The Annals of Statistics
2020, Vol. 48, No. 3, 1432–1451
https://doi.org/10.1214/19-AOS1853

NONPARAMETRIC BAYESIAN ANALYSIS OF THE COMPOUND POISSON PRIOR FOR SUPPORT BOUNDARY RECOVERY

BY MARKUS REISS[1] AND JOHANNES SCHMIDT-HIEBER[2]

[1] *Institut für Mathematik, Humboldt-Universität zu Berlin,* mreiss@math.hu-berlin.de
[2] *Department of Applied Mathematics, University of Twente,* a.j.schmidt-hieber@utwente.nl

Given data from a Poisson point process with intensity $(x, y) \mapsto n\mathbf{1}(f(x) \leq y)$, frequentist properties for the Bayesian reconstruction of the support boundary function f are derived. We mainly study compound Poisson process priors with fixed intensity proving that the posterior contracts with nearly optimal rate for monotone support boundaries and adapts to Hölder smooth boundaries. We then derive a limiting shape result for a compound Poisson process prior and a function space with increasing parameter dimension. It is shown that the marginal posterior of the mean functional performs an automatic bias correction and contracts with a faster rate than the MLE. In this case, $(1 - \alpha)$-credible sets are also asymptotic $(1 - \alpha)$-confidence intervals. As a negative result, it is shown that the frequentist coverage of credible sets is lost for linear functions f outside the function class.

1. Introduction. The estimation of support boundary functions does not only have numerous applications, but also poses intriguing mathematical questions; see Gijbels et al. [17], Chernozhoukov and Hong [6] as well as Korostelev and Tsybakov [22] for some overview. Here, we consider the fundamental observation model of a Poisson point process (PPP) N on $[0, T] \times \mathbb{R}$, $T > 0$, with intensity

$$\lambda(x, y) = \lambda_f(x, y) = n\mathbf{1}\big(f(x) \leq y\big). \tag{1}$$

We thus observe points $(X_i, Y_i)_{i \geq 1}$ on the epigraph of the boundary function $f : [0, T] \to \mathbb{R}$. The goals is to recover the support boundary f nonparametrically; see Figure 1. In a similar way as the Gaussian white noise model is the continuous analogue of nonparametric regression with centered errors, support boundary recovery occurs as the continuous limit of nonparametric regression with one-sided errors; see Meister and Reiß [25] for related asymptotic equivalence results. The fundamental difference is the model geometry: the Hellinger distance for the Gaussian white noise model is induced by the L^2-norm, whereas for support boundary recovery the Hellinger geometry comes from the L^1-norm and the laws are not mutually absolutely continuous. As a consequence, not only convergence rates differ, but also the asymptotic distributions of estimators are nonclassical. Moreover, the maximum-likelihood estimator (MLE) is often not efficient and in parametric settings Bayesian methods are advocated. At a methodological level, we explore here to what extent this remains true for non and semiparametric problems. This is particularly interesting because for many function classes a nonparametric MLE exists in the PPP model. In the related problem of boundary detection in images under Gaussian noise, the Hellinger distance is also of L^1-type (cf. Li and Ghosal [23] for posterior contraction results) but the observation laws are mutually absolutely continuous and a nonparametric MLE usually does not exist.

Received September 2018; revised February 2019.
MSC2010 subject classifications. Primary 62C10, 62G05; secondary 60G55.
Key words and phrases. Frequentist Bayes analysis, posterior contraction, Bernstein–von Mises theorem, Poisson point process, boundary detection, compound Poisson process, subordinator prior.

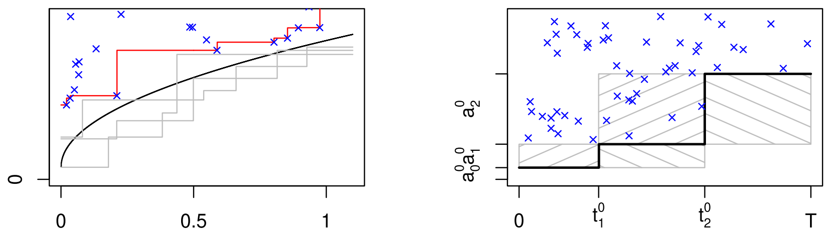

FIG. 1. *Two simulated data examples for the PPP model with true boundary* (*black*) *and observations* (*blue*). *Left*: *MLE* (*red*), *posterior draws* (*gray*). *Right*: *shaded gray areas related to Definition* 4.1 *below.*

A general goal is to understand the performance of compound Poisson processes (CPP) as nonparametric priors. CPPs are probabilistically well understood, are easy to sample and can be equivalently understood as piecewise constant priors, where the jump locations are uniform, the jump sizes are i.i.d. random and the number of jumps is chosen by a Poisson hyperprior. For binary regression, CPP priors were studied by Coram and Lalley [8], establishing nonparametric consistency, and they are often recommended in practice, for example, as priors for monotone functions in Holmes and Heard [18] with applications to gene expression data. We prove below that under CPP priors, optimal posterior contraction rates (sometimes up to logarithmic factors) are attained for Hölder functions and for monotone functions. They even adapt automatically to the unknown Hölder smoothness. Given that the jump intensity remains fixed, this shows how powerful and versatile simple CPP priors are. The derivation of the contraction rates is based on the general theory developed in the companion paper [27]. The theory for monotone functions extends to subordinator priors, that is, monotone Lévy processes, which have been studied in survival analysis by Kim and Lee [20], but not yet in the context of nonparametric posterior contraction rates.

Going beyond rate results, most effort is required to study limiting shape results for the function f and its mean $\vartheta = \int f$, a basic semiparametric functional. Concerning the frequentist approach, the nonparametric MLE $\widehat{f}^{\mathrm{MLE}}$ exists for Hölder balls with smoothness index $\beta \le 1$ and monotone functions, possibly constrained to be piecewise constant, and achieves each time the minimax estimation rate. For functionals such as ϑ, however, the MLE $\widehat{\vartheta}^{\mathrm{MLE}} = \int \widehat{f}^{\mathrm{MLE}}$ converges usually with a suboptimal rate. A rate-optimal estimator can be obtained if we subtract a term that scales with the number of observations lying on the boundary of the MLE and consider

$$\widehat{\vartheta} = \int \widehat{f}^{\mathrm{MLE}} - \frac{\text{number of data points } (X_i, Y_i) \text{ on boundary of } \widehat{f}^{\mathrm{MLE}}}{n}; \tag{2}$$

see Reiß and Selk [29]. This bias correction accounts for the fact that $\widehat{f}^{\mathrm{MLE}}$ overshoots the true boundary function f considerably. In the case of a constant function f and for more general parametric setups, Bayes estimators correct the bias of the MLE by distributing the posterior mass correctly below $\widehat{f}^{\mathrm{MLE}}$; cf. Kleijn and Knapik [21].

It is therefore natural to ask whether a nonparametric Bayesian approach also performs this correction automatically. Here, we show that the answer is positive if the model is well specified. For piecewise constant and monotone support boundaries under CPP priors, the posterior concentrates around $\widehat{\vartheta}$ with the optimal contraction rate. Optimal frequentist estimation of piecewise constant and monotone functions in Gaussian noise has attracted a lot of attention recently; see Gao et al. [12] and the references discussed there. Furthermore, we obtain intervals which are simultaneously asymptotic $(1-\alpha)$-credible and $(1-\alpha)$-confidence intervals of rate-optimal length. The Bayesian approach clearly outperforms the MLE in this case.

As a negative example, we consider a linear support boundary f. The posterior contracts around the true support boundary f with the optimal rate, but the bias correction of the marginal posterior for ϑ is of incorrect order. In this case, credible sets have asymptotically no frequentist coverage. Conceptionally, we study a Bayesian model selection procedure for increasing parameter dimensions, where the hyperprior on the number of jumps determines the model dimension. For linear and exponential family models, such Bernstein–von Mises results have been obtained by Ghosal [13, 14] and by Bontemps [1] for Gaussian regression. Panov and Spokoiny [26] explore the scope of Bernstein–von Mises phenomenon for regular i.i.d. models of growing dimension and find a critical dimension related to ours; see the discussion in Section C.1 of the supplementary material. A bias problem for functional estimation by adaptive Bayesian methods has been exhibited by Castillo and Rousseau [2] and Rivoirard and Rousseau [30], which bears some similarity with ours, but at a parametric $\sqrt{n}$-rate.

Related to CPP priors are many popular piecewise constant prior prescriptions. First of all, there are priors on regression trees, such as Bayesian CART (Denison et al. [9]) and BART (Chipman et al. [7]). Regression trees subdivide the space of covariates and then put a constant value on each of the cells. These priors are henceforth supported on piecewise constant functions. Posterior contraction for BART has been derived only recently by Rockova and van der Pas [31]. For density estimation, histogram priors are well studied. Scricciolo [34] considers random histograms with fixed bin width and the number of bins a hyperprior. It is shown that near optimal contraction rates are obtained if the true density is either Hölder with index at most one or piecewise constant.

So far, only little theory has been developed for nonparametric Bayes under constraints on the shape like monotonicity or convexity. Exceptions are Salomond [32] for monotone densities and Mariucci et al. [24] for log-concave densities. In both cases, mixtures of Dirichlet processes are taken as priors. To the best of our knowledge, the present paper is the first one that derives Bernstein–von Mises-type results under monotonicity constraints.

In Section 2, contraction rates for compound Poisson process and subordinator priors are investigated. In an interlude, Section 3 discusses a general description of the asymptotic posterior shape, in which the results thereafter can be embedded. Bernstein-von Mises-type theorems and results on the frequentist coverage of credible sets for CPP priors can be found in Section 4. Before, it is recommended to read Appendix C in the supplement where the prototypical case of random histogram priors with fixed jump times is treated. Most proofs are deferred to the supplementary material [28].

Notation. We write $N = \sum_i \delta_{(X_i, Y_i)}$ for a random point measure on $[0, 1] \times \mathbb{R}$ and denote the support points by $(X_i, Y_i)_i$. Whenever N is observed, it is natural to call the support points observations. Moreover, we use the standard terminology $\mathbf{1}_A := \mathbf{1}(\cdot \in A)$, $(x)_+ := \max(x, 0)$ and $\|\cdot\|_p$ for the $L^p([0, 1])$-norm.

2. Posterior contraction.

Bayes formula. Let us first recall the Bayes formula for the PPP model as derived in [27]. Let (Θ, d) be a Polish space equipped with its Borel σ-algebra and d a stronger metric than the L^1-norm. For $f_0 \in L^1([0, T])$, a prior Π on Θ and a Borel set $B \subset \Theta$, Lemma 2.2 in [27] gives an explicit Bayes formula under the law P_{f_0}:

$$\Pi(B \mid N) = \frac{\int_B e^{n\int_0^T f} \mathbf{1}(\forall i : f(X_i) \le Y_i)\, d\Pi(f)}{\int_\Theta e^{n\int_0^T f} \mathbf{1}(\forall i : f(X_i) \le Y_i)\, d\Pi(f)} \tag{3}$$
$$= \frac{\int_B e^{-n\int_0^T (f_0-f)_+} \frac{dP_{f\vee f_0}}{dP_{f_0}}(N)\, d\Pi(f)}{\int_\Theta e^{-n\int_0^T (f_0-f)_+} \frac{dP_{f\vee f_0}}{dP_{f_0}}(N)\, d\Pi(f)} \quad P_{f_0}\text{-a.s.}$$

The default is $T = 1$ but in Section 4 it is convenient to work with $T > 1$.

Compound Poisson process prior. We study posterior contraction for compound Poisson process priors defined on the space $\Theta = D[0, 1]$ of càdlàg functions, equipped with the Skorokhod topology. A compound Poisson process Y on $[0, 1]$ can be written as $Y_t = \sum_{i=1}^{N_t} \Delta_i$ with a Poisson process $(N_t)_{t\ge 0}$ of intensity $\lambda > 0$ and an i.i.d. sequence (Δ_i) of random variables, independent of the Poisson process. We denote the distribution of Δ_1 by G. We randomize the starting value $X_0 = \Delta_0$ according to a distribution H and consider

$$X_t = \Delta_0 + \sum_{i=1}^{N_t} \Delta_i = \sum_{i=0}^{N_t} \Delta_i, \tag{4}$$

with $\Delta_0 \sim H$ independent of $(\Delta_i)_{i\ge 1}$ and $(N_t)_{t\ge 0}$.

A CPP can equivalently be viewed as a hierarchical prior on f in the spirit of [3, 4]. The hierarchical CPP construction picks in a first step a model dimension prior $\pi \sim \mathrm{Pois}(\lambda)$. The order statistics property of a Poisson process ([10], page 186) says that conditionally on the event that the CPP jumps K times on $[0, 1]$, the ordered jump locations $(t_1, \dots, t_K)$, $t_0 := 0 \le t_1 \le \cdots \le t_K \le 1$, have the same distribution as the order statistic of K independent $U([0, 1])$-random variables. The Lebesgue density of $(t_1, \dots, t_K)|K$ is therefore $K!\mathbf{1}(0 \le t_1 \le t_2 \le \cdots \le t_K \le 1)$. The last step is then to assign the starting value a_0 and the jump sizes $a_1, \dots, a_K$. Assuming that the distributions G, H have Lebesgue densities g and h, respectively, we can write the CPP prior in closed form as a prior on K, $\mathbf{t}$ and $\mathbf{b}$

$$(K, \mathbf{t}, \mathbf{a}) \mapsto e^{-\lambda}\lambda^K h(a_0) \prod_{j=1}^K g(a_j)\mathbf{1}(0 < t_1 < t_2 < \cdots < t_K < 1) \tag{5}$$

generating random càdlàg functions $f = \sum_{j=0}^K a_j \mathbf{1}_{[t_j,1]}$ with $t_0 := 0$.

Since λ is fixed, for most draws of the prior the number of jumps will be of order λ. As we show below, the CPP prior puts still enough mass around functions with an increasing number of jumps to ensure nearly optimal posterior contraction rates for Hölder functions. Let us also mention that the CPP prior randomizes over the jump points and should therefore be able to adapt to local smoothness.

Function classes. We denote by $\mathcal{C}^\beta(R)$ the ball of β-Hölder functions $f : [0, 1] \to \mathbb{R}$ with Hölder norm $\|f\|_{\mathcal{C}^\beta}$ bounded by R. The CPP prior allows to build in monotonicity as prior knowledge by choosing a positive jump distribution. We define the space of monotone functions, which are bounded by R, as

$$\mathcal{M}(R) := \{f : f \text{ monotone increasing and } -R \le f(0) \le f(1) \le R\}.$$

The following result is proved in Appendix A.1.

THEOREM 2.1. *Consider the CPP prior* (4) *with a positive and continuous Lebesgue density h on $\mathbb{R}$. If there are constants $\gamma, L > 0$ such that $\mathbb{P}(|\Delta_i| \ge s) \le L^{-1}e^{-Ls^\gamma}$ for all $s \ge 0$, then there exist positive constants M and c such that*:

(i) *if g is positive and continuous on $\mathbb{R}^+$*,

$$\sup_{f_0\in\mathcal{M}(R)} E_{f_0}\Big[\Pi\Big(f : \|f - f_0\|_1 \geq M\sqrt{\frac{\log n}{n}} \;\Big|\; N\Big)\Big] \leq e^{-c\sqrt{n\log n}};$$

(ii) *if g is positive and continuous on $\mathbb{R}$ and $\beta \leq 1$*,

$$\sup_{f_0\in\mathcal{C}^\beta(R)} E_{f_0}\Big[\Pi\Big(f : \|f - f_0\|_1 \geq M\Big(\frac{\log n}{n}\Big)^{\beta/(1+\beta)} \;\Big|\; N\Big)\Big] \leq e^{-cn(\frac{\log n}{n})^{\frac{\beta}{1+\beta}}}.$$

For bounded piecewise constant functions with K_n jumps at fixed jump times, the posterior contraction rate under quite general CPP priors is similarly found to be $\frac{K_n}{n}\log n$ whenever $n^\varepsilon \lesssim K_n \lesssim n^{1-\varepsilon}$ for some $\varepsilon > 0$. Much finer limiting shape results for a similar class, however, will be obtained in the next section.

In all cases, the rate is optimal up to logarithmic factors. This follows from the lower bound of Theorem 4.2 in [19] for Hölder balls. The argument can also be extended from $\mathcal{C}^1(R')$ to $\mathcal{M}(R)$ by adding a multiple of the identity to each test functions. Thus, $n^{-1/2}$ is also a lower bound for the rate over monotone functions. Compound Poisson processes thus furnish a very versatile prior adapting to unknown smoothness and possibly monotonicity.

The proof is based on a Ghosal–Ghosh–van der Vaart-type result from [27]. To check the conditions, we derive lower bounds on the one-sided small ball probabilities of the CPP prior for the function classes considered above. These bounds could be used to derive contraction rates for other nonparametric models, also.

Subordinators. CPPs form the subclass of Lévy processes with finite jump intensity. Allowing also for infinitely many jumps, subordinators, that is, Lévy processes with monotone sample paths, generate a rich class of monotone function priors. We consider only subordinators without drift, characterized by their characteristic function

$$\phi_t(u) = \mathbb{E}[e^{iuY_t}] = \exp\Big(t\int_{\mathbb{R}^+}(e^{iux} - 1)\nu(dx)\Big), \quad t \geq 0,$$

where the Lévy measure ν is a σ-finite measure on $\mathbb{R}^+$, satisfying $\int_{\mathbb{R}^+}(x\wedge 1)\nu(dx) < \infty$. Its intensity is $\lambda = \nu(\mathbb{R}^+) \in [0,\infty]$ and in the finite intensity case a subordinator is just a compound Poisson process of intensity λ with jump distribution $G = \nu/\lambda$.

Among subordinators of infinite intensity prominent examples are the Gamma and inverse Gaussian processes; see [33] for a comprehensive treatment. Dirichlet processes belong to the most frequently used priors in nonparametric Bayesian methods and can be viewed as time-changed and normalized Gamma processes; see [16], Section 4.2.3. Subordinators as priors have been studied in the context of survival models by [20]. There the target of estimation is the cumulative hazard function, which can be estimated at the parametric rate $n^{-1/2}$. Subordinators as priors for monotone estimation problems in regression or density-type models do not seem to have been analyzed yet so that the result below can be of independent interest.

The randomly initialized subordinator prior. As priors, we consider randomly initialized subordinators of the form

$$X_t = Y_0 + Y_t, \quad \text{with } (Y_t)_{t\geq 0} \text{ a subordinator and } Y_0 \sim H \text{ independent of } (Y_t)_{t>0},$$

where H is assumed to have a positive and continuous Lebesgue density on $\mathbb{R}$. Moreover, we suppose that the Lévy measure ν has a Lebesgue density which by some slight abuse of notation is called $\nu(x)$ and is assumed to be continuous and positive on $\mathbb{R}^+$.

THEOREM 2.2. *Consider the randomly initialized subordinator prior. If there exist constants* $\gamma, L > 0$ *such that* $\nu(x) \le Lx^{-3/2}$ *for all* $x > 0$ *and* $\int_s^\infty \nu(x)\,dx \le Le^{-L^{-1}s^\gamma}$ *for all* $s \ge 1$, *then there are constants* $M, c > 0$ *such that*

$$\sup_{f_0 \in \mathcal{M}(R)} E_{f_0}\left[\Pi\left(f : \|f - f_0\|_1 \ge M\sqrt{\frac{\log n}{n}} \,\Big|\, N\right)\right] \le e^{-c\sqrt{n \log n}}.$$

The theorem is proved in Appendix A.2.

3. On the generalized Bernstein–von Mises phenomenon. Before we move on and study the posterior limit for the CPP prior, we briefly discuss the extension of the Bernstein–von Mises theorem beyond regular models. The classical Bernstein–von Mises theorem assumes a parametric model $(P_\vartheta^n : \vartheta \in \Theta)$ that is differentiable in quadratic mean and has nonsingular Fisher information $I_{\vartheta,n}$. Then, for a continuous and positive prior, the posterior can be approximated in total variation distance by

$$\mathcal{N}(\widehat{\vartheta}_n^{\mathrm{MLE}}, I_{\vartheta_0,n}^{-1})$$

if the i.i.d. data are generated from $P_{\vartheta_0}^n$, $\vartheta_0 \in \Theta$; see [36], Section 10.2 for a precise statement. It can also be easily seen that if we observe $Y_i = \vartheta_0 + \varepsilon_i$, $i = 1, \dots, n$, with independent $\varepsilon_i \sim \mathrm{Exp}(1)$, then $\widehat{\vartheta}_n^{\mathrm{MLE}} = \min(Y_1, \dots, Y_n) \sim \vartheta_0 + \varepsilon$ with $\varepsilon \sim \mathrm{Exp}(n)$. For a continuous and positive prior, we obtain in the limit the posterior $(\widehat{\vartheta}_n^{\mathrm{MLE}} - \widetilde{\varepsilon})|\widehat{\vartheta}^{\mathrm{MLE}}$ with $\widetilde{\varepsilon} \sim \mathrm{Exp}(n)$ and $\widetilde{\varepsilon}$ independent of ε; see [21].

This suggests that a generalized Bernstein–von Mises theorem should be of the following form: If there exists a MLE $\widehat{\vartheta}_n^{\mathrm{MLE}}$ such that

$$\widehat{\vartheta}_n^{\mathrm{MLE}} = \vartheta_0 + \varepsilon_n(\vartheta_0), \tag{6}$$

with $\varepsilon_n(\vartheta_0)$ some random variable, then, under standard assumptions on the prior, the posterior should be close to the conditional distribution of

$$\left(\widehat{\vartheta}_n^{\mathrm{MLE}} - \widetilde{\varepsilon}_n(\vartheta_0)\right)\big|\widehat{\vartheta}_n^{\mathrm{MLE}}, \tag{7}$$

where $\widetilde{\varepsilon}_n(\vartheta_0)$ has the same distribution as $\varepsilon_n(\vartheta)$ but is independent of it. This unifies both cases above and extends the general insight gained in [15] to not mutually absolutely continuous distributions. For problems with increasing model dimension, we can additionally build in a model selection prior such that the posterior concentrates on smaller models. If the posterior puts asymptotically all mass on one model, then (6) and (7) have to be replaced by the corresponding expressions in this model; see [3], Section 2.4 for an example. The posterior limit distributions that occur in the subsequent chapters are exactly of this form.

4. Limiting shape of the posterior. We consider the CPP prior and study support boundaries that are piecewise constant and monotone. This function class has received a lot of attention recently in nonparametric statistics; see [5, 12]. Due to the imposed monotonicity, the nonparametric MLE exists and we believe that this is crucial for the posterior to have a tractable limit distribution; see also Section 3. For the model size, we show that the full posterior concentrates on the true number of jumps under minimal signal strength assumptions. The randomness of the jump locations and the function values on each piece induce the randomness of the limiting shape.

A prior class which is easier to analyze and reveals already main features of the CPP results are random histogram priors. They consist of piecewise constant functions with fixed jump times and number K_n of jumps possibly tending to infinity. In Appendix C of the supplementary material, we show a limiting shape result for random histogram priors and study the bias correction for estimation of a functional.

4.1. *The limiting shape of the full posterior.* We first derive the limiting shape of the full posterior and then study the marginal distribution for functionals.

Model. The likelihood taken over all increasing functions on $[0, T]$ is unbounded. This is caused by functions that have an extremely steep jump close to the right boundary of the observation interval $[0, T]$. Similar boundary phenomena are well known in the nonparametric maximum likelihood theory under shape constraints. The unboundedness of the likelihood causes the Bayes formula to be extremely sensitive to values close to the right boundary. Since we are interested in a framework that avoids these extreme spikes at the boundary, we consider the PPP model (1) with $T > 1$, assuming that the true function is constant on the interval $[1, T]$. For jump functions, this is the same as saying that all jumps occur before time one.

Function class. We consider piecewise constant, right-continuous functions that are monotone increasing, assuming that all jumps occur up to time one:

$$\mathcal{M}(K, R) := \left\{ f = \sum_{\ell=0}^{K} a_\ell \mathbf{1}_{[t_\ell, T]} : 0 \le a_\ell \le R, 0 \le t_1 \le \cdots \le t_K \le 1 \right\}.$$

A discretized version of this class has been recently studied in [12]. For a generic function in $\mathcal{M}(K, R)$, we write $f = \sum_{\ell=0}^{K} a_\ell \mathbf{1}(\cdot \ge t_\ell)$ with ordered jump locations $0 =: t_0 \le t_1 \le \cdots \le t_K \le 1 < t_{K+1} := T$. We assume that there is a minimal signal strength. Without such a constraint, one cannot exclude the case that the number of true jumps is consistently underestimated; see for instance [11], Section 2.1. Typically, conditions of this type occur when there is an underlying model selection problem, compare with the β-min conditions for high-dimensional problems.

DEFINITION 4.1. A function $f_0 \in \mathcal{M}(K_n, R)$ belongs to the subclass $\mathcal{M}_S(K_n, R)$ if and only if for all $k = 1, \dots, K_n$,

$$a_k^0(t_{k+1}^0 - t_k^0) \wedge a_k^0(t_k^0 - t_{k-1}^0) \ge 2K_n \log(eK_n)\frac{\log^3 n}{n},$$
$$a_k^0 \ge \frac{2\log n}{\sqrt{n}}, (t_{k+1}^0 - t_k^0) \ge \frac{2}{\sqrt{n}},$$

and the two last inequalities also hold for $k = 0$.

REMARK 4.2. Since $\sum_{k=0}^{K_n-1}(t_{k+1}^0 - t_k^0) \le 1$, the last condition implies $K_n = \mathcal{O}(n^{1/2})$. In view of $\max_k a_k^0 \le R$, the first condition even implies $K_n^2 \log(eK_n) \le Rn/\log^3(n)$, in particular $K_n = o(n^{1/2})$.

The expressions $a_i^0(t_{i+1}^0 - t_i^0)$ and $a_i^0(t_i^0 - t_{i-1}^0)$ are the areas in Figure 1(right). Let us briefly discuss the imposed lower bound on these areas. The PPP has intensity n on the epigraph of the support boundary. In order to ensure that each of the K_n sets contain at least one support point of the PPP, all of them need to have an area of at least order $\log(K_n)/n$. One might therefore wonder whether the factor K_n in the lower bound for the areas is necessary to ensure strong model selection. We shall see that the posterior has to choose among a huge number of models; cf. the proof of Proposition 4.3. To find the correct model might therefore indeed require a larger lower bound on the areas.

Prior. By assumption all jumps occur before time one. We therefore draw the prior from a CPP on $[0, 1]$ and then extend it continuously to a prior on $[0, T]$ by appending a constant function on $(1, T]$. The Lebesgue density of $(t_1, \dots, t_K)|K$ is $K!\mathbf{1}(0 \le t_1 \le t_2 \le \cdots \le t_K \le 1)$; see Section 2. To model the monotonicity, the jump distribution should be supported on the positive real line. It turns out that there is one natural prior on the jump sizes. The construction is as follows: choose the random starting value of the CPP according to $a_0 \sim \text{Exp}(1)$ and independently draw i.i.d. jump sizes $a_\ell \sim \Gamma(2, 1)$ for $\ell = 1, \dots, K$. With

$$g_K(\mathbf{a}) = e^{-\sum_{k=0}^K a_k} \prod_{k=1}^K a_k, \quad \mathbf{a} = (a_0, \dots, a_K) \in \mathbb{R}_+^{K+1} \tag{8}$$

the prior (5) takes the more specific form

$$(K, \mathbf{t}, \mathbf{a}) \mapsto e^{-\lambda} \lambda^K g_K(\mathbf{a}) \mathbf{1}(0 \le t_1 \le t_2 \le \cdots \le t_K \le 1). \tag{9}$$

We can also rewrite the prior as a prior on functions of the form $f = \sum_{k=0}^K b_k \mathbf{1}_{[t_k, t_{k+1})}$. Under this reparametrization, we obtain $g_K(\mathbf{b}) = e^{-b_K} \cdot \prod_{k=1}^K (b_k - b_{k-1})_+$.

Since $f(0) = a_0$, this means in particular that all paths generated by the prior are nonnegative. To put different priors on a_0 and a_ℓ, $\ell \ge 1$, turns out to be natural. For this specific choice, the marginal posterior of any a_k follows approximately an exponential distribution. This is a crucial property that allows us to derive tight bounds for the numerator and denominator in the Bayes formula; compare the proofs of Lemma D.1 and Lemma D.4 for more details.

MLE. Over all monotone functions on $[0, T]$, $T > 1$, that are constant on $[1, T]$, there exists a nonparametric MLE $\widehat{f}^{\text{MLE}}$ (unique almost surely). Existence follows from the general theory because the class of monotone functions is closed under the maximum; see [29]. Almost surely, the MLE is piecewise constant with finitely many jumps and bounded. This implies in particular, that $\widehat{f}^{\text{MLE}}$ is also the MLE over all piecewise constant monotone functions with jumps on $[0, 1]$. Furthermore, $f \le \widehat{f}^{\text{MLE}}$ for all piecewise constant and monotone functions f satisfying $f(X_i) \le Y_i$ for all i. Denoting the number of jumps by M, we write

$$\widehat{f}^{\text{MLE}}(t) = \sum_{\ell=0}^M \widehat{a}_\ell^{\text{MLE}} \mathbf{1}(t \ge \widehat{t}_\ell^{\text{MLE}}), \quad t \in [0, T]$$

with $0 =: \widehat{t}_0^{\text{MLE}} < \widehat{t}_1^{\text{MLE}} < \cdots < \widehat{t}_M^{\text{MLE}} \le 1$. This MLE should not be confused with the monotone MLE on $[0, T]$ without the restriction that the functions are constant on $[1, T]$.

Construction of the majorant process $\widetilde{f}$. We consider two sequences of observation points that are close to the true jump points of the unknown regression function f_0. Recall that $t_0^0 = 0$ and $t_{K_n+1}^0 = T$. For $k = 0, 1, \dots, K_n$ consider

$$(X_k^*, Y_k^*) := \operatorname*{arg\,min}_{(X_i, Y_i) \text{ observation point}} \{Y_i : X_i \in [t_k^0, t_{k+1}^0)\} \tag{10}$$

and with

$$R_k := \{(X_i, Y_i) \text{ observation} : X_i \in [t_{k-1}^0, t_k^0), Y_i \le f_0(t_k^0)\}$$

for $k = 1, \dots, K_n$

$$(X_k', Y_k') := \begin{cases} \operatorname*{arg\,max}_{(X_i, Y_i)} \{X_i : (X_i, Y_i) \in R_k\}, & \text{if } R_k \ne \varnothing, \\ (t_{k-1}^0, f_0(t_{k-1}^0)), & \text{otherwise.} \end{cases} \tag{11}$$

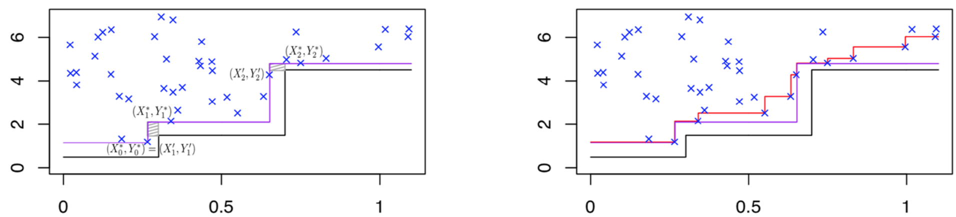

FIG. 2. *Left*: *Data example with true boundary* (*black*), *the function* $\widetilde{f}$ (*purple*) *and the sequences* (X_k^*, Y_k^*), (X_k', Y_k'). *Right*: *If none of the observations fall into the gray areas*, *then the sequences* (X_k^*, Y_k^*), (X_k', Y_k') *lie on the MLE over monotone functions* (*red*).

We also set $X_0' := 0$ and $X_{K_n+1}' := T$. With probability one, the sequences are unique; see also Figure 2. The assigned values for the case $R_k = \varnothing$ do not affect the asymptotic analysis, but are convenient choices giving the guarantee that the subsequent formulas are well-defined. By construction and the properties of the PPP, we have for $k = 1, \ldots, K_n$ that

$$\begin{aligned} Y_k^* - f_0(t_k^0) &\sim \mathrm{Exp}\big(n(t_{k+1}^0 - t_k^0)\big), \\ t_k^0 - X_k' &\sim \mathrm{Exp}\big(na_k^0\big) \wedge \big(t_k^0 - t_{k-1}^0\big). \end{aligned} \tag{12}$$

Here, $\mathrm{Exp}(\beta) \wedge t$ denotes a truncated exponential distribution with density $\frac{\beta}{1-\exp(-\beta t)} e^{-\beta x} \times \mathbf{1}_{[0,t]}(x)$. The definition of Y_k^* is based on the set $[t_k^0, t_{k+1}^0) \times \mathbb{R}$. For different k, the sets are disjoint and the random variables Y_k^* are independent. The same argument shows that X_k', $k = 1, \ldots, K_n$ is a sequence of independent random variables and Y_k^*, X_ℓ' are independent if $k \neq \ell - 1$.

The key object for the limiting shape result of the posterior is the process

$$\widetilde{f} = \sum_{k=0}^{K_n} Y_k^* \mathbf{1}_{[X_k', X_{k+1}')}, \tag{13}$$

a realization of which is displayed in Figure 2. Since $\widetilde{f} \geq f_0$, we call $\widetilde{f}$ also the *majorant process* (of f_0). Observe that the majorant process is piecewise constant with K_n jumps. The distribution of $\widetilde{f}$ can essentially be deduced from (12). As the support boundary is unknown, the majorant process cannot be computed from the data alone. As proved in Appendix D.4, $\widetilde{f}$ coincides asymptotically with the MLE over monotone functions with the correct number of K_n jumps.

PROPOSITION 4.3. *If* $\widehat{f}_{K_n}^{\mathrm{MLE}}$ *denotes the MLE in the space* $\mathcal{M}(K_n, \infty)$, *then*

$$\inf_{f_0 \in \mathcal{M}_S(K_n, R)} P_{f_0}\big(\widetilde{f} = \widehat{f}_{K_n}^{\mathrm{MLE}}\big) \to 1.$$

In particular, $\inf_{f_0 \in \mathcal{M}_S(K_n,R)} P_{f_0}(\widetilde{f}$ *is monotone*$) \to 1$.

For the construction, note that $\widehat{f}_{K_n}^{\mathrm{MLE}}$ is obtained as the monotone and piecewise constant function f with at most K_n jumps that maximizes $\int f$ under the constraint $f(X_i) \leq Y_i$ for all observations (X_i, Y_i). The upper jump points lie on the monotone MLE (corresponding to $K_n = \infty$), which is described explicitly in [29].

Limit distribution. We now describe the sequence of distributions that asymptotically approximates the posterior. For convenience, we ignore the dependence on n and refer to this sequence as the limit distribution. Working conditionally on the sequences $(X'_k)_k$ and $(Y^*_k)_k$, the limit distribution $\Pi^\infty_{f_0,n}$ is the distribution on the Skorokhod space $D([0,T])$ of

$$f = \sum_{k=0}^{K_n} (Y^*_k - E^*_k)\mathbf{1}_{[X'_k+E'_k, X'_{k+1}+E'_{k+1})} \tag{14}$$

with independent $E^*_k \sim \mathrm{Exp}(n(X'_{k+1} - X'_k))$ and $E'_k \sim \mathrm{Exp}(n(Y^*_k - Y^*_{k-1})) \wedge (X'_{k+1} - X'_k)$, $k \le K_n$, and $E'_0 := E'_{K_n+1} := 0$.

The limit distribution is obtained from the majorant process $\widetilde{f}$ by moving each jump location independently to the right by a (truncated) exponential distribution with scale parameter $n(Y^*_k - Y^*_{k-1})$. Moreover, the function value on each piece is decreased by another independently generated exponential random variable. In view of Proposition 4.3, it follows that the limit is of the generalized form discussed in Section 3. In Appendix D.3, we show the following.

THEOREM 4.4 (Limiting shape result for CPP prior). *Let $K_n \le n^{1/2-\delta}$ for some $\delta > 0$. Then for the prior (8) and $\Pi^\infty_{f_0,n}$ as defined in (14)*

$$\lim_{n\to\infty} \sup_{f_0 \in \mathcal{M}_S(K_n,R)} E_{f_0}\big[\big\|\Pi(\cdot \mid N) - \Pi^\infty_{f_0,n}\big\|_{\mathrm{TV}}\big] = 0.$$

Since we work with one specific prior, we call this a limiting shape result instead of a Bernstein–von Mises theorem. Using (14), one can show that the posterior contracts with rate K_n/n. We conjecture that the MLE only achieves the slower rate $K_n \log n/n$. One of the heuristic reasons is that the MLE overshoots the true model dimension K_n by choosing a model with order $K_n \log n$ many jumps; see Figure 2 and Lemma E.3. It is conceivable that each of the additional jumps introduces an error of size $1/n$ which then gives the rate $K_n \log n/n$. A similar phenomenon occurs in the nonparametric regression model; see Proposition 2.1 in [12].

The proof is nonstandard. It follows immediately from the likelihood that the posterior only puts mass on paths that lie below the monotone MLE $\widehat{f}^{\mathrm{MLE}}$. Let f be a piecewise constant function with K jumps such that there exists a function $f_>$ with $K-1$ jumps and such that $f \le f_> \le \widehat{f}^{\mathrm{MLE}}$. Interestingly, the posterior puts negligible mass on the union over all such functions and all K. The remaining paths have more structure. We use this to introduce a parametrization from which we can derive sufficiently sharp bounds over the corresponding integrals in the Bayes formula. The proof also requires many properties of the monotone MLE which might be of independent interest and are collected in Appendix E.

4.2. *Posterior coverage for a functional.* For the functional $\vartheta = \int_0^T f$, we have under the limit distribution $\Pi^\infty_{f_0,n}$,

$$\begin{aligned} \vartheta = \int_0^T \widetilde{f} - \sum_{k=0}^{K_n} E^*_k (X'_{k+1} - X'_k) - \sum_{k=1}^{K_n} E'_k (Y^*_k - Y^*_{k-1}) \\ - \sum_{k=0}^{K_n} E^*_k (E'_{k+1} - E'_k). \end{aligned} \tag{15}$$

We show the convergence to a normal distribution in Appendix D.5 of the supplementary material. Given two probability measures P, Q on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, let us consider the Kolmogorov–Smirnov distance

$$\|P - Q\|_{\mathrm{KS}} := \sup_{x\in\mathbb{R}} \big|P\big((-\infty, x]\big) - Q\big((-\infty, x]\big)\big|.$$

THEOREM 4.5. *Consider the prior* (8). *Then, for any sequence $K_n \to \infty$ with $K_n \le n^{1/2-\delta}$ for some $\delta > 0$,*

$$\sup_{f_0 \in \mathcal{M}_S(K_n,R)} E_{f_0}\left[\left\| \Pi(\vartheta \in \cdot \mid N) - \mathcal{N}\left(\int_0^T \widetilde{f} - \frac{2K_n+1}{n}, \frac{2K_n+1}{n^2}\right)\right\|_{\mathrm{KS}}\right] \to 0.$$

The asymptotic $(1-\alpha)$-credible interval

$$\overline{I}(\alpha) = \left[\int_0^T \widetilde{f} - \frac{2K_n+1}{n} - \frac{\sqrt{2K_n+1}}{n} q_{1-\alpha/2}, \int_0^T \widetilde{f} - \frac{2K_n+1}{n} + \frac{\sqrt{2K_n+1}}{n} q_{1-\alpha/2}\right] \tag{16}$$

with the $q_{1-\alpha/2}$-quantile of $N(0,1)$ is moreover an honest asymptotic confidence set:

$$\sup_{f_0 \in \mathcal{M}_S(K_n,R)} \left| P_{f_0}\left(\int f_0 \in \overline{I}(\alpha)\right) - (1-\alpha)\right| \to 0.$$

By Lemma 4.3, the majorant process $\widetilde{f}$ in the limit distribution can be replaced by $\widehat{f}_{K_n}^{\mathrm{MLE}}$. The result is formulated in terms of the Kolmogorov–Smirnov distance, which suffices to describe asymptotic probabilities for credible intervals. It is not clear whether a total variation version holds as well because point masses enter into the proof argument and are difficult to control.

The observations that lie on the majorant process are (X_k', Y_k'), $k = 1, \dots, K_n$ and (X_k^*, Y_k^*), $k = 0, \dots, K_n$. This means that $2K_n + 1$ observations lie on the boundary of $\widetilde{f}$ (almost surely). The bias correction term $(2K_n+1)/n$ is consequently of the same form as for the bias-corrected MLE in [29]. We can now argue as in Corollary C.2 to construct a $(1-\alpha)$-credible interval that is also an asymptotic $(1-\alpha)$-confidence interval and shrinks with the correct rate $O(\sqrt{K_n}/n)$.

The proof of Theorem 4.5 can be adapted to treat other functionals. For linear functionals $\vartheta = \int f(u)w(u)\,du$ with a continuous function w, a much more complicated limit is obtained, involving $w(t_k^0)$ as well as the local averages $\int_{t_k^0}^{t_{k+1}^0} w(u)\,du/(t_{k+1}^0 - t_k^0)$. We omit the details.

4.3. *A negative result on posterior coverage for functionals.* We ask for the coverage of credible sets if the support boundary function is not piecewise constant. For the specific choice,

$$f_0(x) = (x + 1/2) \wedge 3/2, \qquad x \in [0, T], \tag{17}$$

of the support boundary function it is shown that the credible sets for $\vartheta_0 = \int f_0$ do not have asymptotic coverage under a CPP prior. Notice that f_0 is constant on $[1, T]$.

Class of priors. Consider a (generalized) CPP prior. Given the number of jumps K, the jump heights $\mathbf{a} = (a_0, a_1, \dots, a_K)$ are assumed to be independent, but not necessarily identically distributed and the prior is of the form

$$g_K(\mathbf{a}) = \prod_{k=0}^{K} g_k(a_k). \tag{18}$$

For the marginal prior on the individual jumps, we assume that there exist constants $c > 0$, $\gamma \ge 0$, such that

$$g_k(x) \ge c x^{\gamma}, \qquad \forall x \in [0, 1], k \ge 0. \tag{19}$$

In particular, this is satisfied by the prior (8) with $\gamma = 1$ and $c = e^{-1}$.

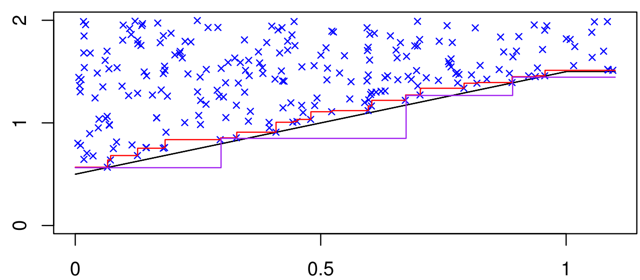

FIG. 3. *The argument for the lower bound with monotone MLE* (*red*), *true function* (*black*) *and a function* $\widehat{f}^{\mathrm{MLE}}_{K,\mathbf{s}}$ *with few jumps* (*purple*). *The posterior puts asymptotically all mass on paths with much fewer jumps than the monotone MLE. This creates a downwards bias of the posterior for the marginal posterior of the integral* $\int_0^1 f$.

The first result shows that under P_{f_0} the posterior concentrates on models with size $\sqrt{n/\log n}$. This is of a slightly smaller order than the MLE, which has of the order $\sqrt{n}$ many jumps. This causes then a downwards bias of the posterior; compare Figure 3. Interestingly, a similar phenomenon occurs in the Gaussian white noise model; cf. Proposition 2 in [2].

PROPOSITION 4.6. *Consider a CPP prior with jump distribution satisfying* (18) *and* (19) *and* f_0 *from* (17). *Then there exists* $c_* > 0$ *such that*

$$E_{f_0}\Big[\Pi\Big(K \ge c_*\sqrt{\frac{n}{\log n}} \,\Big|\, N\Big)\Big] \to 0.$$

Proposition 4.6 is proved in Appendix B.1. The next theorem, proved in Appendix B.2, shows that the entire posterior mass lies asymptotically below the true value. The distance $\sqrt{\log(n)/n}$ is much larger than the optimal estimation rate $n^{-3/4}$ obtained for the mean of monotone functions in [29]. The main argument is that for piecewise constant functions with $K \le c_*(n/\log n)^{1/2}$ jumps the best approximation of the linear function f_0 has order $(n/\log n)^{-1/2}$ in L^1-norm, whereas the monotone MLE has approximation rate $n^{-1/2}$ and forms an upper bound for f_0 and the posterior mass simultaneously.

THEOREM 4.7. *For* f_0 *from* (17), *there exists* $\widetilde{c} > 0$ *such that for the marginal posterior on the functional* $\vartheta = \int_0^1 f$

$$E_{f_0}\Big[\Pi\Big(\vartheta \ge \int_0^1 f_0(x)\,dx - \widetilde{c}\sqrt{\frac{\log n}{n}} \,\Big|\, N\Big)\Big] \to 0.$$

We conjecture that the negative result continues to hold if f_0 is a piecewise constant function with at least $\sqrt{n}$ jumps because the posterior will put all asymptotic mass on models of dimension $O(\sqrt{n/\log n})$, underestimating the number of true jumps by at least a logarithmic factor.

APPENDIX A: PROOFS FOR SECTION 2

Denote by $N(\varepsilon, \mathcal{F}, d)$ the ε-covering number of $\mathcal{F} \subset L^1([0,1])$ with respect to the distance d. The one-sided bracketing number $N_{[}(\delta, \mathcal{F})$ is the smallest number M of functions $\ell_1, \dots, \ell_M \in L^1([0,1])$ such that for any $f \in \mathcal{F}$ there exists $j \in \{1, \dots, M\}$ with $\ell_j \le f$ (almost everywhere) and $\int (f - \ell_j) \le \delta$. The functions ℓ_j are not required to be in $\mathcal{F}$.

THEOREM A.1 (Theorem 2.3 and Corollary 2.6 in [27]). *If for some $\Theta_n \subset \Theta$, some rate $\varepsilon_n \to 0$ and constants $C, C', C'' \geq 1$, $A > 0$:*

$$\text{(i)} \quad N_{[}(\varepsilon_n, \Theta_n) \leq C'' e^{C' n\varepsilon_n};$$

$$\text{(ii)} \quad \Pi\big(f : \|f - f_0\|_1 \leq A\varepsilon_n, f \leq f_0\big) \geq e^{-Cn\varepsilon_n};$$

$$\text{(iii)} \quad \Pi\big(\Theta_n^c\big) \leq C'' e^{-(C+A+1)n\varepsilon_n},$$

then there exists a constant M such that

$$E_{f_0}\big[\Pi\big(f : \|f - f_0\|_1 \geq M\varepsilon_n \mid N\big)\big] \leq 3C'' e^{-n\varepsilon_n}.$$

A.1. Proof of Theorem 2.1. It is convenient to use the notation $\mathbb{P}(X \in A) := \Pi(f \in A)$ to prove generic properties of the compound Poisson process X defined in (4).

LEMMA A.2. *Consider the CPP prior (4) with a positive and continuous Lebesgue density h on $\mathbb{R}$.*

(i) *If g is positive and continuous on $\mathbb{R}^+$, there exists a positive constant $c = c(R)$, such that for any $0 < \varepsilon \leq R \wedge \frac{1}{2}$,*

$$\inf_{f \in \mathcal{M}(R)} \mathbb{P}\big(\|X - f\|_1 \leq 2\varepsilon, X \leq f\big) \geq e^{-2\lambda}(1 \wedge \lambda)^{4R/\varepsilon} \varepsilon^{c\varepsilon^{-1}};$$

(ii) *If g is positive and continuous on $\mathbb{R}$, then for $0 < \beta \leq 1$ there exists a positive constant $c = c(\beta, R)$ such that for any $0 < \varepsilon \leq \frac{R \wedge 1}{2}$,*

$$\inf_{f \in \mathcal{C}^\beta(R)} \mathbb{P}\big(\|X - f\|_\infty \leq \varepsilon\big) \geq e^{-2\lambda}(1 \wedge \lambda)^{(4R/\varepsilon)^{1/\beta}} \varepsilon^{c\varepsilon^{-1/\beta}};$$

PROOF OF (i). For fixed $f \in \mathcal{M}(R)$, we construct a deterministic step function f_- with $f_- \leq f$ and $\|f_- - f\|_1 \leq \varepsilon$. It is then enough to show that for any $0 < \varepsilon \leq R \wedge 1/2$,

$$\mathbb{P}\big(\|X - f_-\|_1 \leq \varepsilon, X \leq f_-\big) \geq e^{-2\lambda}(1 \wedge \lambda)^{4R/\varepsilon} \varepsilon^{c/\varepsilon}. \tag{20}$$

If $\varepsilon \leq R$, there exists δ such that $\varepsilon/(4R) \leq \delta \leq \varepsilon/(2R)$ and $N := 1/\delta$ is a positive integer. Let $r(j, \delta) := f(j\delta) - f((j-1)\delta)$ for $j \geq 1$. Define the step functions

$$f_- := \sum_{j=0}^{N-1} f(j\delta)\mathbf{1}_{[j\delta,(j+1)\delta)} = f(0) + \sum_{j=1}^{N-1} r(j, \delta)\mathbf{1}_{[j\delta,1]}$$

and $f_+ := \sum_{j=1}^{N} f(j\delta)\mathbf{1}_{[(j-1)\delta,j\delta)}$. Since f is monotone increasing, $f_- \leq f \leq f_+$ and $\|f - f_-\|_1 \leq \|f_+ - f_-\|_1 = \delta(f(1) - f(0)) \leq \varepsilon$. By the assumptions on g and h, $c_0 := \inf_{-R-1 \leq x \leq R} h(x) \wedge \inf_{0 \leq y \leq R+1} g(y)$ is positive. Let D be the event that

$$k = N - 1, \qquad f(0) - \varepsilon \leq a_0 \leq f(0) - \frac{\varepsilon}{2},$$

$$r(j, \delta) \leq a_j \leq r(j, \delta) + \frac{\varepsilon\delta}{2}, \qquad t_j \in \left[j\delta, j\delta + \frac{\varepsilon\delta}{2}\right]$$

holds for all $j = 1, \ldots, N - 1$. Then, due to (5) and $e^{-\lambda}/\lambda \geq e^{-2\lambda}$,

$$\begin{aligned}\mathbb{P}\big(\|X - f_-\|_1 \leq \varepsilon, X \leq f_-\big) &\geq \mathbb{P}(X \in D) \geq e^{-\lambda}\lambda^{N-1}\left(c_0\frac{\varepsilon\delta}{2}\right)^N \left(\frac{\varepsilon\delta}{2}\right)^{N-1} \\ &\geq e^{-2\lambda}(1 \wedge \lambda)^{4R/\varepsilon}\left(\sqrt{c_0}\frac{\varepsilon^2}{4R}\right)^{2/\delta}.\end{aligned}$$

This yields (20) and proves (i).

PROOF OF (ii). The argument is very similar to (i). Let now δ be such that $(\varepsilon/(4R))^{1/\beta} \le \frac{1}{2}(\varepsilon/(2R))^{1/\beta} \le \delta \le (\varepsilon/(2R))^{1/\beta}$ and $N := 1/\delta$ is a positive integer. With $r(0,\delta) := f(0)$ and $r(j,\delta) := f(j\delta) - f((j-1)\delta)$ for $j \ge 1$, define

$$f_- := \sum_{j=0}^{N-1} f(j\delta)\mathbf{1}_{[j\delta,(j+1)\delta)} = \sum_{j=0}^{N-1} r(j,\delta)\mathbf{1}_{[j\delta,1]}.$$

Now, $\delta \le (\varepsilon/(2R))^{1/\beta}$ and $f \in \mathcal{C}^\beta(R)$ give $\|f - f_-\|_\infty \le \varepsilon/2$. It is thus enough to prove $\mathbb{P}(\|X - f_-\|_\infty \le \varepsilon/2) \ge e^{-2\lambda}(1 \wedge \lambda)^{(4R/\varepsilon)^{1/\beta}} \varepsilon^{c\varepsilon^{-1/\beta}}$. By assumption, g and h are continuous and positive and, therefore, $c_0 := \inf_{-2R-1 \le x \le 2R} g(x) \wedge h(x)$ is positive. Due to (5), $|r(j,\delta)| \le 2R$ and $e^{-\lambda}/\lambda \ge e^{-2\lambda}$,

$$\begin{aligned}
&\mathbb{P}(\|X - f_-\|_\infty \le \varepsilon/2) \\
&\quad \ge \Pi\biggl(k = N-1, r(j,\delta) - \frac{\varepsilon\delta}{4} \le a_j \le r(j,\delta), \\
&\qquad t_j \in \biggl[j\delta, j\delta + \frac{\varepsilon\delta}{4}\biggr], j = 0, \dots, N-1\biggr) \\
&\quad \ge e^{-\lambda}\lambda^{N-1}\biggl(c_0\frac{\varepsilon\delta}{4}\biggr)^N \biggl(\frac{\varepsilon\delta}{4}\biggr)^{N-1} \\
&\quad \ge e^{-2\lambda}(1 \wedge \lambda)^{2(2R/\varepsilon)^{1/\beta}} \biggl(\frac{\sqrt{c_0}}{8}(2R)^{-1/\beta}\varepsilon^{\frac{\beta+1}{\beta}}\biggr)^{2/\delta}.
\end{aligned}$$

Choosing $c = c(\beta, R)$ large enough, the result follows. $\square$

LEMMA A.3. *Consider the randomly initialized CPP (4) and assume that there are constants $\gamma, L > 0$ such that $\mathbb{P}(|\Delta_i| \ge s) \le L^{-1}e^{-Ls^\gamma}$ for all $s \ge 0$. Then for any $M > 0$, any $\varepsilon > 0$, and any $K > 1$ there exists a Borel set Θ and constants C', C'' that only depend on M, L, γ, such that*

$$\mathbb{P}(X \notin \Theta) \le C''K^{-MK} \quad \textit{and} \quad N_{[}(\varepsilon, \Theta, \|\cdot\|_1) \le C''\biggl(\frac{K}{\varepsilon}\biggr)^{C'K}.$$

PROOF. If $N \sim \mathrm{Pois}(\lambda)$, $K \ge 1$ and $M \ge \max(2\lambda e, 1)$, then, using Stirling's formula,

$$\begin{aligned}
\mathbb{P}(N \ge MK) &= e^{-\lambda}\sum_{k=\lceil MK\rceil}^\infty \frac{\lambda^k}{k!} \le \sum_{k=\lceil MK\rceil}^\infty \biggl(\frac{\lambda e}{k}\biggr)^k \\
&\le \sum_{k=\lceil MK\rceil}^\infty \biggl(\frac{1}{2K}\biggr)^k \le K^{-MK}.
\end{aligned} \tag{21}$$

With $t := ((MK+1)L^{-1}\log K)^{1/\gamma}$ and the assumption on the tail behavior of the jump heights, we obtain

$$\begin{aligned}
\mathbb{P}\Bigl(\{N \ge MK\} \cup \Bigl\{\max_{i=0,\dots,N} |\Delta_i| \ge t\Bigr\}\Bigr) &\le \mathbb{P}(N \ge MK) + MK\mathbb{P}(|\Delta_1| \ge t) \\
&\le (1 + M/L)K^{-MK}.
\end{aligned} \tag{22}$$

Define Θ as the space of piecewise constant functions f with $|f(0)| \le t$, maximal jump size bounded by t and less than MK jumps. By the computations above, $\mathbb{P}(X \notin \Theta) \le (1 + M/L)K^{-MK}$.

Next, we compute the bracketing number of Θ with respect to the L^1-norm. Let r_ε be such that $\varepsilon/(4MKt) \le r_\varepsilon \le \varepsilon/(2MKt)$ and $1/r_\varepsilon$ is an integer. Define $x_j := jr_\varepsilon$ for $0 \le j < 1/r_\varepsilon$. In y-direction, consider the grid points $y_\ell := \ell\varepsilon/2$, $\ell = -S_\varepsilon, \dots, S_\varepsilon$ with $S_\varepsilon = \lfloor 2MKt/\varepsilon \rfloor$. Let $\Theta^0 \subset \Theta$ be the space of piecewise constant functions in Θ with all jumps locations on the grid points x_j, and function values in the discrete set $\{y_\ell : \ell = -S_\varepsilon, \dots, S_\varepsilon\}$. We prove that for any function $f \in \Theta$, there exists a function $h \in \Theta^0$ such that $h \le f$ and $\|h - f\|_1 \le \varepsilon$. Consider

$$h = \sum_{j=1}^{1/r_\varepsilon} \max\Big\{ y_\ell : y_\ell \le \min_{x \in [x_{j-1}, x_j]} f(x) \Big\} \mathbf{1}_{[x_{j-1}, x_j)}.$$

Obviously, $h \in \Theta^0$ and $h \le f$. Let us show $\|h - f\|_1 \le \varepsilon$. Observe that $\|h - \widetilde{h}\|_\infty \le \varepsilon/2$ with $\widetilde{h} = \sum_{j=1}^{1/r_\varepsilon} \min_{x \in [x_{j-1}, x_j)} f(x) \mathbf{1}_{[x_{j-1}, x_j)}$. If f jumps k times on the interval $[x_{j-1}, x_j)$ then $\sup_{x \in [x_{j-1}, x_j)} |f(x) - \widetilde{h}(x)| \le kt$. Since the total number of jumps is bounded by MK, $\|f - \widetilde{h}\|_1 \le MKtr_\varepsilon = \varepsilon/2$ implying $\|f - h\|_1 \le \varepsilon$. There are at most $\binom{1/r_\varepsilon}{\ell}(2S_\varepsilon + 1)^{\ell+1}$ functions in Θ^0 with ℓ jumps. The cardinality of Θ^0 is therefore bounded by

$$\begin{aligned} \sum_{\ell=0}^{MK} \binom{1/r_\varepsilon}{\ell} (2S_\varepsilon + 1)^{\ell+1} &\le \sum_{\ell=0}^{MK} r_\varepsilon^{-\ell} (2S_\varepsilon + 1)^{\ell+1} \le 2 r_\varepsilon^{-MK} (2S_\varepsilon + 1)^{MK+1} \\ &\le C'' \left(\frac{K}{\varepsilon}\right)^{C'K} \end{aligned}$$

for suitable constants C' and C''. $\square$

PROOF OF THEOREM 2.1. For all both cases, we apply Lemma A.2 and Lemma A.3 to verify the conditions of Theorem A.1. For (i), we choose $\varepsilon = (\log n/n)^{\beta/(\beta+1)}$ and $K = (n/\log n)^{1/(\beta+1)}$ in Lemma A.3. (ii) can be proved in the same way with $\beta = 1$. $\square$

A.2. Proof of Theorem 2.2.

PROPOSITION A.4. *Consider the randomly initialized subordinator prior. If ν satisfies $\nu(x) \le Cx^{-3/2}$ for all x, then there exists a positive constant $c > 0$ such that*

$$\inf_{f_0 \in \mathcal{M}(R)} P\big(\|X - f_0\|_1 \le 3\varepsilon, X \le f_0\big) \ge \varepsilon^{c\varepsilon^{-1}} \quad \textit{for all } \varepsilon \in (0, 1/2).$$

PROOF. We shall use the following small ball probability of an α-stable subordinator around zero:

$$\lim_{\varepsilon \to 0} \varepsilon^{\alpha/(1-\alpha)} \log\big(P\big(\|X\|_\infty \le \varepsilon\big)\big) \in (-\infty, 0),$$

which follows from Proposition 1 in [35] noting that for nondecreasing functions starting in zero the 1-variation equals the supremum norm. This result shows that the α-stable subordinators satisfy the small ball probability in L^∞ with rate $e^{-c\varepsilon^{-1}}$ if and only if $\alpha \le 1/2$.

Introducing $\nu_>(x) = (\nu(x) \wedge \nu(1))\mathbf{1}(0 \le x \le 1) + \nu(x)\mathbf{1}(x > 1)$ and $\nu_< = \nu - \nu_>$, we can decompose X as $X_0 + X^< + X^>$ with two independent Lévy processes $X^<$, $X^>$ having Lévy densities $\nu_<$, $\nu_>$, respectively. The small jump process $X^<$ is a subordinator whose Lévy density is smaller than $\nu_{1/2}(x) = Cx^{-3/2}\mathbf{1}(x > 0)$, the Lévy density of a stable subordinator $X^{(1/2)}$ of index $\alpha = 1/2$. We can thus couple $X^<$ and $X^{(1/2)}$ such that $X_t^< \le X_t^{(1/2)}$ holds for all $t \ge 0$ a.s. By the above result, this gives

$$\log\big(P\big(\|X^<\|_\infty \le \varepsilon\big)\big) \gtrsim -\varepsilon^{-1}.$$

Because of $\lambda := \int \nu_> \leq \nu(1) + \int_1^\infty \nu < \infty$, the process $X^>$ is a CPP with jump distribution $G = \nu_>/\lambda$. If $f_0 \in \mathcal{M}(R)$ and $\varepsilon \leq R$, then $f_0 - \varepsilon \in \mathcal{M}(2R)$ and by Lemma A.2(i),

$$\inf_{f_0 \in \mathcal{M}(R)} P\big(\|X_0 + X^> - (f_0 - \varepsilon)\|_1 \leq 2\varepsilon, X \leq f_0 - \varepsilon\big) \geq e^{-2\lambda}(1 \wedge \lambda)^{8R/\varepsilon} \varepsilon^{c\varepsilon^{-1}}.$$

By independence, we conclude for $X = X_0 + X^< + X^>$:

$$\begin{aligned}
&\log\big(P\big(\|X - f_0\|_1 \leq 3\varepsilon, X \leq f_0\big)\big) \\
&\quad \geq \log\big(P\big(\|X_0 + X^> - (f_0 - \varepsilon)\|_1 \leq 2\varepsilon, X_0 + X^> \leq f_0 - \varepsilon, X^< \leq \varepsilon\big)\big) \\
&\quad \gtrsim -\varepsilon^{-1}\log(\varepsilon^{-1}) - \varepsilon^{-1}.
\end{aligned}$$

This gives the result. $\square$

LEMMA A.5. *Consider the randomly initialized subordinator prior. Assume that there are constants $\gamma, L > 0$ such that $\nu(x) \leq Lx^{-3/2}$ for all x and $\int_s^\infty \nu(x) + h(x) + h(-x)\,dx \leq L^{-1}e^{-Ls^\gamma}$ for all $s \geq 1$. Then for any $M, A > 0$ there exist Borel sets $(\Theta_n)_n$ and constants C', C'', such that for all sufficiently large n,*

$$\mathbb{P}(X \notin \Theta_n) \leq C'' e^{-M\sqrt{n \log n}} \quad \textit{and} \quad N_{[}\big(A(\sqrt{\log n/n}), \Theta_n, \|\cdot\|_1\big) \leq C'' e^{C'\sqrt{n \log n}}.$$

PROOF. Let $\delta = 1/(2M\sqrt{n \log n})$. We can decompose the subordinator in $X = X_0 + X_< + X_>$, where $X_<$ and $X_>$ are subordinators with Lévy densities $\nu_<(x) = \nu(x)\mathbf{1}(x \leq \delta)$ and $\nu_> = \nu - \nu_<$, respectively. Observe that by the Lévy–Khintchine formula, extended to the moment-generating function,

$$\begin{aligned}
P\big(X_<(1) > 1\big) &\leq \frac{E[e^{\delta^{-1}X_<(1)}]}{e^{\delta^{-1}}} = \exp\left(\int_0^\delta (e^{x/\delta} - 1)\nu(x)\,dx - \frac{1}{\delta}\right) \\
&\leq \exp\left(\int_0^\delta (e - 1)\frac{x}{\delta}\nu(x)\,dx - \frac{1}{\delta}\right) \\
&\leq \exp\left(\frac{2L(e-1)\delta^{1/2} - 1}{\delta}\right) \\
&\leq e^{-M\sqrt{n \log n}}
\end{aligned}$$

for all sufficiently large n. The process $X_>$ is a CPP with intensity $\lambda = \int_\delta^\infty \nu(x) \leq 2L\delta^{-1/2}$ and jump density $\nu_>(x)/\lambda$. If $N \sim \mathrm{Pois}(\lambda)$ denotes the number of jumps of $X_>$ on [0, 1], we find by (21), $P(N \geq \max(2\lambda e, 1)m) \leq m^{-m}$. Let $\Delta_0 := X_0$ and denote the jump heights of the CPP $X_>$ by Δ_i, $i = 1, \ldots$. Let $c_0 := \inf_{x \in [1,2]} \nu(x)$ and observe that $c_0 > 0$ because ν is continuous and positive. Arguing as for (22), with $t := 1 \vee (L^{-1}(m+1)\log m)^{1/\gamma}$,

$$\begin{aligned}
\mathbb{P}\Big(\max_{i=0,\ldots,N} |\Delta_i| \geq t\Big) &\leq \mathbb{P}\big(|\Delta_0| \geq t\big) + m\max(2\lambda e, 1)\frac{\int_t^\infty \nu}{\lambda} + m^{-m} \\
&\leq \left(2 + \frac{m}{L}\max(2e, 1/c_0)\right)e^{-Lt^\gamma} + m^{-m} \\
&\leq \left(\frac{1}{L}\max(2e, 1/c_0) + 3\right)m^{-m}.
\end{aligned}$$

Put $m = 4M\sqrt{n/\log n}$ and define $\Theta_n^>$ as the space of piecewise constant functions f with $|f(0)| \leq t$, less than m jumps, minimal jump size δ and maximal jump size bounded by t. For all sufficiently large n,

$$m^{-m} \leq e^{-2M\sqrt{n/\log n}(\log n - \log\log n)} \leq e^{-M\sqrt{n \log n}}.$$

From the computations above, $P(X_> \notin \Theta_n^>) \le \text{const.} \times e^{-M\sqrt{n\log n}}$. Let $\Theta_{\text{mon},\delta} = \{g : g \text{monotone}, g \le 1 \text{ and all jumps are } \le \delta\}$ and $\Theta_n = \{f = g + h : g \in \Theta_{\text{mon},\delta}, h \in \Theta_n^>\}$ then also $\mathbb{P}(X \notin \Theta_n) \le \text{const.} \times e^{-M\sqrt{n\log n}}$ due to the uniqueness of the decomposition $f = g + h$ in Θ_n.

Notice that

$$N_{[}(\varepsilon, \Theta_n, \|\cdot\|_1) \le N_{[}(\varepsilon/2, \Theta_{\text{mon},0}, \|\cdot\|_1) N_{[}(\varepsilon/2, \Theta_n^>, \|\cdot\|_1).$$

It is well known ([37], 2.7.5 Theorem) that $N_{[}(\varepsilon/2, \Theta_{\text{mon},0}, \|\cdot\|_1) \le e^{K/\varepsilon}$ for some constant K. A bound for the second factor follows from the proof of Lemma A.3 with $K_n = m$. This completes the proof. $\square$

PROOF OF THEOREM 2.2. Using Lemma A.2 with $\varepsilon = \sqrt{\log n/n}$ and A.5 yield the conditions of Theorem A.1 for contraction rate $\sqrt{\log n/n}$. $\square$

APPENDIX B: PROOFS FOR SECTION 4.3

B.1. Proof of Proposition 4.6. The Bayes formula (3) gives for any $m \ge 0$,

$$\Pi(K \ge m \mid N) \le \frac{\int_{K\ge m} e^{-n\int(f_0 - f)_+} \frac{dP_{f\vee f_0}}{dP_{f_0}}(N)\, d\Pi(f)}{e^{-\sqrt{n\log n}}\Pi(X : \|X - f_0\|_1 \le \sqrt{\log n/n}, X \le f_0)}$$

with X a CPP with intensity λ. Bounding $e^{-n\int(f_0-f)_+} \le 1$ and taking expectation with respect to f_0 yields

$$E_{f_0}[\Pi(K \ge m \mid N)] \le \frac{e^{\sqrt{n\log n}}\Pi(K \ge m)}{\Pi(X : \|X - f_0\|_1 \le \sqrt{\log n/n}, X \le f_0)}. \tag{23}$$

If $m \ge 1$, we find by Stirling's approximation $m^m e^{-m} \le \sqrt{2\pi} m^{m+1/2} e^{-m} \le m! \le m^m$ and since K follows under the prior a Poisson distribution with intensity λ,

$$\Pi(K \ge m) \le e^{-\lambda}\frac{\lambda^m}{m!}\sum_{\ell=0}^{\infty}\frac{\lambda^\ell}{\ell!} = \frac{\lambda^m}{m!} \le \lambda^m e^{m - m\log m}$$

as well as $\Pi(K = m) \ge \lambda^m e^{-\lambda - m\log m}$. The latter inequality will be used to derive a lower bound for the denominator. For any $K \ge 1$,

$$\begin{aligned} M_K := \Bigg\{ X = \sum_{k=0}^{K} a_k \mathbf{1}(\cdot \ge t_k) : t_k \in \left[\frac{2k-1}{2K}, \frac{k}{K}\right), \\ f_0(t_{k+1}) - \frac{3}{2K} \le \sum_{\ell=0}^{k} a_\ell \le f_0(t_k) \Bigg\} \\ \subset \left\{ X : \|X - f_0\|_\infty \le \frac{3}{2K}, X \le f_0 \right\}, \end{aligned}$$

where $k = 0, \dots, K$ (except for $t_0 := 0$) and $t_{K+1} := 1$. On M_K, for any $k = 1, \dots, K$,

$$\sum_{\ell=0}^{k-1} a_\ell \le f_0(t_{k-1}) \le \frac{k-1}{K} + \frac{1}{2} \le f_0(t_{k+1}) - \frac{3}{2K} \le \sum_{\ell=0}^{k} a_\ell,$$

and subtracting $\sum_{\ell=0}^{k-1} a_\ell$ on both sides yields $a_k \ge 0$. The difference between the upper bound and the lower bound for $\sum_{\ell=0}^{k} a_\ell$ in the definition of M_K is $f_0(t_k) - f_0(t_{k+1}) + 3/(2K) \le 1/K$. Each of the a_k ranges therefore over an interval of length $\ge 1/K$ in $[0, 1]$. For $K_n :=$

$\lceil \sqrt{n/\log n} \rceil$, this gives with (19) the lower bound,

$$\begin{aligned}
\Pi\Big(X : \|X - f_0\|_1 \le \sqrt{\frac{\log n}{n}}, X \le f_0\Big) \\
&\ge \frac{\Pi(K = K_n)}{(2K_n)^{K_n}} \prod_{k=0}^{K_n} \inf_{\eta_k \in [0, 1-1/K_n]} g_k\Big(\Big[\eta_k, \eta_k + \frac{1}{K_n}\Big]\Big) \\
&\ge \frac{\lambda^{K_n} e^{-\lambda - K_n \log K_n}}{(2K_n)^{K_n}} \Big(\frac{c}{(\gamma+1)K_n^{\gamma+1}}\Big)^{K_n+1},
\end{aligned}$$

where we used that $x \mapsto x^\gamma$ is monotone for the last inequality. Consequently, there exists a constant $C = C(\lambda, c, \gamma)$, such that with (23),

$$E_{f_0}\big[\Pi(K \ge m \mid N)\big] \le e^{\lambda + A\sqrt{n \log n} + m \log \lambda - m \log m + m}.$$

Choosing $m = c^* \sqrt{n/\log n}$ with c^* large enough, the right-hand side converges to zero.

B.2. Proof of Theorem 4.7.

LEMMA B.1. *If $f_0(x) = ax + b$ for $a > 0$, $b \in \mathbb{R}$, then*

$$\inf_{f \in \mathcal{M}(K,\infty)} \int_0^1 \big|f_0(x) - f(x)\big|\, dx \ge \frac{a}{4K}.$$

PROOF. For any real c and $r < s$, we have $\int_r^s |f_0(x) - c|\, dx \ge a(s-r)^2/4$, and hence

$$\begin{aligned}
\inf_{f \in \mathcal{M}(K,\infty)} \int_0^1 \big|f_0(x) - f(x)\big|\, dx &\ge \inf_{0 =: t_0 \le t_1 \le \cdots \le t_K := 1} \sum_{k=1}^K \inf_{c_k \in \mathbb{R}} \int_{t_{k-1}}^{t_k} \big|f_0(x) - c_k\big|\, dx \\
&\ge \frac{a}{4} \inf_{0 =: t_0 \le t_1 \le \cdots \le t_K := 1} \sum_{k=1}^K (t_k - t_{k-1})^2 \ge \frac{a}{4K},
\end{aligned}$$

where we use Jensen's inequality for the last step. □

LEMMA B.2. *For $f_0 = (\frac{1}{2} + \cdot) \wedge \frac{3}{2}$ and any sequence $M_n \to \infty$,*

$$P_{f_0}\Big(\int_0^1 \big(\widehat{f}^{\mathrm{MLE}}(x) - f_0(x)\big)\, dx \ge \frac{M_n}{\sqrt{n}}\Big) \to 0.$$

PROOF. By Markov inequality,

$$P_{f_0}\Big(\int_0^1 \big(\widehat{f}^{\mathrm{MLE}}(x) - f_0(x)\big)\, dx \ge \frac{M_n}{\sqrt{n}}\Big) \le \frac{\sqrt{n}}{M_n} \int_0^1 E_{f_0}\big[\widehat{f}^{\mathrm{MLE}}(x) - f_0(x)\big]\, dx.$$

The proof of Theorem 3.9 in [29], specifically the last equation display of the proof and replacing $[0, 1]$ by $[0, T]$ with $\varepsilon = T - 1$, yields $\int_0^1 E_{f_0}[\widehat{f}^{\mathrm{MLE}}(x) - f_0(x)]\, dx = O(n^{-1/2})$, and thus the result. □

PROOF OF THEOREM 4.7. Lemma B.2 shows that it is enough to prove the existence of a positive constant c', such that

$$\begin{aligned}
E_{f_0}\Big[\Pi\Big(\vartheta \ge \int_0^1 f_0(x)\, dx - \widetilde{c}\sqrt{\frac{\log n}{n}} \;\Big|\; N\Big) \\
\cdot \mathbf{1}\Big(\int_0^1 \big(\widehat{f}^{\mathrm{MLE}}(x) - f_0(x)\big)\, dx \le c'\sqrt{\frac{\log n}{n}}\Big)\Big] \to 0.
\end{aligned} \tag{24}$$

By Proposition 4.6, we know that the posterior concentrates on models with $K_n \le c^* \sqrt{n/\log n}$ for some positive constant c^*. Applying Lemma B.1, this means that the posterior puts asymptotically all mass on paths f with

$$\int_0^1 |f_0(x) - f(x)|\,dx \geq \frac{1}{8c^*}\sqrt{\frac{\log n}{n}}.$$

Since the posterior also puts only mass on functions f with $f \leq \widehat{f}^{\mathrm{MLE}}$, the posterior puts asymptotically all mass on ϑ with

$$\begin{aligned}\vartheta &= \int_0^1 f_0(x)\,dx + \int_0^1 \bigl(f(x) - f_0(x)\bigr)\,dx \\ &\leq \int_0^1 f_0(x)\,dx + 2\int_0^1 \bigl(\widehat{f}^{\mathrm{MLE}}(x) - f_0(x)\bigr)\,dx - \int_0^1 |f(x) - f_0(x)|\,dx \\ &\leq \int_0^1 f_0(x)\,dx + 2\int_0^1 \bigl(\widehat{f}^{\mathrm{MLE}}(x) - f_0(x)\bigr)\,dx - \frac{1}{8c^*}\sqrt{\frac{\log n}{n}}.\end{aligned}$$

Choosing $c' = \frac{1}{32c^*}$ in (24) yields the assertion for $\widetilde{c} = \frac{1}{8c^*} - 2c' = \frac{1}{16c^*}$. $\square$

Acknowledgments. We are very grateful for the comments and remarks made by the Associate Editor and two expert referees which helped to improve the article.

This work was supported by the DFG research unit FOR1735 *Structural Inference in Statistics: Adaptation and Efficiency*.

SUPPLEMENTARY MATERIAL

Supplement to "Nonparametric Bayesian analysis of the compound Poisson prior for support boundary recovery" (DOI: 10.1214/19-AOS1853SUPP; .pdf). The remaining proofs are given in the supplement. The supplement contains also analogous results for random histogram priors.

REFERENCES

[1] BONTEMPS, D. (2011). Bernstein–von Mises theorems for Gaussian regression with increasing number of regressors. *Ann. Statist.* **39** 2557–2584. MR2906878 https://doi.org/10.1214/11-AOS912

[2] CASTILLO, I. and ROUSSEAU, J. (2015). A Bernstein–von Mises theorem for smooth functionals in semiparametric models. *Ann. Statist.* **43** 2353–2383. MR3405597 https://doi.org/10.1214/15-AOS1336

[3] CASTILLO, I., SCHMIDT-HIEBER, J. and VAN DER VAART, A. (2015). Bayesian linear regression with sparse priors. *Ann. Statist.* **43** 1986–2018. MR3375874 https://doi.org/10.1214/15-AOS1334

[4] CASTILLO, I. and VAN DER VAART, A. (2012). Needles and straw in a haystack: Posterior concentration for possibly sparse sequences. *Ann. Statist.* **40** 2069–2101. MR3059077 https://doi.org/10.1214/12-AOS1029

[5] CHATTERJEE, S., GUNTUBOYINA, A. and SEN, B. (2015). On risk bounds in isotonic and other shape restricted regression problems. *Ann. Statist.* **43** 1774–1800. MR3357878 https://doi.org/10.1214/15-AOS1324

[6] CHERNOZHUKOV, V. and HONG, H. (2004). Likelihood estimation and inference in a class of nonregular econometric models. *Econometrica* **72** 1445–1480. MR2077489 https://doi.org/10.1111/j.1468-0262.2004.00540.x

[7] CHIPMAN, H. A., GEORGE, E. I. and MCCULLOCH, R. E. (2010). BART: Bayesian additive regression trees. *Ann. Appl. Stat.* **4** 266–298. MR2758172 https://doi.org/10.1214/09-AOAS285

[8] CORAM, M. and LALLEY, S. P. (2006). Consistency of Bayes estimators of a binary regression function. *Ann. Statist.* **34** 1233–1269. MR2278357 https://doi.org/10.1214/009053606000000236

[9] DENISON, D. G. T., MALLICK, B. K. and SMITH, A. F. M. (1998). A Bayesian CART algorithm. *Biometrika* **85** 363–377. MR1649118 https://doi.org/10.1093/biomet/85.2.363

[10] EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. (2003). *Modelling Extremal Events*: *For Insurance and Finance*. *Applications of Mathematics* (*New York*) **33**. Springer, New York. MR1458613 https://doi.org/10.1007/978-3-642-33483-2

[11] FRICK, K., MUNK, A. and SIELING, H. (2014). Multiscale change point inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 495–580. MR3210728 https://doi.org/10.1111/rssb.12047

[12] GAO, C., HAN, F. and ZHANG, C.-H. (2017). On estimation of isotonic piecewise constant signals. Preprint. Available at arXiv:1705.06386. *Ann. Statist.* (to appear).

[13] GHOSAL, S. (1999). Asymptotic normality of posterior distributions in high-dimensional linear models. *Bernoulli* **5** 315–331. MR1681701 https://doi.org/10.2307/3318438

[14] GHOSAL, S. (2000). Asymptotic normality of posterior distributions for exponential families when the number of parameters tends to infinity. *J. Multivariate Anal.* **74** 49–68. MR1790613 https://doi.org/10.1006/jmva.1999.1874

[15] GHOSAL, S., GHOSH, J. K. and SAMANTA, T. (1995). On convergence of posterior distributions. *Ann. Statist.* **23** 2145–2152. MR1389869 https://doi.org/10.1214/aos/1034713651

[16] GHOSAL, S. and VAN DER VAART, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. *Cambridge Series in Statistical and Probabilistic Mathematics* **44**. Cambridge Univ. Press, Cambridge. MR3587782 https://doi.org/10.1017/9781139029834

[17] GIJBELS, I., MAMMEN, E., PARK, B. U. and SIMAR, L. (1999). On estimation of monotone and concave frontier functions. *J. Amer. Statist. Assoc.* **94** 220–228. MR1689226 https://doi.org/10.2307/2669696

[18] HOLMES, C. C. and HEARD, N. A. (2003). Generalized monotonic regression using random change points. *Stat. Med.* **22** 623–638.

[19] JIRAK, M., MEISTER, A. and REISS, M. (2014). Adaptive function estimation in nonparametric regression with one-sided errors. *Ann. Statist.* **42** 1970–2002. MR3262474 https://doi.org/10.1214/14-AOS1248

[20] KIM, Y. and LEE, J. (2004). A Bernstein–von Mises theorem in the nonparametric right-censoring model. *Ann. Statist.* **32** 1492–1512. MR2089131 https://doi.org/10.1214/009053604000000526

[21] KLEIJN, B. and KNAPIK, B. (2012). Semiparametric posterior limits under local asymptotic exponentiality. Preprint. Available at arXiv:1210.6204.

[22] KOROSTELËV, A. P. and TSYBAKOV, A. B. (1993). *Minimax Theory of Image Reconstruction*. *Lecture Notes in Statistics* **82**. Springer, New York. MR1226450 https://doi.org/10.1007/978-1-4612-2712-0

[23] LI, M. and GHOSAL, S. (2017). Bayesian detection of image boundaries. *Ann. Statist.* **45** 2190–2217. MR3718166 https://doi.org/10.1214/16-AOS1523

[24] MARIUCCI, E., RAY, K. and SZABO, B. (2017). A Bayesian nonparametric approach to log-concave density estimation. Preprint. Available at arXiv:1703.09531.

[25] MEISTER, A. and REISS, M. (2013). Asymptotic equivalence for nonparametric regression with non-regular errors. *Probab. Theory Related Fields* **155** 201–229. MR3010397 https://doi.org/10.1007/s00440-011-0396-x

[26] PANOV, M. and SPOKOINY, V. (2015). Finite sample Bernstein–von Mises theorem for semiparametric problems. *Bayesian Anal.* **10** 665–710. MR3420819 https://doi.org/10.1214/14-BA926

[27] REISS, M. and SCHMIDT-HIEBER, J. (2017). Posterior contraction rates for support boundary recovery. Preprint. Available at arXiv:1703.08358.

[28] REISS, M. and SCHMIDT-HIEBER, J. (2019). Supplement to "Nonparametric Bayesian analysis of the compound Poisson prior for support boundary recovery." https://doi.org/10.1214/19-AOS1853SUPP.

[29] REISS, M. and SELK, L. (2017). Efficient estimation of functionals in nonparametric boundary models. *Bernoulli* **23** 1022–1055. MR3606758 https://doi.org/10.3150/15-BEJ768

[30] RIVOIRARD, V. and ROUSSEAU, J. (2012). Bernstein–von Mises theorem for linear functionals of the density. *Ann. Statist.* **40** 1489–1523. MR3015033 https://doi.org/10.1214/12-AOS1004

[31] ROCKOVA, V. and VAN DER PAS, S. (2017). Posterior concentration for Bayesian regression trees and their ensembles. Preprint. Available at arXiv:1708.08734. *Ann. Statist.* (to appear).

[32] SALOMOND, J.-B. (2014). Concentration rate and consistency of the posterior distribution for selected priors under monotonicity constraints. *Electron. J. Stat.* **8** 1380–1404. MR3263126 https://doi.org/10.1214/14-EJS929

[33] SATO, K. (2013). *Lévy Processes and Infinitely Divisible Distributions*. *Cambridge Studies in Advanced Mathematics* **68**. Cambridge Univ. Press, Cambridge. MR3185174

[34] SCRICCIOLO, C. (2007). On rates of convergence for Bayesian density estimation. *Scand. J. Stat.* **34** 626–642. MR2368802 https://doi.org/10.1111/j.1467-9469.2006.00540.x

[35] SIMON, T. (2004). Small ball estimates in p-variation for stable processes. *J. Theoret. Probab.* **17** 979–1002. MR2105744 https://doi.org/10.1007/s10959-004-0586-x

[36] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

[37] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. *Springer Series in Statistics*. Springer, New York. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2