equal to

$$\begin{aligned}
&\mathrm{P}\left(\mathrm{P}\left(C\operatorname*{arg\,min}_{t\in\mathbb{R}}\left\{W_1(t))+W_2(t)+t^2\right\}\geq 0\Big|W_1\right)\leq 1-\gamma\right)\\
&\qquad\qquad=\mathrm{P}\left(\mathrm{P}\left(\Delta^*_{W_1,W_2}\geq 0\Big|W_1\right)\leq 1-\gamma\right).
\end{aligned}$$

Similarly $\mathrm{P}_0\left(f_0(x_0)\geq Q_{n,1-\gamma}\right)=1-\mathrm{P}_0\left(f_0(x_0)<Q_{n,1-\gamma}\right)$, and therefore $\mathrm{P}_0\left(f_0(x_0)\geq Q_{n,1-\gamma}\right)$ converges to

$$1-\mathrm{P}\left(\mathrm{P}(\Delta^*_{W_1,W_2}\geq 0\Big|W_1)<\gamma\right)=\mathrm{P}\left(\mathrm{P}(\Delta^*_{W_1,W_2}\geq 0\Big|W_1)\geq\gamma\right).$$

□

We shall prove Theorem 2.4.1 next. The proof of Lemma 2.4.2 has been provided in Section 2.7.

Proof of Theorem 2.4.1. (a) Let $f_0 \in \mathscr{F}(K)$. The Type 1 error can be evaluated as

$$\mathrm{E}_0\Pi(\|f-f^*\|_{1,G} > M_0 n^{-1/3}|D_n) \leq \mathrm{E}_0\Pi(\|f-f_0\|_{1,G} > M_0 n^{-1/3}|D_n),$$

which goes to zero as a consequence of Theorem 2.2.2 and $J \asymp n^{1/3}$.

(b) Let $f_0 \notin \mathscr{F}$ be fixed and integrable. Using the properties of the projection, $d(f_0,\mathscr{F}) = \|f_0 - f_0^*\|_{1,G}$ is bounded by $\|f_0 - f^*\|_{1,G}$ which, by the triangle inequality, is further bounded above by

$$\|f_0 - f\|_{1,G} + \|f - f^*\|_{1,G} = \|f - f_0\|_{1,G} + d(f,\mathscr{F}).$$

We therefore have $d(f,\mathscr{F}) \geq d(f_0,\mathscr{F}) - \|f-f_0\|_{1,G}$, and $\Pi(d(f,\mathscr{F}) \leq \tau_n \big| D_n) \leq \Pi(\|f_0 - f\|_{1,G} + \tau_n \geq d(f_0,\mathscr{F}) \big| D_n)$.

Let $\theta_{0j} = \int_{I_j} f_0 dG/G(I_j)$, $1 \leq j \leq J$. Then as shown in the proof of Theorem 2.2.2, $\Pi(\|f - f_{0J}\|_{1,G} > M_n\sqrt{J/n} \big| D_n) \to^{P_0} 0$. Therefore for M_0 large enough and $J \asymp n^{1/3}$, $\Pi(\|f - f_{0J}\|_{1,G} > M_0 n^{-1/3} \big| D_n) \to^{P_0} 0$. Next, since f_0 is integrable, by the martingale convergence theorem, $\|f_0 - f_{0J}\|_{1,G} \to 0$. The Type 2 error $\mathrm{E}_0\Pi\big(d(f,\mathscr{F}) \leq \tau_n | D_n\big)$ can therefore be bounded by

$$\begin{aligned}
&\mathrm{E}_0\Pi(\|f-f_0\|_{1,G} + \tau_n \geq d(f_0,\mathscr{F})|D_n) \\
&\leq \mathrm{E}_0\Pi\big(\|f-f_{0J}\|_{1,G} + \|f_{0J} - f_0\|_{1,G} + M_0 n^{-1/3} \geq d(f_0,\mathscr{F}) \big| D_n\big),
\end{aligned}$$

which goes to zero.

(c) Let $f_0 \notin \mathscr{F}$ and $f_0 \in \mathscr{H}(\alpha, L)$. With the same definition of f_{0J} as in part (b) and arguments similar to those used in proving Theorem 2.2.2, we have that $\|f_0 - f_{0J}\|_{1,G} \lesssim J^{-\alpha}$, and so for $J \asymp n^{1/3}$ and M_0 large, $\Pi(\|f - f_0\|_{1,G} > M_0 n^{-\alpha/3}|D_n) \to^{P_0} 0$, uniformly for all $f_0 \in \mathscr{H}(\alpha, L)$. Let A_n denote the event $\{\|f-f_0\|_{1,G} \leq M_0 n^{-\alpha/3}$. On A_n, we have that

$$\begin{aligned}
&\sup\big\{\mathrm{E}_0\Pi(d(f,\mathscr{F}) \leq \tau_n|D_n) : d(f_0,\mathscr{F}) > \rho_n(\alpha)\big\} \\
&\leq \sup\{\mathrm{E}_0\Pi(\|f-f_{0J}\|_{1,G} + n^{-\alpha/3} + M_0 n^{-1/3} > n^{-\alpha/3}(\log n)^T | D_n) : \\
&\qquad\qquad\qquad\qquad\qquad\qquad d(f_0,\mathscr{F}) > \rho_n(\alpha)\} \\
&\leq \sup\{\mathrm{E}_0\Pi(\|f-f_{0J}\|_{1,G} > M_0 n^{-\alpha/3}(\log n)^T | D_n) : d(f_0,\mathscr{F}) > \rho_n(\alpha)\},
\end{aligned}$$

which goes to zero since $\alpha \in (0,1)$ and $\Pi(\|f - f_{0J}\|_{1,G} > M_0 n^{-1/3}|D_n) \to^{P_0} 0$ for $J \asymp n^{1/3}$. □

2.7 Proofs of results in Section 2.6

Proof of Proposition 2.2.1. (a) Let $f_0 \in \mathscr{F}(K)$. We first show that there exists $\boldsymbol{\theta}_{0J} = (\theta_{01}, \ldots, \theta_{0J})$ such that $n^{-1}\|\boldsymbol{F}_0 - \boldsymbol{B}\boldsymbol{\theta}_{0J}\|_2^2 \lesssim J^{-1}$ for deterministic $\boldsymbol{X}$, and $n^{-1}\mathrm{E}_G\|\boldsymbol{F}_0 - \boldsymbol{B}\boldsymbol{\theta}_{0J}\|_2^2 \lesssim J^{-1}$ for random $\boldsymbol{X}$.

On a set with $\min\{N_j : 1 \le j \le J\} > 0$, let $\theta_{0j} = N_j^{-1}\sum_{i:X_i \in I_j} f_0(X_i)$. Using the monotonicity of f_0, we write $n^{-1}\|\boldsymbol{F}_0 - \boldsymbol{B}\boldsymbol{\theta}_{0J}\|_2^2$ as

$$\begin{aligned} n^{-1}\sum_{j=1}^{J}\sum_{i:X_i\in I_j}(f_0(X_i)-\theta_{0j})^2 &\le n^{-1}\sum_{j=1}^{J}\sum_{i:X_i\in I_j}(f_0(j/J)-f_0((j-1)/J))^2 \\ &= n^{-1}\sum_{j=1}^{J} N_j\big(f_0(j/J)-f_0((j-1)/J)\big)^2. \end{aligned} \tag{2.7.1}$$

For deterministic $\boldsymbol{X}$ following (2.1.1), (2.7.1) is bounded above by

$$\begin{aligned} &\sup_{1\le j\le J}(N_j/n)\sum_{j=1}^{J}\big(f_0(j/J)-f_0((j-1)/J)\big)^2 \\ &\lesssim J^{-1}\left(\sum_{j=1}^{J}\big[f_0(j/J)-f_0((j-1)/J)\big]\right)^2 \qquad (2.7.2) \\ &= J^{-1}(f_0(1)-f_0(0))^2. \qquad (2.7.3) \end{aligned}$$

For random $\boldsymbol{X}$ satisfying (2.1.2) and the fact that $N_j \sim \mathrm{Bin}(n; G(I_j))$, the expectation of (2.7.1) under G equals to

$$\begin{aligned} \sum_{j=1}^{J} G(I_j)\big(f_0(j/J)-f_0((j-1)/J)\big)^2 &\lesssim J^{-1}\left(\sum_{j=1}^{J}\big[f_0(j/J)-f_0((j-1)/J)\big]\right)^2 \\ &= J^{-1}(f_0(1)-f_0(0))^2. \end{aligned} \tag{2.7.4}$$

For the rest of the proof, we shall assume that $\boldsymbol{X}$ is fixed, satisfying (2.1.1). The random case can be dealt with similarly, by taking expectation with respect to G and using (2.1.2). We shall imitate the proof of Proposition 4.1 (a) of Yoo and Ghosal (2016) in the situation where f_0 is monotone. Define $\boldsymbol{U} = (\boldsymbol{B}\boldsymbol{\Lambda}\boldsymbol{B}^T + \boldsymbol{I}_n)^{-1}$. We write $|\mathrm{E}_0(\widehat{\sigma}_n^2) - \sigma_0^2|$ as

$$\begin{aligned} &|n^{-1}\sigma_0^2\mathrm{tr}(\boldsymbol{U}) - \sigma_0^2| + n^{-1}(\boldsymbol{F}_0 - \boldsymbol{B}\boldsymbol{\zeta})^T\boldsymbol{U}(\boldsymbol{F}_0 - \boldsymbol{B}\boldsymbol{\zeta}) \\ &\lesssim n^{-1}[\mathrm{tr}(\boldsymbol{I}_n - \boldsymbol{U}) + (\boldsymbol{F}_0 - \boldsymbol{B}\boldsymbol{\theta}_{0J})^T\boldsymbol{U}(\boldsymbol{F}_0 - \boldsymbol{B}\boldsymbol{\theta}_{0J}) \end{aligned}$$

$$+(\boldsymbol{B}\boldsymbol{\theta}_{0J}-\boldsymbol{B}\boldsymbol{\zeta})^T\boldsymbol{U}(\boldsymbol{B}\boldsymbol{\theta}_{0J}-\boldsymbol{B}\boldsymbol{\zeta})]. \tag{2.7.5}$$

The first term in (2.7.5) is bounded by $n^{-1}\mathrm{tr}(\boldsymbol{P}_{\boldsymbol{B}}) = n^{-1}J$. By construction of $\boldsymbol{\theta}_{0J}$, the second term in (2.7.5) is bounded by

$$n^{-1}\|\boldsymbol{U}\|_{(2,2)}\|\boldsymbol{F}_0-\boldsymbol{B}\boldsymbol{\theta}_{0J}\|_2^2 \le n^{-1}\|\boldsymbol{F}_0-\boldsymbol{B}\boldsymbol{\theta}_{0J}\|_2^2 \lesssim J^{-1}. \tag{2.7.6}$$

For the last term in (2.7.5), note that $(\Lambda+(\boldsymbol{B}^T\boldsymbol{B})^{-1})^{-1}$ is a $J\times J$ diagonal matrix with diagonal entries $(\lambda_j^2+N_j^{-1})^{-1}$, $1\le j\le J$. By the assumption on the prior, $\|\boldsymbol{\theta}_{0J}-\boldsymbol{\zeta}\|_\infty^2 = O(1)$, and therefore the third term of (2.7.5) is bounded above by

$$n^{-1}\left(\boldsymbol{\theta}_{0J}-\boldsymbol{\zeta}\right)^T(\Lambda+(\boldsymbol{B}^T\boldsymbol{B})^{-1})^{-1}\left(\boldsymbol{\theta}_{0J}-\boldsymbol{\zeta}\right) = n^{-1}\sum_{j=1}^{J}\frac{(\theta_{0j}-\zeta_j)^2}{(\lambda_j^2+N_j^{-1})} \lesssim J/n.$$

Combining the three bounds thus obtained into (2.7.5), we obtain $|\mathrm{E}_0(\widehat{\sigma}_n^2)-\sigma_0^2| \lesssim J/n+J^{-1}$.

We next bound $\mathrm{Var}_0(n\hat{\sigma}_n^2)$ up to a constant multiple by

$$\begin{aligned} n^{-2}[&(\boldsymbol{F}_0-\boldsymbol{B}\boldsymbol{\theta}_{0J})^T\boldsymbol{U}^2(\boldsymbol{F}_0-\boldsymbol{B}\boldsymbol{\theta}_{0J}) \\ &+(\boldsymbol{B}\boldsymbol{\theta}_{0J}-\boldsymbol{B}\boldsymbol{\zeta})^T\boldsymbol{U}^2(\boldsymbol{B}\boldsymbol{\theta}_{0J}-\boldsymbol{B}\boldsymbol{\zeta})+\mathrm{Var}_0(\boldsymbol{\varepsilon}^T\boldsymbol{U}\boldsymbol{\varepsilon})]. \end{aligned} \tag{2.7.7}$$

By an argument similar to that used to show (2.7.6), the first term in (2.7.7) is bounded by

$$n^{-2}\|\boldsymbol{U}\|_{(2,2)}^2\|\boldsymbol{F}_0-\boldsymbol{B}\boldsymbol{\theta}_{0J}\|_2^2 \le n^{-1}(n^{-1}\|\boldsymbol{F}_0-\boldsymbol{B}\boldsymbol{\theta}_{0J}\|_2^2) \lesssim n^{-1}J^{-1}.$$

The second term in (2.7.7) is bounded by

$$n^{-2}(\boldsymbol{\theta}_{0J}-\boldsymbol{\zeta})^T\boldsymbol{B}^T\boldsymbol{B}(\boldsymbol{\theta}_{0J}-\boldsymbol{\zeta}) \le n^{-2}\sum_{j=1}^{J}N_j(\theta_{0j}-\zeta_j)^2 \lesssim n^{-2}\sum_{j=1}^{J}N_j = n^{-1}.$$

By Theorem B.4.3 and Assumption A on the errors, the last term in (2.7.7) is $O(n^{-1})$. We therefore obtain that $\mathrm{Var}_0(\hat{\sigma}_n^2) \lesssim n^{-1}$, and the mean square error is

$$\mathrm{E}_0(\widehat{\sigma}_n^2-\sigma_0^2)^2 \lesssim n^{-1}+J^2n^{-2}+J^{-2}, \tag{2.7.8}$$

which implies the first assertion.

(b) We observe that

$$\begin{aligned} \mathrm{E}(\sigma^2|D_n) &= \beta_2(\beta_1+n-2)^{-1}+n(\beta_1+n-2)^{-1}\widehat{\sigma}_n^2, \\ \mathrm{Var}(\sigma^2|D_n) &= 4(\beta_1+n-4)^{-1}(\beta_2(\beta_1+n-2)^{-1}+n(\beta_1+n-2)^{-1}\widehat{\sigma}_n^2)^2. \end{aligned}$$

Applying Markov's inequality, the posterior for σ^2 is seen to concentrate around $\widehat{\sigma}_n^2$ at the rate $n^{-1/2}$, so the assertion follows from the plug-in Bayes estimator's rate.

□

Proof of Lemma 2.4.2. The proof follows using arguments exactly similar to those in Theorem 2.4.1, with the following additional result about the contraction of f for $f_0 \in \mathscr{F}$ or $f_0 \in \mathscr{H}(\alpha, L)$, under the prior of Type 3.

Lemma 2.7.1 (Contraction rates under the Type 3 prior on f)**.** *Let the prior on f be of Type 3. Let t_1, t_2 be as defined in (2.4.2), and $\bar{t}_1 = \max\{t_1, 1\}$.*

(a) Let $f_0 \in \mathscr{F}(K)$. Then for any $\delta > 0$, $\epsilon_n = n^{-1/4}(\log n)^{\bar{t}_1/4 - t_2/2 + 1 + \delta}$, ϵ_n is a contraction rate for f under the empirical L_1-distance d.

(b) Let $f_0 \in \mathscr{H}(\alpha, L)$. Then for any $\delta > 0$, $t' = \alpha\bar{t}_1/(1+2\alpha) - t_2/2 + 1 + \delta$, and $\epsilon_n = n^{-\alpha/(1+2\alpha)}(\log n)^{t'}$, ϵ_n is a contraction rate for f under the empirical L_1-distance d.

The proof follows from the assumptions on the prior and Theorem 10.21 of Ghosal and van der Vaart (2017). In part (a), the conditions of the theorem are met for $\bar{J}_n \asymp (n/\log^{\bar{t}_1} n)^{1/2}$, $J_n \gg n^{1/2}(\log n)^{1/2\bar{t}_1 - t_2}$ and $\bar{\epsilon}_n \asymp (n/\log^{\bar{t}_1} n)^{1/4}$. For part (b), the theorem holds for $\bar{J}_n \asymp (n/\log^{\bar{t}_1} n)^{1/(1+2\alpha)}$, $J_n \gg n^{1/(1+2\alpha)}(\log n)^{2\alpha/(1+2\alpha)\bar{t}_1 - t_2}$ and $\bar{\epsilon}_n \asymp (n/\log^{\bar{t}_1} n)^{\alpha/(1+2\alpha)}$. □

CHAPTER

3

RATES, COVERAGE AND TESTS FOR BAYESIAN MONOTONE DENSITY ESTIMATION

Let $X_1, \ldots, X_n$ be independent and identically distributed observations on $(0, \infty)$. Consider the following model

$$X_1, \ldots, X_n \sim G, \tag{3.0.1}$$

where G has a density $g > 0$ which is a monotone non-increasing function on $[0, \infty)$, with $g(0) < \infty$. In the monotone density scenario, consistency of the constrained point estimator was established by Grenander (1957) and its asymptotic distribution was evaluated by Prakasa Rao (1969). The rate of convergence of the constrained nonparametric maximum likelihood estimator at a point x_0 is equal to $n^{-1/3}$ and the limit distribution is equal to a scaled Chernoff's distribution (see Groeneboom and Wellner (2001)). Grenander (1957) and Balabdaoui and Wellner (2007) established the global L_1-convergence of $\hat{g}$ at the minimax rate $n^{-1/3}$, when the support of the distribution is compact. A possible Bayesian approach

to this problem would involve using a random series prior on g and analyzing the behavior of the posterior distribution of g to obtain credible intervals. Monotone non-increasing densities on $[0,\infty)$ can be represented as a mixture of uniform distributions. Using the mixture representation, Salomond (2014) developed posterior contraction rates for g based on a Dirichlet process mixture prior on P, establishing that the posterior of g contracts at g_0 under the L_1 or the Hellinger metric at the optimal rate $n^{-1/3}$ up to a logarithmic factor.

In this chapter, we follow a Bayesian approach by using a finite histogram prior on g. The conjugate posterior is then projected on the space of monotone functions using the algorithm for isotonic regression. We establish contraction rates for the induced distribution of the projection. We develop contraction rates under both these settings and achieve the minimax rate $n^{-1/3}$ for the bounded domain and up to a logarithmic factor for the unbounded domain. For point-wise credible interval, we derive the asymptotic distribution of the sieve maximum likelihood estimator under monotonicity. The point-wise asymptotic distribution of the projection-posterior is then developed, and a credible interval obtained whose asymptotic coverage can be analytically characterized. We then use a recalibration method to obtain the credible interval with the right coverage.

3.1 Prior on g

For G on $(0,1]$, we represent g as a piece-wise constant function on $(0,1]$ with J the number of pieces being deterministic. The knots of g are equally spaced on $(0,1)$. The prior is described as

$$\begin{aligned} &g = J\sum_{j=1}^{J}\theta_j \mathbb{1}_{I_j}, \\ &(\theta_1,\ldots,\theta_J) \sim \mathrm{Dir}(\alpha_1,\ldots,\alpha_J), \\ &\alpha_j \in (b_1,b_2),\ 1\le j\le J \text{ for some } b_1,b_2>0. \end{aligned}$$

For g on an unbounded domain, we have the following setup. Let $T_0=0$, $T_k=2^{k-1}$ for $k\ge 1$. Let $t_0=0$ and

$$t_j = \begin{cases} j/J, & 1\le j\le J, \\ t_{\lfloor j/J\rfloor} + (j/J - \lceil j/J\rceil)(T_{\lfloor j/J\rfloor+1} - T_{\lfloor j/J\rfloor}), & j>J. \end{cases} \tag{3.1.1}$$

Let $\delta_j = t_j - t_{j-1}$, $\delta_0 = 0$. Let L_j be the interval $(t_{j-1}, t_j]$. Let $J' = JK$. The prior on g over the unbounded domain is

$$\begin{aligned} &g = \sum_{j=1}^{JK} \delta_j^{-1} \theta_j \mathbb{1}_{L_j}, \\ &(\theta_1, \ldots, \theta_{JK}) \sim \mathrm{Dir}(\alpha_1, \ldots, \alpha_{JK}), \\ &\alpha_j \in (b_1, b_2),\ 1 \le j \le JK, \text{ for some } b_1, b_2 > 0. \end{aligned}$$

3.2 Posterior contraction rate under monotonicity

We impose monotonicity on g by projecting g on $\mathscr{F}$ using the algorithm for isotonic regression. Let Π_n^* be the distribution induced by the posterior of g defined as

$$\Pi_n^*(g \in B) = \Pi(g : g^* \in B | D_n),$$

for every B measurable with respect to Pi. For a monotone g_0 and g^* denoting the L_1-projection of g onto $\mathscr{F}$, we have, by the property of projection and the triangle inequality.

$$\|g^* - g_0\|_1 \le \|g^* - g\|_1 + \|g - g_0\|_1 \le \|g_0 - g\|_1 + \|g - g_0\|_1 = 2\|g - g_0\|_1. \tag{3.2.1}$$

For g on $[0, \infty)$, we assume that for some $b, r > 0$,

$$g_0(x) \le e^{-bx^r}, \quad \text{for all large } x. \tag{3.2.2}$$

The following theorem provides a contraction rate for g^* in terms of the L_1-metric.

Theorem 3.2.1. *Let $g_0 \in \mathscr{F}$ and let $J \to \infty$ and $J \ll n$. Then the following assertions hold:*

(a) For g on $[0,1]$, $\epsilon_n = \max\{J^{-1}, J^{1/2} n^{-1/2}\}$, and every $M_n \to \infty$,

$$\mathrm{E}_0\, \Pi_n^*\left(\|g - g_0\|_1 > M_n \epsilon_n\right) \to 0.$$

(b) For g on $[0, \infty)$, $\epsilon_n = \max\{J^{-1}(\log n)^{-1}, J^{1/2} n^{-1/2} (\log n)^{(1/2+1/r)}\}$, and every $M_n \to \infty$,

$$\mathrm{E}_0\, \Pi_n^*\left(\|g - g_0\|_1 > M_n \epsilon_n\right) \to 0.$$

In particular, when $J \asymp n^{1/3}$, the projection-posterior contracts at the optimal rate

$\epsilon_n = n^{-1/3}$ for g on a bounded domain, and a nearly optimal rate $\epsilon_n = n^{-1/3}(\log n)^{1/2+1/r}$ for g on an unbounded domain.

3.3 Uncertainty quantification in point-wise estimation

Let $x_0 \in (0,1)$ be such that $g'(x_0)$ exists and $g'(x_0) < 0$. The Grenander estimator of $g(x_0)$ is obtained by maximizing the log likelihood $\sum_{i=1}^n \log g(X_i)$ over all non increasing densities g on $(0,1)$. The estimate of $g(x_0)$ thus obtained is the slope (left derivative) of the least concave majorant of the empirical cdf, at the point x_0.

Now suppose that g is a non-increasing piece-wise constant function on $(0,1)$. In particular, for $J > 1$, let $g = \sum_{j=1}^J p_j \mathbb{1}_{I_j}$, where $(p_1, \ldots, p_J) \in \mathbb{R}^J$ with $p_1 \geq \cdots \geq p_J$, $I_j = [(j-1)/J, j/J)$, and $\int_0^1 g(u)du = 1$. Any density h on $(0,1)$ which is piece-wise constant on each I_j can be transformed to a non-increasing piece-wise constant density on $(0,1)$ by "projecting" h onto the space of monotone non increasing densities on $(0,1)$ in the following way.

Let $h = \sum_{j=1}^J p_j \mathbb{1}_{I_j}$, $\int_0^1 h(u)du = 1$. Let $\bar{h} = \sum_{j=1}^J \bar{p}_j \mathbb{1}_{I_j}$ be obtained from h by minimizing the sum of squares $h = J^{-1}\sum_{j=1}^J (p_j - \bar{p}_j)^2$ under the constraint $\bar{p}_1, \geq \cdots \geq \bar{p}_J$. The algorithm for the constrained minimization problem as described in Section 2.1 of Groeneboom and Jongbloed (2014) gives us that $\bar{p}_j$ is the slope (left derivative) of the least concave majorant of the graph

$$\left\{(0,0), \left(\frac{1}{J}, \frac{p_1}{J}\right), \left(\frac{2}{J}, \frac{p_1+p_2}{J}\right), \ldots, \left(1, \frac{\sum_{j=1}^J p_j}{J}\right)\right\}, \tag{3.3.1}$$

at the point j/J, for $1 \leq j \leq J$. The following lemma states that $\bar{h}$ thus obtained is a density on $(0,1)$, i.e. $\bar{h} > 0$ on $(0,1)$ and $\int_0^1 \bar{h}(u)du = 1$.

Lemma 3.3.1. *Let $(\bar{p}_1, \ldots, \bar{p}_J)$ be obtained as the left derivative of the least concave majorant of the graph described in* (3.3.1). *Then $\bar{h} = \sum_{j=1}^J \bar{p}_j \mathbb{1}_{I_j}$ satisfies $\bar{h} > 0$ on $(0,1)$ and $\int_0^1 \bar{h}(u)du = 1$.*

The proof of the lemma follows from properties of the least concave majorant and arguments similar to those in Section 3.2 of Groeneboom and Jongbloed (2014). Now consider a sieve Θ_J defined as

$$\Theta_J = \left\{ g = \sum_{j=1}^J q_j \mathbb{1}_{I_j} : q_1 \geq \cdots \geq q_J, \quad I_j = \left(\frac{j-1}{J}, \frac{j}{J}\right], \quad 1 \leq j \leq J \right\}.$$

We consider the projection-posterior credible interval for $g(x_0)$ for g on a bounded domain. The posterior samples of g need not be monotone functions, so we impose monotonicity by projecting each sample onto $\mathscr{F}$ using the isotonization algorithm, to obtain the projection $g^* = \sum_{j=1}^{J} \theta_j^* \mathbb{1}_{I_j}$ where θ_j^* equals the left-derivative of the least concave majorant of the graph of

$$\left\{(0,0), \left(\frac{1}{J}, \frac{\theta_1}{J}\right), \left(\frac{2}{J}, \frac{\theta_1+\theta_2}{J}\right), \ldots, \left(\frac{J-1}{J}, \frac{\sum_{j=1}^{J-1}\theta_j}{J}\right), (1,1)\right\}, \tag{3.3.2}$$

at the point $\sum_{k=1}^{j} N_k/n$.

Define the unconditional probability $(\mathrm{P}_0 \times \Pi_n)$ as

$$(\mathrm{P}_0 \times \Pi_n)\big(g : g \in B\big) = \mathrm{E}_0\big(\Pi(g : g \in B|\boldsymbol{X})\big),$$

for every B measurable with respect to Π. The following result gives the unconditional asymptotic distributions of $n^{1/3}(g^*(x_0) - g_0(x_0))$ and $n^{1/3}(\hat{g}_n(x_0) - g_0(x_0))$.

Theorem 3.3.2. *Let $n^{1/3} \ll J \ll n^{2/3}$. Let W_1, W_2 be independent two-sided Brownian motions on $\mathbb{R}$ with $W_1(0) = W_2(0) = 0$, $Z_2 = \arg\max\{W_1(t) + W_2(t) - t^2 : t \in \mathbb{R}\}$, $Z_1 = \arg\max\{W_1(t) - t^2 : t \in \mathbb{R}\}$, and $C = 2b\,(a/b)^{2/3}$ with $a = \sqrt{g_0(x_0)}$ and $b = |g_0'(x_0)|/2$, the following assertions hold.*

(a) For every $z \in \mathbb{R}$, $\mathrm{P}_0(n^{1/3}(\hat{g}_n(x_0) - g_0(x_0)) \le z) \to \mathrm{P}(CZ_1 \le z)$.

(b) For every $z \in \mathbb{R}$, $\big(\mathrm{P}_0 \times \Pi_n\big)(n^{1/3}(g^*(x_0) - g_0(x_0)) \le z) \to \mathrm{P}(CZ_2 \le z)$.

(c) The conditional distribution $\Pi(n^{1/3}(g^*(x_0) - \hat{g}_n(x_0))\big|\boldsymbol{X})$ does not have a limit in probability.

Next, let W_1, W_2 be independent two-sided standard Brownian motions starting at zero, $a = \sqrt{g_0(x_0)}$, and $b = |g_0'(x_0)|/2$. Define stochastic processes F_n^* and $F_{a,b}^*$ on $\mathbb{R}$ as

$$\begin{aligned} F_n^*(z|\boldsymbol{X}) &= \Pi\big(n^{1/3}(g^*(x_0) - g_0(x_0)) \le z \,|\boldsymbol{X}\big), \\ F_{a,b}^*(z|W_1) &= \mathrm{P}\left(2b(a/b)^{2/3} \operatorname*{arg\,max}_{t\in\mathbb{R}} V(t) \le z \,|W_1\right), \end{aligned} \tag{3.3.3}$$

where $V(t) = W_1(t) + W_2(t) + t^2$. For every $n \ge 1$, $\gamma \in [0, 1/2]$, define

$$Q_{n,\gamma} = \inf\big\{z \in \mathbb{R} : \Pi(g^*(x_0) \le z|\boldsymbol{X}) \ge 1 - \gamma\big\},$$

$$I_{n,\gamma} = [Q_{n,1-\frac{\gamma}{2}}, Q_{n,\frac{\gamma}{2}}],$$
$$\Delta^*_{W_1,W_2} = \arg\max_{t\in\mathbb{R}} \left\{ W_1(t) + W_2(t) - t^2 \right\}. \tag{3.3.4}$$

We then have the following result.

Theorem 3.3.3. *Let $F^*_n(\cdot)$, $F^*_{a,b}(\cdot)$, $Q_{n,\gamma}$, $I_{n,\gamma}$ and $\Delta^*_{W_1,W_2}$ be as defined in* (3.3.3)–(3.3.4)*, with $a = \sqrt{g_0(x_0)}$ and $b = |g_0'(x_0)|/2$. Then the following assertions hold.*

(a) For every $z \in \mathbb{R}$, $F^*_n(z|\boldsymbol{X}) \rightsquigarrow F^*_{a,b}(z|W_1)$;

(b) the distribution of $F^*_{a,b}(0|W_1)$ is symmetric about $1/2$;

(c) the limiting coverage of $I_{n,\gamma}$ is characterized as follows:

$$\mathrm{P}_0\big(g_0(x_0) \in I_{n,\gamma}\big) \to \mathrm{P}\left(\frac{\gamma}{2} \leq \mathrm{P}(\Delta^*_{W_1,W_2} \geq 0 \Big| W_1) \leq 1 - \frac{\gamma}{2}\right).$$

Similar to the case of regression (Corollary 2.3.3), a re-calibration of the credibility levels can give a credible interval with the right coverage.

3.4 Simulation

In this section we study the behavior of point-wise credible intervals and test for monotonicity in finite samples. We first compare finite sample performance of Bayesian credible intervals with confidence intervals in terms of their coverages of the true parameter $g_0(x_0)$. We consider the density of Beta(1,3) as the true g_0. We take J to be the greatest integer less than or equal to $n^{1/3}\log(n)$. We vary sample size n across four different values. For each n, we consider 1000 instances of the data $\boldsymbol{X}$. For each instance of data, we generate 1000 posterior samples of $\boldsymbol{\theta}$, isotonize $\boldsymbol{\theta}$ to obtain the corresponding projection $g^*(x_0)$ for each simulated value $g_0(x_0)$ from the unrestricted posterior. We also compute $\hat{g}_n(x_0)$ for every replication. We then obtain the recalibrated $100(1-\alpha)\%$ credible interval using the $A^{-1}(1-\alpha/2)$ and $1-A^{-1}(1-\alpha/2)$ quantiles of the point-wise projection-posterior, and check if it contains $g_0(x_0)$. The coverage is determined by the proportion of times the credible interval contains $g_0(x_0)$. We also compute the $100(1-\alpha)\%$ confidence intervals from $\hat{g}_n(x_0)$ based on the quantiles of Chernoff's distribution and the estimated value of the constant C that appears in Theorem 3.3.2. We also obtain the $100(1-\alpha)\%$ credible interval using the $\alpha/2$ and $1-\alpha/2$ projection-posterior quantiles. Let $C_B(\alpha)$, $C^*_B(\alpha)$ and $C_F(\alpha)$ respectively denote

Table 3.1: Comparison of obtained coverage and average length of Bayesian credible intervals for $g(x_0)$, with that obtained from a confidence interval.

$1-\alpha$	n=500						n=1000					
	$C_B(\alpha)$	$L_B(\alpha)$	$C_B^*(\alpha)$	$L_B^*(\alpha)$	$C_F(\alpha)$	$L_F(\alpha)$	$C_B(\alpha)$	$L_B(\alpha)$	$C_B^*(\alpha)$	$L_B^*(\alpha)$	$C_F(\alpha)$	$L_F(\alpha)$
0.99	0.998	0.44	0.983	0.43	0.992	0.51	0.996	0.36	0.993	0.35	0.984	0.43
0.98	0.985	0.44	0.969	0.40	0.985	0.49	0.991	0.36	0.981	0.31	0.979	0.39
0.95	0.956	0.38	0.935	0.35	0.957	0.42	0.967	0.30	0.951	0.28	0.952	0.34
0.92	0.929	0.34	0.913	0.32	0.928	0.38	0.939	0.27	0.929	0.26	0.918	0.30
0.90	0.911	0.32	0.893	0.30	0.907	0.32	0.929	0.26	0.900	0.27	0.894	0.26
0.86	0.876	0.29	0.853	0.26	0.835	0.30	0.881	0.26	0.857	0.29	0.828	0.26
$1-\alpha$	n=1500						n=2000					
	$C_B(\alpha)$	$L_B(\alpha)$	$C_B^*(\alpha)$	$L_B^*(\alpha)$	$C_F(\alpha)$	$L_F(\alpha)$	$C_B(\alpha)$	$L_B(\alpha)$	$C_B^*(\alpha)$	$L_B^*(\alpha)$	$C_F(\alpha)$	$L_F(\alpha)$
0.99	0.991	0.32	0.985	0.30	0.989	0.38	0.997	0.31	0.988	0.28	0.993	0.34
0.98	0.984	0.31	0.979	0.28	0.984	0.34	0.988	0.29	0.976	0.26	0.984	0.31
0.95	0.967	0.27	0.949	0.25	0.953	0.29	0.968	0.25	0.939	0.23	0.951	0.27
0.92	0.939	0.24	0.916	0.23	0.921	0.26	0.932	0.22	0.917	0.21	0.911	0.24
0.90	0.914	0.23	0.894	0.21	0.905	0.25	0.914	0.21	0.895	0.19	0.889	0.23
0.86	0.881	0.2	0.860	0.19	0.842	0.21	0.879	0.19	0.850	0.17	0.828	0.19

the coverage of $(Q_{n,1-\alpha/2}, Q_{n,\alpha/2})$, $(Q_{n,1-A^{-1}(\alpha/2)}, Q_{n,A^{-1}(\alpha/2)})$, and that of level $(1-\alpha)$ confidence interval, and $L_B(\alpha)$, $L_B^*(\alpha)$ and $L_F(\alpha)$ respectively denote their lengths.

3.5 Proofs

Proof of Theorem 3.3.2. We shall prove parts (b) and (c). Part (a) is given in Theorem A.2.4.

(b) Since g^* is piece-wise constant on each I_j, $g^*(x_0) = \theta^*_{\lceil x_0 J \rceil}$. Let $z \in \mathbb{R}$. We shall evaluate $\big(\mathrm{P}_0 \times \Pi_n\big)(n^{1/3}(g^*(x_0) - g_0(x_0)) \le z)$ as $n \to \infty$. Note that $\big(\mathrm{P}_0 \times \Pi_n\big)(n^{1/3}(g^*(x_0) - g_0(x_0)) \le z) = \big(\mathrm{P}_0 \times \Pi_n\big)(\theta^*_{\lceil x_0 J \rceil} \le g_0(x_0) + n^{-1/3} z)$. Let $c(\cdot)$ denote the graph of

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

with $c(s) = 0$ for $s \le 0$ and $c(s) = \sum_{k=1}^{J} \theta_k$ for $s \ge 1$. Let V_n, G_n be stochastic processes denoting $V_n(s) = \sum_{j=1}^{\lceil sJ \rceil} \theta_j$ and $G_n(s) = \lceil sJ \rceil / J$. Now 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 (B.2), the fact that $V_n(s) = G_n(s) = 0$ for $s \le 0$ and that the location of minimum does not change upon adding a constant term or upon multiplication by a positive constant, we have that $g^*(x_0) \le g_0(x_0) + n^{-1/3} z$ if and only if $\arg\max\{c(s) - (g_0(x_0) + n^{-1/3} z)s : s \ge 0\} \ge \lceil x_0 J \rceil / J$. Hence the

probability of this event can be written as

$$
\begin{aligned}
&\big(\mathrm{P}_0\times\Pi_n\big)\Big(\operatorname*{arg\,max}_{s\in\mathbb{R}}\big\{c(s)-(g_0(x_0)+n^{-1/3}z)s\big\}\geq\lceil x_0J\rceil J^{-1}\Big)\\
&=\big(\mathrm{P}_0\times\Pi_n\big)\Big(\operatorname*{arg\,max}_{s\in\mathbb{R}}\big\{V_n(s)-(g_0(x_0)+n^{-1/3}z)G_n(s)\big\}\geq J^{-1}\lceil x_0J\rceil\Big)\\
&=\big(\mathrm{P}_0\times\Pi_n\big)\Big(\operatorname*{arg\,max}_{s\in\mathbb{R}}\big\{n^{2/3}(V_n(s)-V_n(x_0))\\
&\qquad -n^{2/3}(g_0(x_0)+n^{-1/3}z)(G_n(s)-G_n(x_0))\big\}\geq J^{-1}\lceil x_0J\rceil\big).
\end{aligned}
\tag{3.5.1}
$$

We use a change of variable $s=x_0+n^{-1/3}t$. For notational convenience, we use the notation $\sum_{j=\lceil x_0J\rceil+1}^{\lceil(x_0+n^{-1/3}t)J\rceil}\theta_j$ to denote $\sum_{j=\lceil x_0J\rceil+1}^{\lceil(x_0+n^{-1/3}t)J\rceil}\theta_j$ when $t\geq 0$ and $\sum_{j=\lceil(x_0+n^{-1/3}t)J\rceil+1}^{\lceil x_0J\rceil}\theta_j$ when $t<0$. A similar rule shall be applied for $\sum_{j=\lceil x_0J\rceil+1}^{\lceil(x_0+n^{-1/3}t)J\rceil}J^{-1}$. For $\tau_n=n^{1/3}(J^{-1}\lceil x_0J\rceil-x_0)$, the last expression in (3.5.1) then equals to

$$
\begin{aligned}
&\big(\mathrm{P}_0\times\Pi_n\big)\Big(\operatorname*{arg\,max}_{t\in\mathbb{R}}\big\{n^{2/3}\big(V_n(x_0+n^{-1/3}t)-V_n(x_0)\big)\\
&\qquad -n^{2/3}\big(g_0(x_0)+n^{-1/3}z\big)\big(G_n(x_0+n^{-1/3}t)-G_n(x_0)\big)\big\}\geq\tau_n\big)\\
&=\big(\mathrm{P}_0\times\Pi_n\big)\Bigg(\operatorname*{arg\,max}_{t\in\mathbb{R}}\Bigg\{n^{2/3}\sum_{j=\lceil x_0J\rceil+1}^{\lceil(x_0+n^{-1/3}t)J\rceil}(\theta_j-J^{-1}g_0(x_0))\\
&\qquad\qquad -n^{1/3}zJ^{-1}(\lceil(x_0+n^{-1/3}t)J\rceil-\lceil x_0J\rceil-1)\big\}\geq\tau_n\big).
\end{aligned}
\tag{3.5.2}
$$

Now note that $n^{1/3}zJ^{-1}(\lceil(x_0+n^{-1/3}t)J\rceil-\lceil x_0J\rceil-1)$ is equal to

$$
n^{1/3}zJ^{-1}(x_0J+n^{-1/3}tJ-x_0J+O(1))=zt+O(n^{1/3}J^{-1}),
$$

which goes to zt as $J\gg n^{1/3}$. Using similar arguments, we obtain that $\tau_n\to 0$.

Using Lemma 3.5.1 and 3.5.3, Theorem B.1.6, and Lemma A.2.8, we rewrite (3.5.2) to obtain

$$
\begin{aligned}
&\big(\mathrm{P}_0\times\Pi_n\big)(n^{1/3}(g^*(x_0)-g_0(x_0))\leq z)\\
&\to\mathrm{P}\Big(\operatorname*{arg\,max}_{t\in\mathbb{R}}\big\{aW_1(t)+aW_2(t)-bt^2-zt\big\}\geq 0\Big)\\
&=\mathrm{P}\Big((a/b)^{2/3}\operatorname*{arg\,max}_{t\in\mathbb{R}}\big\{W_1(t)+W_2(t)-t^2\big\}+z/2b\geq 0\Big)\\
&=\mathrm{P}\Big(2b(a/b)^{2/3}\operatorname*{arg\,max}_{t\in\mathbb{R}}\big\{W_1(t)+W_2(t)-t^2\big\}\geq -z\Big)
\end{aligned}
$$

$$= \mathrm{P}\left(2b(a/b)^{2/3} \operatorname*{arg\,max}_{t\in\mathbb{R}} \left\{ W_1(t) + W_2(t) - t^2 \right\} \le z \right),$$

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

$$\left(\mathrm{P}_0 \times \Pi_n\right)\left(n^{1/3}(g^*(x_0) - g_0(x_0)) \le z\right) \to \mathrm{P}\left(C \operatorname*{arg\,max}_{t\in\mathbb{R}} \left\{ W_1(t) + W_2(t) - t^2 \right\} \le z \right).$$

This completes the proof of part (b).

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

$$H_1(t_1, z_1) = \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t_1) J \rceil} \left(\frac{N_j}{n} - \frac{g_0(x_0)}{J} \right) - n^{-1/3} z_1 \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t_1) J \rceil} \frac{1}{J},$$

$$H_2(t_2, z_2) = \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t_2) J \rceil} \left(\theta_j - \frac{g_0(x_0)}{J} \right) - n^{-1/3} z_2 \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t_2) J \rceil} \frac{1}{J}.$$

Then from the proofs of part (a) and (b), and Theorem B.1.7 we obtain that

$$\begin{aligned} &\left(\operatorname*{arg\,max}_{t_1\in\mathbb{R}} \{H_1(t_1, z_1)\}, \operatorname*{arg\,max}_{t_2\in\mathbb{R}} \{H_2(t_2, z_2)\} \right) \\ &\rightsquigarrow \left(\operatorname*{arg\,max}_{t_1\in\mathbb{R}} \left\{ a\, W_1(t_1) - b\, t_1^2 - z_1 t_1 \right\}, \right. \\ &\qquad\qquad \left. \operatorname*{arg\,max}_{t_2\in\mathbb{R}} \left\{ a\, W_1(t_2) + a\, W_2(t_2) - b\, t_2^2 - y\, t_2 \right\} \right). \end{aligned} \tag{3.5.3}$$

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 (3.5.3) and Lemma A.2.8, we get that for all $z \in \mathbb{R}$,

$$\left(\mathrm{P}_0 \times \Pi_n\right)\left(n^{1/3}(g^*(x_0) - \hat{g}_n(x_0)) \le z\right) \to \mathrm{P}\left(CZ_2 - CZ_1 \le z\right).$$

From the proof of part (a) of Theorem 3.3.3, we also obtain that

$$\Pi\left(n^{1/3}(g^*(x_0) - \hat{g}_n(x_0)) \le z | \boldsymbol{X}\right) \rightsquigarrow \mathrm{P}\left(CZ_2 - CZ_1 \le z | W_1\right).$$

If $\Pi(n^{1/3}(g^*(x_0) - \hat{g}_n(x_0)) \le z | \boldsymbol{X})$ converges in probability, then the limit must be the random variable $Q = \mathrm{P}(CZ_2 - CZ_1 \le z | W_1)$. Using Theorem B.4.2 for $W_n = n^{1/3}(g^*(x_0) - \hat{g}_n(x_0))$ and $W_n^* = \Pi(n^{1/3}(g^*(x_0) - \hat{g}_n(x_0)) \le z | \boldsymbol{X})$, we obtain that if W_n^* converges in probability, then

$(Z_2 - Z_1)$ and Q are independent. This is clearly not true, since both the random variables are dependent on W_1. This completes the proof of part (c). □

Lemma 3.5.1. *Let $a = \sqrt{g_0(x_0)}$, $b = |g_0'(x_0)|/2$, and W_1, W_2 be independent two-sided Brownian motions on $\mathbb{R}$ starting at zero. Then for $n^{1/3} \ll J \ll n^{2/3}$, the process*

$$\left\{ n^{2/3} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t) J \rceil} (\theta_j - J^{-1} g_0(x_0)), \quad t \in [-K, K] \right\},$$

converges in $L^\infty([-K,K])$ to $\{a W_1(t) + a W_2(t) - b\, t^2, t \in [-K,K]\}$, for all $K > 0$.

Proof of Lemma 3.5.1. Since $\theta_j | \boldsymbol{X} \sim \text{Dir}(\alpha_1 + N + 1, \ldots, \alpha_J + N_J)$, for every $1 \le j \le J$, we write θ_j as

$$\theta_j = \frac{Z_j}{\sum_{l=1}^J Z_l}, \qquad \text{where} \qquad Z_j \sim \text{Gamma}(\alpha_j + N_j, 1). \tag{3.5.4}$$

Using the representation in (3.5.4), we define

$$\begin{aligned} A_n(\theta, \boldsymbol{X}, t) &= n^{2/3} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t) J \rceil} \frac{Z_j - \mathrm{E}(Z_j)}{\sum_{l=1}^J Z_l}, \\ \bar{A}_n(\boldsymbol{X}, t) &= n^{2/3} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t) J \rceil} \left(\frac{\mathrm{E}(Z_j)}{\sum_{l=1}^J Z_l} - \frac{N_j}{n} \right), \\ B_n(\boldsymbol{X}, t) &= n^{2/3} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t) J \rceil} \left(\frac{N_j}{n} - \frac{g_0(x_0)}{J} \right). \end{aligned} \tag{3.5.5}$$

Therefore, $n^{2/3} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t) J \rceil} (\theta_j - J^{-1} g_0(x_0)) = A_n(\theta, \boldsymbol{X}, t) + \bar{A}_n(\boldsymbol{X}, t) + B_n(\boldsymbol{X}, t)$. We shall show that $\{A_n(\theta, \boldsymbol{X}, t), t \in [-K,K]\}$ and $\{\bar{A}_n(\boldsymbol{X}, t), t \in [-K,K]\}$ converge in $L^\infty([-K,K])$ for all $K > 0$, to $\{a W_2(t) : t \in [-K,K]\}$ and 0 respectively. To prove $A_n(\theta, \boldsymbol{X}, \cdot) \rightsquigarrow a W_2$, we use the following lemma about the tightness of n^{-1} times the numerator of $A_n(\theta, \boldsymbol{X}, \cdot)$.

Lemma 3.5.2. *Let $Y_n(\theta, \boldsymbol{X}, t) = n^{-1/3} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t) J \rceil} (Z_j - \mathrm{E}(Z_j)$ for $t \in [0, K]$. Then $\{Y_n(\theta, \boldsymbol{X}, \cdot)\}$ is tight with respect to the joint probability of $\boldsymbol{\theta}$ and $\boldsymbol{X}$.*

To show that $Y_n(\theta, \boldsymbol{X}, t)$ converges in distribution to $a W_2(t)$ for every $t \in [-K,K]$, note

that since $N_j \sim \text{Bin}(n; G_0(I_j))$ for $1 \le j \le J$, $\sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} N_j$ is equal to

$$n\int_{(\lceil x_0 J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}t)J\rceil/J} g_0(u)du + O_{P_0}\left(\sqrt{n\int_{(\lceil x_0 J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}t)J\rceil/J} g_0(u)du}\right).$$

As g_0 is differentiable in a neighborhood of x_0, the first term of the last display can be written as

$$\begin{aligned} n\int_{(\lceil x_0 J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}t)J\rceil/J} & [g_0(x_0)+g_0'(x_0)(u-x_0)+o((u-x_0)^2)]du \\ &= n^{2/3}g_0(x_0)t + n^{1/3}g_0'(x_0)t^2 + O_{P_0}(nJ^{-1}). \end{aligned} \tag{3.5.6}$$

Now we define s_n^2 as

$$\sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} \text{Var}\left[n^{-1/3}(Z_j - \text{E}(Z_j))\middle|\boldsymbol{X}\right] = n^{-2/3}\sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} \left(\alpha_j + N_j\right). \tag{3.5.7}$$

Using (3.5.6) and the fact that α_j's are uniformly bounded, s_n^2 is of the order

$$\begin{aligned} n^{-2/3}\sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} N_j &= n^{-2/3}\left[n^{2/3}g_0(x_0)t + n^{1/3}g_0'(x_0)t^2 + O_{P_0}(nJ^{-1})\right] \\ &= g_0(x_0)t + o(n^{-1/3}) + O_{P_0}(n^{1/3}J^{-1}), \end{aligned} \tag{3.5.8}$$

which goes to $g_0(x_0)t$ because $J \gg n^{1/3}$. We now verify Lyapunov's condition with $\delta = 2$, to apply Lindeberg's central limit theorem to conclude that $s_n^{-1}Y_n(\theta, \boldsymbol{X}, t) \rightsquigarrow \text{N}(0,1)$, for every $t \in [-K, K]$. Let σ_j denote $\alpha_j + N_j$ for $1 \le j \le J$. As $Z_j \sim \text{Gamma}(\alpha_j + N_j, 1)$, its moment generating function at u is $(1-u)^{-(\alpha_j+N_j)}$. We therefore obtain that

$$\begin{aligned} \text{E}\left[(Z_j - \text{E}(Z_j))^4\middle|\boldsymbol{X}\right] &= \sigma_j(\sigma_j+1)(\sigma_j+2)(\sigma_j+3) - 4\sigma_j^2(\sigma_j+1)(\sigma_j+2) \\ &\qquad + 6\sigma_j^3(\sigma_j+1) - 4\sigma_j^4 + \sigma_j^4 \\ &= 3\sigma_j^2 + 9\sigma_j. \end{aligned}$$

Therefore, $\sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} \mathrm{E}\big[(n^{-1/3}(Z_j-\mathrm{E}(Z_j)))^4\big|\boldsymbol{X}\big]$ is equal to

$$3n^{-4/3}\sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil}(\alpha_j+N_j)^2+9n^{-4/3}\sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil}(\alpha_j+N_j)$$
$$\lesssim n^{-4/3}\left[\sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil}(N_j^2+N_j)\right]$$

By the monotonicity assumption on g_0, and arguments similar to those used in (2.6.7), we get that $\sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} \mathrm{E}\big[(n^{-1/3}(Z_j-\mathrm{E}(Z_j)))^4\big|\boldsymbol{X}\big]$ is bounded by a constant multiple of

$$n^{-4/3}\times n^{-1/3}tJ\times(nJ^{-1}+O_{P_0}(nJ^{-1}))=n^{-2/3}t+O_{P_0}(n^{-2/3}t). \tag{3.5.9}$$

By using (3.5.8) and (3.5.9), we obtain that

$$s_n^{-4}\sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil}\mathrm{E}\big[(n^{-1/3}(Z_j-\mathrm{E}(Z_j)))^4\big|\boldsymbol{X}\big]\to^{P_0}0,$$

and Lyapunov's condition is satisfied for $\delta=2$. Using (3.5.8), we have that $s_n\to^{P_0}\sqrt{g_0(x_0)t}$. Applying Lindeberg's central limit theorem on $Y_n(\boldsymbol{\theta},\boldsymbol{X},\cdot)$, we obtain that

$$\sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} n^{-1/3}(Z_j-\mathrm{E}(Z_j))=s_n\,\frac{\sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} n^{-1/3}(Z_j-\mathrm{E}(Z_j)}{s_n}$$
$$\rightsquigarrow\sqrt{g_0(x_0)t}\,\mathrm{N}(0,1). \tag{3.5.10}$$

For $t_1,t_2\in[-K,K]$, we evaluate the covariance between $Y_n(\theta,\boldsymbol{X},t_1)$ and $Y_n(\theta,\boldsymbol{X},t_2)$ as

$$\mathrm{E}_0\,Y_n(\theta,\boldsymbol{X},t_1)Y_n(\theta,\boldsymbol{X},t_2)=\mathrm{E}_0\mathrm{E}\big[\,Y_n(\theta,\boldsymbol{X},t_1)Y_n(\theta,\boldsymbol{X},t_2)\big|\boldsymbol{X}\big].$$

Let $0<t_1<t_2$. Then the covariance is equal to

$$\mathrm{E}_0\left[\mathrm{E}\sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t_1)J\rceil}(Z_j-\mathrm{E}(Z_j))\sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t_2)J\rceil}(Z_j-\mathrm{E}(Z_j))\big|\boldsymbol{X}\right]$$
$$=n^{4/3}\mathrm{E}_0\left[\sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t_1)J\rceil}\mathrm{Var}(Z_j)\right],$$

the last step following from the fact that Z_j are independent. Using arguments similar to (3.5.7) – (3.5.8), we obtain that

$$\mathrm{Cov}(Y_n(\theta, \boldsymbol{X}, t_1), Y_n(\theta, \boldsymbol{X}, t_2)) \to g_0(x_0)t_1.$$

Using similar arguments, we obtain that the limiting covariance between $Y_n(\theta, \boldsymbol{X}, t_1)$ and $Y_n(\theta, \boldsymbol{X}, t_2)$ asymptotically equals to zero if t_1 and t_2 are of the opposite signs, and $(t_1 \wedge t_2)g_0(x_0)$ if t_1 and t_2 are of same signs. Along with the tightness of $Y_n(\theta, \boldsymbol{X}, \cdot)$ concluded in Lemma 3.5.2, we obtain that

$$\{Y_n(\theta, \boldsymbol{X}, t) : t \in [-K, K]\} \rightsquigarrow \{a W_2(t) : t \in [-K, K]\}, \tag{3.5.11}$$

in $L^\infty([-K, K])$ for all $K > 0$. For the denominator of $A_n(\theta, \boldsymbol{X}, t)$, note that $\sum_{l=1}^J Z_l | \boldsymbol{X} \sim \mathrm{Gamma}(\alpha_0 + n, 1)$, and so $\mathrm{E}_0(\sum_{l=1}^J Z_l) = \alpha_0 + n$, and $\mathrm{Var}_0(\sum_{l=1}^J Z_l) = \alpha_0 + n$. By the assumptions on the prior, $\max\{\alpha_j : 1 \le j \le J\} = O(1)$, and therefore $\sum_{l=1}^J Z_l = n + O(J) + O_{P_0}(n^{1/2})$. Therefore,

$$n^{-1} \sum_{l=1}^{J} Z_l = 1 + O(n^{-1}J) + O_{P_0}(n^{-1/2}) \to^{P_0} 1. \tag{3.5.12}$$

Using (3.5.11) and (3.5.12), we obtain that,

$$\begin{aligned}
&\left\{ \frac{n^{2/3} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t) J \rceil} (Z_j - \mathrm{E}(Z_j))}{\sum_{l=1}^J Z_l}, \quad t \in [-K, K] \right\} \\
&\qquad = \left\{ \frac{n^{-1/3} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t) J \rceil} (Z_j - \mathrm{E}(Z_j))}{n^{-1} \sum_{l=1}^J Z_l}, \quad t \in [-K, K] \right\} \\
&\qquad \rightsquigarrow \{a W_2(t), \quad t \in [-K, K]\},
\end{aligned}$$

in $L^\infty([-K, K])$, for all $K > 0$.

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

$$\begin{aligned}
&n^{2/3} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t) J \rceil} \left(\frac{\alpha_j + N_j}{\sum_{l=1}^J Z_l} - \frac{N_j}{\sum_{l=1}^J Z_l} + \frac{N_j}{\sum_{l=1}^J Z_l} - \frac{N_j}{n} \right) \\
&\le \frac{n^{2/3}}{\sum_{l=1}^J Z_l} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t) J \rceil} \max_{1 \le j \le J} \alpha_j + n^{2/3} \left(\frac{n}{\sum_{l=1}^J Z_l} - 1 \right) \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t) J \rceil} \frac{N_j}{n}.
\end{aligned} \tag{3.5.13}$$

By the assumptions on the prior and (3.5.12), the first term of (3.5.13) is bounded in probability by a constant multiple of $n^{-2/3}JK$. For the second term of the display, note that, by (3.5.6),

$$\sup_{t\in[-K,K]}\left\{n^{2/3}\sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil}\frac{N_j}{n}\right\}=n^{1/3}g_0(x_0)K+O_{P_0}(1).$$

Also, since $\sum_{l=1}^{J}Z_l\sim\text{Gamma}(\alpha_0+n,1)$, we have that $((n/\sum_{l=1}^{J}Z_l)-1)$ is $O_{P_0}(n^{-1/2})$. Taking the supremum over $[-K,K]$ on (3.5.13), we obtain that

$$\begin{aligned}\sup_{t\in[-K,K]}\bar{A}_n(\boldsymbol{X},t)&=o_{P_0}(n^{-2/3}JK)+O_{P_0}(n^{-1/2})(n^{1/3}g_0(x_0)K+O_{P_0}(1))\\&=o(1)+O_{P_0}(n^{-1/6})+O_{P_0}(n^{-1/2}).\end{aligned}$$

Therefore, $\bar{A}_n(\boldsymbol{X},\cdot)\to^{P_0}0$ in $L^\infty([-K,K])$. □

Lemma 3.5.3. *For $n^{1/3}\ll J\ll n^{2/3}$,*

$$\operatorname*{arg\,max}_{t\in\mathbb{R}}\left\{\sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil}\left(\theta_j-\frac{g_0(x_0)}{J}\right)-n^{-1/3}z\sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil}\frac{1}{J}\right\}=O_{P_0}(1).$$

Proof of Lemma 3.5.3. Using the change of variable $r=n^{-1/3}t$, we establish the tightness of $n^{1/3}\hat{h}_n$, where

$$\hat{h}_n=\operatorname*{arg\,max}_{r\in\mathbb{R}}\left\{\sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+r)J\rceil}\left(\theta_j-\frac{g_0(x_0)}{J}\right)-n^{-1/3}z\sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+r)J\rceil}\frac{1}{J}\right\}-x_0.$$

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

$$\begin{aligned}M_n(r)&=\sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+r)J\rceil}\left(\theta_j-\frac{g_0(x_0)}{J}\right)+\frac{zn^{-1/3}}{J}[\lceil(x_0+r)J\rceil-\lceil x_0 J\rceil],\\M(r)&=g_0(x_0+r)-g_0(x_0)-g_0'(x_0)r.\end{aligned}$$

Then $\hat{h}_n=\arg\max\{M_n(r):r\in\mathbb{R}\}-x_0$. We apply Theorem B.1.5 to show that

$$n^{1/3}(\arg\max\{M_n(r):r\in\mathbb{R}\}-x_0)=O_P(1),$$

with $\mathbb{M}_n=-M_n$, $\mathbb{M}=-M$, $\Theta=(-\infty,0]$, $\hat{\theta}_n=\arg\max_{r\in\mathbb{R}}M_n(r)$, $\theta_0=0$ and d the Euclidean

distance on $\mathbb{R}$. With these notations, we have $-M_n(0) = -M(0) = 0$ and $\hat{\theta}_n$ maximizes $-M_n$. The condition $\mathbb{M}(\theta) - \mathbb{M}(\theta_0) \lesssim -d^2(\theta, \theta_0)$ has been verified within the proof of Theorem 3.3.2, part (a). We need to construct functions ϕ_n such that

$$\sqrt{n}\, \mathrm{E}^* \sup_{|r|<\delta} \left| (\mathbb{M}_n - \mathbb{M})(\theta) - (\mathbb{M}_n - \mathbb{M})(\theta_0) \right| \lesssim \phi_n(\delta), \tag{3.5.14}$$

for sufficiently small δ, and $\phi_n(\delta)/\delta^\alpha$ is decreasing in δ for some $\alpha \in (0,2)$. Using the representations in (3.5.5), we write the left-hand side of (3.5.14) as $\sqrt{n}\, \mathrm{E}_0^* \sup\{|\sum_{l=1}^5 H_{ln}(r)| : |r| < \delta\}$, where $H_{ln}, 1 \le l \le 5$ are defined as

$$\begin{aligned}
H_{1n}(r) &= \sqrt{n} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0+r)J \rceil} \left(\frac{Z_j - \mathrm{E}(Z_j)}{\sum_{l=1}^J Z_l} \right), \\
H_{2n}(r) &= \sqrt{n} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0+r)J \rceil} \left(\frac{\mathrm{E}(Z_j)}{\sum_{l=1}^J Z_l} - \frac{N_j}{n} \right), \\
H_{3n}(r) &= \sqrt{n} \mathbb{P}_n [\mathbb{1}\{X \in (\lceil x_0 J \rceil / J, \lceil (x_0 + r) J \rceil / J]\} \\
&\qquad\qquad - \mathbb{1}\{X \in (x_0, x_0 + r]\}], \\
H_{4n}(r) &= \sqrt{n} [(\mathbb{P}_n - \mathrm{E}_0) \mathbb{1}\{X \in (x_0, x_0 + r]\}], \\
H_{5n}(r) &= \frac{z n^{1/6}}{J} [\lceil (x_0 + r) J \rceil - \lceil x_0 J \rceil].
\end{aligned} \tag{3.5.15}$$

By arguments similar to those for proving Lemma 3.5.2, for

$$\rho_n(r_1, r_2) = [\mathrm{E}(H_{1n}(r_1) - H_{1n}(r_2))^2 | (\boldsymbol{X})]^{1/2},$$

and $\rho(r_1, r_2) = |r_1 - r_2|^{1/2}$, for sufficiently small δ,

$$\sqrt{n} \mathrm{E}_0^* \sup_{|r|<\delta} \left| \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0+r)J \rceil} \left(\frac{Z_j - \mathrm{E}(Z_j)}{\sum_{l=1}^J Z_l} \right) \right| \lesssim \int_0^{\sqrt{\delta'}/2} \epsilon^{-1/4} \sqrt{\log(1/\epsilon)}\, d\epsilon,$$

which is bounded by a constant multiple of $\delta^{1/4}$. For H_{2n}, we have, using arguments similar to (3.5.13), for sufficiently small δ,

$$\sqrt{n} \mathrm{E}_0^* \sup_{|r|<\delta} \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0+r)J \rceil} \left| \frac{\alpha_j + N_j}{\sum_{l=1}^J Z_l} - \frac{N_j}{n} \right| \lesssim n^{-1/2} J \delta,$$

which is bounded by a constant multiple of $n^{1/6}\delta$, as $J \ll n^{2/3}$. Using the last two displays

and the proof of Lemma 3.5.5 provided in Appendix A, we have that for sufficiently small δ, (3.5.14) holds for $\phi_n(\delta) = C[\delta^{1/4} + \delta^{1/2} + n^{1/6}\delta + n^{-1/3}]$ for some $C > 0$. The rest of the proof follows similarly to that of Lemma 3.5.5, with $r_n = n^{1/3}$ and the condition of uniform consistency of $g^*(x)$ in a neighborhood of x_0 satisfied by the consistency of the constrained MLE near x_0 as shown by Prakasa Rao (1969), and the fact that

$$\Pi\left(\sup_{r\in[0,1]} \sum_{j=1}^{\lceil rJ \rceil} \left| \theta_j - \frac{N_j}{n} \right| > \epsilon \middle| \boldsymbol{X} \right) \to^{P_0} 0.$$

The difference in the slope of the greatest convex minorants of (3.3.1) and (3.3.2) therefore goes to zero in P_0-probability, uniformly in a neighborhood of x_0, and the uniform consistency of $g^*(x)$ in a neighborhood of x_0 follows. Therefore $n^{1/3}\hat{h}_n$ is $O_{P_0}(1)$. □

Proof of Theorem 3.3.3. (a) Let $z \in \mathbb{R}$. Let $A_n(\theta, \boldsymbol{X}, t)$, $\bar{A}_n(\boldsymbol{X}, t)$, and $B_n(\boldsymbol{X}, t)$ be as defined in (3.5.5). By using similar reasoning as in the proof of part (b) of Theorem 3.3.2, we can rewrite $F_n^*(z|\boldsymbol{X})$ as

$$\begin{aligned}
&\Pi\big(n^{1/3}(g^*(x_0) - g_0(x_0)) \le z \big| \boldsymbol{X}\big) \\
&\quad = \Pi\Bigg(\operatorname*{arg\,max}_{t\in\mathbb{R}} \Big\{ A_n(\theta, \boldsymbol{X}, t) + \bar{A}_n(\boldsymbol{X}, t) + B_n(\boldsymbol{X}, t) \\
&\qquad\qquad - n^{1/3} z \sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3} t) J \rceil} \frac{1}{J} \Big\} \ge \tau_n \Big| \boldsymbol{X} \Bigg).
\end{aligned}$$

In the proof of Lemma 3.5.1 it has been shown that $\sup\{|\bar{A}_n(\boldsymbol{X}, t)| : t \in [-K, K]\} \to^{P_0} 0$, and the conditional distribution of $\{A_n(\theta, \boldsymbol{X}, t) : t \in [-K, K]\}$ given $\boldsymbol{X}$ converges weakly to $\{a W_2(t) : t \in [-K, K]\}$ in $L^\infty([-K, K])$ for all $K > 0$. Also, from the proof of part (a) of Theorem 3.3.2, the process $\{B_n(\boldsymbol{X}, t) : t \in [-K, K]\}$ converges to $\{a W_1 : t \in [-K, K]\}$ in $L^\infty([-K, K])$. As W_1 and W_2 are independent, the posterior distribution of $A_n(\theta, \boldsymbol{X}, \cdot) + B_n(\boldsymbol{X}, \cdot)$ given $\boldsymbol{X}$ converges weakly to the conditional distribution of $a(W_1 + W_2)$ given W_1. We therefore obtain that for every $T \in \mathbb{R}$ and all $K > 0$,

$$\begin{aligned}
&\big\{ \Pi(A_n(\theta, \boldsymbol{X}, t) + \bar{A}_n(\boldsymbol{X}, t) + B_n(\boldsymbol{X}, t) - z\,t \le T \big| \boldsymbol{X}), \quad t \in [-K, K] \big\} \\
&\qquad \rightsquigarrow \big\{ \mathrm{P}\left(a W_1(t) + a W_2(t) + b\,t^2 - z\,t \le T \middle| W_1 \right), \quad t \in [-K, K] \big\},
\end{aligned}$$

in $L^\infty([-K, K])$. Adapting the arguments used in the proof of Lemma 3.5.3 to the posterior distribution of $\hat{h}_n$ given $\boldsymbol{X}$, we have that the conditional distribution of $\operatorname{arg\,max}\{A_n(\theta, \boldsymbol{X}, t) +$

$\bar{A}_n(\boldsymbol{X},t)+B_n(\boldsymbol{X},t)-zt\}$ given $\boldsymbol{X}$ is tight with respect to P_0. By the Argmax theorem and part (a) of Lemma A.2.8 for the conditional distribution, we get that

$$\begin{aligned}
&\Pi\big(n^{1/3}(g^*(x_0)-g_0(x_0))\le z\big|\boldsymbol{X}\big)\\
&\quad\rightsquigarrow \mathrm{P}\left(\operatorname*{arg\,max}_{t\in\mathbb{R}}\big\{aW_1(t)+aW_2(t)-bt^2-zt\big\}\ge 0\Big|W_1\right)\\
&\quad=\mathrm{P}\left((a/b)^{2/3}\operatorname*{arg\,max}_{t\in\mathbb{R}}\big\{W_1(t)+W_2(t)-t^2\big\}+z/(2b)\ge 0\Big|W_1\right)\\
&\quad\overset{d}{=}\mathrm{P}\left(C\operatorname*{arg\,max}_{t\in\mathbb{R}}\big\{W_1(t)+W_2(t)-t^2\big\}\le z\Big|W_1\right).
\end{aligned}$$

The last step follows from the fact that $\operatorname{arg\,max}\big\{W_1(t)+W_2(t)-t^2:t\in\mathbb{R}\big\}$ is symmetric about zero. This completes the proof of part (a).

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

$$\begin{aligned}
&\mathrm{P}\left(\operatorname*{arg\,max}_{t\in\mathbb{R}}\{W_1(t)+W_2(t)-t^2\}\ge 0\Big|W_1\right)\\
&\qquad=\mathrm{P}\left(\operatorname*{arg\,max}_{t\in\mathbb{R}}\{W_1(-t)+W_2(-t)-(-t)^2\}\le 0\Big|W_1\right)\\
&\qquad\overset{d}{=}\mathrm{P}\left(\operatorname*{arg\,max}_{t\in\mathbb{R}}\{W_1'(t)+W_2(t)-t^2\}\le 0\Big|W_1'\right).
\end{aligned}$$

Therefore $\mathrm{P}(\Delta^*_{W_1,W_2}\ge 0\big|W_1)\overset{d}{=}\mathrm{P}(\Delta^*_{W_1,W_2}\le 0\big|W_1)$, and $\mathrm{P}(\Delta^*_{W_1,W_2}\ge 0\big|W_1)$ is symmetric about $1/2$.

(c) By the definition of posterior quantile $Q_{n,\gamma}$, we have that $g_0(x_0)\le Q_{n,\gamma}$ if and only if $\Pi(g^*(x_0)\le g_0(x_0)|\boldsymbol{X})\le 1-\gamma$. Therefore,

$$\begin{aligned}
\mathrm{P}_0\big(g_0(x_0)\le Q_{n,\gamma}\big)&=\mathrm{P}_0\big(\Pi(g^*(x_0)\le g_0(x_0)|\boldsymbol{X})\le 1-\gamma\big)\\
&=\mathrm{P}_0\big(\Pi(n^{1/3}(g^*(x_0)-g_0(x_0))\le 0|\boldsymbol{X})\le 1-\gamma\big)\\
&=\mathrm{P}_0\big(F_n^*(0\big|\boldsymbol{X})\le 1-\gamma\big)\\
&\to\mathrm{P}\big(g^*_{a,b}(0\big|W_1)\le 1-\gamma\big),
\end{aligned}$$

where the last step follows from part (a), substituting $z=0$. Now rewriting $g^*_{a,b}(0\big|W_1)$ in terms of $\Delta^*_{W_1,W_2}$, we have that $\mathrm{P}\big(g^*_{a,b}(0\big|W_1)\le 1-\gamma\big)$ is equal to

$$\mathrm{P}\left(\mathrm{P}\left(C\operatorname*{arg\,max}_{t\in\mathbb{R}}\big\{W_1(t))+W_2(t)-t^2\big\}\ge 0\Big|W_1\right)\le 1-\gamma\right)$$

$$= \mathrm{P}\left(\mathrm{P}\left(\Delta^*_{W_1,W_2} \geq 0 \middle| W_1\right) \leq 1-\gamma\right).$$

Similarly $\mathrm{P}_0\left(g_0(x_0) \geq Q_{n,1-\gamma}\right) = 1 - \mathrm{P}_0\left(g_0(x_0) < Q_{n,1-\gamma}\right)$, and therefore as $n \to \infty$,

$$\mathrm{P}_0\left(g_0(x_0) \geq Q_{n,1-\gamma}\right) \to 1 - \mathrm{P}\left(\mathrm{P}(\Delta^*_{W_1,W_2} \geq 0 \middle| W_1) < \gamma\right) = \mathrm{P}\left(\mathrm{P}(\Delta^*_{W_1,W_2} \geq 0 \middle| W_1) \geq \gamma\right).$$

□

Let U_n, G_n be stochastic processes denoting $U_n(s) = \sum_{j=1}^{\lceil sJ \rceil}(N_j/n)$ and $G_n(s) = \sum_{j=1}^{\lceil sJ \rceil} J^{-1}$. From the switch relation (B.2), we evaluate $\mathrm{P}(n^{1/3}(\hat{g}_n(x_0) - g_0(x_0)) \leq x)$ for $z \in \mathbb{R}$ as

$$\begin{aligned}
&\mathrm{P}(n^{1/3}(\hat{g}_n(x_0) - g_0(x_0)) \leq z) \\
&= \mathrm{P}(\hat{g}_n(x_0) \leq g_0(x_0) + n^{1/3}z) \\
&= \mathrm{P}\left(\operatorname*{arg\,max}_{s \geq 0}\left\{U_n(s) - (g_0(x_0) + n^{-1/3}z)G_n(s)\right\} \geq \sum_{k=1}^{\lceil x_0 J \rceil} J^{-1}\right) \\
&= \mathrm{P}\left(\operatorname*{arg\,max}_{s \in \mathbb{R}}\left\{U_n(s) - (g_0(x_0) + n^{-1/3}z)G_n(s)\right\} \geq \tau_n\right) \\
&= \mathrm{P}\Bigg(\operatorname*{arg\,max}_{s \in \mathbb{R}}\Big\{n^{2/3}(U_n(s) - U_n(x_0)) - \\
&\qquad\qquad n^{2/3}(g_0(x_0) + n^{-1/3}z)(G_n(s) - G_n(x_0))\Big\} \geq \tau_n\Bigg).
\end{aligned}$$

The last expression in the display then equals to

$$\begin{aligned}
&\mathrm{P}\Bigg(\operatorname*{arg\,max}_{t \in \mathbb{R}}\Big\{n^{2/3}(U_n(x_0 + n^{-1/3}t) - U_n(x_0)) \\
&\qquad\qquad -n^{2/3}(g_0(x_0) + n^{-1/3}z)(G_n(x_0 + n^{-1/3}t) - G_n(x_0))\Big\} \geq \tau_n\Bigg) \\
&= \mathrm{P}\Bigg(\operatorname*{arg\,max}_{t \in \mathbb{R}}\Bigg\{n^{2/3}\sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3}t)J \rceil} \frac{N_j}{n} \\
&\qquad\qquad -n^{2/3}(g_0(x_0) + n^{-1/3}z)\sum_{j=\lceil x_0 J \rceil + 1}^{\lceil (x_0 + n^{-1/3}t)J \rceil} \frac{1}{J}\Bigg\} \geq \tau_n\Bigg).
\end{aligned}$$

In the proof of part (b), it was shown that $\tau_n \to 0$ and $n^{1/3}zJ^{-1}[\lceil (x_0 + n^{-1/3}t)J \rceil - \lceil x_0 J \rceil - 1] \to zt$. We shall therefore evaluate

$$\mathrm{P}(\hat{g}_n(x_0) \leq g_0(x_0) + n^{-1/3}z)$$

$$= \mathrm{P}\left(\operatorname*{arg\,max}_{t\in\mathbb{R}} \left\{ n^{2/3} \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} \left(\frac{N_j}{n} - \frac{g_0(x_0)}{J} \right) - zt \right\} \geq 0 \right).$$

From Lemma 3.5.4 and 3.5.5, Theorem B.1.6, and Lemma A.2.8, we have that

$$\begin{aligned}
\mathrm{P}(n^{1/3}(\hat{g}_n(x_0) - g_0(x_0)) \leq z) &\to \mathrm{P}\left(\operatorname*{arg\,max}_{t\in\mathbb{R}} \left\{ a W_1(t) - b t^2 - zt \right\} \geq 0 \right) \\
&= \mathrm{P}\left((a/b)^{2/3} \operatorname*{arg\,max}_{t\in\mathbb{R}} \left\{ W_1(t) - t^2 \right\} + z/2b \geq 0 \right) \\
&= \mathrm{P}\left(2b(a/b)^{2/3} \operatorname*{arg\,max}_{t\in\mathbb{R}} \left\{ W_1(t) - t^2 \right\} \geq -z \right) \\
&= \mathrm{P}\left(2b(a/b)^{2/3} \operatorname*{arg\,max}_{t\in\mathbb{R}} \left\{ W_1(t) - t^2 \right\} \leq z \right),
\end{aligned}$$

the last step following from the fact that $\operatorname{arg\,max}\left\{ W_1(t) - t^2 : t \in \mathbb{R} \right\}$ is symmetric about zero. Substituting $C = 2b(a/b)^{2/3}$ in the last expression of the display we get $\mathrm{P}(n^{1/3}(\hat{g}_n(x_0) - g_0(x_0)) \leq z) \to \mathrm{P}\left(C \operatorname{arg\,max}\left\{ W_1(t) - t^2 : t \in \mathbb{R} \right\} \leq z \right)$ as $n \to \infty$.

Lemma 3.5.4. *Let W_1 be a two sided Brownian motion on $\mathbb{R}$ starting at zero. Let $n^{1/3} \ll J \ll n^{2/3}$. Then for every $K > 0$,*

$$\left\{ n^{2/3} \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} \left(\frac{N_j}{n} - \frac{g_0(x_0)}{J} \right) : t \in [-K, K] \right\} \rightsquigarrow \left\{ a W_1(t) - b t^2 : t \in [-K, K] \right\},$$

in $L^\infty([-K, K])$.

Proof of Lemma 3.5.4. We shall use Theorem B.1.1 with $f_{n,t}$ defined as

$$\begin{aligned}
(n^{1/6} g_0(x_0)) \left(\mathbb{1}\{X \leq \lceil (x_0 + n^{-1/3}t)J \rceil / J\} - \mathbb{1}\{X \leq \lceil x_0 J \rceil / J + 1/J\} \right) \\
- (n^{1/6} g_0(x_0)) J^{-1} \left[\lceil (x_0 + n^{-1/3}t)J \rceil - \lceil x_0 J \rceil - 1 \right].
\end{aligned}$$

We shall assume $t \geq 0$ for the proof. The steps for $t \leq 0$ follow in an exactly similar manner. For $t \geq 0$ we write $\sqrt{n} f_{n,t}$ as

$$\begin{aligned}
n^{2/3}(\mathbb{P}_n - \mathrm{P}) \left(\mathbb{1}\{X \leq \lceil (x_0 + n^{-1/3}t)J \rceil / J\} - \mathbb{1}\{X \leq \lceil x_0 J \rceil / J + 1/J\} \right) \\
+ n^{2/3} \mathrm{P} \left(\mathbb{1}\{X \leq \lceil (x_0 + n^{-1/3}t)J \rceil / J\} - \mathbb{1}\{X \leq \lceil x_0 J \rceil / J + 1/J\} \right) \\
= \sqrt{n}(\mathbb{P}_n - \mathrm{P}) f_{n,t} + \sqrt{n}\, \mathrm{P} f_{n,t}.
\end{aligned}$$

For every $K>0$, let F_n be defined as

$$(n^{1/6}g_0(x_0))\Big|\mathbb{1}\{X\le\lceil(x_0+n^{-1/3}K)J\rceil/J\}-\mathbb{1}\{X\le\lceil x_0J\rceil/J+1/J\}\Big| \\ -(n^{1/6}g_0(x_0))J^{-1}\big[\lceil(x_0+n^{-1/3}K)J\rceil-\lceil x_0J\rceil-1\big].$$

Then $|f_{n,t}|\le F_n$ for every $t\in[0,K]$, and therefore F_n is an envelope for the class of functions $\mathscr{F}_n=\{f_{n,t}:t\in[0,K]\}$. We shall verify the assumptions F_n needs to satisfy for Theorem 2.11.22 to hold.

We first show that $\mathrm{P}^*F_n^2$ is bounded. We have

$$\mathrm{P}^*F_n^2=n^{1/3}\int_{(\lceil x_0J\rceil+1)/J}^{\lceil(x_0+n^{-1/3}K)J\rceil/J}(g_0(s)-g_0(x_0))ds.$$

As g_0 is differentiable at x_0 with $g_0'(x_0)<0$, $g_0(x)$ is bounded by a constant $B>0$ in a δ-neighborhood of x_0 for some $\delta>0$. For n large enough $|\lceil(x_0+n^{-1/3}K)J\rceil/J-x_0|<\delta$. Using Taylor's expansion of g around x_0 we write $g_0(s)-g_0(x_0)=g_0'(x_0)(s-x_0)+O((s-x_0)^2)$ and for large n have,

$$\begin{aligned}\mathrm{P}^*F_n^2&\le n^{1/3}\,B\int_{(\lceil x_0J\rceil+1)/J}^{\lceil(x_0+n^{-1/3}K)J\rceil/J}(g_0(s)-g_0(x_0))ds\\&\lesssim n^{1/3}\big[g_0'(x_0)\big(\lceil(x_0+n^{-1/3}K)J\rceil/J-(\lceil x_0J\rceil+1)/J\big)+O(n^{-2/3})\big]\\&=n^{1/3}\big[g_0'(x_0)n^{-1/3}K+O(J^{-1})+O(n^{-2/3})\big]\\&=Kg_0'(x_0)+o(1),\end{aligned}$$

which is bounded. Next we show that $\mathrm{P}^*F_n^2\mathbb{1}\{F_n>\eta\sqrt{n}\}$ goes to zero for every $\eta>0$. For large n, $|\lceil(x_0+n^{-1/3}K)J\rceil/J-x_0|<\delta$. We write $\mathrm{P}^*F_n^2\mathbb{1}\{F_n>\eta\sqrt{n}\}$ as

$$\begin{aligned}\int_{(\lceil x_0J\rceil+1)/J}^{\lceil(x_0+n^{-1/3}K)J\rceil/J}&n^{1/3}\mathbb{1}\{|g_0(s)-g_0(x_0)|>\eta n^{1/3}\}ds\\&\le n^{1/3}\mathbb{1}\{|g_0(0)-g_0(1)|>\eta n^{1/3}\}\int_{(\lceil x_0J\rceil+1)/J}^{\lceil(x_0+n^{-1/3}K)J\rceil/J}ds\\&\lesssim\mathbb{1}\{|g_0(0)-g_0(1)|>\eta n^{1/3}\}O(1),\end{aligned}$$

which goes to zero as $\mathbb{1}\{|g_0(0)-g_0(1)|>\eta n^{1/3}\}\to 0$, since $g_0(0)<\infty$.

Next, let δ_n be a sequence of positive real numbers decreasing to zero. Let $0\le t\le s$,

such that $|s-t|<\delta_n$. Then as $n\to\infty$,

$$\begin{aligned}
\mathrm{P}(f_{n,s}-f_{n,t})^2 &= n^{1/3}\int_{(\lceil (x_0+n^{-1/3}t)J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}s)J\rceil/J}(g_0(u)-g_0(x_0))du\\
&\lesssim n^{1/3}J^{-1}\left[\lceil (x_0+n^{-1/3}s)J\rceil-\lceil (x_0+n^{-1/3}t)J\rceil-1\right]\\
&= n^{1/3}\left[n^{-1/3}(s-t)+O(J^{-1})\right]\\
&= |s-t|+o(1)\\
&\le \delta_n+o(1),
\end{aligned}$$

which goes to zero. We shall now find the limit of the covariance. Let $s,t>0$. We have

$$\begin{aligned}
\mathrm{P}f_{n,s}f_{n,t}-\mathrm{P}f_{n,s}\mathrm{P}f_{n,t} &= \left[\mathrm{P}(f_{n,s\wedge t}^2)-(\mathrm{P}f_{n,s\wedge t})^2\right]\\
&\quad+\left[\mathrm{P}(f_{n,s\wedge t}(f_{n,s\vee t}-f_{n,s\wedge t}))-\mathrm{P}(f_{n,s\wedge t})\mathrm{P}((f_{n,s\vee t}-f_{n,s\wedge t}))\right]\\
&= \mathrm{P}(f_{n,s\wedge t}^2)-(\mathrm{P}f_{n,s\wedge t})^2,
\end{aligned}$$

as $f_{n,s\wedge t}$ and $f_{n,s\vee t}-f_{n,s\wedge t}$ are independent. Expanding $\mathrm{P}(f_{n,s\wedge t}^2)$ and using Taylor's expansion of g_0 and g around x_0, we get

$$\begin{aligned}
\mathrm{P}(f_{n,s\wedge t}^2) &= n^{1/3}\int_{(\lceil x_0J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}(s\wedge t))J\rceil/J}(g_0(u)-g_0(x_0))du\\
&= n^{1/3}\int_{(\lceil x_0J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}(s\wedge t))J\rceil/J}\{g_0'(x_0)(u-x_0)+O(|u-x_0|^2)\}du\\
&= n^{1/3}\left[g_0(x_0)n^{-1/3}(s\wedge t)+O(n^{-2/3}+J^{-1})\right]\\
&= g_0(x_0)(s\wedge t)+o(1).
\end{aligned}$$

Therefore, $\mathrm{Cov}(f_{n,s},f_{,t})\to(s\wedge t)g_0(x_0)$ as $n\to\infty$. The uniform entropy condition of the theorem holds by Lemma B.1.2. The assumptions of Theorem B.1.1 are therefore satisfied and we have that for every $K>0$,

$$\left\{\sqrt{n}(\mathbb{P}_n-\mathrm{P})f_{n,t}:t\in[-K,K]\right\}\rightsquigarrow\left\{\sqrt{g_0(x_0)}W_1(t):t\in[-K,K]\right\},$$

where W_1 is a two sided standard Brownian motion starting at zero. It remains to show that

$\sqrt{n}\mathrm{P}f_{n,t} \to b\,t^2$ in P_0-probability. Using Taylor's expansion of g_0 and g around x_0, we get

$$\begin{aligned}\sqrt{n}\mathrm{P}f_{n,t} &= n^{2/3}\int_{([x_0J]+1)/J}^{[(x_0+n^{-1/3}t)J]/J} \{g_0(u)-g_0(x_0)\}du \\ &= n^{2/3}\int_{([x_0J]+1)/J}^{[(x_0+n^{-1/3}t)J]/J} \{g_0'(x_0)(u-x_0)+O((u-x_0)^2)\}du \\ &= g_0'(x_0)t^2/2+o(1).\end{aligned}$$

Therefore,

$$\left\{n^{2/3}\sum_{j=[x_0J]+1}^{[(x_0+n^{-1/3}t)J]}(N_j/n-g_0(x_0)/J): t\in[-K,K]\right\} \rightsquigarrow \left\{a\,W_1(t)-b\,t^2 : t\in[-K,K]\right\},$$

for every $K>0$. □

Lemma 3.5.5. *Let $n^{1/3}\ll J\ll n^{2/3}$. Then*

$$\operatorname*{arg\,max}_{t\in\mathbb{R}}\left\{n^{2/3}\sum_{j=[x_0J]+1}^{[(x_0+n^{-1/3}t)J]}\left(\frac{N_j}{n}-\frac{g_0(x_0)}{J}\right)-\frac{n^{1/3}}{g_0(x_0)}z\sum_{j=[x_0J]+1}^{[(x_0+n^{-1/3}t)J]}\frac{1}{J}\right\}=O_{P_0}(1).$$

Proof of Lemma 3.5.5. For $r\geq 0$, define $K_n(r)$ and $K(r)$ by

$$\begin{aligned}K_n(r) &= \mathbb{P}_n[(\mathbb{1}\{X\leq[(x_0+r)J]/J\}-\mathbb{1}\{X\leq[x_0J]/J\})] \\ &\quad + z\,n^{-1/3}\mathbb{P}_n[\mathbb{1}\{X\leq[(x_0+r)J]/J\}-\mathbb{1}\{X\leq[x_0J]/J\}]; \\ K(r) &= G_0(x_0+r)-G_0(x_0)-g_0'(x_0)r.\end{aligned}$$

Then $\hat{h}_n$ equals $\operatorname{argmin}\{K_n(r) : r\in\mathbb{R}\}-x_0$. We shall employ Theorem B.1.5 to show that $n^{1/3}(\operatorname{argmin}\{K_n(r) : r\in\mathbb{R}\}-x_0)$ is $O_P(1)$, with $\mathbb{M}_n=-K_n$, $\mathbb{M}=-K$, $\Theta=[0,\infty)$, $\hat{\theta}_n=\operatorname{argmin}_{r\in\mathbb{R}}K_n(r)$, $\theta_0=0$ and d the Euclidean distance on $\mathbb{R}$. With these notations, we have that $-K_n(0)=-K(0)=0$ and $\hat{\theta}_n$ maximizes $-K_n$. We shall now verify the assumptions of the theorem.

The condition $\mathbb{M}(\theta)-\mathbb{M}(\theta_0)\lesssim -d^2(\theta,\theta_0)$ translates to $K(r)\gtrsim r^2$ in the present context. Using Taylor's expansion of g_0 around x_0, we have

$$g_0(s)-g_0(x_0)=g_0'(x_0)(s-x_0)+o(|s-x_0|).$$

We verify that $K(r) \gtrsim r^2$ by writing $K(r)$ as

$$\mathrm{E}_0[\mathbb{1}\{X \in (x_0, x_0 + r]\}] = \int_{x_0}^{x_0+r} g_0(s)ds,$$

which is bounded below by a constant multiple of r^2. Next we shall find functions ϕ_n such that $\phi_n(\delta)/\delta^\alpha$ is decreasing in δ for some $\alpha \in (0,2)$, and for sufficiently small δ,

$$\sqrt{n}\mathrm{E}_0^* \sup_{|r|<\delta} |K_n(r) - K(r)| \lesssim \phi_n(\delta). \tag{3.5.16}$$

The left-hand side of (A.2.3) equals to $\mathrm{E}_0^* \sup\{|\sum_{l=3}^5 H_{ln}(r)| : |r| < \delta\}$, with $H_{ln}(r)$ as defined in (3.5.15). Using the fact that for sufficiently small δ, g_0 is differentiable in a δ-neighborhood of x_0, we get that $\mathrm{E}_0^* \sup\{|H_{3n}(r)| : |r| < \delta\}$ is bounded above by

$$\begin{aligned}
\mathrm{E}_0 \sqrt{n}\mathbb{P}_n \mathbb{1}\left\{X \in \left(x_0, \frac{\lceil x_0 J \rceil}{J}\right] \bigcup \left(x_0 + \delta, \frac{\lceil (x_0+\delta)J \rceil}{J}\right]\right\} & \\
\leq \sqrt{n} \int_{x_0}^{(\lceil x_0 J \rceil)/J} |g_0(u) - g_0(x_0)|du & \\
+ \sqrt{n} \int_{x_0+r}^{(\lceil (x_0+r)J \rceil)/J} |g_0(u) - g_0(x_0)|du & \\
\lesssim \sqrt{n}(J^{-1})^2 g_0'(x_0), &
\end{aligned}$$

which is bounded above by a constant multiple of $n^{-1/6}$, as $J \gg n^{1/3}$. Next, we define M_δ and a class of functions $\mathscr{M}_\delta$ as

$$\begin{aligned}
M_\delta &= \mathbb{1}\{X \in (x_0, x_0 + \delta]\} \\
\mathscr{M}_\delta &= \{\mathbb{1}\{X \in (x_0, x_0 + r]\} : |r| < \delta\}.
\end{aligned}$$

Then M_δ is an envelope for $\mathscr{M}_\delta$, and from the discussion on Page 291 of van der Vaart and Wellner (1996), we have that

$$\mathrm{E}_0^* \sup\{|H_{4n}(r)| : |r| < \delta\} \lesssim J(1, \mathscr{M}_\delta)(\mathrm{E}_0^* M_\delta^2)^{1/2},$$

where $J(\eta, \mathscr{M}_\delta)$ for $\eta > 0$ is the uniform entropy integral

$$\sup_Q \int_0^\eta \sqrt{1 + \log \mathscr{N}(\epsilon \|M_\delta\|_{2,Q}, \mathscr{M}_\delta, \|\cdot\|_{2,Q})}d\epsilon$$

Using arguments similar to those in Page 216, within the proof of Theorem 3.6.2 of Banerjee (2000), we get that $J(1,\mathscr{M}_\delta)$ is a finite constant. Also, using the fact that $(g_0(s)-g_0(x_0))^2$ is bounded by some $B>0$ in a neighborhood of x_0 and arguments similar to those used for proving $\mathrm{E}_0^* F_n^2$ is bounded in the proof of Lemma 3.5.4, we obtain that $\mathrm{E}_0^* M_\delta^2 \lesssim \delta$, for sufficiently small $\delta > 0$. Therefore $\mathrm{E}_0^* \sup\{|H_{4n}(r)| : |r| < \delta\} \lesssim \sqrt{\delta}$. Next we bound $\mathrm{E}_0^* \sup\{|H_{5n}(r)| : |r| < \delta\}$ as

$$\frac{z n^{1/6}}{J}[\lceil (x_0+r)J\rceil - \lceil x_0 J\rceil] \lesssim n^{1/6}|z|(\delta + O(J^{-1})),$$

which is bounded by a constant multiple of $n^{1/6}|z|\delta$, as $J \gg n^{1/3}$. Therefore for some $C>0$, (A.2.3) is satisfied for $\phi_n = C(n^{-1/6} + \delta^{1/2} + n^{1/6}|z|\delta)$. For $r_n = n^{1/3}$, we have that $r_n^2 \phi_n(r_n^{-1}) \asymp n^{1/2}|z| \lesssim \sqrt{n}$. The condition $\mathbb{M}_n(\hat{\theta}_n) \geq \mathbb{M}_n(\theta_0) - O_{P_0}(r_n^{-2})$ has been verified within the proof of Lemma 3.5.4. □

CHAPTER

4

BAYESIAN INFERENCE ON MONOTONE REGRESSION QUANTILES WITH TWO-STAGE ACCELERATED RATE

Let $(Y_1, X_1), \ldots, (Y_n, X_n)$ be data observed from

$$Y_i = f(X_i) + \varepsilon_i,\ i = 1, \ldots, n, \tag{4.0.1}$$

where f is monotone increasing on $[0,1]$, and ε_i are independent and identically distributed (i.i.d.) Gaussian errors with mean 0 and finite variance σ^2. We want to estimate a regression quantile $\mu = f^{-1}(y_0)$, where μ_0, the true value of μ is such that $\mu_0 \in (0,1)$ is such that $f_0'(\mu_0) > 0$. In this chapter, we shall use the projection-posterior f^* to obtain a credible set for $f^{-1}(y_0)$ and establish the asymptotic coverage of the corresponding credible interval. Further, we shall develop a Bayesian procedure when samples are collected in two stages. A fraction of sampling budget is first used to get observations, and in the second stage, the remaining sampling points are selected depending on the posterior distribution obtained in the first stage. By representing f as a linear function on this interval and by using Gaussian priors on

the coefficients, we shall show that such an "active learning mechanism" can improve the contraction rate of the posterior distribution of the monotone regression quantile $f^{-1}(y_0)$.

4.1 Assumptions, preliminaries and priors

Let $\mu_0 \in (0,1)$, f_0 be the true regression function, and let f_0 belong to the class $\mathscr{F}$ of monotone increasing functions on $[0,1]$ satisfying

1. $f'(x) > 0$, for all x in a neighborhood of μ_0.
2. for some $L > 0$ and $\kappa > 1/2$,

$$|f'(x_1) - f'(x_2)| \leq L|x_1 - x_2|^{\kappa}, \tag{4.1.1}$$

for all x_1, x_2 in a neighborhood of μ_0.

The design points satisfy the assumption in (2.1.1) or (2.1.2), depending on whether the design is deterministic or random. We have the following the assumption on the true distribution P_0.

Assumption B. *Under the true distribution P_0, $Y_i = f_0(X_i) + \varepsilon_i$, such that ε_i are i.i.d. sub-Gaussian with mean 0 and variance σ_0^2 for $i = 1, \ldots, n$.*

We construct a prior on f by representing f as a piece-wise constant function on $[0,1]$ with J the number of pieces. The knots of f are equally-spaced on $[0,1]$. We represent the model in (4.0.1) as $\boldsymbol{Y} = \boldsymbol{B}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$. For a deterministic J, let $I_j = ((j-1)/J, j/J]$, $1 \leq j \leq J$. We represent f as $f(\cdot) = \sum_{j=1}^{J} \theta_j \mathbb{1}_{I_j}(\cdot)$, and use the prior $\boldsymbol{\theta} \sim \mathrm{N}_J(\boldsymbol{\zeta}, \sigma^2\boldsymbol{\Lambda})$ where $\|\boldsymbol{\zeta}\|_\infty$ is bounded, and $\boldsymbol{\Lambda}$ is a $J \times J$ diagonal matrix with diagonal entries $\lambda_1^2, \ldots, \lambda_J^2$, with $B_1 < \lambda_j < B_2$ for some $B_1, B_2 > 0$. The error variance σ^2 may be estimated by maximizing the marginal likelihood of σ or by endowing σ^2 with an inverse-gamma prior with parameters $(\beta_1/2, \beta_2/2)$ with $\beta_1 > 4, \beta_2 > 0$.

Observe that $\boldsymbol{Y}|\sigma \sim \mathrm{N}_n\big(\boldsymbol{B}\boldsymbol{\zeta}, \sigma^2(\boldsymbol{B}\boldsymbol{\Lambda}\boldsymbol{B}^T + \boldsymbol{I}_n)\big)$. Maximizing the corresponding log-likelihood with respect to σ^2, we obtain

$$\widehat{\sigma}_n^2 = n^{-1}(\boldsymbol{Y} - \boldsymbol{B}\boldsymbol{\zeta})^T(\boldsymbol{B}\boldsymbol{\Lambda}\boldsymbol{B}^T + \boldsymbol{I}_n)^{-1}(\boldsymbol{Y} - \boldsymbol{B}\boldsymbol{\zeta}). \tag{4.1.2}$$

The plug-in posterior distribution of f is obtained by substituting $\widehat{\sigma}_n$ for σ.

For a monotone increasing F on $[0,1]$, define the left-continuous inverse function $F^{-1}:\mathbb{R}\to[0,1]$ as

$$F^{-1}(y)=\inf\{x:F(x)\geq y\}. \tag{4.1.3}$$

For $f=\sum_{j=1}^{J}\theta_j\mathbb{1}_{I_j}$, let f^* denote $\sum_{j=1}^{J}\theta_j^*\mathbb{1}_{I_j}$, where $(\theta_1^*,\ldots,\theta_J^*)$ is obtained as the minimizer of $\sum_{j=1}^{J}(\theta_j-\eta_j)^2$ over $\boldsymbol{\eta}=(\eta_1,\ldots,\eta_j)\in\mathbb{R}^J$ subject to $\eta_1\leq\cdots\leq\eta_J$. In other words, f^* is the projection of a posterior sample on the monotone class. The distribution of f^* is what we have defined in the previous chapters as the 'projection-posterior'. Let $y_0=f_0(\mu_0)$, $h=f^{*-1}$, and $h_0=f_0^{-1}$.

4.2 Credible set for $f^{-1}(y_0)$ and its coverage

Before establishing the results about the coverage of a credible interval for μ, first note that the maximum marginal likelihood estimator for σ^2 in the plug-in Bayes approach or the marginal posterior distribution of σ^2 in the fully Bayes approach, are consistent uniformly for $f_0\in\mathscr{F}$, as a result of Proposition 2.2.1.

Let $n^{1/3}\ll J\ll n^{2/3}$. Let W_1 and W_2 be two independent standard Brownian motions on $\mathbb{R}$ starting at zero. As obtained for the point-wise estimator of the regression function, the unconditional and conditional distribution given the data, of $n^{1/3}(h(y_0)-h_0(y_0))$ can be described as given by the following theorem.

Theorem 4.2.1. *Let* $Z_2=\arg\min\{W_1(t)+W_2(t)+t^2:t\in\mathbb{R}\}$, $Z_1=\arg\min\{W_1(t)+t^2:t\in\mathbb{R}\}$, *and* $C=(a/b)^{2/3}$ *with* $a=\sqrt{\sigma_0^2/g(\mu_0)}$ *and* $b=f_0'(\mu_0)/2$. *Let the design points be random, satisfying* (2.1.2). *Then under Assumption B on the errors, the following assertions hold.*

(a) For every $z\in\mathbb{R}$, $(\mathrm{P}_0\times\Pi)(n^{1/3}(h(y_0)-h_0(y_0))\leq z)\to\mathrm{P}(CZ_2\leq z)$.

(b) For every $z\in\mathbb{R}$, $\Pi(n^{1/3}(h(y_0)-h_0(y_0))\leq z|D_n)\rightsquigarrow\mathrm{P}(CZ_2\leq z|W_1)$.

Part (b) of Theorem 4.2.1 gives us a method of constructing a point-wise credible interval for the level set at y_0. For every $n\geq1$, $\gamma\in[0,1]$, define

$$\begin{aligned}&Q_{n,\gamma}=\inf\left\{z\in\mathbb{R}:\Pi(h(y_0)\leq z|D_n)\geq1-\gamma\right\},\\&I_{n,\gamma}=\left[Q_{n,1-\frac{\gamma}{2}},Q_{n,\frac{\gamma}{2}}\right],\quad\Delta^*_{W_1,W_2}=\operatorname*{arg\,min}_{t\in\mathbb{R}}\left\{W_1(t)+W_2(t)+t^2\right\}.\end{aligned} \tag{4.2.1}$$

Then the limiting coverage of $I_{n,\gamma}$ can be characterized by following theorem.

Theorem 4.2.2. *Let $Q_{n,\gamma}$, $I_{n,\gamma}$ and $\Delta^*_{W_1,W_2}$ be as defined in* (4.2.1)*, with $a = \sqrt{\sigma_0^2/g(\mu_0)}$ and $b = f_0'(\mu_0)/2$. Then for $\boldsymbol{X}$ satisfying* (2.1.2) *and under Assumption B on the errors, the limiting coverage of $I_{n,\gamma}$, $\gamma \in [0,1]$, is given by*

$$\mathrm{P}_0(h_0(y_0) \in I_{n,\gamma}) \to \mathrm{P}\left(\mathrm{P}(\Delta^*_{W_1,W_2} \le 0 | W_1) \in \left(\frac{\gamma}{2}, 1-\frac{\gamma}{2}\right)\right).$$

4.3 Two-stage sampling and accelerated contraction rate

In this section, we show that we can accelerate the posterior contraction rate for $f^{-1}(y_0)$ using an appropriate two-stage sampling method. For a total sampling size of n, we first obtain $n_1 < n$ samples to compute the first stage posterior distributions. The remaining $n_2 = n - n_1$ samples are then obtained by sampling points uniformly from a credible region constructed from the projection-posterior of μ. We choose the length of the credible interval small enough such that the regression function f can be approximated in this region by a linear function. By using normal priors on the coefficients of the linear function, we use the second-stage samples to obtain the posterior of $f^{-1}(y_0)$ and show that the second stage posterior contracts to the true value at a faster rate.

We first describe the two-stage sampling procedure. Let $p \in (0,1)$. In the first stage, we choose $n_1 \in \mathbb{N}$ design points $\{\tilde{\mu}_i : i = 1, \ldots, n_1\}$ in $(0,1)$ such that $n_1/n \to p$ as $n \to \infty$, to obtain the first-stage data $D_{1,n} = \{(\tilde{\mu}_i, \tilde{Y}_i) : i = 1, \ldots n_1\}$ for the model in (4.0.1). Typically, $p = 1/2$ is chosen to obtain equal splitting, but other values of p can also be used depending on the sampling configuration and practical considerations such as financial constraints. By using the prior scheme described in Section 4.1, we obtain a first stage posterior of $f^{-1}(y_0)$ from the projection-posterior f^* using the inversion formula (4.1.3). This then allows us to construct the interval

$$\mathscr{C}_n = (\tilde{\mu} - \delta_n, \tilde{\mu} + \delta_n),$$

for a chosen sequence $\delta_n \to 0$, where $\tilde{\mu}$ is the first stage posterior mean of $f^{-1}(y_0)$. We then sample $n_2 = n - n_1$ locations uniformly from the second stage sampling interval $\mathscr{C}_n$ and observe the second stage samples $D_{2,n} = \{(X_i, Y_i) : i = 1, \ldots n_2\}$.

Next, we center the second stage design points at 0 by $z_i = X_i - \tilde{\mu}$, $i = 1, \ldots, n_2$, so that z_i, $i = 1, \ldots, n_2$ are i.i.d. uniform samples from the interval $\mathscr{Q}_n = (-\delta_n, \delta_n)$. Note that the centered design points z_i are independent of the errors $\boldsymbol{\varepsilon}$. At the second stage, we represent f as a linear function on $\mathscr{Q}_n$. In other words, we let $f(z) = \theta_0 + \theta_1 z$ for $z \in \mathscr{Q}_n$. Let $\boldsymbol{\theta}$ denote

the vector (θ_0, θ_1) and $\boldsymbol{Z}$ denote the $n_2 \times 2$ matrix whose i-th row is $(1, z_i)$.

On $\boldsymbol{\theta}$, we use the prior $\boldsymbol{\theta}|\sigma^2 \sim \mathrm{N}_2(\xi, \sigma^2 \boldsymbol{V})$, where ξ is a 2×1 vector whose entries do not depend on n, and for some pre-fixed large v_1 and v_2,

$$\boldsymbol{V} = \begin{pmatrix} v_1^2 & 0 \\ 0 & v_2^2 \end{pmatrix}.$$

Let $\boldsymbol{Y} = (Y_1, \ldots, Y_{n_2})$ be the vector of second stage Y-observations. The posterior $\Pi(\theta|D_{2,n}, \sigma^2)$ is then

$$\mathrm{N}_2\big[(\boldsymbol{Z}^T\boldsymbol{Z} + \boldsymbol{V}^{-1})^{-1}(\boldsymbol{Z}^T\boldsymbol{Y} + \boldsymbol{V}^{-1}\xi),\ \sigma^2(\boldsymbol{Z}^T\boldsymbol{Z} + \boldsymbol{V}^{-1})^{-1}\big].$$

The plug-in posterior of σ^2 is obtained by replacing σ^2 with $\tilde{\sigma}_*^2 = (n_1\tilde{\sigma}^2 + n_2\tilde{\sigma}^2)/n$, where $\tilde{\sigma}_1^2 = n_1^{-1}(\boldsymbol{Y} - \boldsymbol{B}\boldsymbol{\zeta})^T(\boldsymbol{B}\boldsymbol{\Lambda}\boldsymbol{B}^T + \boldsymbol{I}_{n_1})^{-1}(\boldsymbol{Y} - \boldsymbol{B}\boldsymbol{\zeta})$ is the plug-in estimate of σ^2 based on the first stage samples, and $\tilde{\sigma}_2^2 = n_2^{-1}(\boldsymbol{Y} - \boldsymbol{Z}\xi)^T(\boldsymbol{Z}\boldsymbol{V}\boldsymbol{Z}^T + \boldsymbol{I}_{n_2})^{-1}(\boldsymbol{Y} - \boldsymbol{Z}\xi)$ is the same estimate based on the second stage samples.

For the fully Bayes approach, we use the first stage posterior of σ^2 as a prior for the second stage. We use the $\mathrm{IG}(\beta_1/2, \beta_2/2)$ prior on σ^2 with $\beta_1 > 4, \beta_2 > 0$. We then use the resulting posterior $\mathrm{IG}[(\beta_1 + n_1)/2, (\beta_2 + n_1\tilde{\sigma}_1^2)/2]$ as prior for the second stage, yielding $\mathrm{IG}[(\beta_1 + n)/2, (\beta_2 + n\tilde{\sigma}_*^2)/2]$ as the second stage posterior for σ^2. The following proposition shows that the second stage plug-in Bayes estimator and the fully Bayes posterior of σ^2 are consistent.

Proposition 4.3.1. *(Second stage error variance). For $f_0 \in \mathscr{F}(K)$,*

(a) the second stage plug-in Bayes estimator $\tilde{\sigma}_*^2$ converges to σ_0^2 in P_0-probability at the rate $\max\{n^{-1/2}, \delta_n^2\}$.

(b) If the inverse-gamma posterior from the first stage is used as a prior for the second stage, then the second stage posterior of σ^2 contracts to σ_0^2 at the same rate.

Let us define the regression quantile of y_0 with respect to the $f_{\boldsymbol{\theta}}$ inside the centered region $\mathscr{Q}_n$ as $x = \tilde{\mu} + x_{\tilde{\mu}}$, with $y_0 = f_{\boldsymbol{\theta}}(x_{\tilde{\mu}})$. The following theorem establishes the contraction rate of $f^{-1}(y_0)$ for $f_0 \in \mathscr{F}(K)$.

Theorem 4.3.2. *Let $\delta_n = n^{-1/3}(\log n)^t$ for some $t > 0$, and $n^{1/3} \ll J \ll n^{2/3}$. Then for $\epsilon_n = n^{-1/2}$ and every $M_n \to \infty$,*

$$\mathrm{E}_0\Pi\big[|f^{-1}(y_0) - f_0^{-1}(y_0)| > M_n\epsilon_n \big| D_n\big] \to 0.$$

Therefore, if we perform a Bayesian two-stage procedure with the length of the second stage sampling interval a logarithmic factor of the first stage rate $n^{-1/3}$, we accelerate the contraction rate for $f^{-1}(y_0)$ from $n^{-1/3}$ to $n^{-1/2}$.

4.4 Simulation results

4.4.1 Coverage of the first stage credible interval

We first study the behavior of credible intervals for $\mu = f^{-1}(y_0)$ based on the projection-posterior f^*. We first compare finite sample performance of Bayesian credible intervals with confidence intervals in terms of their coverages of $\mu_0 = f_0^{-1}(y_0)$, using $f_0(x) = x^2 + x/5$ for $x \in [0,1]$, $\mu_0 = 0.3$, G the uniform distribution on $[0,1]$, and $\sigma_0 = 0.1$. We take J to be the greatest integer less than or equal to $n^{1/3}\log n$. We vary sample size n across four different values. For each n, we consider 1000 instances of the data D_n. For each instance, we generate 1000 posterior samples of $\boldsymbol{\theta}$ and obtain the corresponding projection f^* from $\boldsymbol{\theta}$ and compute the inverse f^{*-1}. We also compute the inverse $\hat{f}_n^{-1}(y_0)$ from the sieve-MLE $\hat{f}_n$ for every replication. We then obtain the recalibrated $100(1-\alpha)\%$ credible interval using the $A^{-1}(1-\alpha/2)$ and $1-A^{-1}(1-\alpha/2)$ quantiles of $f^{*-1}(y_0)$ (see Table 2.1 for the values of $A^{-1}(1-\alpha)$), and we check if it contains μ_0. The coverage is determined by the proportion of times the credible interval contains μ_0. We also compute the $100(1-\alpha)\%$ confidence intervals from $\hat{f}_n^{-1}(y_0)$ based on the quantiles of Chernoff's distribution and the estimated value of the constant C that appears in Theorem 4.2.1. We also obtain the $100(1-\alpha)\%$ credible interval using the $\alpha/2$ and $1-\alpha/2$ projection-posterior quantiles. In the results, $C_B(\alpha)$, $C_B^*(\alpha)$ and $C_F(\alpha)$ respectively denote the coverage of $(Q_{n,1-\alpha/2}, Q_{n,\alpha/2})$, $(Q_{n,1-A^{-1}(\alpha/2)}, Q_{n,A^{-1}(\alpha/2)})$, and that of level $(1-\alpha)$ confidence interval, and $L_B(\alpha)$, $L_B^*(\alpha)$ and $L_F(\alpha)$ respectively denote their lengths.

4.4.2 Comparison of the single and two-stage estimation errors

We compare the performance of our two-stage Bayesian procedure with the single stage Bayesian procedure. We use the same true regression function $f_0(x) = x^2 + x/5$ for $x \in [0,1]$, $\mu_0 = 0.3$, G the uniform distribution on $[0,1]$, and $\sigma_0 = 0.1$. We take J to be the greatest integer less than or equal to $n^{1/3}\log n$. We use equal splitting of the sample size across the two stages. The true y_0 is given by 0.15. In the first stage, we observe f_0 on n_1 uniformly sampled design points. We then sample n_2 points from the interval $(\tilde{\mu} - \delta_n, \tilde{\mu} + \delta_n)$, with

Table 4.1: Comparison of obtained coverage and average length of Bayesian credible intervals for $f^{-1}(y_0)$, with that obtained from a confidence interval.

$1-\alpha$	n=800						n=1000					
	$C_B(\alpha)$	$L_B(\alpha)$	$C_B^*(\alpha)$	$L_B^*(\alpha)$	$C_F(\alpha)$	$L_F(\alpha)$	$C_B(\alpha)$	$L_B(\alpha)$	$C_B^*(\alpha)$	$L_B^*(\alpha)$	$C_F(\alpha)$	$L_F(\alpha)$
0.99	0.992	0.41	0.986	0.37	0.992	0.47	0.99	0.33	0.976	0.3	0.992	0.37
0.98	0.988	0.37	0.976	0.34	0.988	0.43	0.984	0.3	0.972	0.27	0.976	0.33
0.95	0.962	0.32	0.95	0.29	0.962	0.36	0.96	0.26	0.928	0.24	0.956	0.28
0.92	0.946	0.28	0.936	0.27	0.926	0.33	0.92	0.23	0.902	0.22	0.912	0.25
0.9	0.934	0.27	0.91	0.25	0.904	0.31	0.902	0.22	0.874	0.2	0.884	0.24
0.86	0.884	0.24	0.85	0.22	0.854	0.26	0.852	0.19	0.812	0.18	0.826	0.21
$1-\alpha$	n=1200						n=1500					
	$C_B(\alpha)$	$L_B(\alpha)$	$C_B^*(\alpha)$	$L_B^*(\alpha)$	$C_F(\alpha)$	$L_F(\alpha)$	$C_B(\alpha)$	$L_B(\alpha)$	$C_B^*(\alpha)$	$L_B^*(\alpha)$	$C_F(\alpha)$	$L_F(\alpha)$
0.99	0.996	0.29	0.984	0.26	0.992	0.32	0.998	0.26	0.994	0.23	0.996	0.29
0.98	0.988	0.26	0.982	0.24	0.978	0.29	0.994	0.24	0.976	0.22	0.984	0.26
0.95	0.972	0.23	0.948	0.21	0.94	0.25	0.97	0.2	0.95	0.19	0.948	0.22
0.92	0.941	0.2	0.932	0.19	0.922	0.22	0.946	0.18	0.936	0.18	0.922	0.2
0.9	0.93	0.19	0.906	0.18	0.902	0.21	0.934	0.17	0.916	0.16	0.888	0.19
0.86	0.888	0.17	0.868	0.16	0.826	0.18	0.908	0.16	0.874	0.15	0.828	0.16

$\delta_n = n^{-1/3}(\log n)^{1/6}$. We observe the data on the second stage samples. We center each of the n_2 design points at the origin by subtracting each of them by $\tilde{\mu}$. We then use a linear function with normal coefficients as prior for f over the second stage sampling interval, with $\boldsymbol{\theta}|\sigma^2 \sim \mathrm{N}_2(\mathbf{0}, \sigma^2 \boldsymbol{V})$, where $\boldsymbol{V}$ is a 2×2 diagonal matrix with the diagonal entries equal to 10. We use the plug-in Bayes method to estimate σ^2 by the weighted average of the first and second stage estimates $\tilde{\sigma}_1^2$ and $\tilde{\sigma}_2^2$. For each n, we repeat the experiment 1000 times. At each instance of the data, we compute the first stage error $|\tilde{\mu}-\mu_0|$ and the second stage error $|(y_0-\tilde{\theta}_0)/\tilde{\theta}_1+\tilde{\mu}-\mu_0|$, where $\tilde{\theta}_0$ and $\tilde{\theta}_1$ are the second stage posterior means of θ_0 and θ_1 respectively. We obtain the boxplots of these 1000 errors for both the procedures.

We observe that the Bayesian two-stage procedure has considerably lower absolute error than the corresponding single stage procedure, with significant improvements evident in larger samples sizes.

4.5 Proofs

In order to prove Theorem 4.2.1, we shall need the following result on the asymptotic distribution of f^* in a $n^{-1/3}$-neighborhood of μ_0.

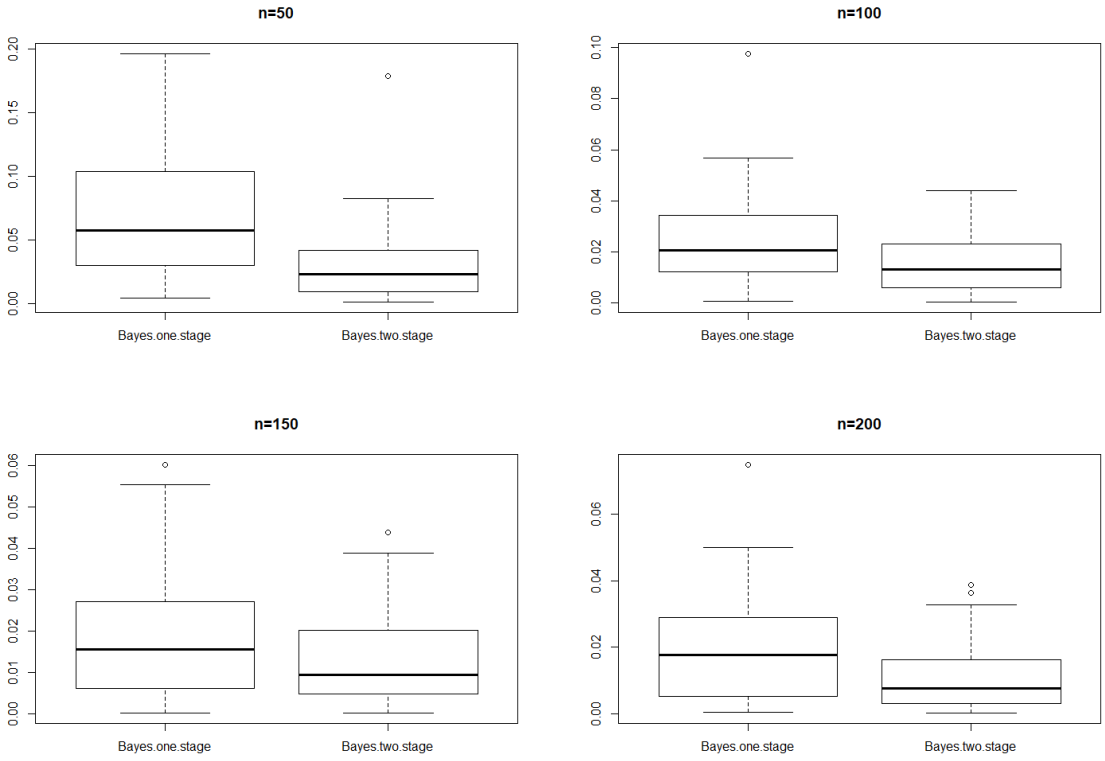

Figure 4.1: Comparison of the mean absolute errors based on the Bayesian one-stage and two-stage procedures for $n = 50, 100, 150, 200$.

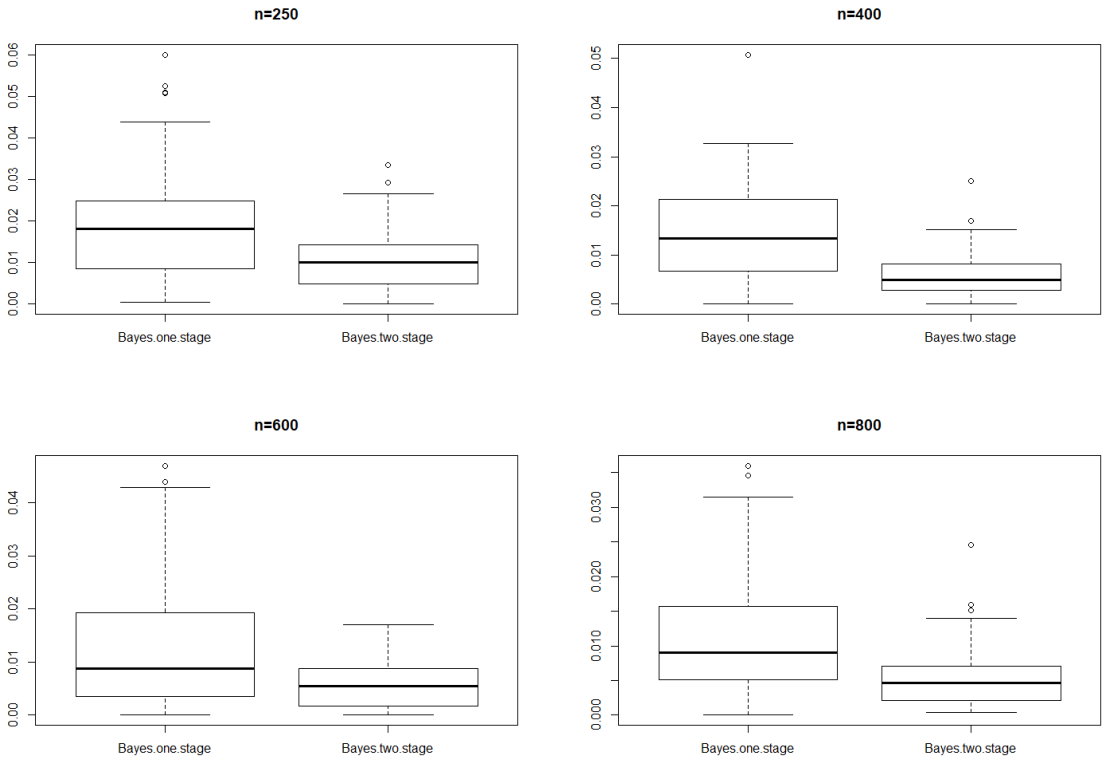

Figure 4.2: Comparison of the mean absolute errors based on the Bayesian one-stage and two-stage procedures for $n = 250, 400, 600, 800$.

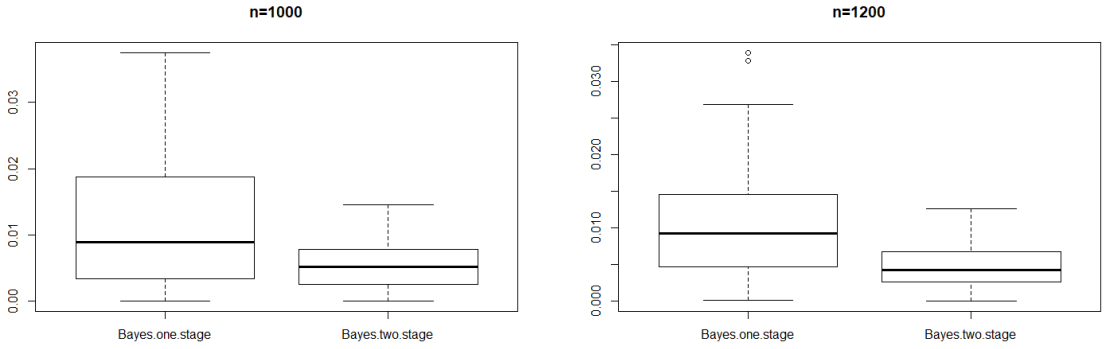

Figure 4.3: Comparison of the mean absolute errors based on the Bayesian one-stage and two-stage procedures for $n = 1000, 1200$.

Lemma 4.5.1. *Let W_1, W_2, a, b be as defined in Theorem 4.2.1. Then for $x \in \mathbb{R}$ and $\lambda \in \mathbb{R}$,*

$$\Pi\left(n^{1/3}\left(f^*(\mu_0 + n^{-1/3}x) - f_0(\mu_0)\right) < \lambda \middle| D_n\right) \rightsquigarrow \mathrm{P}\left(\left(\frac{a}{b}\right)^{2/3} \operatorname*{argmin}_{t\in\mathbb{R}} \left\{W_1(t) + W_2(t) + t^2\right\} + \frac{\lambda}{2b} > x \middle| W_1\right).$$

Proof. Let $v = \mu_0 + n^{-1/3}x$. Since f^* is piece-wise constant on each I_j, $f^*(v) = \theta^*_{\lceil vJ \rceil}$. Let $z \in \mathbb{R}$. Let $c(\cdot)$ denote the graph of

$$\left\{(0,0), \left(\frac{N_1}{n}, \frac{N_1}{n}\theta_1\right), \left(\sum_{k=1}^{2}\frac{N_k}{n}, \sum_{k=1}^{2}\frac{N_k}{n}\theta_k\right), \ldots, \left(\sum_{k=1}^{J}\frac{N_k}{n}, \sum_{k=1}^{J}\frac{N_k}{n}\theta_k\right)\right\},$$

with $c(s) = 0$ for $s \le 0$ and $c(s) = \sum_{k=1}^{J}(N_k/n)\theta_k$ for $s \ge 1$. Let V_n, G_n be stochastic processes denoting $V_n(s) = \sum_{j=1}^{\lceil sJ \rceil}(N_j/n)\theta_j$ and $G_n(s) = \sum_{j=1}^{\lceil sJ \rceil} N_j/n$. Now since $\theta^*_{\lceil \mu_0 J \rceil}$ is the left-derivative of the greatest convex minorant of $c(\cdot)$ at the point $\sum_{k=1}^{\lceil \mu_0 J \rceil} N_k/n$, by the switch relation (B.2.1), we have that

$$\left\{f^*(v) < f_0(\mu_0) + n^{-1/3}\lambda\right\} = \left\{\operatorname*{argmin}_{s\ge 0}\{c(s) - (f_0(\mu_0) + n^{-1/3}\lambda)s\} > \sum_{k=1}^{\lceil vJ \rceil}\frac{N_k}{n}\right\}.$$

Since the function inside the argmin is piece-wise linear in s, the minimum occurs on the set $\{0, N_1/n, (N_1+N_2)/n, \ldots, 1\}$, and therefore

$$\operatorname*{argmin}_{s\in\mathbb{R}}\left\{c(s) - (f_0(\mu_0) + n^{-1/3}\lambda)s\right\} = \operatorname*{argmin}_{s\in\mathbb{R}}\left\{V_n(s) - (f_0(\mu_0) + n^{-1/3}\lambda)G_n(s)\right\}.$$

Therefore,

$$\begin{aligned}&\Pi\left(f^*(v) < f_0(\mu_0) + n^{-1/3}\lambda \middle| D_n\right)\\ &\quad = \Pi\left(\operatorname*{argmin}_{s\in\mathbb{R}}\left\{V_n(s) - (f_0(\mu_0) + n^{-1/3}\lambda)G_n(s)\right\} > \sum_{k=1}^{\lceil vJ \rceil}\frac{N_k}{n} \middle| D_n\right). \qquad (4.5.1)\end{aligned}$$

Using the fact that $V_n(s) = G_n(s) = 0$ for $s \le 0$, and that the location of the minimum does not change upon adding a constant term or upon multiplication by a positive constant, (4.5.1) can be written as

$$\Pi\left(\operatorname*{argmin}_{s\in\mathbb{R}}\left\{\frac{n^{2/3}}{g(\mu_0)}(V_n(s) - V_n(\mu_0))\right.\right.$$

$$-\frac{n^{2/3}}{g(\mu_0)}(f_0(\mu_0)+n^{-1/3}\lambda)(G_n(s)-G_n(\mu_0))\Big\} > \sum_{k=1}^{\lceil vJ\rceil}\frac{N_k}{n}\Big|D_n\Big). \tag{4.5.2}$$

Using the change of variable $s=\mu_0+n^{-1/3}t$ and denoting $\tau_n = n^{1/3}(\sum_{k=1}^{\lceil vJ\rceil}(N_k/n)-\mu_0)$, we write (4.5.2) as

$$\begin{aligned}
&\Pi\Bigg(\operatorname*{arg\,min}_{t\in\mathbb{R}}\Bigg\{\frac{n^{2/3}}{g(\mu_0)}\big(V_n(\mu_0+n^{-1/3}t)-V_n(\mu_0)\big)\\
&\qquad -\frac{n^{2/3}}{g(\mu_0)}\big(f_0(\mu_0)+n^{-1/3}\lambda\big)\big(G_n(\mu_0+n^{-1/3}t)-G_n(\mu_0)\big)\Bigg\} > \tau_n\Big|D_n\Bigg)\\
&=\Pi\Bigg(\operatorname*{arg\,min}_{t\in\mathbb{R}}\Bigg\{\frac{n^{2/3}}{g(\mu_0)}\sum_{j=\lceil\mu_0 J\rceil+1}^{\lceil(\mu_0+n^{-1/3}t)J\rceil}\frac{N_j}{n}\theta_j\\
&\qquad -\frac{n^{2/3}}{g(\mu_0)}\big(f_0(\mu_0)+n^{-1/3}\lambda\big)\sum_{j=\lceil\mu_0 J\rceil+1}^{\lceil(\mu_0+n^{-1/3}t)J\rceil}\frac{N_j}{n}\Bigg\} > \tau_n\Big|D_n\Bigg).
\end{aligned}\tag{4.5.3}$$

Now note that $\sum_{k=1}^{\lceil vJ\rceil}N_k$ is distributed as $\mathrm{Bin}(n;G(\lceil vJ\rceil/J))$. Using (2.1.2) on G, we get that

$$G\left(\frac{\lceil(\mu_0+n^{-1/3}x)J\rceil}{J}\right)-\mu_0=\mu_0+n^{-1/3}x+O(J^{-1})-\mu_0=n^{-1/3}x+O(J^{-1}).$$

Therefore

$$\mathrm{E}\,\tau_n = n^{1/3}\left(G\left(\frac{\lceil(\mu_0+n^{-1/3}x)J\rceil}{J}\right)-\mu_0\right)=x+O(n^{1/3}J^{-1}).$$

Similarly, $\mathrm{Var}\,\tau_n \lesssim n^{-1/3}$. Therefore $\tau_n \to^{P_0} x$, as $n^{1/3}\ll J\ll n^{2/3}$. The left-hand side of the expression inside the probability in (4.5.3) does not involve v, and has been shown in the proof of part (a) of Theorem 2.3.2, to converge conditionally given D_n to $\operatorname{arg\,min}\{aW_1(t)+aW_2(t)+bt^2-\lambda t : t\in\mathbb{R}\}$. Using part (a) of Lemma A.2.7 for the conditional distribution, we get that

$$\begin{aligned}
\Pi\big(n^{1/3}(f^*(v)-f_0(\mu_0))<\lambda\big|D_n\big) &\rightsquigarrow \mathrm{P}\left(\operatorname*{arg\,min}_{t\in\mathbb{R}}\big\{aW_1(t)+aW_2(t)+bt^2-\lambda t\big\} > x\Big|W_1\right)\\
&\stackrel{d}{=}\mathrm{P}\left(\left(\frac{a}{b}\right)^{2/3}\operatorname*{arg\,min}_{t\in\mathbb{R}}\big\{W_1(t)+W_2(t)+t^2\big\}+\frac{\lambda}{2b} > x\Big|W_1\right).
\end{aligned}$$

□

Proof of Theorem 4.2.1. Let $z \in \mathbb{R}$. We have that

$$\begin{aligned}\Pi\big(n^{1/3}(h(y_0)-h_0(y_0)) \le z \big| D_n\big) &= \Pi\big(n^{1/3}(f^{*-1}(y_0)-f_0^{-1}(y_0)) \le z \big| D_n\big) \\ &= \Pi\big(f^{*-1}(y_0) \le \mu_0 + n^{-1/3}z \big| D_n\big). \end{aligned} \tag{4.5.4}$$

By the definition of f^{*-1}, we have that $f^{*-1}(u) \le t$ if and only if $f^*(t) \ge u$. We can write (4.5.4) as

$$\begin{aligned}\Pi\big(y_0 \le f^*(\mu_0 + n^{-1/3}z)\big| D_n\big) &= \Pi\big(f^*(\mu_0 + n^{-1/3}z) \ge f_0(\mu_0)\big| D_n\big) \\ &= \Pi\big(n^{1/3}(f^*(\mu_0 + n^{-1/3}z) - f_0(\mu_0)) \ge 0 \big| D_n\big) \\ &= 1 - \Pi\big(n^{1/3}(f^*(\mu_0 + n^{-1/3}z) - f_0(\mu_0)) < 0 \big| D_n\big). \end{aligned} \tag{4.5.5}$$

Using Lemma 4.5.1, with $\lambda = 0$ and $x = z$, the weak limit of (4.5.5) is equal to

$$\begin{aligned}1 - \mathrm{P}\left((a/b)^{2/3} \operatorname*{arg\,min}_{t\in\mathbb{R}} \big\{W_1(t) + W_2(t) + t^2\big\} > z \big| W_1\right) \\ = \mathrm{P}\left((a/b)^{2/3} \operatorname*{arg\,min}_{t\in\mathbb{R}} \big\{W_1(t) + W_2(t) + t^2\big\} \le z \big| W_1\right).\end{aligned}$$

□

Proof of Theorem 4.2.2. By the definition of posterior quantile $Q_{n,\gamma}$, we have that $h_0(y_0) \le Q_{n,\gamma}$ if and only if $\Pi(h(y_0) \le h_0(y_0)|D_n) \le 1-\gamma$. For every $z \in \mathbb{R}$, define

$$F_n^*(z\big|D_n) = \Pi(n^{1/3}(h(y_0)-h_0(y_0)) \le z|D_n).$$

Also, let

$$F_{a,b}^*(z\big|W_1) = \mathrm{P}\left(\left(\frac{a}{b}\right)^{2/3} \operatorname*{arg\,min}_{t\in\mathbb{R}} \{W_1(t) + W_2(t) + t^2\} \le z|W_1\right).$$

Therefore,

$$\begin{aligned}\mathrm{P}_0\big(h_0(y_0) \le Q_{n,\gamma}\big) &= \mathrm{P}_0\big(\Pi(h(y_0) \le h_0(y_0)|D_n) \le 1-\gamma\big) \\ &= \mathrm{P}_0\big(\Pi(n^{1/3}(h(y_0)-h_0(y_0)) \le 0|D_n) \le 1-\gamma\big) \\ &= \mathrm{P}_0\big(F_n^*(0\big|D_n) \le 1-\gamma\big) \\ &\to \mathrm{P}\big(F_{a,b}^*(0\big|W_1) \le 1-\gamma\big),\end{aligned}$$

where the last step follows from part (b) of Theorem 4.2.1, substituting $z = 0$. Now rewriting $F^*_{a,b}(0 \,|\, W_1)$ in terms of $\Delta^*_{W_1,W_2}$, with $C = (a/b)^{2/3}$, we have that $\mathrm{P}\left(F^*_{a,b}(0 \,|\, W_1) \le 1-\gamma\right)$ is equal to

$$\mathrm{P}\left(\mathrm{P}\left(C \operatorname*{arg\,min}_{t\in\mathbb{R}} \left\{W_1(t)) + W_2(t) + t^2\right\} \ge 0 \,\middle|\, W_1\right) \le 1-\gamma\right) \\ = \mathrm{P}\left(\mathrm{P}\left(\Delta^*_{W_1,W_2} \ge 0 \,\middle|\, W_1\right) \le 1-\gamma\right).$$

Similarly $\mathrm{P}_0\left(h_0(y_0) \ge Q_{n,1-\gamma}\right) = 1 - \mathrm{P}_0\left(h_0(y_0) < Q_{n,1-\gamma}\right)$, and therefore as $n \to \infty$,

$$\mathrm{P}_0\left(h_0(y_0) \ge Q_{n,1-\gamma}\right) \to 1 - \mathrm{P}\left(\mathrm{P}(\Delta^*_{W_1,W_2} \ge 0 \,|\, W_1) < \gamma\right) = \mathrm{P}\left(\mathrm{P}(\Delta^*_{W_1,W_2} \ge 0 \,|\, W_1) \ge \gamma\right).$$

This establishes the proof. □

We shall state and prove some preliminary results before giving the proofs of Proposition 4.3.1 and Theorem 4.3.2. Define $f_{0,\tilde{\mu}}(x-\tilde{\mu}) = f_0(x)$ to be the shifted true function. Let $\boldsymbol{\theta}_0 = (\theta_{00}, \theta_{01})$ be a random vector such $f_{\boldsymbol{\theta}_0}(x-\tilde{\mu}) = T_{\mu_0} f_0(x)$, where $f_{\boldsymbol{\theta}}(z) = \theta_0 + \theta_1 z$, and $T_{\mu_0} f_0(x)$ is the Taylor polynomial of order 1 by expanding f_0 around μ_0, so that

$$\theta_{00} + \theta_{01}(x-\tilde{\mu}) = f_0(\mu_0) + f_0'(\mu_0)(x-\mu_0). \tag{4.5.6}$$

From the last display we obtain that $\theta_{00} = f_0(\mu_0) + f_0'(\mu_0)(\tilde{\mu}-\mu_0)$ and $\theta_{01} = f_0'(\mu_0)$. Therefore $\boldsymbol{\theta}$ is a random vector, as the first stage estimator $\tilde{\mu}$ is random. The centered design points are i.i.d. uniform samples and so by Lemma 4.5.2, $\boldsymbol{Z}^T\boldsymbol{Z}$ is invertible with probability tending to 1 as $n \to \infty$. Moreover, in the computation of the posterior distribution, the invertibility is not an issue because of the regularization factor in the form of the prior covariance matrix $\boldsymbol{V}$. By our choice of $\boldsymbol{V}$, $\boldsymbol{Z}^T\boldsymbol{Z} + \boldsymbol{V}^{-1}$ is always invertible.

As a consequence of Theorem 4.2.2, $|\tilde{\mu}-\mu_0| = O_P(n^{-1/3})$. Let $\boldsymbol{F}_0 = (f_0(x_1), \ldots, f_0(x_{n_2}))^T$ where $\{x_1, \ldots, x_{n_2}\}$ are the original(unshifted) second stage design points. Note that $(\boldsymbol{Z}\boldsymbol{\theta}_0)_i = f_{\boldsymbol{\theta}_0}(x_i - \tilde{\mu})$ and by Taylor's expansion of f_0 around μ_0, we have that

$$\begin{aligned} \|\boldsymbol{F}_0 - \boldsymbol{Z}\boldsymbol{\theta}_0\|_\infty &= \max_{1\le i\le n_2} |f_0(x_i) - \theta_{00} - \theta_{01}(x_i - \tilde{\mu})| \\ &= \max_{1\le i\le n_2} |f_0(x_i) - f_0(\mu_0) - f_0'(\mu_0)(\tilde{\mu}-\mu_0) - f_0'(\mu_0)(x_i - \tilde{\mu})| \\ &\lesssim \max_{1\le i\le n_2} \left[|f_0'(x_i-\mu_0) - f_0'(\mu_0)(x_i-\mu_0)| + O(|x_i-\mu_0|^{1+\kappa})\right] \\ &\lesssim \delta_n^{1+\kappa} \end{aligned} \tag{4.5.7}$$

$$\ll \delta_n^2. \tag{4.5.8}$$

The last inequality is due to the fact that $|x_i - \tilde{\mu}| \leq \delta_n$ and $|\tilde{\mu} - \mu_0| = O_{P_0}(n^{-1/3}) = o_{P_0}(\delta_n)$, and by our choice of δ_n, $\delta_n \to 0$ and $\delta_n \gg n^{-1/3}$. For matrices $\boldsymbol{Z}^T\boldsymbol{Z}$ and $\boldsymbol{V}$, we enumerate their rows and columns starting from 0 and ending at 1. Note that by our choices of the first and second stage sample sizes, $n_1 \asymp n \asymp n_2$.

We shall first upper bound the entries of $(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}$ and obtain bounds for the posterior mean and variance of $\boldsymbol{\theta}$. Centering the design points around zero ensures that $(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}$ is free of $\tilde{\mu}$. The following lemma describes the asymptotic behavior of the entries of $(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}$ when the second stage samples are collected under uniform random sampling.

Lemma 4.5.2. *As $n \to \infty$, $\boldsymbol{Z}^T\boldsymbol{Z}$ is invertible with probability tending to 1. Also, for $j, k = 0, 1$, we have that $[(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}]_{j,k} = O_P\big(n^{-1}\delta_n^{-(j+k)}\big)$.*

Proof. As $z_i \sim \text{Uniform}[-\delta_n, \delta_n]$, we have that

$$\boldsymbol{Z}^T\boldsymbol{Z} = n_2 \begin{bmatrix} a_{00} & a_{01}\delta_n \\ a_{10}\delta_n & a_{11}\delta_n^2 \end{bmatrix} := n_2 \boldsymbol{\Delta}\boldsymbol{A}\boldsymbol{\Delta}, \tag{4.5.9}$$

with $\boldsymbol{A} = (\!(a_{ij})\!)$, $a_{ij} = n_2^{-1}\sum_{k=1}^{n_2} U_k^i U_k^j$, where U_1, U_2 are i.i.d. Uniform$[-1,1]$, and $\boldsymbol{\Delta}$ is a 2×2 diagonal matrix with diagonal entries δ_n^j, $j = 0, 1$. Define $\mathbb{U} = (U^0, U^1)^T$ for $U \sim$ Uniform$[-1,1]$. By the law of large numbers, each entry of $\boldsymbol{A}$ converges in probability to the corresponding entry of $\mathrm{E}\mathbb{U}\mathbb{U}^T$, and therefore

$$\mathrm{E}\mathbb{U}\mathbb{U}^T - \epsilon\boldsymbol{I} \leq \boldsymbol{A} \leq \mathrm{E}\mathbb{U}\mathbb{U}^T + \epsilon\boldsymbol{I},$$

for a sufficiently small $\epsilon > 0$. As the entries of $\mathrm{E}\mathbb{U}\mathbb{U}^T$ are mixed moments of $U \sim$ Uniform$[-1,1]$, $\mathrm{E}\mathbb{U}\mathbb{U}^T$ is positive definite, and $\mathrm{E}\mathbb{U}\mathbb{U}^T - \epsilon\boldsymbol{I}$ being invertible when ϵ is smaller than the minimum eigenvalue of $\mathrm{E}\mathbb{U}\mathbb{U}^T$. Therefore for a sufficiently small $\epsilon > 0$,

$$n_2^{-1}\boldsymbol{\Delta}^{-1}(\mathrm{E}\mathbb{U}\mathbb{U}^T + \epsilon\boldsymbol{I})^{-1}\boldsymbol{\Delta}^{-1} \leq (\boldsymbol{Z}^T\boldsymbol{Z})^{-1} \leq n_2^{-1}\boldsymbol{\Delta}^{-1}(\mathrm{E}\mathbb{U}\mathbb{U}^T - \epsilon\boldsymbol{I})^{-1}\boldsymbol{\Delta}^{-1}.$$

Let $u_{ij} = (\mathrm{E}\mathbb{U}\mathbb{U}^T)^{-1}$. Since $n_2 \geq c\,n$ for some constant $c > 0$, using the last display, $[(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}]_{aa}$ is $O_P\big(n_2^{-1}u_{aa}\delta_n^{-(a+a)}\big) = O_P\big(n^{-1}\delta_n^{-(a+a)}\big)$ for $a = 0, 1$. Using the Cauchy-Schwarz inequality, $[(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}]_{ab}$ for $a, b = 0, 1$, is bounded by

$$\sqrt{[(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}]_{aa}[(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}]_{bb}} = O_P\big(n^{-1}\delta_n^{-(a+b)}\big).$$

□

Next, we prove Proposition 4.3.1.

Proof of Proposition 4.3.1. We shall adapt the proof of Proposition 5.1 (a) of Yoo and Ghosal (pear). Define $\boldsymbol{U} = (\boldsymbol{Z}\boldsymbol{V}\boldsymbol{Z}^T + \boldsymbol{I}_n)^{-1}$. We write $|\mathrm{E}_0(\tilde{\sigma}_2^2|\boldsymbol{\theta}_0) - \sigma_0^2|$ as

$$\begin{aligned}&|n^{-1}\sigma_0^2\mathrm{tr}(\boldsymbol{U}) - \sigma_0^2| + n^{-1}(\boldsymbol{F}_0 - \boldsymbol{Z}\boldsymbol{\xi})^T\boldsymbol{U}(\boldsymbol{F}_0 - \boldsymbol{Z}\boldsymbol{\xi}) \\ &\quad \lesssim n^{-1}[\mathrm{tr}(\boldsymbol{I}_n - \boldsymbol{U}) + (\boldsymbol{F}_0 - \boldsymbol{Z}\boldsymbol{\theta}_0)^T\boldsymbol{U}(\boldsymbol{F}_0 - \boldsymbol{Z}\boldsymbol{\theta}_0) + (\boldsymbol{Z}\boldsymbol{\theta}_0 - \boldsymbol{Z}\boldsymbol{\xi})^T\boldsymbol{U}(\boldsymbol{Z}\boldsymbol{\theta}_0 - \boldsymbol{Z}\boldsymbol{\xi})], \quad (4.5.10)\end{aligned}$$

where we have used the Cauchy-Schwarz and the geometric-arithmetic inequalities to get $(x+y)^T\boldsymbol{G}(x+y) \le 2x^T\boldsymbol{G}x + 2y^T\boldsymbol{G}y$. Let $\boldsymbol{P}_{\boldsymbol{Z}} = \boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T$ be the orthogonal projection matrix. Using the binomial inverse theorem (Theorem B.4.1) twice on $\boldsymbol{U}$, we get that

$$\left(\boldsymbol{Z}\boldsymbol{V}\boldsymbol{Z}^T + \boldsymbol{I}_n\right)^{-1} = \boldsymbol{I}_n - \boldsymbol{Z}\left(\boldsymbol{Z}^T\boldsymbol{Z} + \boldsymbol{V}^{-1}\right)\boldsymbol{Z}^T = \boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{Z}} + \boldsymbol{M}, \quad (4.5.11)$$

where $\boldsymbol{M} = \boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}[\boldsymbol{V} + (\boldsymbol{Z}^T\boldsymbol{Z})^{-1}]^{-1}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T \ge \boldsymbol{0}$. Therefore the first term in (4.5.10) is bounded by $n^{-1}\mathrm{tr}(\boldsymbol{P}_{\boldsymbol{Z}}) \lesssim n^{-1}$. Since $\boldsymbol{Z}\boldsymbol{V}\boldsymbol{Z}^T \ge 0$, $\boldsymbol{U} \le \boldsymbol{I}_n$, and so the second term of (4.5.10) is bounded by

$$n^{-1}\|\boldsymbol{U}\|_{(2,2)}\|\boldsymbol{F}_0 - \boldsymbol{Z}\boldsymbol{\theta}_0\|_2^2 \le \|\boldsymbol{F}_0 - \boldsymbol{Z}\boldsymbol{\theta}_0\|_\infty^2 \lesssim \delta_n^4. \quad (4.5.12)$$

By (4.5.11) and $(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{Z}})\boldsymbol{Z} = \boldsymbol{0}$, the last term in (4.5.10) is

$$n^{-1}(\boldsymbol{\theta}_0 - \boldsymbol{\xi})^T[\boldsymbol{V} + (\boldsymbol{Z}^T\boldsymbol{Z})^{-1}]^{-1}(\boldsymbol{\theta}_0 - \boldsymbol{\xi}) \le n^{-1}\sum_{j=0}^{1}\delta_n^j(\theta_{0,j} - \xi_j)^2 = O_{P_0}(n^{-1}),$$

as $\delta_n \to 0$, $\theta_{0,j} = O_{P_0}(1)$ and $\xi_j = O(1)$ for $j = 0, 1$, by the assumptions on the prior. Combining the three bounds thus obtained into (4.5.10), we obtain $|\mathrm{E}_0(\tilde{\sigma}_2^2|\boldsymbol{\theta}_0) - \sigma_0^2| \lesssim n^{-1} + \delta_n^4$.

Next, by writing $\boldsymbol{Y} = \boldsymbol{F}_0 + \boldsymbol{\varepsilon}$, we have

$$n\tilde{\sigma}_2^2 = (\boldsymbol{F}_0 - \boldsymbol{Z}\boldsymbol{\xi})^T\boldsymbol{U}^2(\boldsymbol{F}_0 - \boldsymbol{Z}\boldsymbol{\xi}) + 2(\boldsymbol{F}_0 - \boldsymbol{Z}\boldsymbol{\xi})^T\boldsymbol{U}\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}^T\boldsymbol{U}\boldsymbol{\varepsilon}.$$

As $\boldsymbol{\theta}_0$ and $\boldsymbol{\varepsilon}$ are independent, $\mathrm{Var}(\tilde{\sigma}_2^2|\boldsymbol{\theta}_0)$ is bounded up to a constant multiple by

$$n^{-2}[(\boldsymbol{F}_0 - \boldsymbol{Z}\boldsymbol{\theta}_0)^T\boldsymbol{U}^2(\boldsymbol{F}_0 - \boldsymbol{Z}\boldsymbol{\theta}_0) + (\boldsymbol{Z}\boldsymbol{\theta}_0 - \boldsymbol{Z}\boldsymbol{\xi})^T\boldsymbol{U}^2(\boldsymbol{Z}\boldsymbol{\theta}_0 - \boldsymbol{Z}\boldsymbol{\xi}) + \mathrm{Var}_0(\boldsymbol{\varepsilon}^T\boldsymbol{U}\boldsymbol{\varepsilon})]. \quad (4.5.13)$$

By an argument similar to that used to show (4.5.12), the first term in (4.5.13) is bounded by

$$n^{-2}\|\boldsymbol{U}\|_{(2,2)}^2\|\boldsymbol{F}_0-\boldsymbol{Z}\boldsymbol{\theta}_0\|_2^2 \le n^{-1}\|\boldsymbol{F}_0-\boldsymbol{Z}\boldsymbol{\theta}_0\|_\infty^2 \lesssim n^{-1}\delta_n^4.$$

Using (4.5.11), idempotency of $\boldsymbol{I}_n-\boldsymbol{P}_{\boldsymbol{Z}}$ and $(\boldsymbol{I}_n-\boldsymbol{P}_{\boldsymbol{Z}})\boldsymbol{Z}-\boldsymbol{0}$, the second term in (4.5.13) is bounded above by

$$n^{-2}(\boldsymbol{\theta}_0-\xi)^T\boldsymbol{Z}^T\boldsymbol{Z}(\boldsymbol{\theta}_0-\xi).$$

Using arguments from the proof of Lemma 4.5.2, the last two terms of (4.5.13) can be shown to be bounded by a constant multiple of n^{-1}. We therefore obtain that $\mathrm{Var}(\tilde{\sigma}_2^2|\boldsymbol{\theta}_0)=O_P(n^{-1})$, and the mean square error is

$$\mathrm{E}_0((\tilde{\sigma}_2^2-\sigma_0^2)^2 \lesssim n^{-1}+\delta_n^4, \tag{4.5.14}$$

which implies the first assertion.

(b) Note that by posterior conjugacy of the inverse-gamma distribution, $\mathrm{E}(\sigma^2|D_n)\lesssim n^{-1}+\tilde{\sigma}_*^2$, and $\mathrm{Var}(\sigma^2|D_n)\lesssim n^3+n^{-1}\tilde{\sigma}_*^4$. By the Markov's inequality, the second stage posterior of σ^2 concentrates near the plug-in Bayes estimator $\tilde{\sigma}_*^2$, and therefore inherits the rate of $\tilde{\sigma}_*^2$ as established in part (a). □

For the remaining part of this section, we shall use $\mathscr{K}_n$ to denote a shrinking neighborhood of σ_0 such that $\Pi(\mathscr{K}_n|D_n)\to 0$. We shall now show evaluate the second stage contraction rate of $\boldsymbol{\theta}$ at $\boldsymbol{\theta}_0$.

Lemma 4.5.3. *For any $\sigma\in\mathscr{K}_n$, $\delta_n \gg n^{-1/3}$ and $\delta_n\to 0$,*

$$\mathrm{E}[|\theta_0-\theta_{00}|\big|D_n,\sigma]=O_{P_0}(\max\{n^{-1/2},\delta_n^2,n^{-1}\delta_n^{-1}\}),$$
$$\mathrm{E}[|\theta_1-\theta_{01}|\big|D_n,\sigma]=O_{P_0}(\max\{n^{-1/2}\delta_n^{-1},\delta_n,n^{-1}\delta_n^{-2}\}).$$

In particular, for $\delta_n \asymp n^{-1/3}(\log n)^t$ for any choice of $t>0$, the contraction rates of θ_0 and θ_1 at θ_{00} and θ_{01} are $n^{-1/2}$ and $n^{-1/6}(\log n)^{-t}$ respectively.

Proof. Let $0\le h\le 1$. By the assumption on the prior, $\boldsymbol{V}>\boldsymbol{0}$. We have by Lemma 4.5.2 that

$$\sup_{\sigma\in\mathscr{K}_n}\mathrm{Var}(\theta_h|D_n,\sigma)=[\sigma_0^2+o(1)]\left[\left(\boldsymbol{Z}^T\boldsymbol{Z}+\boldsymbol{V}^{-1}\right)^{-1}\right]_{hh}\lesssim\left[\left(\boldsymbol{Z}^T\boldsymbol{Z}\right)^{-1}\right]\lesssim n^{-1}\delta_n^{-2h}. \tag{4.5.15}$$

The conditional posterior mean of $\boldsymbol{\theta}$ has the bias

$$\begin{aligned}\mathrm{E}(\boldsymbol{\theta}|D_n,\sigma)-\boldsymbol{\theta}_0 &= \left(\boldsymbol{Z}^T\boldsymbol{Z}+\boldsymbol{V}^{-1}\right)^{-1}\left(\boldsymbol{Z}^T\boldsymbol{Y}+\boldsymbol{V}^{-1}\boldsymbol{\xi}\right)-\boldsymbol{\theta}_0 \\ &= \left(\boldsymbol{Z}^T\boldsymbol{Z}+\boldsymbol{V}^{-1}\right)^{-1}\left[\boldsymbol{Z}^T\boldsymbol{\varepsilon}+\boldsymbol{Z}^T(\boldsymbol{F}_0-\boldsymbol{Z}\boldsymbol{\theta}_0)+\boldsymbol{V}^{-1}\left(\boldsymbol{\xi}-\boldsymbol{\theta}_0\right)\right]. \qquad (4.5.16)\end{aligned}$$

The h-th diagonal entry of the first term in (4.5.16) has expectation equal to zero, and using arguments similar to (4.5.15), its variance is equal to

$$\sigma_0^2\left[\left(\boldsymbol{Z}^T\boldsymbol{Z}+\boldsymbol{V}^{-1}\right)^{-1}\boldsymbol{Z}^T\boldsymbol{Z}\left(\boldsymbol{Z}^T\boldsymbol{Z}+\boldsymbol{V}^{-1}\right)^{-1}\right]_{hh} \lesssim n^{-1}\delta_n^{-2h}.$$

Therefore, $\left[\left(\boldsymbol{Z}^T\boldsymbol{Z}+\boldsymbol{V}^{-1}\right)^{-1}\boldsymbol{Z}^T\boldsymbol{\varepsilon}\right]_h$ is $O_{P_0}\left(n^{-1/2}\delta_n^{-h}\right)$ for $h=0,1$. Next, let $\beta_{hj}=\left[\left(\boldsymbol{Z}^T\boldsymbol{Z}+\boldsymbol{V}^{-1}\right)^{-1}\right]_{hj}$, and $\gamma_i=\left(\boldsymbol{F}_0-\boldsymbol{Z}\boldsymbol{\theta}_0\right)_i$, for $h,j=0,1$ and $i=1,\ldots,n_2$. Using the Cauchy-Schwarz inequality and the fact that $\boldsymbol{V}^{-1}>\boldsymbol{0}$, we have that

$$\begin{aligned}\left[\left(\boldsymbol{Z}^T\boldsymbol{Z}+\boldsymbol{V}^{-1}\right)^{-1}\right]_{hj} &\le \sqrt{\left[\left(\boldsymbol{Z}^T\boldsymbol{Z}+\boldsymbol{V}^{-1}\right)^{-1}\right]_{hh}\left[\left(\boldsymbol{Z}^T\boldsymbol{Z}+\boldsymbol{V}^{-1}\right)^{-1}\right]_{jj}} \\ &\le \sqrt{\left[\left(\boldsymbol{Z}^T\boldsymbol{Z}\right)^{-1}\right]_{hh}\left[\left(\boldsymbol{Z}^T\boldsymbol{Z}\right)^{-1}\right]_{jj}} \\ &\lesssim n_2^{-1}\delta_n^{-(h+j)}.\end{aligned}$$

Therefore the h-th entry of $\left[\left(\boldsymbol{Z}^T\boldsymbol{Z}+\boldsymbol{V}^{-1}\right)^{-1}\boldsymbol{Z}^T\left(\boldsymbol{F}_0-\boldsymbol{Z}\boldsymbol{\theta}_0\right)\right]$ is

$$\beta_{h0}\sum_{i=1}^{n_2}\gamma_i+\beta_{h1}\sum_{i=1}^{n_2}z_i\gamma_i \lesssim \delta_n^{-h}\delta_n^2,$$

where the last inequality follows from the fact that $|z_i|\le\delta_n$ as $z_i\in\mathcal{Q}_n$ for $i=1,\ldots,n_2$, and (4.5.7). The second term of (4.5.16) is therefore bounded above by a constant multiple of δ_n^{2-h}. For the last term, note that $|\theta_{0j}|=O_{P_0}(1)$ as $|\tilde{\mu}-\mu_0|=o_{P_0}(1)$, and $|\xi_j|=O(1)$ by the assumption on the prior, for $j=0,1$. Therefore, $|\theta_{0j}-\xi_h|\le|\theta_{0j}|+|\xi_j|=O_{P_0}(1)$. Then using Lemma 4.5.2, for $h=0,1$, we obtain that

$$\left[\left(\boldsymbol{Z}^T\boldsymbol{Z}\right)^{-1}\boldsymbol{V}^{-1}\left(\boldsymbol{\xi}-\boldsymbol{\theta}_0\right)\right]_h=\beta_{h0}\sum_{j=0}^{1}\left(\boldsymbol{V}^{-1}\right)_{0j}(\xi_j-\theta_{0j})+\beta_{h1}\sum_{j=0}^{1}\left(\boldsymbol{V}^{-1}\right)_{1j}(\xi_j-\theta_{1j})\lesssim n^{-1}\delta_n^{-h}.$$

Combining the bounds thus obtained into (4.5.16), we obtain that

$$\mathrm{E}[|\theta_h-\theta_{0h}|\big|D_n,\sigma]=O_{P_0}(\max\{n^{-1/2}\delta_n^{-h},\delta_n^{2-h},n^{-1}\delta_n^{-h}\}),$$

for $h = 0, 1$, and the proof follows. □

Proof of Theorem 4.3.2. First, using the fact that $\theta_{00} = f_0(\mu_0) + f_0'(\mu_0)(\tilde{\mu} - \mu_0)$ and $\theta_{01} = f_0'(\mu_0)$, note that

$$\begin{aligned} f^{-1}(y_0) - f_{\boldsymbol{\theta}_0}^{-1}(y_0) &= \frac{y_0 - \theta_0}{\theta_1} - \frac{y_0 - \theta_{00}}{\theta_{01}} \\ &= (f_0(\mu_0) - \theta_{00})\left(\frac{1}{\theta_1} - \frac{1}{\theta_{01}}\right) - (\theta_0 - \theta_{00})\left(\frac{1}{\theta_1} - \frac{1}{\theta_{01}}\right) - \frac{\theta_0 - \theta_{00}}{\theta_{01}} \\ &= -f_0'(\mu_0)(\tilde{\mu} - \mu_0)\left(\frac{1}{\theta_1} - \frac{1}{\theta_{01}}\right) - (\theta_0 - \theta_{00})\left(\frac{1}{\theta_1} - \frac{1}{\theta_{01}}\right) - \frac{\theta_0 - \theta_{00}}{f_0'(\mu_0)} \end{aligned} \tag{4.5.17}$$

For the chosen $\delta_n \asymp n^{-1/3}(\log n)^t$ for some $t > 0$, using Theorem 4.2.1 and Lemma 4.5.3, we obtain that $\mathrm{E}(|f^{-1}(y_0) - f_{\boldsymbol{\theta}_0}^{-1}(y_0)| \big| D_n)$ is of the order

$$n^{-1/3} n^{-1/6} (\log n)^{-t} + n^{-1/2} n^{-1/6} (\log n)^{-t} + n^{-1/2} \lesssim n^{-1/2}.$$

Also, note that $f_{\boldsymbol{\theta}_0}^{-1}(y_0) - f_0^{-1}(y_0) = (f_0(\mu_0) - \theta_{00})/\theta_{01} + (\tilde{\mu} - \mu_0) = 0$. Therefore, for every $M_n \to \infty$,

$$\begin{aligned} \Pi\left(\left|f^{-1}(y_0) - f_0^{-1}(y_0)\right| > M_n n^{-1/2} \big| D_n\right) &\leq \Pi\left(\left|f^{-1}(y_0) - f_{\boldsymbol{\theta}_0}^{-1}(y_0)\right| > M_n n^{-1/2} \big| D_n\right) \\ &\to 0. \end{aligned}$$

□

REFERENCES

Akakpo, N., Balabdaoui, F., and Durot, C. (2014). Testing monotonicity via local least concave majorants. *Bernoulli*, 20(2):514–544.

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T., and Silverman, E. (1955). An empirical distribution function for sampling with incomplete information. *Ann. Math. Statist.*, 26:641–647.

Balabdaoui, F. and Wellner, J. A. (2007). Estimation of a k -monotone density: Limit distribution theory and the spline connection. *Ann. Statist.*, 35(6):2536–2564.

Banerjee, M. (2000). *Likelihood Ratio Inference in Regular and Non-regular Problems*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)–University of Washington.

Baraud, Y., Huet, S., and Laurent, B. (2005). Testing convex hypotheses on the mean of a Gaussian vector. Application to testing qualitative hypotheses on a regression function. *Ann. Statist.*, 33(1):214–257.

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.* John Wiley & Sons, London-New York-Sydney. Wiley Series in Probability and Mathematical Statistics.

Barlow, R. E. and Brunk, H. D. (1972). The isotonic regression problem and its dual. *J. Amer. Statist. Assoc.*, 67:140–147.

Barron, A., Schervish, M. J., and Wasserman, L. (1999). The consistency of posterior distributions in nonparametric problems. *Ann. Statist.*, 27(2):536–561.

Belitser, E. (2017). On coverage and local radial rates of credible sets. *Ann. Statist.*, 45(3):1124–1151.

Belitser, E. and Nurushev, N. (2015). Needles and straw in a haystack: robust confidence for possibly sparse sequences. *arXiv e-prints*, page arXiv:1511.01803.

Belitser, E. and Nurushev, N. (2019). Robust inference for general framework of projection structures. *arXiv e-prints*, page arXiv:1904.01003.

Belitser, E. N. and Ghosal, S. (2017). Empirical bayes oracle uncertainty quantification for regression.

Bhaumik, P. and Ghosal, S. (2015). Bayesian two-step estimation in differential equation models. *Electron. J. Stat.*, 9(2):3124–3154.

Billingsley, P. (1999). *Convergence of Probability Measures*. Wiley, New York, 2nd edition.

Bowman, A. W., Jones, M. C., and Gijbels, I. (1998). Testing monotonicity of regression. *J. Comput. Graph. Statist.*, 7(4):489–500.

Brunk, H. D. (1970). Estimation of isotonic regression. In *Nonparametric Techniques in Statistical Inference (Proc. Sympos., Indiana Univ., Bloomington, Ind., 1969)*, pages 177–197. Cambridge Univ. Press, London.

Castillo, I. and Nickl, R. (2013). Nonparametric Bernstein-von Mises theorems in Gaussian white noise. *Ann. Statist.*, 41(4):1999–2028.

Cox, D. D. (1993). An analysis of Bayesian inference for nonparametric regression. *Ann. Statist.*, 21(2):903–923.

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.

Durot, C. (2002). Sharp asymptotics for isotonic regression. *Probab. Theory Related Fields*, 122(2):222–240.

Ferguson, T. S. (1983). Bayesian density estimation by mixtures of normal distributions. In *Recent Advances in Statistics*, pages 287–302. Academic Press, New York.

Freedman, D. (1999). On the Bernstein-von Mises theorem with infinite-dimensional parameters. *Ann. Statist.*, 27(4):1119–1140.

Ghosal, S., Ghosh, J. K., and Ramamoorthi, R. V. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *Ann. Statist.*, 27(1):143–158.

Ghosal, S., Ghosh, J. K., and van der Vaart, A. W. (2000a). Convergence rates of posterior distributions. *Ann. Statist.*, 28(2):500–531.

Ghosal, S., Sen, A., and van der Vaart, A. W. (2000b). Testing monotonicity of regression. *Ann. Statist.*, 28(4):1054–1082.

Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*, volume 44 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

Grenander, U. (1957). On the theory of mortality measurement. *Skand. Aktuarietidskr., 39:125âĂŞ153*.

Groeneboom, P. and Jongbloed, G. (2014). *Nonparametric Estimation under Shape Constraints*, volume 38 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, New York. Estimators, algorithms and asymptotics.

Groeneboom, P. and Wellner, J. A. (2001). Computing Chernoff's distribution. *J. Comput. Graph. Statist.*, 10(2):388–400.

Hall, P. and Heckman, N. E. (2000). Testing for monotonicity of a regression mean by calibrating for linear functions. *Ann. Statist.*, 28(1):20–39.

Jones, P., Parker, D., Osborn, T., and Briffa, K. (2016). Global and hemispheric temperature anomalies. Land and marine instrumental records. *Trends: A Compendium of Data on Global Change. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A.*

Knapik, B. T., van der Vaart, A. W., and van Zanten, J. H. (2011). Bayesian inverse problems with Gaussian priors. *Ann. Statist.*, 39(5):2626–2657.

Knapik, B. T., van der Vaart, A. W., and van Zanten, J. H. (2013). Bayesian recovery of the initial condition for the heat equation. *Comm. Statist. Theory Methods*, 42(7):1294–1313.

Kosorok, M. R. (2008). Bootstrapping the Grenander estimator. In *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*, volume 1 of *Inst. Math. Stat. (IMS) Collect.*, pages 282–292. Inst. Math. Statist., Beachwood, OH.

Leahu, H. (2011). On the Bernstein-von Mises phenomenon in the Gaussian white noise model. *Electron. J. Stat.*, 5:373–404.

Lin, L. and Dunson, D. B. (2014). Bayesian monotone regression using Gaussian process projection. *Biometrika*, 101(2):303–317.

Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates. I. Density estimates. *Ann. Statist.*, 12(1):351–357.

Martin, R. (2017). Empirical priors and posterior concentration rates for a monotone density. *arXiv e-prints*, page arXiv:1706.08567.

Neelon, B. and Dunson, D. B. (2004). Bayesian isotonic regression and trend analysis. *Biometrics*, 60(2):398–406.

Ning, B. and Ghosal, S. (2018). Bayesian linear regression for multivariate responses under group sparsity. *arXiv e-prints*, page arXiv:1807.03439.

Prakasa Rao, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A*, 31:23–36.

Ray, K. (2017). Adaptive Bernstein–von Mises theorems in Gaussian white noise. *Ann. Statist.*, 45(6):2511–2536.

Robertson, T. and Wright, F. T. (1973). Multiple isotonic median regression. *Ann. Statist.*, 1:422–432.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*, volume 12 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

Salomond, J.-B. (2014). Adaptive Bayes test for monotonicity. In *The Contribution of Young Researchers to Bayesian Statistics*, volume 63 of *Springer Proc. Math. Stat.*, pages 29–33. Springer, Cham.

Salomond, J.-B. (2018). Testing un-separated hypotheses by estimating a distance. *Bayesian Anal.*, 13(2):461–484.

Schwartz, L. (1965). On Bayes procedures. *Z. Verw. Gebiete*, 4:10–26.

Sen, B., Banerjee, M., and Woodroofe, M. (2010). Inconsistency of bootstrap: the Grenander estimator. *Ann. Statist.*, 38(4):1953–1977.

Serra, P. and Krivobokova, T. (2017). Adaptive empirical Bayesian smoothing splines. *Bayesian Anal.*, 12(1):219–238.

Shively, T. S., Sager, T. W., and Walker, S. G. (2009). A Bayesian approach to non-parametric monotone function estimation. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 71(1):159–175.

Sniekers, S. and van der Vaart, A. (2015a). Adaptive Bayesian credible sets in regression with a Gaussian process prior. *Electron. J. Stat.*, 9(2):2475–2527.

Sniekers, S. and van der Vaart, A. (2015b). Credible sets in the fixed design model with Brownian motion prior. *J. Statist. Plann. Inference*, 166:78–86.

Szabó, B., van der Vaart, A. W., and van Zanten, J. H. (2015). Rejoinder to discussions of "Frequentist coverage of adaptive nonparametric Bayesian credible sets". *Ann. Statist.*, 43(4):1463–1470.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Process With Applications to Statistics*. Springer-Verlag New York, Inc.

Williamson, R. E. (1956). Multiply monotone functions and their Laplace transforms. *Duke Math. J.*, 23:189–207.

Yoo, W. W. and Ghosal, S. (2016). Supremum norm posterior contraction and credible sets for nonparametric multivariate regression. *Ann. Statist.*, 44(3):1069–1102.

Yoo, W. W. and Ghosal, S. (to appear). Bayesian mode and maximum estimation and accelerated rates of contraction. *Bernoulli*.

APPENDICES

APPENDIX

A

AUXILIARY RESULTS

In this appendix we state and prove some results that have been derived in the course of obtaining the main results of this dissertation, and are of independent interest as well.

A.1 Approximation of a monotone function by step functions

Lemma A.1.1. *Let $p \geq 1$. For $f_0 \in \mathscr{F}(K)$ and $J > 1$, there exists $f_{0J} \in \mathscr{F}(K)$ such that $f_{0J}(\cdot) = \sum_{j=1}^{J} \theta_{0j} \mathbb{1}_{[\xi_{0,j-1}, \xi_{0,j})}(\cdot)$ with $0 = \xi_{0,0} < \xi_{0,1} < \cdots < \xi_{0,J-1} < \xi_{0,J} = 1$, $\{\xi_{0,1}, \ldots, \xi_{0,J-1}\} \subset \{X_1, \ldots, X_n\}$ and f_{0J} satisfying*

$$\{n^{-1} \sum_{i=1}^{n} \left| f_0(X_i) - f_{0J}(X_i) \right|^p\}^{1/p} \lesssim J^{-1}. \tag{A.1.1}$$

Proof. We shall adapt the proof of Theorem 2.7.5 of van der Vaart and Wellner (1996) to the context of the empirical distance with the knots chosen from the design points. We shall show that for $f_0 \in \mathscr{F}(K)$ and given $\epsilon \in (0,1)$, there exists $J \geq 1$ such that $J \lesssim K/\epsilon$, $\{\xi_1, \ldots, \xi_{J-1}, \xi_J = 1\}$ such that $\{\xi_1, \ldots, \xi_{J-1}\} \subset \{X_1, \ldots, X_n\}$ and $0 = \xi_0 < \xi_1 < \cdots < \xi_{J-1} < \xi_J =$

1, and $(\theta_{01},\ldots,\theta_{0J})\in\mathbb{R}^J$ such that $f_{0J}(\cdot)=\sum_{j=1}^{J}\theta_{0j}\mathbb{1}\{[\xi_{j-1},\xi_j)\}(\cdot)$ satisfies

$$n^{-1}\sum_{i=1}^{n}\left|f_0(X_i)-f_{0J}(X_i)\right|^p\lesssim\epsilon^p.$$

Let $\epsilon\in(0,1)$, $c=2^{-1/p}$. Consider the coarsest partition $\mathscr{P}_0=\{0=\xi_0^0<\xi_1^0=1\}$. At any stage $k\geq 0$, let $J(k)$ denote the number of intervals obtained from the partition $\mathscr{P}_k$ of $[0,1]$. Now define ϵ_k for stage $k\geq 0$ as

$$\epsilon_k=\max_{1\leq j\leq J(k)}\left(f_0(\xi_j^k)-f_0(\xi_{j-1}^k)\right)\left(N_j^k/n\right)^{1/p},$$

where $N_j^k=\sum_{i=1}^{n}\mathbb{1}\{X_i\in[\xi_{j-1}^k,\xi_j^k)\}$. In order to obtain the partition $\mathscr{P}_{k+1}$ from $\mathscr{P}_k$, let $\tilde{X}_{(j)}^k$ be the median of X_i's contained in $[\xi_{j-1}^k,\xi_j^k)$. For an even number of $X_1,\ldots,X_n$, either of the two middle values may be chosen. Also, denote by $\mathscr{S}_k$ the collection of all $j\in\{1,\ldots,J(k)\}$ such that

$$\left(f_0(\xi_j^k)-f_0(\xi_{j-1}^k)\right)\left(N_j^k/n\right)^{1/p}>c\epsilon_k.$$

We then define $\mathscr{P}_{k+1}$ as $\mathscr{P}_{k+1}=\mathscr{P}_k\bigcup\{\bigcup_{j\in\mathscr{S}_k}\{\tilde{X}_{(j)}^k\}\}$. Denote the elements of $\mathscr{P}_{k+1}$ as $\{\xi_j^{k+1}:0\leq j\leq J,\ 0=\xi_0^{k+1}<\xi_1^{k+1}<\cdots<\xi_{J-1}^{k+1}<\xi_J^{k+1}=1\}$. We continue partitioning $[0,1]$ on the design points $X_1,\ldots,X_n$ in the same manner until at some stage k, $\epsilon_k^p\leq\epsilon^{p+1}$. Also, note that by the definitions of ϵ_k and $\mathscr{P}_k$, we have that $\epsilon_{k+1}\leq c\epsilon_k\leq 2\epsilon_{k+1}$. We now show that $J(k)$ is bounded above by a constant multiple of $\epsilon_k^{-2/3}$. First, denote by s_k the number of intervals of $\mathscr{P}_k$ that have been subdivided to obtain $\mathscr{P}_{k+1}$. In other words, s_k is the cardinality of $\mathscr{S}_k$. Using Hölder's inequality, we have

$$\begin{aligned}
s_k(c\epsilon_k)^{p/(p+1)}&\leq\sum_{j\in\mathscr{S}_k}\left(f_0(\xi_j^k)-f_0(\xi_{j-1}^k)\right)^{p/(p+1)}\left(N_j^k/n\right)^{1/(p+1)}\\
&\leq\sum_{j=1}^{J(k)}\left(f_0(\xi_j^k)-f_0(\xi_{j-1}^k)\right)^{p/(p+1)}\left(N_j^k/n\right)^{1/(p+1)}\\
&\leq\left\{\sum_{j=1}^{J(k)}\left(f_0(\xi_j^k)-f_0(\xi_{j-1}^k)\right)\right\}^{p/(p+1)}\left\{\sum_{j=1}^{J(k)}N_j^k/n\right\}^{1/(p+1)}\\
&\leq(f_0(1)-f_0(0))^{p/(p+1)}\cdot 1\\
&\lesssim K^{p/(p+1)}.
\end{aligned}$$

Therefore $s_k(c\,\epsilon_k)^{p/(p+1)} \lesssim K^{p/(p+1)}$. Next we have

$$\begin{aligned}
\sum_{l=1}^{k} J(l) &= k + \sum_{l=1}^{k} l\, s_{k-l} \\
&\lesssim k + \sum_{l=1}^{k} l (c\,\epsilon_{k-l})^{-p/(p+1)} \\
&\leq k + \sum_{l=1}^{k} l (c\,c^{-l}\,\epsilon_k)^{-p/(p+1)} \\
&\lesssim \left\{ 2 \sum_{l=1}^{k} l\, c^{2(l-1)/3} \right\} \epsilon_k^{-p/(1+p)} \\
&\lesssim \epsilon_k^{-p/(1+p)}.
\end{aligned}$$

We now construct the approximating functions f_{0k} for every $k \geq 0$. Let $f_{00} = 0$ on $[0,1]$. Given $f_{0(k-1)}$, define f_{0k} as

$$f_{0k}(x) = f_{0(k-1)}(x) + l_j^k \epsilon_k (N_j^k/n)^{-1/p}, \quad x \in [\xi_{j-1}^k, \xi_j^k),$$

where $l_j^k \geq 0$ is the largest integer such that $f_{0k} \leq f_0$. Next, from the construction of f_{0k} we have that

$$\left(f_0(\xi_j^k) - f_{0k}(\xi_{j-1}^k)\right)\left(N_j^k/n\right)^{1/p} \leq \epsilon_k.$$

From the definition of ϵ_k we have

$$\left(f_0(x) - f_0(\xi_j^k)\right)\left(N_j^k/n\right)^{1/p} \leq \epsilon_k, \quad x \in [\xi_{j-1}^k, \xi_j^k).$$

The last two equations give us that

$$\left(f_0(x) - f_{0k}(\xi_{j-1}^k)\right)\left(N_j^k/n\right)^{1/p} \leq 2\epsilon_k, \quad x \in [\xi_{j-1}^k, \xi_j^k).$$

We therefore have

$$n^{-1} \sum_{i=1}^{n} \left| f_0(X_i) - f_{0k}(X_i) \right|^p = n^{-1} \sum_{j=1}^{J} \sum_{i : X_i \in [\xi_{j-1}^k, \xi_j^k)} \left| f_0(X_i) - f_{0k}(\xi_{j-1}^k) \right|^p$$

$$\lesssim n^{-1}\sum_{j=1}^{J}\sum_{i:X_i\in[\xi_{j-1}^k,\xi_j^k)}\epsilon_k^p(N_j^k/n)^{-1}$$
$$= J(k)\epsilon_k^p,$$

which is bounded by a constant multiple of $\epsilon_k^p\epsilon_k^{-p/(p+1)}=\epsilon_k^{p^2/(p+1)}$. For the final partition k, $J(k)\lesssim\epsilon_k^{-p/(p+1)}$, and $\epsilon_k^p\le\epsilon^{p+1}$. Therefore

$$n^{-1}\sum_{i=1}^{n}\left|f_0(X_i)-f_{0k}(X_i)\right|^p\lesssim\epsilon_k^{p^2/(p+1)}\le\epsilon^p.$$

To show $J(k)\lesssim\epsilon^{-1}$, we note that since k is the first time $\epsilon_k^p\le\epsilon^{p+1}$, we have that $\epsilon_{k-1}^p>\epsilon^{p+1}$. Hence $J(k-1)\lesssim\epsilon_{k-1}^{-p/(p+1)}\le\epsilon^{-1}$, and therefore $J(k)\le 2J(k-1)\lesssim\epsilon^{-1}$. □

A.2 Asymptotic behavior of the point-wise sieve-MLE

For the model (2.0.1), let $\hat{f}_n$ denote the sieve-maximum likelihood estimator of f, subject the monotonicity constraint on f, obtained by minimizing the error sum of squares $\sum_{i=1}^n(Y_i-\hat{f}_n(X_i))^2$ as

$$\sum_{j=1}^{J}\sum_{i:X_i\in I_j}(Y_i-q_j)^2=\sum_{j=1}^{J}\sum_{i:X_i\in I_j}(Y_i-\bar{Y}_j)^2+\sum_{j=1}^{J}N_j(\bar{Y}_j-q_j)^2.$$

The solution is obtained by minimizing the weighted sum of squares $\sum_{j=1}^J(N_j/n)(\bar{Y}_j-q_j)^2$ subject to the ordering constraint $q_1\le\cdots\le q_J$. From the algorithm for weighted isotonic regression, q_j is the left-derivative of the greatest convex minorant of the graph of (A.2.1) at $\sum_{k=1}^j N_k/n$. Then for x_0 such that $f_0'(x_0)>0$, the following theorem gives the asymptotic distribution of $\hat{f}_n(x_0)$.

Theorem A.2.1. *Let $n^{1/3}\ll J\ll n^{2/3}$. Let W_1 be a two-sided Brownian motion on $\mathbb{R}$ with $W_1(0)=0$, $Z_1=\operatorname{argmin}\{W_1(t)+t^2:t\in\mathbb{R}\}$, and $C=2b\,(a/b)^{2/3}$ with $a=\sqrt{\sigma_0^2/g(x_0)}$ and $b=f_0'(x_0)/2$. Let $\boldsymbol{X}$ be random, satisfying* (2.1.2)*. Then under Assumption A on the errors, for every $z\in\mathbb{R}$,*

$$\mathrm{P}_0(n^{1/3}(\hat{f}_n(x_0)-f_0(x_0))\le z)\to\mathrm{P}(CZ_1\le z).$$

Proof. From the algorithm for obtaining the isotonic regression estimate for $f(x_0)$ as described in Section 2.3, the sieve-maximum likelihood estimator $\hat{f}_n(x_0)$ is the left derivative

of the greatest convex minorant of the graph

$$\left\{(0,0),\left(\frac{N_1}{n},\frac{N_1}{n}\bar{Y}_1\right),\left(\sum_{k=1}^{2}\frac{N_k}{n},\sum_{k=1}^{2}\frac{N_k}{n}\bar{Y}_k\right),\ldots,\left(\sum_{k=1}^{J}\frac{N_k}{n},\sum_{k=1}^{J}\frac{N_k}{n}\bar{Y}_k\right)\right\}, \tag{A.2.1}$$

at the point $\sum_{k=1}^{\lceil x_0 J\rceil} N_k/n$. Let U_n, G_n be stochastic processes denoting $U_n(s)=\sum_{j=1}^{\lceil sJ\rceil}(N_j/n)\bar{Y}_j$ and $G_n(s)=\sum_{j=1}^{\lceil sJ\rceil} N_j/n$. Using the switch relation (B.2.1), we evaluate $\mathrm{P}_0(n^{1/3}(\hat{f}_n(x_0)-f_0(x_0))\le z)$ for $z\in\mathbb{R}$ as

$$\begin{aligned}
&\mathrm{P}_0(\hat{f}_n(x_0)\le f_0(x_0)+n^{1/3}z)\\
&\quad=\mathrm{P}_0\left(\operatorname*{arg\,min}_{s\ge 0}\left\{U_n(s)-(f_0(x_0)+n^{-1/3}z)G_n(s)\right\}\ge\sum_{k=1}^{\lceil x_0J\rceil}\frac{N_k}{n}\right)\\
&\quad=\mathrm{P}_0\left(\operatorname*{arg\,min}_{s\in\mathbb{R}}\left\{U_n(s)-(f_0(x_0)+n^{-1/3}z)G_n(s)\right\}\ge\sum_{k=1}^{\lceil x_0J\rceil}\frac{N_k}{n}\right)\\
&\quad=\mathrm{P}_0\left(\operatorname*{arg\,min}_{s\in\mathbb{R}}\left\{\frac{n^{2/3}}{g(x_0)}(U_n(s)-U_n(x_0))-\right.\right.\\
&\qquad\left.\left.\frac{n^{2/3}}{g(x_0)}(f_0(x_0)+n^{-1/3}z)(G_n(s)-G_n(x_0))\right\}\ge\sum_{k=1}^{\lceil x_0J\rceil}\frac{N_k}{n}\right).
\end{aligned}$$

We use a change of variable $s=x_0+n^{-1/3}t$. Similar to the proof of part (b), we use the notation $\sum_{j=\lceil x_0J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil}(N_j/n)\bar{Y}_j$ for $\sum_{j=\lceil (x_0+n^{-1/3}t)J\rceil+1}^{\lceil x_0J\rceil}(N_j/n)\bar{Y}_j$ when $t<0$. A similar rule shall be applied for $\sum_{j=\lceil x_0J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} N_j/n$. For $\tau_n=n^{1/3}(\sum_{k=1}^{\lceil x_0J\rceil}(N_k/n)-x_0)$, the last expression in the display then equals to

$$\begin{aligned}
&\mathrm{P}_0\left(\operatorname*{arg\,min}_{t\in\mathbb{R}}\left\{\frac{n^{2/3}}{g(x_0)}(U_n(x_0+n^{-1/3}t)-U_n(x_0))\right.\right.\\
&\qquad\left.\left.-\frac{n^{2/3}}{g(x_0)}(f_0(x_0)+n^{-1/3}z)(G_n(x_0+n^{-1/3}t)-G_n(x_0))\right\}\ge\tau_n\right)\\
&=\mathrm{P}_0\left(\operatorname*{arg\,min}_{t\in\mathbb{R}}\left\{\frac{n^{2/3}}{g(x_0)}\sum_{j=\lceil x_0J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil}\frac{N_j}{n}\bar{Y}_j\right.\right.\\
&\qquad\left.\left.-\frac{n^{2/3}}{g(x_0)}(f_0(x_0)+n^{-1/3}z)\sum_{j=\lceil x_0J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil}\frac{N_j}{n}\right\}\ge\tau_n\right).
\end{aligned}$$

In the proof of part (b), it was shown that $\tau_n \to 0$, and

$$\sup_{t\in[-K,K]} \left| n^{1/3} z \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} \frac{N_j}{n} - zt \right| \to^{P_0} 0.$$

We shall therefore evaluate

$$\mathrm{P}_0\left(\operatorname*{arg\,min}_{t\in\mathbb{R}} \left\{ \frac{n^{2/3}}{g(x_0)} \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} \frac{N_j}{n}(\bar{Y}_j - f_0(x_0)) - zt \right\} \geq 0 \right).$$

From Lemma A.2.2, Lemma A.2.3, Theorem B.1.6, and part (b) of Lemma A.2.7, we have that

$$\begin{aligned}
&\mathrm{P}_0(n^{1/3}(\hat{f}_n(x_0) - f_0(x_0)) \leq z) \\
&\quad \to \mathrm{P}\left(\operatorname*{arg\,min}_{t\in\mathbb{R}} \left\{ a W_1(t) + b t^2 - zt \right\} \geq 0 \right) \\
&\quad = \mathrm{P}\left(\left(\frac{a}{b}\right)^{2/3} \operatorname*{arg\,min}_{t\in\mathbb{R}} \left\{ W_1(t) + t^2 \right\} + z/(2b) \geq 0 \right) \\
&\quad = \mathrm{P}\left(2b\left(\frac{a}{b}\right)^{2/3} \operatorname*{arg\,min}_{t\in\mathbb{R}} \left\{ W_1(t) + t^2 \right\} \geq -z \right) \\
&\quad = \mathrm{P}\left(2b\left(\frac{a}{b}\right)^{2/3} \operatorname*{arg\,min}_{t\in\mathbb{R}} \left\{ W_1(t) + t^2 \right\} \leq z \right),
\end{aligned}$$

where the last step follows from the fact that $\operatorname{arg\,min}\left\{ W_1(t) + t^2 : t \in \mathbb{R} \right\}$ is symmetric about zero. Substituting $C = 2b(a/b)^{2/3}$ in the last expression of the display, we get that

$$\mathrm{P}_0(n^{1/3}(\hat{f}_n(x_0) - f_0(x_0)) \leq z) \to \mathrm{P}\left(C \operatorname{arg\,min}\left\{ W_1(t) + t^2 : t \in \mathbb{R} \right\} \leq z \right).$$

□

Lemma A.2.2. *Let $a = \sigma_0^2/g(x_0)$, $b = f_0'(x_0)/2$, and W_1 be a two-sided Brownian motion on $\mathbb{R}$ starting at zero. Let $n^{1/3} \ll J \ll n^{2/3}$. Then for all $K > 0$,*

$$\left\{ \frac{n^{2/3}}{g(x_0)} \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} \frac{N_j}{n}(\bar{Y}_j - f_0(x_0)), \quad t \in [-K,K] \right\},$$

converges to $\{a W_1(t) + b t^2 : t \in [-K,K]\}$ in $L^\infty([-K,K])$.

Proof of Lemma A.2.2. We shall use Theorem B.1.1 with $f_{n,t}$ defined as

$$\frac{n^{1/6}}{g(x_0)}\big(Y-f_0(x_0)\big)\left(\mathbb{1}\left\{X\le\frac{\lceil(x_0+n^{-1/3}t)J\rceil}{J}\right\}-\mathbb{1}\left\{X\le\frac{\lceil x_0J\rceil}{J}+\frac{1}{J}\right\}\right).$$

We shall assume $t\ge 0$ in the proof. Proof for the case $t\le 0$ follows similarly. For $t\ge 0$ we have

$$\begin{aligned}
&\frac{n^{2/3}}{g(x_0)}\sum_{j=\lceil x_0J\rceil+1}^{\lceil(x_0+n^{-1/3}t)J\rceil}\frac{N_j}{n}(\bar{Y}_j-f_0(x_0))\\
&=\frac{n^{2/3}}{g(x_0)}(\mathbb{P}_n-\mathrm{E}_0)(Y-f_0(x_0))\left(\mathbb{1}\left\{\frac{\lceil x_0J+1\rceil}{J}<X\le\frac{\lceil(x_0+n^{-1/3}t)J\rceil}{J}\right\}\right)\\
&\quad+\frac{n^{2/3}}{g(x_0)}\mathrm{E}_0\left(\mathbb{1}\left\{X\le\frac{\lceil(x_0+n^{-1/3}t)J\rceil}{J}\right\}-\mathbb{1}\left\{X\le\frac{\lceil x_0J+1\rceil}{J}\right\}\right)\\
&=\sqrt{n}(\mathbb{P}_n-\mathrm{E}_0)f_{n,t}+\sqrt{n}\,\mathrm{E}_0f_{n,t}.
\end{aligned}$$

Let $K>0$. Let F_n be defined as

$$\frac{n^{1/6}}{g(x_0)}\big|Y-f_0(x_0)\big|\left|\mathbb{1}\left\{X\le\frac{\lceil(x_0+n^{-1/3}K)J\rceil}{J}\right\}-\mathbb{1}\left\{X\le\frac{\lceil x_0J\rceil}{J}+\frac{1}{J}\right\}\right|.$$

Then $|f_{n,t}|\le F_n$ for every $t\in[0,K]$, and therefore F_n is an envelope for the class of functions $\mathscr{F}_n=\{f_{n,t}:t\in[0,K]\}$. We shall verify the assumptions F_n needs to satisfy for Theorem 2.11.22 to hold.

We first show that $\mathrm{E}_0^*F_n^2$ is bounded. We have

$$\begin{aligned}
\mathrm{E}_0^*F_n^2&=\frac{n^{1/3}}{g^2(x_0)}\int_{(\lceil x_0J\rceil+1)/J}^{\lceil(x_0+n^{-1/3}K)J\rceil/J}\mathrm{E}\big[(Y-f_0(x_0))^2\big|X=s\big]g(s)ds\\
&=\frac{n^{1/3}}{g^2(x_0)}\int_{(\lceil x_0J\rceil+1)/J}^{\lceil(x_0+n^{-1/3}K)J\rceil/J}\big[\sigma_0^2+(f_0(s)-f_0(x_0))^2\big]g(s)d\,s.
\end{aligned}$$

As f_0 is differentiable at x_0, it is continuous in a neighborhood of x_0 and so $\{\sigma_0^2+(f_0(s)-f_0(x_0))^2\}$ is bounded by a constant $B>0$ in a δ-neighborhood of x_0 for some $\delta>0$. For n large enough $|\lceil(x_0+n^{-1/3}K)J\rceil/J-x_0|<\delta$. Using Taylor's expansion of g around x_0 we write

$g(s) = g(x_0)(s-x_0) + O((s-x_0)^2)$, and for large n have

$$\begin{aligned}
\mathrm{E}_0^* F_n^2 &\le \frac{n^{1/3}}{g^2(x_0)}\, B \int_{(\lceil x_0 J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}K)J\rceil/J} g(s)ds \\
&\lesssim \frac{n^{1/3}}{g^2(x_0)}\Big[g(x_0)\big(\lceil (x_0+n^{-1/3}K)J\rceil/J - (\lceil x_0 J\rceil+1)/J\big) + O(n^{-2/3})\Big] \\
&= \frac{n^{1/3}}{g^2(x_0)}\Big[g(x_0)n^{-1/3}K + O(J^{-1}) + O(n^{-2/3})\Big] \\
&= \big(K/g(x_0)\big) + o(1).
\end{aligned}$$

Next we show that $\mathrm{E}_0^* F_n^2 \mathbb{1}\{F_n > \eta\sqrt{n}\}$ goes to zero for every $\eta > 0$. Using the fact that $\mathrm{E}_0^* F_n^2$ is bounded, and the Markov's inequality, we have that

$$\begin{aligned}
&\mathrm{E}_0^* F_n^2 \mathbb{1}\{F_n > \eta\sqrt{n}\} \\
&= \int_{(\lceil x_0 J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}K)J\rceil/J} \mathrm{E}_0\Big[n^{1/3}[f_0(s)-f_0(x_0)+\varepsilon]^2 \mathbb{1}\{|f_0(s)-f_0(x_0)+\varepsilon| > \eta n^{1/3}\}\Big] g(s)ds \\
&\le \int_{(\lceil x_0 J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}K)J\rceil/J} \mathrm{E}_0\Big[n^{1/3}[f_0(s)-f_0(x_0)+\varepsilon]^2 \mathbb{1}\{|f_0(s)-f_0(x_0)| > \eta n^{1/3}/2\}\Big] g(s)ds \\
&\quad + \int_{(\lceil x_0 J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}K)J\rceil/J} \mathrm{E}_0\Big[n^{1/3}[f_0(s)-f_0(x_0)+\varepsilon]^2 \mathbb{1}\{|\varepsilon| > \eta n^{1/3}/2\}\Big] g(s)ds \\
&\le \int_{(\lceil x_0 J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}K)J\rceil/J} \mathrm{E}_0\Big[n^{1/3}[f_0(s)-f_0(x_0)+\varepsilon]^2 \mathbb{1}\{|f_0(K)-f_0(x_0)| > \eta n^{1/3}/2\}\Big] g(s)ds \\
&\quad + \int_{(\lceil x_0 J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}K)J\rceil/J} \mathrm{E}_0\Big[n^{1/3}[f_0(s)-f_0(x_0)+\varepsilon]^2 \mathbb{1}\{|\varepsilon| > \eta n^{1/3}/2\}\Big] g(s)ds \\
&\le n^{1/3}\mathbb{1}\{|f_0(K)-f_0(x_0)| > \eta n^{1/3}/2\} \int_{(\lceil x_0 J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}K)J\rceil/J} \mathrm{E}_0\big[[f_0(s)-f_0(x_0)+\varepsilon]^2\big] g(s)ds \\
&\qquad + n^{1/3}(f_0(1)-f_0(0))^2 \int_{(\lceil x_0 J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}K)J\rceil/J} \mathrm{E}_0\big[\varepsilon^2 \mathbb{1}\{|\varepsilon| > \eta n^{1/3}/2\big] g(s)ds \\
&\lesssim o(1) + \mathrm{E}_0\big[\varepsilon^2 \mathbb{1}\{|\varepsilon| > \eta n^{1/3}/2\big] n^{1/3} \int_{(\lceil x_0 J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}K)J\rceil/J} g(s)ds,
\end{aligned}$$

which goes to zero by the assumption on the error ε and the fact that $|f_0(1)-f_0(0)|$ is bounded.

Next, let $\delta_n \downarrow 0$ and $0 \le t \le s$, such that $|s-t| < \delta_n$. Then again using the fact that $\{\sigma_0^2 + (f_0(s)-f_0(x_0))^2\}$ is bounded by a constant $B > 0$ in a neighborhood of x_0, we evaluate

$\mathrm{E}_0(f_{n,s}-f_{n,t})^2$ as

$$\begin{aligned}
&\frac{n^{1/3}}{g^2(x_0)}\int_{(\lceil (x_0+n^{-1/3}t)J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}s)J\rceil/J}\mathrm{E}_0\big[(Y-f_0(x_0))^2\big|X=u\big]g(u)du\\
&\quad=\frac{n^{1/3}}{g^2(x_0)}\int_{(\lceil (x_0+n^{-1/3}t)J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}s)J\rceil/J}\{\sigma_0^2+(f_0(u)-f_0(x_0))^2\}g(u)du\\
&\quad\lesssim n^{1/3}J^{-1}\big[\lceil (x_0+n^{-1/3}s)J\rceil-\lceil (x_0+n^{-1/3}t)J\rceil-1\big]\\
&\quad=n^{1/3}\big[n^{-1/3}(s-t)+O(J^{-1})\big]\\
&\quad\le\delta_n+o(1).
\end{aligned}$$

We shall now find the limit of the covariance. As the observations $\boldsymbol{Y}$ given $\boldsymbol{X}$ are independent, so for s and t of the opposite signs, $\mathrm{Cov}(f_{n,s},f_{,t})$ is clearly equal to zero. So let $s,t>0$. We then have

$$\begin{aligned}
&\mathrm{E}_0f_{n,s}f_{n,t}-\mathrm{E}_0f_{n,s}\mathrm{E}_0f_{n,t}\\
&=\big[\mathrm{E}_0(f_{n,s\wedge t}^2)-(\mathrm{E}_0f_{n,s\wedge t})^2\big]\\
&\qquad+\big[\mathrm{E}_0(f_{n,s\wedge t}(f_{n,s\vee t}-f_{n,s\wedge t}))-\mathrm{E}_0(f_{n,s\wedge t})\mathrm{E}_0((f_{n,s\vee t}-f_{n,s\wedge t}))\big]\\
&=\mathrm{E}_0(f_{n,s\wedge t}^2)-(\mathrm{E}_0f_{n,s\wedge t})^2,
\end{aligned}$$

as $f_{n,s\wedge t}$ and $f_{n,s\vee t}-f_{n,s\wedge t}$ are independent. Expanding $\mathrm{E}_0(f_{n,s\wedge t}^2)$ and using Taylor's expansion of f_0 and g around x_0, $\mathrm{E}_0(f_{n,s\wedge t}^2)$ equals to

$$\begin{aligned}
&\frac{n^{1/3}}{g^2(x_0)}\int_{(\lceil x_0J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}(s\wedge t))J\rceil/J}\mathrm{E}\big[(Y-f_0(x_0))^2\big|X=u\big]g(u)du\\
&\quad=\frac{n^{1/3}}{g^2(x_0)}\int_{(\lceil x_0J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}(s\wedge t))J\rceil/J}\{\sigma_0^2+(f_0(u)-f_0(x_0))^2\}g(u)du\\
&\quad=\frac{n^{1/3}}{g^2(x_0)}\int_{(\lceil x_0J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}(s\wedge t))J\rceil/J}\{\sigma_0^2+(f_0'(x_0))^2(u-x_0)^2\\
&\qquad\qquad+O(|u-x_0|^3)\}g(u)du\\
&\quad=\frac{n^{1/3}}{g^2(x_0)}\big[\sigma_0^2g(x_0)n^{-1/3}(s\wedge t)+O(n^{-2/3}+J^{-1})\big]\\
&\quad=\big(\sigma_0^2/g(x_0)\big)(s\wedge t)+o(1).
\end{aligned}$$

Also, the limit of $\mathrm{E}_0 f_{n,s\wedge t}$ can be evaluated in a similar way as

$$\begin{aligned}
\mathrm{E}_0 f_{n,s\wedge t} &= \frac{n^{1/3}}{g^2(x_0)} \int_{([x_0 J]+1)/J}^{\lceil (x_0+n^{-1/3}(s\wedge t))J\rceil/J} \mathrm{E}_0\big[Y - f_0(x_0)\big| X = u\big] g(u)du \\
&= \frac{n^{1/3}}{g^2(x_0)} \int_{([x_0 J]+1)/J}^{\lceil (x_0+n^{-1/3}(s\wedge t))J\rceil/J} \{f_0(u) - f_0(x_0)\} g(u)du \\
&\lesssim n^{1/3} \int_{([x_0 J]+1)/J}^{\lceil (x_0+n^{-1/3}(s\wedge t))J\rceil/J} (u - x_0) g(u)du \\
&= n^{1/3}\big[O(n^{-2/3}) + O(J^{-1})\big],
\end{aligned}$$

which goes to zero, as $n^{1/3} \ll J \ll n^{2/3}$. Therefore $\mathrm{Cov}(f_{n,s}, f_{n,t}) \to \sigma_0^2/g(x_0)(s\wedge t)$ as $n\to\infty$. The uniform entropy condition of the theorem holds from Lemma B.1.2. The assumptions of Theorem B.1.1 are therefore satisfied and we have that the process $t \mapsto \sqrt{n}(\mathbb{P}_n - \mathrm{E}_0) f_{n,t}$ converges to $\sqrt{\sigma_0^2/g(x_0)}W_1$ in $L^\infty([-K,K])$, where W_1 is a two-sided standard Brownian motion starting at zero. Next, using Taylor's expansion of f_0 and g around x_0, we have that $\sqrt{n}\mathrm{E}_0 f_{n,t}$ equals to

$$\begin{aligned}
&\frac{n^{2/3}}{g(x_0)} \int_{([x_0 J]+1)/J}^{\lceil (x_0+n^{-1/3}t)J\rceil/J} \{f_0(u) - f_0(x_0)\} g(u)du \\
&\quad= \frac{n^{2/3}}{g(x_0)} \int_{([x_0 J]+1)/J}^{\lceil (x_0+n^{-1/3}t)J\rceil/J} \{f_0'(x_0)(u-x_0) + O((u-x_0)^2)\} g(u)du \\
&\quad= \frac{n^{2/3}}{g(x_0)} \int_{([x_0 J]+1)/J}^{\lceil (x_0+n^{-1/3}t)J\rceil/J} \big\{f_0'(x_0) g(x_0)(u-x_0) + O((u-x_0)^2)\big\} du \\
&\quad= \frac{1}{2} f_0'(x_0) t^2 + o(1).
\end{aligned}$$

Therefore, $\{(n^{2/3}/g(x_0)) \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} (N_j/n)(\bar{Y}_j - f_0(x_0)) : t \in [-K,K]\}$ converges weakly to $\{a W_1(t) + b\, t^2 : t \in [-K,K]\}$ in $L^\infty([-K,K])$. □

Lemma A.2.3. *Let $n^{1/3} \ll J \ll n^{2/3}$. Then*

$$\operatorname*{arg\,min}_{t\in\mathbb{R}} \left\{ \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} \frac{N_j}{n} (\bar{Y}_j - f_0(x_0)) - n^{-1/3} z \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} \frac{N_j}{n} \right\} = O_{P_0}(1).$$

Proof of Lemma A.2.3. Using the change of variable $r = n^{-1/3}t$, we establish the tightness

of $n^{1/3}\hat{h}_n$, where

$$\hat{h}_n = \arg\min_{r\in\mathbb{R}} \left\{ \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+r)J\rceil} \frac{N_j}{n}(\bar{Y}_j - f_0(x_0)) - n^{1/3} z \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+r)J\rceil} \frac{N_j}{n} \right\} - x_0.$$

For $r \geq 0$, define $K_n(r)$ and $K(r)$ by

$$\begin{aligned} K_n(r) &= \mathbb{P}_n\big[(Y - f_0(x_0))(\mathbb{1}\{X \leq \lceil (x_0+r)J\rceil/J\} - \mathbb{1}\{X \leq \lceil x_0 J\rceil/J\})\big] \\ &\quad + z n^{-1/3}\mathbb{P}_n[\mathbb{1}\{X \leq \lceil (x_0+r)J\rceil/J\} - \mathbb{1}\{X \leq \lceil x_0 J\rceil/J\}]; \\ K(r) &= \mathrm{E}_0\big[(Y - f_0(x_0))(\mathbb{1}\{X \leq x_0 + r\} - \mathbb{1}\{X \leq x_0\})\big]. \end{aligned}$$

Then $\hat{h}_n$ is equal to $\arg\min\{K_n(r) : r \in \mathbb{R}\} - x_0$. We shall employ Theorem B.1.5 to show that $n^{1/3}(\arg\min\{K_n(r) : r \in \mathbb{R}\} - x_0)$ is $O_P(1)$, with $\mathbb{M}_n = -K_n$, $\mathbb{M} = -K$, $\Theta = [0,\infty)$, $\hat{\theta}_n = \arg\min_{r\in\mathbb{R}} K_n(r)$, $\theta_0 = 0$ and d the Euclidean distance on $\mathbb{R}$. With these notations, we have that $-K_n(0) = -K(0) = 0$ and $\hat{\theta}_n$ maximizes $-K_n$. We shall now verify the assumptions of the theorem.

The condition $\mathbb{M}(\theta) - \mathbb{M}(\theta_0) \lesssim -d^2(\theta,\theta_0)$ translates to $K(r) \gtrsim r^2$ in the present context. Using Taylor's expansion of f_0 and g around x_0, we have

$$\begin{aligned} &f_0(s) - f_0(x_0) = f_0'(x_0)(s - x_0) + o(|s - x_0|) \\ &g(s) = g(x_0) + [g(s) - g(x_0)], \qquad |g(s) - g(x_0)| \leq B_1, \quad B_1 > 0. \end{aligned}$$

We verify that $K(r) \gtrsim r^2$ by writing $K(r)$ as

$$\begin{aligned} \mathrm{E}_0\big[(Y - f_0(x_0))\mathbb{1}\{X \in (x_0, x_0 + r]\}\big] &= \int_{x_0}^{x_0+r} \mathrm{E}_0(Y - f_0(x_0)\big|X = s)g(s)ds \\ &= \int_{x_0}^{x_0+r} (f_0(s) - f_0(x_0))g(s)ds, \end{aligned}$$

which is bounded below by a constant multiple of r^2. Next we shall find functions ϕ_n such that $\phi_n(\delta)/\delta^\alpha$ is decreasing in δ for some $\alpha \in (0,2)$, and for sufficiently small δ,

$$\sqrt{n}\,\mathrm{E}_0^* \sup_{|r|<\delta} |K_n(r) - K(r)| \lesssim \phi_n(\delta). \tag{A.2.2}$$

The left-hand side of (A.2.2) equals to $\mathrm{E}_0^* \sup\{|\sum_{l=3}^5 H_{ln}(r)| : |r| < \delta\}$, with $H_{ln}(r)$ as defined in (2.6.19). Using the fact that for sufficiently small δ, f_0 is differentiable in a δ-neighborhood

of x_0, we get that $\mathrm{E}_0^* \sup\{|H_{3n}(r)| : |r| < \delta\}$ is bounded above by

$$\begin{aligned}
\mathrm{E}_0 \sqrt{n}\mathbb{P}_n |Y - f_0(x_0)| \mathbb{1}\left\{X \in \left(x_0, \frac{\lceil x_0 J \rceil}{J}\right] \bigcup \left(x_0 + \delta, \frac{\lceil (x_0+\delta) J \rceil}{J}\right]\right\} \\
&\leq \sqrt{n} \int_{x_0}^{(\lceil x_0 J \rceil)/J} |f_0(u) - f_0(x_0)| g(u) du \\
&\qquad + \sqrt{n} \int_{x_0+\delta}^{(\lceil (x_0+\delta) J \rceil)/J} |f_0(u) - f_0(x_0)| g(u) du \\
&\lesssim \sqrt{n} (J^{-1})^2 f_0'(x_0),
\end{aligned}$$

which is bounded above by a constant multiple of $n^{-1/6}$, as $J \gg n^{1/3}$. Next, we define M_δ and a class of functions $\mathscr{M}_\delta$ as

$$\begin{aligned}
M_\delta &= |Y - f_0(x_0)| \mathbb{1}\{X \in (x_0, x_0 + \delta]\} \\
\mathscr{M}_\delta &= \left\{(Y - f_0(x_0)) \mathbb{1}\{X \in (x_0, x_0 + r]\} : |r| < \delta\right\}.
\end{aligned}$$

Then M_δ is an envelope for $\mathscr{M}_\delta$, and from the discussion on Page 291 of van der Vaart and Wellner (1996), we have that

$$\mathrm{E}_0^* \sup\{|H_{4n}(r)| : |r| < \delta\} \lesssim J(1, \mathscr{M}_\delta)(\mathrm{E}_0^* M_\delta^2)^{1/2},$$

where $J(\eta, \mathscr{M}_\delta)$ for $\eta > 0$ is the uniform entropy integral

$$\sup_Q \int_0^\eta \sqrt{1 + \log \mathscr{N}(\epsilon \|M_\delta\|_{2,Q}, \mathscr{M}_\delta, \|\cdot\|_{2,Q})} d\epsilon.$$

Using arguments similar to those in Page 216, within the proof of Theorem 3.6.2 of Banerjee (2000), we get that $J(1, \mathscr{M}_\delta)$ is a finite constant. Also, using the fact that $\{\sigma_0^2 + (f_0(s) - f_0(x_0))^2\}$ is bounded by some $B > 0$ in a neighborhood of x_0 and arguments similar to those used for proving $\mathrm{E}_0^* F_n^2$ is bounded in the proof of Lemma A.2.2, we obtain that $\mathrm{E}_0^* M_\delta^2 \lesssim \delta$, for sufficiently small $\delta > 0$. Therefore $\mathrm{E}_0^* \sup\{|H_{4n}(r)| : |r| < \delta\} \lesssim \sqrt{\delta}$. Next we bound $\mathrm{E}_0^* \sup\{|H_{5n}(r)| : |r| < \delta\}$ as

$$\begin{aligned}
n^{1/6}|z| \mathrm{E}^* \sup_{|r|<\delta} \mathbb{P}_n \mathbb{1}\{X \in (x_0, x_0 + r]\} &\leq n^{1/6}|z| \mathrm{E}^* \mathbb{P}_n \mathbb{1}\{X \in (x_0, x_0 + \delta]\} \\
&\asymp n^{1/6}|z|(\delta + O(J^{-1})),
\end{aligned}$$

which is bounded by a constant multiple of $n^{1/6}|z|\delta$, as $J \gg n^{1/3}$. Therefore for some $C > 0$, (A.2.2) is satisfied for $\phi_n = C(n^{-1/6} + \delta^{1/2} + n^{1/6}|z|\delta)$. For $r_n = n^{1/3}$, we have that $r_n^2 \phi_n(r_n^{-1}) \asymp n^{1/2}|z| \lesssim \sqrt{n}$. The condition $\mathbb{M}_n(\hat{\theta}_n) \geq \mathbb{M}_n(\theta_0) - O_{P_0}(r_n^{-2})$ has been verified within the proof of Lemma A.2.2. The final condition that $\hat{\theta}_n \to^{P_0^*} \theta_0$ requires the uniform consistency of $\hat{f}_n(x)$ in a neighborhood of x_0, which is satisfied by the uniform consistency of the MLE near x_0, as established by Brunk (1970), and the fact that

$$\sup_{r \in [0,1]} \mathbb{P}_n Y[\mathbb{1}\{X \in (0, \lceil rJ \rceil / J]\} - \mathbb{1}\{X \in (0, r]\}] \to^{P_0} 0.$$

The assertion in the last display establishes that the difference of the slopes of the corresponding greatest convex minorants converges to zero in probability, uniformly in a neighborhood of x_0. Therefore the difference between the MLE and $\hat{f}_n$ converges to zero in probability uniformly in a neighborhood of x_0. We therefore obtain that $\hat{\hat{f}}_n(x) \to^{P_0} f_0(x)$ uniformly in a neighborhood of x_0, and the proof follows. □

For the model (3.0.1), let $h = \sum_{j=1}^J p_j \mathbb{1}_{I_j}$, $\int_0^1 h(u)du = 1$. Then let $\bar{h} = \sum_{j=1}^J \bar{p}_j \mathbb{1}_{I_j}$ be obtained from h by minimizing the sum of squares $h = J^{-1} \sum_{j=1}^J (p_j - \bar{p}_j)^2$ under the constraint $\bar{p}_1, \geq \cdots \geq \bar{p}_J$. The algorithm for the constrained minimization problem as described in Section 2.1 of Groeneboom and Jongbloed (2014) gives us that $\bar{p}_j$ is the slope (left derivative) of the least concave majorant of the graph (3.3.1) at the point j/J, for $1 \leq j \leq J$. Then for x_0 such that $g_0'(x_0) < 0$, the following theorem gives the asymptotic distribution of $\hat{g}_n(x_0)$.

Theorem A.2.4. *Let $n^{1/3} \ll J \ll n^{2/3}$. Let W_1 be a two-sided Brownian motion on $\mathbb{R}$ with $W_1(0) = 0$, $Z_1 = \arg\max\{W_1(t) - t^2 : t \in \mathbb{R}\}$, and $C = 2b\,(a/b)^{2/3}$ with $a = \sqrt{g_0(x_0)}$ and $b = |g_0'(x_0)|/2$. Then for every $z \in \mathbb{R}$,*

$$\mathrm{P}_0(n^{1/3}(\hat{g}_n(x_0) - g_0(x_0)) \leq z) \to \mathrm{P}(CZ_1 \leq z).$$

Proof. Let U_n, G_n be stochastic processes denoting $U_n(s) = \sum_{j=1}^{\lceil sJ \rceil} (N_j/n)$ and $G_n(s) = \sum_{j=1}^{\lceil sJ \rceil} J^{-1}$. From the switch relation (B.2), we evaluate $\mathrm{P}(n^{1/3}(\hat{g}_n(x_0) - g_0(x_0)) \leq x)$ for $z \in \mathbb{R}$ as

$$\begin{aligned}
&\mathrm{P}(n^{1/3}(\hat{g}_n(x_0) - g_0(x_0)) \leq z) \\
&= \mathrm{P}(\hat{g}_n(x_0) \leq g_0(x_0) + n^{1/3} z) \\
&= \mathrm{P}\left(\arg\max_{s \geq 0} \left\{ U_n(s) - (g_0(x_0) + n^{-1/3} z) G_n(s) \right\} \geq \sum_{k=1}^{\lceil x_0 J \rceil} J^{-1} \right) \\
&= \mathrm{P}\left(\arg\max_{s \in \mathbb{R}} \left\{ U_n(s) - (g_0(x_0) + n^{-1/3} z) G_n(s) \right\} \geq \tau_n \right)
\end{aligned}$$

$$= \mathrm{P}\Bigg(\operatorname*{arg\,max}_{s\in\mathbb{R}} \big\{ n^{2/3}(U_n(s)-U_n(x_0)) - n^{2/3}(g_0(x_0)+n^{-1/3}z)(G_n(s)-G_n(x_0)) \big\} \geq \tau_n \Bigg).$$

The last expression in the display then equals to

$$\begin{aligned}
&\mathrm{P}\Bigg(\operatorname*{arg\,max}_{t\in\mathbb{R}} \big\{ n^{2/3}(U_n(x_0+n^{-1/3}t)-U_n(x_0)) \\
&\qquad -n^{2/3}(g_0(x_0)+n^{-1/3}z)(G_n(x_0+n^{-1/3}t)-G_n(x_0)) \big\} \geq \tau_n \Bigg) \\
&= \mathrm{P}\Bigg(\operatorname*{arg\,max}_{t\in\mathbb{R}} \Bigg\{ n^{2/3} \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} \frac{N_j}{n} \\
&\qquad -n^{2/3}(g_0(x_0)+n^{-1/3}z) \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} \frac{1}{J} \Bigg\} \geq \tau_n \Bigg).
\end{aligned}$$

In the proof of part (b), it was shown that $\tau_n \to 0$ and $n^{1/3} z J^{-1}[\lceil (x_0+n^{-1/3}t)J\rceil - \lceil x_0 J\rceil - 1] \to z\,t$. We shall therefore evaluate

$$\begin{aligned}
&\mathrm{P}(\hat{g}_n(x_0) \leq g_0(x_0)+n^{-1/3}z) \\
&= \mathrm{P}\Bigg(\operatorname*{arg\,max}_{t\in\mathbb{R}} \Bigg\{ n^{2/3} \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} \left(\frac{N_j}{n} - \frac{g_0(x_0)}{J} \right) - z\,t \Bigg\} \geq 0 \Bigg).
\end{aligned}$$

From Lemma A.2.5 and A.2.6, Theorem B.1.6, and Lemma A.2.8, we have that

$$\begin{aligned}
\mathrm{P}(n^{1/3}(\hat{g}_n(x_0)-g_0(x_0)) \leq z) &\to \mathrm{P}\left(\operatorname*{arg\,max}_{t\in\mathbb{R}} \big\{ a\,W_1(t) - b\,t^2 - z\,t \big\} \geq 0 \right) \\
&= \mathrm{P}\left((a/b)^{2/3} \operatorname*{arg\,max}_{t\in\mathbb{R}} \big\{ W_1(t) - t^2 \big\} + z/2b \geq 0 \right) \\
&= \mathrm{P}\left(2b(a/b)^{2/3} \operatorname*{arg\,max}_{t\in\mathbb{R}} \big\{ W_1(t) - t^2 \big\} \geq -z \right) \\
&= \mathrm{P}\left(2b(a/b)^{2/3} \operatorname*{arg\,max}_{t\in\mathbb{R}} \big\{ W_1(t) - t^2 \big\} \leq z \right),
\end{aligned}$$

the last step following from the fact that $\operatorname{arg\,max}\big\{ W_1(t) - t^2 : t \in \mathbb{R} \big\}$ is symmetric about zero. Substituting $C = 2b(a/b)^{2/3}$ in the last expression of the display we get $\mathrm{P}(n^{1/3}(\hat{g}_n(x_0) - g_0(x_0)) \leq z) \to \mathrm{P}\big(C \operatorname{arg\,max}\big\{ W_1(t) - t^2 : t \in \mathbb{R} \big\} \leq z \big)$ as $n \to \infty$.

□

Lemma A.2.5. *Let W_1 be a two sided Brownian motion on $\mathbb{R}$ starting at zero. Let $n^{1/3} \ll J \ll$*

$n^{2/3}$. *Then for every* $K > 0$,

$$\left\{ n^{2/3} \sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil} \left(\frac{N_j}{n} - \frac{g_0(x_0)}{J} \right) : t \in [-K,K] \right\} \rightsquigarrow \left\{ a\,W_1(t) - b\,t^2 : t \in [-K,K] \right\},$$

in $L^\infty([-K,K])$.

Proof of Lemma A.2.5. We shall use Theorem B.1.1 with $f_{n,t}$ defined as

$$\begin{aligned}(n^{1/6}g_0(x_0))\big(\mathbb{1}\{X \le \lceil (x_0+n^{-1/3}t)J\rceil/J\} - \mathbb{1}\{X \le \lceil x_0 J\rceil/J + 1/J\}\big) \\ -(n^{1/6}g_0(x_0))J^{-1}\big[\lceil (x_0+n^{-1/3}t)J\rceil - \lceil x_0 J\rceil - 1\big].\end{aligned}$$

We shall assume $t \ge 0$ for the proof. The steps for $t \le 0$ follow in an exactly similar manner. For $t \ge 0$ we write $\sqrt{n}f_{n,t}$ as

$$\begin{aligned}n^{2/3}(\mathbb{P}_n - \mathrm{P})\big(\mathbb{1}\{X \le \lceil (x_0+n^{-1/3}t)J\rceil/J\} - \mathbb{1}\{X \le \lceil x_0 J\rceil/J + 1/J\}\big) \\ +n^{2/3}\mathrm{P}\big(\mathbb{1}\{X \le \lceil (x_0+n^{-1/3}t)J\rceil/J\} - \mathbb{1}\{X \le \lceil x_0 J\rceil/J + 1/J\}\big) \\ = \sqrt{n}(\mathbb{P}_n - \mathrm{P})f_{n,t} + \sqrt{n}\,\mathrm{P}\,f_{n,t}.\end{aligned}$$

For every $K > 0$, let F_n be defined as

$$\begin{aligned}(n^{1/6}g_0(x_0))\big|\mathbb{1}\{X \le \lceil (x_0+n^{-1/3}K)J\rceil/J\} - \mathbb{1}\{X \le \lceil x_0 J\rceil/J + 1/J\}\big| \\ -(n^{1/6}g_0(x_0))J^{-1}\big[\lceil (x_0+n^{-1/3}K)J\rceil - \lceil x_0 J\rceil - 1\big].\end{aligned}$$

Then $|f_{n,t}| \le F_n$ for every $t \in [0,K]$, and therefore F_n is an envelope for the class of functions $\mathscr{F}_n = \{f_{n,t} : t \in [0,K]\}$. We shall verify the assumptions F_n needs to satisfy for Theorem 2.11.22 to hold.

We first show that $\mathrm{P}^* F_n^2$ is bounded. We have

$$\mathrm{P}^* F_n^2 = n^{1/3} \int_{(\lceil x_0 J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}K)J\rceil/J} (g_0(s) - g_0(x_0))ds.$$

As g_0 is differentiable at x_0 with $g_0'(x_0) < 0$, $g_0(x)$ is bounded by a constant $B > 0$ in a δ-neighborhood of x_0 for some $\delta > 0$. For n large enough $|\lceil (x_0+n^{-1/3}K)J\rceil/J - x_0| < \delta$. Using Taylor's expansion of g around x_0 we write $g_0(s) - g_0(x_0) = g_0'(x_0)(s-x_0) + O((s-x_0)^2)$ and

for large n we have that

$$\begin{aligned}
\mathrm{P}^* F_n^2 &\leq n^{1/3}\, B \int_{(\lceil x_0 J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}K)J\rceil/J} (g_0(s)-g_0(x_0))ds \\
&\lesssim n^{1/3}\Big[g_0'(x_0)\Big(\lceil (x_0+n^{-1/3}K)J\rceil/J-(\lceil x_0 J\rceil+1)/J\Big)+O(n^{-2/3})\Big] \\
&= n^{1/3}\Big[g_0'(x_0)n^{-1/3}K+O(J^{-1})+O(n^{-2/3})\Big] \\
&= K g_0'(x_0)+o(1),
\end{aligned}$$

which is bounded. Next we show that $\mathrm{P}^* F_n^2 \mathbb{1}\{F_n > \eta\sqrt{n}\}$ goes to zero for every $\eta > 0$. For large n, $|\lceil (x_0+n^{-1/3}K)J\rceil/J - x_0| < \delta$. We write $\mathrm{P}^* F_n^2 \mathbb{1}\{F_n > \eta\sqrt{n}\}$ as

$$\begin{aligned}
&\int_{(\lceil x_0 J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}K)J\rceil/J} n^{1/3}\mathbb{1}\{|g_0(s)-g_0(x_0)| > \eta n^{1/3}\}ds \\
&\qquad\leq n^{1/3}\mathbb{1}\{|g_0(0)-g_0(1)| > \eta n^{1/3}\}\int_{(\lceil x_0 J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}K)J\rceil/J} ds \\
&\qquad\lesssim \mathbb{1}\{|g_0(0)-g_0(1)| > \eta n^{1/3}\}O(1),
\end{aligned}$$

which goes to zero as $\mathbb{1}\{|g_0(0)-g_0(1)| > \eta n^{1/3}\} \to 0$, since $g_0(0) < \infty$.

Next, let δ_n be a sequence of positive real numbers decreasing to zero. Let $0 \leq t \leq s$, such that $|s-t| < \delta_n$. Then as $n \to \infty$,

$$\begin{aligned}
\mathrm{P}(f_{n,s}-f_{n,t})^2 &= n^{1/3}\int_{(\lceil (x_0+n^{-1/3}t)J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}s)J\rceil/J} (g_0(u)-g_0(x_0))du \\
&\lesssim n^{1/3}J^{-1}\Big[\lceil (x_0+n^{-1/3}s)J\rceil-\lceil (x_0+n^{-1/3}t)J\rceil-1\Big] \\
&= n^{1/3}\Big[n^{-1/3}(s-t)+O(J^{-1})\Big] \\
&= |s-t|+o(1) \\
&\leq \delta_n + o(1),
\end{aligned}$$

which goes to zero. We shall now find the limit of the covariance. Let $s, t > 0$. We have

$$\begin{aligned}
\mathrm{P} f_{n,s} f_{n,t} - \mathrm{P} f_{n,s}\mathrm{P} f_{n,t} &= \Big[\mathrm{P}(f_{n,s\wedge t}^2)-(\mathrm{P} f_{n,s\wedge t})^2\Big] \\
&\quad+\Big[\mathrm{P}(f_{n,s\wedge t}(f_{n,s\vee t}-f_{n,s\wedge t}))-\mathrm{P}(f_{n,s\wedge t})\mathrm{P}((f_{n,s\vee t}-f_{n,s\wedge t}))\Big] \\
&= \mathrm{P}(f_{n,s\wedge t}^2)-(\mathrm{P} f_{n,s\wedge t})^2,
\end{aligned}$$

as $f_{n,s\wedge t}$ and $f_{n,s\vee t}-f_{n,s\wedge t}$ are independent. Expanding $\mathrm{P}(f^2_{n,s\wedge t})$ and using Taylor's expansion of g_0 and g around x_0, we get

$$\begin{aligned}
\mathrm{P}(f^2_{n,s\wedge t}) &= n^{1/3}\int_{(\lceil x_0 J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}(s\wedge t))J\rceil/J} (g_0(u)-g_0(x_0))du \\
&= n^{1/3}\int_{(\lceil x_0 J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}(s\wedge t))J\rceil/J} \{g_0'(x_0)(u-x_0)+O(|u-x_0|^2)\}du \\
&= n^{1/3}\left[g_0(x_0)n^{-1/3}(s\wedge t)+O(n^{-2/3}+J^{-1})\right] \\
&= g_0(x_0)(s\wedge t)+o(1).
\end{aligned}$$

Therefore, $\mathrm{Cov}(f_{n,s},f_{,t})\to(s\wedge t)g_0(x_0)$ as $n\to\infty$. The uniform entropy condition of the theorem holds by Lemma B.1.2. The assumptions of Theorem B.1.1 are therefore satisfied and we have that for every $K>0$,

$$\left\{\sqrt{n}(\mathbb{P}_n-\mathrm{P})f_{n,t} : t\in[-K,K]\right\}\rightsquigarrow\left\{\sqrt{g_0(x_0)}W_1(t) : t\in[-K,K]\right\},$$

where W_1 is a two sided standard Brownian motion starting at zero. It remains to show that $\sqrt{n}\mathrm{P}f_{n,t}\to b\,t^2$ in P_0-probability. Using Taylor's expansion of g_0 and g around x_0, we get

$$\begin{aligned}
\sqrt{n}\mathrm{P}f_{n,t} &= n^{2/3}\int_{(\lceil x_0 J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}t)J\rceil/J} \{g_0(u)-g_0(x_0)\}du \\
&= n^{2/3}\int_{(\lceil x_0 J\rceil+1)/J}^{\lceil (x_0+n^{-1/3}t)J\rceil/J} \{g_0'(x_0)(u-x_0)+O((u-x_0)^2)\}du \\
&= g_0'(x_0)t^2/2+o(1).
\end{aligned}$$

Therefore,

$$\left\{n^{2/3}\sum_{j=\lceil x_0 J\rceil+1}^{\lceil (x_0+n^{-1/3}t)J\rceil}(N_j/n-g_0(x_0)/J) : t\in[-K,K]\right\}\rightsquigarrow\left\{a\,W_1(t)-b\,t^2 : t\in[-K,K]\right\},$$

in $L^\infty([-K,K])$ for every $K>0$. □

Lemma A.2.6. *Let $n^{1/3}\ll J\ll n^{2/3}$. Then*

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

Proof of Lemma A.2.6. For $r \geq 0$, define $K_n(r)$ and $K(r)$ by

$$\begin{aligned} K_n(r) &= \mathbb{P}_n[(\mathbb{1}\{X \leq \lceil (x_0+r)J \rceil / J\} - \mathbb{1}\{X \leq \lceil x_0 J \rceil / J\})] \\ &\quad + z n^{-1/3} \mathbb{P}_n[\mathbb{1}\{X \leq \lceil (x_0+r)J \rceil / J\} - \mathbb{1}\{X \leq \lceil x_0 J \rceil / J\}]; \\ K(r) &= G_0(x_0+r) - G_0(x_0) - g_0'(x_0) r. \end{aligned}$$

Then $\hat{h}_n$ equals $\operatorname{argmin}\{K_n(r) : r \in \mathbb{R}\} - x_0$. We shall employ Theorem B.1.5 to show that $n^{1/3}(\operatorname{argmin}\{K_n(r) : r \in \mathbb{R}\} - x_0)$ is $O_P(1)$, with $\mathbb{M}_n = -K_n$, $\mathbb{M} = -K$, $\Theta = [0, \infty)$, $\hat{\theta}_n = \operatorname{argmin}_{r \in \mathbb{R}} K_n(r)$, $\theta_0 = 0$ and d the Euclidean distance on $\mathbb{R}$. With these notations, we have that $-K_n(0) = -K(0) = 0$ and $\hat{\theta}_n$ maximizes $-K_n$. We shall now verify the assumptions of the theorem.

The condition $\mathbb{M}(\theta) - \mathbb{M}(\theta_0) \lesssim -d^2(\theta, \theta_0)$ translates to $K(r) \gtrsim r^2$ in the present context. Using Taylor's expansion of g_0 around x_0, we have

$$g_0(s) - g_0(x_0) = g_0'(x_0)(s - x_0) + o(|s - x_0|).$$

We verify that $K(r) \gtrsim r^2$ by writing $K(r)$ as

$$\mathrm{E}_0[\mathbb{1}\{X \in (x_0, x_0 + r]\}] = \int_{x_0}^{x_0+r} g_0(s) ds,$$

which is bounded below by a constant multiple of r^2. Next we shall find functions ϕ_n such that $\phi_n(\delta)/\delta^\alpha$ is decreasing in δ for some $\alpha \in (0,2)$, and for sufficiently small δ,

$$\sqrt{n} \mathrm{E}_0^* \sup_{|r| < \delta} |K_n(r) - K(r)| \lesssim \phi_n(\delta). \tag{A.2.3}$$

The left-hand side of (A.2.3) equals to $\mathrm{E}_0^* \sup\{|\sum_{l=3}^5 H_{ln}(r)| : |r| < \delta\}$, with $H_{ln}(r)$ as defined in (3.5.15). Using the fact that for sufficiently small δ, g_0 is differentiable in a δ-neighborhood of x_0, we get that $\mathrm{E}_0^* \sup\{|H_{3n}(r)| : |r| < \delta\}$ is bounded above by

$$\begin{aligned} \mathrm{E}_0 \sqrt{n} \mathbb{P}_n \mathbb{1}\left\{ X \in \left(x_0, \frac{\lceil x_0 J \rceil}{J} \right] \bigcup \left(x_0 + \delta, \frac{\lceil (x_0+\delta) J \rceil}{J} \right] \right\} \\ \leq \sqrt{n} \int_{x_0}^{(\lceil x_0 J \rceil)/J} |g_0(u) - g_0(x_0)| du \\ + \sqrt{n} \int_{x_0+r}^{(\lceil (x_0+r) J \rceil)/J} |g_0(u) - g_0(x_0)| du \end{aligned}$$

$$\lesssim \sqrt{n}(J^{-1})^2 g_0'(x_0),$$

which is bounded above by a constant multiple of $n^{-1/6}$, as $J \gg n^{1/3}$. Next, we define M_δ and a class of functions $\mathscr{M}_\delta$ as

$$M_\delta = \mathbb{1}\{X \in (x_0, x_0+\delta]\}$$
$$\mathscr{M}_\delta = \{\mathbb{1}\{X \in (x_0, x_0+r]\} : |r| < \delta\}.$$

Then M_δ is an envelope for $\mathscr{M}_\delta$, and from the discussion on Page 291 of van der Vaart and Wellner (1996), we have that

$$\mathrm{E}_0^* \sup\{|H_{4n}(r)| : |r| < \delta\} \lesssim J(1, \mathscr{M}_\delta)(\mathrm{E}_0^* M_\delta^2)^{1/2},$$

where $J(\eta, \mathscr{M}_\delta)$ for $\eta > 0$ is the uniform entropy integral

$$\sup_Q \int_0^\eta \sqrt{1 + \log \mathscr{N}(\epsilon \|M_\delta\|_{2,Q}, \mathscr{M}_\delta, \|\cdot\|_{2,Q})} d\epsilon$$

Using arguments similar to those in Page 216, within the proof of Theorem 3.6.2 of Banerjee (2000), we get that $J(1, \mathscr{M}_\delta)$ is a finite constant. Also, using the fact that $(g_0(s) - g_0(x_0))^2$ is bounded by some $B > 0$ in a neighborhood of x_0 and arguments similar to those used for proving $\mathrm{E}_0^* F_n^2$ is bounded in the proof of Lemma A.2.5, we obtain that $\mathrm{E}_0^* M_\delta^2 \lesssim \delta$, for sufficiently small $\delta > 0$. Therefore $\mathrm{E}_0^* \sup\{|H_{4n}(r)| : |r| < \delta\} \lesssim \sqrt{\delta}$. Next we bound $\mathrm{E}_0^* \sup\{|H_{5n}(r)| : |r| < \delta\}$ as

$$\frac{z n^{1/6}}{J} [\lceil (x_0 + r)J \rceil - \lceil x_0 J \rceil] \lesssim n^{1/6}|z|(\delta + O(J^{-1})),$$

which is bounded by a constant multiple of $n^{1/6}|z|\delta$, as $J \gg n^{1/3}$. Therefore for some $C > 0$, (A.2.3) is satisfied for $\phi_n = C(n^{-1/6} + \delta^{1/2} + n^{1/6}|z|\delta)$. For $r_n = n^{1/3}$, we have that $r_n^2 \phi_n(r_n^{-1}) \asymp n^{1/2}|z| \lesssim \sqrt{n}$. The condition $\mathbb{M}_n(\hat{\theta}_n) \geq \mathbb{M}_n(\theta_0) - O_{P_0}(r_n^{-2})$ has been verified within the proof of Lemma A.2.5. □

Lemma A.2.7. *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) $\operatorname{arg\,min}\{a W_1(h) + a W_2(h) + b h^2 + c h : h \in \mathbb{R}\}$
$\stackrel{d}{=} (a/b)^{2/3} \operatorname{arg\,min}\{W_1(g) + W_2(g) + g^2 : g \in \mathbb{R}\} - c/(2b);$

(b) $\operatorname{arg\,min}\left\{a W_1(h) + b h^2 + c h : h \in \mathbb{R}\right\}$
$$\stackrel{d}{=} (a/b)^{2/3} \operatorname{arg\,min}\left\{W_1(g) + g^2 : g \in \mathbb{R}\right\} - c/(2b).$$

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

$$\begin{aligned}
&\Big(\operatorname{arg\,min}\left\{a W_1(h_1) + b h_1^2 + c_1 h_1 : h_1 \in \mathbb{R}\right\}, \\
&\qquad\qquad \operatorname{arg\,min}\left\{a W_1(h_2) + a W_2(h_2) + b h_2^2 + c_2 h_2 : h_2 \in \mathbb{R}\right\}\Big) \\
&\stackrel{d}{=} \Big((a/b)^{2/3} \operatorname{arg\,min}\left\{W_1(h_1) + h_1^2 : h_1 \in \mathbb{R}\right\} - c_1/(2b), \\
&\qquad (a/b)^{2/3} \operatorname{arg\,min}\left\{W_1(h_2) + W_2(h_2) + h_2^2 : h_2 \in \mathbb{R}\right\} - c_2/(2b)\Big).
\end{aligned}$$

Proof. (a) We use the transformation $g \mapsto \psi(g) = (a/b)^{2/3} g - c/(2b)$. Then the range of g is $\mathbb{R}$. For $\hat{g} = \operatorname{arg\,min}\{a W_1(\psi(g)) + b(\psi(g))^2 + c\,\psi(g) : g \in \mathbb{R}\}$, let $\hat{h} = (a/b)^{2/3}\hat{g} + c/(2b)$. We then have,

$$\begin{aligned}
&\operatorname*{arg\,min}_{g\in\mathbb{R}}\{a W_1(\psi(g)) + a W_2(\psi(g)) + b(\psi(g))^2 + c\,\psi(g)\} \\
&= \operatorname*{arg\,min}_{g\in\mathbb{R}}\left\{a W_1(\psi(g)) + a W_2(\psi(g)) + b\left(\frac{a}{b}\right)^{4/3} g^2 - c\left(\frac{a}{b}\right)^{2/3} g \right. \\
&\qquad\qquad \left. + \frac{c^2}{4b^2} + c\left(\frac{a}{b}\right)^{2/3} g - \frac{c^2}{2b}\right\} \\
&= \operatorname*{arg\,min}_{g\in\mathbb{R}}\left\{a W_1\left(\left(\frac{a}{b}\right)^{2/3} g - \frac{c}{2b}\right) + a W_2\left(\left(\frac{a}{b}\right)^{2/3} g - \frac{c}{2b}\right) + b\left(\frac{a}{b}\right)^{4/3} g^2\right\}.
\end{aligned}$$

Using the facts that $W_1(\sigma h - \mu) \stackrel{d}{=} W_1(\sigma h) - W_1(\mu)$, and that $W_1(\sigma h) \stackrel{d}{=} \sqrt{\sigma} W_1(h)$, $\hat{g}$ can be evaluated as

$$\begin{aligned}
\hat{g} &= \operatorname*{arg\,min}_{g\in\mathbb{R}}\left\{a W_1\left(\left(\frac{a}{b}\right)^{2/3} g - \frac{c}{2b}\right) \right. \\
&\qquad\qquad \left. + a W_2\left(\left(\frac{a}{b}\right)^{2/3} g - \frac{c}{2b}\right) + b\left(\frac{a}{b}\right)^{4/3} g^2\right\} \\
&\stackrel{d}{=} \operatorname*{arg\,min}_{g\in\mathbb{R}}\left\{a W_1\left(\left(\frac{a}{b}\right)^{2/3} g\right) - a W_1\left(\frac{c}{2b}\right) \right. \\
&\qquad\qquad \left. + a W_2\left(\left(\frac{a}{b}\right)^{2/3} g\right) - a W_2\left(\frac{c}{2b}\right) + b\left(\frac{a}{b}\right)^{4/3} g^2\right\} \\
&= \operatorname*{arg\,min}_{g\in\mathbb{R}}\left\{a W_1\left(\left(\frac{a}{b}\right)^{2/3} g\right) + a W_2\left(\left(\frac{a}{b}\right)^{2/3} g\right) + b\left(\frac{a}{b}\right)^{4/3} g^2\right\} \\
&\stackrel{d}{=} \operatorname*{arg\,min}_{g\in\mathbb{R}}\left\{a\left(\frac{a}{b}\right)^{1/3}[W_1(g) + W_2(g)] + \frac{a^{4/3}}{b^{1/3}} g^2\right\}
\end{aligned}$$

$$= \arg\min_{g\in\mathbb{R}} \left(\frac{a^{4/3}}{b^{1/3}}\right)\{W_1(g)+W_2(g)+g^2\}$$
$$= \arg\min_{g\in\mathbb{R}}\{W_1(g)+W_2(g)+g^2\}.$$

Therefore, $\hat{g} \stackrel{d}{=} \arg\min\{W_1(g)+W_2(g)+g^2 : g\in\mathbb{R}\}$ and

$$\hat{h} \stackrel{d}{=} \left(\frac{a}{b}\right)^{2/3} \arg\min_{h\in\mathbb{R}}\{W_1(h)+W_2(h)+h^2\} + \frac{c}{2b}.$$

(b) The proof follows using exactly the same arguments as in part (a).

(c) By the independence of W_1 and W_2, we have that

$$\begin{aligned}&\big(W_1(\sigma h_1-\mu_1)+W_2(\sigma h_1-\mu_1), W_1(\sigma h_2-\mu_2)\big)\\ &\stackrel{d}{=}\big(\sqrt{\sigma}(W_1(h_1)+W_2(h_1))-W_1(\mu_1)-W_2(\mu_2), \sqrt{\sigma}W_1(h_2)-W_1(\mu_2)\big).\end{aligned}$$

The proof follows by using the transformations $h_i \mapsto (a/b)^{2/3}h_i - c_i/(2b)$ for $i=1,2$, and the arguments similar to those used in part (a). □

Lemma A.2.8. *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) $\arg\max\{aW_1(h)+aW_2(h)-bh^2+ch : h\in\mathbb{R}\}$
$\stackrel{d}{=}(a/b)^{2/3}\arg\max\{W_1(g)+W_2(g)-g^2 : g\in\mathbb{R}\}-c/(2b)$;

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

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

$$\begin{aligned}&\big(\arg\max\{aW_1(h_1)-bh_1^2+c_1h_1 : h_1\in\mathbb{R}\},\\ &\qquad\qquad \arg\max\{aW_1(h_2)+aW_2(h_2)-bh_2^2+c_2h_2 : h_2\in\mathbb{R}\}\big)\\ &\stackrel{d}{=}\big((a/b)^{2/3}\arg\max\{W_1(h_1)-h_1^2 : h_1\in\mathbb{R}\}-c_1/(2b),\\ &\qquad (a/b)^{2/3}\arg\max\{W_1(h_2)+W_2(h_2)-h_2^2 : h_2\in\mathbb{R}\}-c_2/(2b)\big).\end{aligned}$$

The proof follows using arguments similar to the proof of Lemma A.2.7.

APPENDIX

B

MISCELLANEOUS RESULTS

This appendix states some results on empirical processes, convex minorants, and general theory of posterior contraction. The results have been used in various derivations throughout the dissertation.

B.1 Results on empirical processes

Let $X_1, X_2, \ldots$ be a sequence of independent random elements with the common law P on a measurable space $(\chi, \mathscr{A})$ and let $x \mapsto f_{n,t}(x)$ be functions from χ to $\mathbb{R}$ indexed by $n \in \mathbb{N}$ and a fixed, totally bounded semimetric space (T, ρ). Let $\mathscr{F}_n$ be classes of functions changing with n, defined as

$$\mathscr{F}_n = \{f_{n,t} : t \in T\}.$$

Let F_n be an envelope function for $\mathscr{F}_n$. A natural choice of F_n is $\sup\{\{|f_{n,t}| : t \in T\}$. Assume that the following hold true:

$$\begin{aligned} \mathrm{P}^* F_n^2 &= O(1), \\ \mathrm{P}^* F_n^2 \mathbb{1}\{F_n > \eta\sqrt{n}\} &\to 0, \qquad \text{for every } \eta > 0, \end{aligned} \tag{B.1.1}$$

$$\sup_{\rho(s,t)<\delta_n} \mathrm{P}(f_{n,s}-f_{n,t})^2 \to 0, \qquad \text{for every } \delta_n \downarrow 0.$$

The following theorem establishes a central limit theorem for the processes $\{\mathbb{G}_n f_{n,t} : t \in T\}$ under an entropy condition and the above mentioned assumptions.

Theorem B.1.1. *(Theorem 2.11.22, van der Vaart and Wellner (1996)). For each n, let the classes $\mathscr{F}_{n,\delta} = \{f_{n,s} - f_{n,t} : \rho(s,t) < \delta\}$ and $\mathscr{F}_{n,\delta}^2$ be P-measurable for every $\delta > 0$. Suppose that the conditions of* (B.1.1) *hold, and*

$$\sup_{\mathrm{Q}} \int_0^{\delta_n} \sqrt{\log N\left(\epsilon \|F_n\|_{Q,2}, \mathscr{F}_n, L_2(Q)\right)} d\epsilon \to 0, \qquad \textit{for every } \delta_n \downarrow 0.$$

Then the sequence $\{\mathbb{G}_n f_{n,t} : t \in T\}$ is asymptotically tight in $L^\infty(T)$ and converges in distribution to a Gaussian process, provided the sequence of covariance functions $\mathrm{P} f_{n,s} f_{n,t} - \mathrm{P} f_{n,s} \mathrm{P} f_{n,t}$ converges point-wise on $T \times T$.

Lemma B.1.2. *(Example 2.11.24, van der Vaart and Wellner (1996)). The uniform-entropy condition of Theorem B.1.1 is satisfied if, for each n, the set of functions $\mathscr{F}_n = \{f_{n,t} : t \in T\}$ is a VC- class with VC-index bounded by some constant independent of n.*

The following lemma provides a maximal inequality for sub-Gaussian processes in terms of an entropy integral.

Lemma B.1.3. *(Corollary 2.2.8, van der Vaart and Wellner (1996)). Let $\{X_t : t \in T\}$ be a separable sub-Gaussian process. Then for every $\delta > 0$,*

$$\mathrm{E} \sup_{d(s,t)\le\delta} |X_s - X_t| \le K \int_0^{\delta} \sqrt{\log D(\epsilon, d)} d\epsilon,$$

for a universal constant K. In particular, if $\{X_t : t \in T\}$ is a separable, zero-mean Gaussian process, then it is sub-Gaussian with respect to its standard deviation semimetric $\rho(s,t) = \sigma(X_s - X_t)$, and therefore for every $\delta > 0$,

$$\mathrm{E} \sup_{\rho(s,t)<\delta} |X_s - X_t| \lesssim \int_0^{\delta} \sqrt{\log N(\epsilon, T, \rho)} d\epsilon,$$

where $N(\epsilon, T, \rho)$ is the covering number of T with respect to ρ.

Lemma B.1.4. *(Dudley's result, page 252 of Billingsley (1999)). There is a constant $A>0$ such that for $0<\epsilon<1$, there exists an ϵ-net in $[0,1]$ with $N(\epsilon)$ many elements satisfying*

$$\log N(\epsilon) \leq A\sqrt{\frac{1}{\epsilon}}\log\frac{1}{\epsilon}.$$

Let $\{\mathbb{M}_n(h): h\in\Theta\}$ be a sequence of stochastic processes indexed by a semimetric space Θ equipped with the semimetric d. Consider the estimators $\hat{\theta}_n$ that nearly maximize maps $\theta\mapsto\mathbb{M}_n(\theta)$. Also, let $\mathbb{M}(h): h\in\Theta$ be a deterministic function and let θ_0 be a point of the maximum for $\theta\mapsto\mathbb{M}(\theta)$. The following theorem establishes that the continuity modulus of the difference $\mathbb{M}_n-\mathbb{M}$ can provide an upper bound for the rate of convergence of $\hat{\theta}_n$.

Theorem B.1.5. *(Theorem 3.2.5, van der Vaart and Wellner (1996)). Let for every θ in a neighborhood of θ_0,*

$$\mathbb{M}(\theta)-\mathbb{M}(\theta_0)\lesssim -d^2(\theta,\theta_0). \tag{B.1.2}$$

Let for sufficiently small δ and every n,

$$\mathrm{E}^*\sup_{d(\theta,\theta_0)<\delta}\left|(\mathbb{M}_n-\mathbb{M})(\theta)-(\mathbb{M}_n-\mathbb{M})(\theta_0)\right|\lesssim\frac{\phi_n(\delta)}{\sqrt{n}}, \tag{B.1.3}$$

for some ϕ_n such that $\delta\mapsto\phi_n(\delta)/\delta^\alpha$ is decreasing for some $\alpha<2$, with α not depending on n. Let r_n be a real sequence such that for every n,

$$r_n^2\phi_n(r_n^{-1})\leq\sqrt{n}. \tag{B.1.4}$$

Let $\hat{\theta}_n$ converge in outer probability to θ_0 and let $\mathbb{M}_n(\hat{\theta}_n)\geq\mathbb{M}_n(\theta_0)-O_P(r_n^{-2})$. Then,

$$r_n\, d(\hat{\theta}_n,\theta_0)=O_{P^*}(1).$$

Also, if (B.1.2)–(B.1.4) *are satisfied for every θ and δ, then the consistency of $\hat{\theta}_n$ is unnecessary.*

Theorem B.1.6. *(Argmax continuous mapping, Theorem 3.2.2 van der Vaart and Wellner (1996)). Let $\mathbb{M}_n,\mathbb{M}$ be stochastic processes indexed by a metric space H such that $\mathbb{M}_n\rightsquigarrow\mathbb{M}$ in $L^\infty(K)$ for all compact $K\subset H$. Suppose that almost all sample paths of $\mathbb{M}$ are upper semicontinuous and possess a unique maximium at $\hat{h}$, which as a random map in $\hat{h}$ is tight.*

If the sequence $\hat{h}_n$ is uniformly tight and satisfies

$$\mathbb{M}_n(\hat{h}_n) \geq \sup_h \mathbb{M}_n(h) - o_P(1),$$

then $\hat{h}_n \rightsquigarrow \hat{h}$ in H.

Let $\mathbb{R}_+$, $\mathbb{R}_-$ and $\mathbb{R}_0$ denote the positive real line, the negative real line, and the whole real line respectively. Let $C_{\max}(\mathbb{R})$ be the collection of all continuous functions f from $\mathbb{R}$ to $\mathbb{R}$ for which $f(t) \to -\infty$ as $|t| \to \infty$, and f achieves its maximum at a unique point in $\mathbb{R}$. Let $C_{\min}(\mathbb{R})$ be defined similarly, with $-\infty$ and maximum replaced by ∞ and minimum respectively. Also, $C_{\max}(\mathbb{R}_+)$ is defined similarly to $C_{\max}(\mathbb{R})$, with the domains of the functions restricted to $\mathbb{R}_+$.

The following theorem is an extension of the Argamx theorem to the multivariate case.

Theorem B.1.7. *(Multivariate version of the Argamx theorem, Lemma 3.6.10, Banerjee (2000)). Let $(Y_{n1}, Y_{n2}, \ldots, Y_{nk})$ be random maps into $B_{\text{loc}}(\mathbb{R})^k$ equipped with the product topology, where $B_{\text{loc}}(\mathbb{R})$ is equipped with the topology of uniform convergence on compact sets. Let $(U_{n1}, U_{n2}, \ldots, U_{nk})$ be random maps into $\mathbb{R}^k$ and let $(Y_{01}, Y_{02}, \ldots, Y_{0k})$ be a random map into $B_{loc}(\mathbb{R})^k$ such that*

$$\mathrm{P}((Y_{01}, Y_{02}, \ldots, Y_{0k}) \in C_{\max}(\mathbb{R})^k)) = 1.$$

Let ϵ_j be either $+, -$ or 0 for every $1 \leq j \leq k$. Suppose that

(i) $(Y_{n1}, \ldots, Y_{nk}) \rightsquigarrow (Y_{01}, \ldots, Y_{0k})$;

(ii) $U_{n1}, \ldots, U_{nk}$ *are* $O_P(1)$ *and* U_{nj} *belongs to* $\mathbb{R}_{\epsilon_j}$, *for all* $1 \leq j \leq k$;

(iii) $Y_{nj}(U_{nj}) \geq \{\sup_{t \in \mathbb{R}_{\epsilon_j}} Y_{nj}(t)\} - \alpha_{nj}$, *where* $\alpha_{nj} = o_P(1)$, $1 \leq j \leq k$.

Assuming that for each i, the process Y_{0i} restricted to $\mathbb{R}_{\epsilon_i}$ is almost surely in $C_{\max}(\mathbb{R}_{\epsilon_i})$, let $U_i = \arg\max_{t \in \mathbb{R}_{\epsilon_i}} Y_{0i}(t)$. We then have that

$$(U_{n1}, \ldots, U_{nk}) \rightsquigarrow (U_1, \ldots U_k).$$

A corresponding result for the minimizers is easily deduced from Theorem B.1.7 by replacing max by min everywhere in the theorem, and by changing condition (iii) to

$$Y_{nj}(U_{nj}) \leq \{\inf_{t \in \mathbb{R}_{\epsilon_j}} Y_{nj}(t)\} + \alpha_{nj},$$

where $\alpha_{nj} = o_P(1)$, $1 \leq j \leq k$.

B.2 Switch relations

For deriving most of the results in sections 2.3, 3.3 and 4.2, we have used what are known as “switch relations" pertaining to lower or upper semi-continuous functions. For a lower semi-continuous function Φ on an interval I with Φ^* its greatest convex minorant, and Φ^{*l} denoting the left derivative of Φ^*, we have for every $t \in I$, $v \in \mathbb{R}$,

$$\left\{\Phi^{*l}(t) > v\right\} = \left\{\arg\min_{s\in I}\{\Phi(s) - vs\} < t\right\}, \tag{B.2.1}$$

where ‘arg min’ selects the maximum of the minimizers when multiple minimizers exist. Similarly, 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^*, we have for every $t \in I$, $v \in \mathbb{R}$,

$$\left\{\Phi^{*l}(t) \le v\right\} = \left\{\arg\max_{s\in I}\{\Phi(s) - vs\} \le t\right\}, \tag{B.2.2}$$

where ‘arg max’ selects the maximum of the maximizers when multiple maximizers exist. See Lemma 3.2 of Groeneboom and Jongbloed (2014) and its discussions on pages 56 and 64, for details regarding the switch relations.

B.3 General theory of posterior contraction

In this section we state the result on general theory of posterior contraction, that we have used to prove Theorem 2.2.3.

Let us denote our experiments as $(\mathbb{X}^{(n)}, \mathscr{X}^{(n)}, P_\theta^{(n)} : \theta \in \Theta_n)$, where Θ_n are parameter spaces, $X^{(n)}$ are the observations, and $\theta_{n,0} \in \Theta_n$ are the true parameters. For each n, let d_n and e_n be two semimetrics on Θ_n with the property: there exist universal constants $\xi, K > 0$ such that for every $\epsilon > 0$ and every $\theta_{n,1} \in \Theta_n$, with $d_n(\theta_{n,1}, \theta_{n,0}) > \epsilon$, there exists a test ϕ_n such that

$$\mathrm{P}_{\theta_{n,0}}^{(n)} \phi_n \le e^{-Kn\epsilon^2}, \qquad \sup_{\theta\in\Theta_n : e_n(\theta,\theta_{n1}) < \xi\epsilon} \mathrm{P}_\theta^{(n)}(1-\phi_n) \le e^{-Kn\epsilon^2}. \tag{B.3.1}$$

For K denoting the Kullback-Leibler divergence, define

$$B_{n,0}(\theta_{n,0}, \epsilon) = \left\{\theta \in \Theta_n : K(p_{\theta_{n,0}}^{(n)}; p_\theta^{(n)}) \le n\epsilon^2\right\}. \tag{B.3.2}$$

Theorem B.3.1. *(Theorem 8.19, Ghosal and van der Vaart (2017)). Let d_n and e_n be semimetrics on Θ_n for which there exist tests satisfying* (B.3.1)*, and let $\Theta_{n,1} \subset \Theta_n$, $\Theta_{n,2} = \Theta_n \backslash \Theta_{n,1}$. If for positive constants ϵ_n, with $n\epsilon_n^2 \geq 1$, for every sufficiently large $j \in \mathbb{N}$,*

(i) $\Pi_n(B_{n,0}(\theta_{n,0}, \epsilon_n)) \geq e^{-Cn\epsilon_n^2}$,

(ii) $\log N(\xi\epsilon_n, \Theta_{n,1}, e_n) \leq n\epsilon_n^2$,

(iii) $\Pi_n(\Theta_{n,2}) = o(e^{-D_n n\epsilon_n^2})$*, for some $D_n \to \infty$,*

then $\Pi_n(\theta \in \Theta : d_n(\theta, \theta_{n,0} \geq M_n\epsilon_n | X^{(n)}) \to 0$, in $P_{\theta_{n,0}}^{(n)}$-probability, for every $M_n \to \infty$.

B.4 Miscellaneous results

Theorem B.4.1. *(Binomial Inverse Theorem). For square matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ with $\boldsymbol{A}$ nonsingular,*

$$(\boldsymbol{A}+\boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\left(\boldsymbol{I} + \boldsymbol{B}\boldsymbol{A}^{-1}\right)^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}.$$

Theorem B.4.2. *(Lemma 3.1, Sen et al. (2010)). Let W_n and W_n^* be random vectors in $\mathbb{R}^l$ and $\mathbb{R}^k$ respectively; let Q and Q^* denote distributions on the Borel sets of $\mathbb{R}^l$ and $\mathbb{R}^k$, and let $\mathscr{F}_n$ be sigma-fields for which W_n is $\mathscr{F}_n$-measurable. If the distribution of W_n converges to Q and the conditional distribution of W_n^* given $\mathscr{F}_n$ converges in probability to Q^*, then the joint distribution of (W_n, W_n^*) converges to the product measure $Q \times Q^*$.*

Theorem B.4.3. *(Lemma 8.10, Yoo and Ghosal (2016)). Let $\boldsymbol{A}$ be an $n \times n$ symmetric positive definite matrix. Assume that $\|\boldsymbol{A}\|_{(2,2)} \leq C$ for some constant $C > 0$. Let $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$ be such that ε_i are i.i.d. mean* 0*, variance σ_0^2 with finite fourth moment for $i = 1, \ldots, n$. Then* $\mathrm{Var}(\boldsymbol{\varepsilon}^T\boldsymbol{A}\boldsymbol{\varepsilon}) = O(n)$.