Bernstein von Mises theorems for statistical inverse problems II
Compound Poisson processes

Nickl, Richard; Söhl, Jakob

DOI
10.1214/19-EJS1609

Publication date
2019

Document Version
Final published version

Published in
Electronic Journal of Statistics

Electronic Journal of Statistics
Vol. 13 (2019) 3513–3571
ISSN: 1935-7524
https://doi.org/10.1214/19-EJS1609

Bernstein–von Mises theorems for statistical inverse problems II: compound Poisson processes*

Richard Nickl

Statistical Laboratory
Department of Pure Mathematics
and Mathematical Statistics
University of Cambridge, CB3 0WB, Cambridge, UK
e-mail: r.nickl@statslab.cam.ac.uk

Jakob Söhl

Delft Institute of Applied Mathematics
Faculty of Electrical Engineering,
Mathematics and Computer Science
TU Delft, Van Mourik Broekmanweg 6,
2628 XE, Delft, The Netherlands
e-mail: j.soehl@tudelft.nl

Abstract: We study nonparametric Bayesian statistical inference for the parameters governing a pure jump process of the form

$$Y_t = \sum_{k=1}^{N(t)} Z_k, \quad t \geq 0,$$

where $N(t)$ is a standard Poisson process of intensity λ, and Z_k are drawn i.i.d. from jump measure μ. A high-dimensional wavelet series prior for the Lévy measure $\nu = \lambda\mu$ is devised and the posterior distribution arises from observing discrete samples $Y_\Delta, Y_{2\Delta}, \dots, Y_{n\Delta}$ at fixed observation distance Δ, giving rise to a nonlinear inverse inference problem. We derive contraction rates in uniform norm for the posterior distribution around the true Lévy density that are optimal up to logarithmic factors over Hölder classes, as sample size n increases. We prove a functional Bernstein–von Mises theorem for the distribution functions of both μ and ν, as well as for the intensity λ, establishing the fact that the posterior distribution is approximated by an infinite-dimensional Gaussian measure whose covariance structure is shown to attain the information lower bound for this inverse problem. As a consequence posterior based inferences, such as nonparametric credible sets, are asymptotically valid and optimal from a frequentist point of view.

Keywords and phrases: Bayesian nonlinear inverse problems, compound Poisson processes, Lévy processes, asymptotics of nonparametric Bayes procedures.

Received May 2019.

*We would like to thank Kolyan Ray for helpful comments leading to improvements of the paper. We are grateful to the associate editor and an anonymous referee for valuable remarks on the manuscript. RN acknowledges support by the European Research Council (ERC) under grant agreement No. 647812.

Contents

1 Introduction . . . 3514
2 Model and prior . . . 3516
 2.1 Basic definitions . . . 3516
 2.2 Likelihood, prior and posterior . . . 3517
3 Main results . . . 3519
 3.1 Supremum norm contraction rates . . . 3519
 3.2 Information geometry of the jump process model . . . 3520
  3.2.1 LAN-expansion of the log-likelihood ratio process . . . 3520
  3.2.2 Derivation of the (right-)inverse of the score operator . 3521
  3.2.3 The adjoint score operator . . . 3522
  3.2.4 Inverse information operator and least favourable directions . . . 3523
  3.2.5 Cramér–Rao information lower bound . . . 3524
 3.3 A multi-scale Bernstein–von Mises theorem . . . 3525
 3.4 Bernstein–von Mises theorem for functionals of the Lévy measure 3526
 3.5 Concluding remarks . . . 3528
4 Proofs of the main theorems . . . 3529
 4.1 Asymptotics for the localised posterior distribution . . . 3529
 4.2 Change of measure in the posterior . . . 3530
 4.3 Proof of Theorem 2 . . . 3532
 4.4 Proof of Theorem 7 . . . 3536
5 Proof of Proposition 10 . . . 3540
6 Proof of Proposition 11 . . . 3547
 6.1 Directional derivatives of the likelihood function . . . 3547
 6.2 Likelihood expansion . . . 3553
 6.3 Simplification of Assumption 25 . . . 3564
7 Proof of Proposition 3 . . . 3566
8 Proof of Proposition 21 . . . 3567
9 Proof of Proposition 23 . . . 3568
10 Proof of Lemma 24 . . . 3568
References . . . 3569

1. Introduction

While the Bayesian approach to inverse problems is widely used in scientific and statistical practice, very little theory is available that explains why Bayesian algorithms should be trusted to provide objective solutions of inverse problems in the presence of statistical noise, particularly in infinite-dimensional, non-linear cases which naturally arise in applications, see [32, 11]. In the recent contributions [28, 24, 21] proof techniques were developed that can be used to derive theoretical guarantees for posterior-based inference, based on suitably chosen priors, in various settings, including inverse problems arising with diffusion processes, X-ray tomography or elliptic partial differential equations. A main idea of [24, 21] is that a careful analysis of the 'Fisher information operator' inducing the statistical observation scheme combined with tools from Bayesian nonparametrics [6, 7] can be used to derive sharp results about the frequentist behaviour of posterior distributions in general inverse problems.

The analysis of the 'information operator' depends highly on the particular problem at hand, and in the present article we continue this line of investigation in a statistical inverse problem very different from the ones considered in [28, 24, 21], namely in the problem of recovering parameters of a stochastic jump process from discrete observations. Statistically speaking, the inverse problem is a 'missing observations' problem that arises from the fact that we do not observe all the jumps and need to 'decompound' the effect of possibly seeing an accumulation of jumps without knowing how many have occurred. This has been studied from a non-Bayesian perspective for certain classes of Lévy processes by several authors, we mention here the seminal papers [3, 2, 37, 22] – see also [1] for various further references – and [26, 33, 27, 10] relevant for the results obtained in the present paper. A typical estimation method used in several of these articles is based on spectral regularisation techniques built around the fact that the Lévy measure identifying all parameters of the jump process can be expressed in the Fourier domain by the Lévy-Khintchine formula (see (3) below).

Given the sophistication of the non-linear estimators proposed so far in the 'decompounding problem' just described, one may wonder if a 'principled' Bayesian approach that just places a standard high-dimensional random series prior on the unknown Lévy measure can at all return valid posterior inferences, for example in the sense of frequentist's coverage of credible sets, in such a measurement scheme. In the present article we provide some answers to this question in the prototypical setting where one observes discrete increments of a compound Poisson processes at fixed observation distance $\Delta > 0$. To lift some of the technicalities occurring in the proofs we restrict ourselves to periodic and hence compactly supported processes, and – to avoid identifiability problems arising in the periodic case – to small enough Δ. We show that the posterior distribution optimally recovers all parameters of the jump process, both in terms of convergence rates for the Lévy density ν and in terms of efficient inference for the intensity of the Poisson process and the distribution function of the jump measure μ. For the latter we obtain functional Bernstein–von Mises theorems which are the Bayesian analogues of the 'Donsker-type' central limit theorems obtained in [26], [10] for frequentist regularisation estimators. Just as in [24], our proofs are inspired by techniques put forward in [6, 7, 4, 8, 5] in 'direct' problems. However, due to the different structure of the jump process model, our proofs need to depart from those in [24] in various ways, perhaps most notably since we have to consider a prior with a larger support ellipsoid, and hence need to prove initial contraction rates for our posterior distribution by quite different methods than is commonly done, see Section 5. The inversion of the information operator in the jump process setting also poses some surprising subtleties that nicely reveal finer properties of the inference problem at hand – our explicit construction of the inverse information operator in Section 3.2 also gives new, more direct proofs of the semi-parametric lower bounds obtained in [33] (whose lower bounds admittedly hold in a more general setting than ours). Finally we should mention that substantial work – using tools from empirical process theory – is required in our setting when linearising the likelihood function to obtain quantitative LAN-expansions since, in contrast to [24], our observation scheme is far from Gaussian. In this sense the techniques we develop here are relevant also beyond compound Poisson processes, although, as argued above, the theory for non-linear inverse problems is largely constrained by any specific case one is studying.

The paper is structured as follows: In Section 2 we give basic definitions and describe the model and prior. In Section 3 we state the contraction rates in supremum norm, the Cramér–Rao lower bound as well as the Bernstein–von Mises theorems in multi-scale spaces and for functionals of the Lévy measure. Section 4 contains the proof of the contraction rates and of the multi-scale Bernstein–von Mises theorem. Sections 5-10 contain the remaining proofs.

2. Model and prior

2.1. Basic definitions

Let $(N(t) : t \geqslant 0)$ be a standard Poisson process of intensity $\lambda > 0$. Let μ be a probability measure on $(-1/2, 1/2]$ such that $\mu(\{0\}) = 0$, and let $Z_1, Z_2, \dots$ be an i.i.d. sequence of random variables drawn from μ. In what follows we view $I = (-1/2, 1/2]$ as a compact group under addition modulo 1. Then the (periodic) compound Poisson process taking values in $(-1/2, 1/2]$ is defined as

$$Y_t = \sum_{k=1}^{N(t)} Z_k, \quad t \geqslant 0, \tag{1}$$

where $Y_0 = 0$ almost surely, by convention. The process $(Y_t : t \geqslant 0)$ is a pure jump Lévy process on $I = (-1/2, 1/2]$ with Lévy measure $d\nu = \lambda d\mu$. We observe this process at fixed observation distance Δ, namely $Y_\Delta, Y_{2\Delta}, \dots, Y_{n\Delta}$, and define the increments of the process

$$X_1 = Y_\Delta, X_2 = Y_{2\Delta} - Y_\Delta, \dots, X_n = Y_{n\Delta} - Y_{(n-1)\Delta}. \tag{2}$$

The X_k's are i.i.d. random variables drawn from the infinitely divisible distribution $\mathbb{P}_\nu = \mathbb{P}_{\nu,\Delta}$ which has characteristic function (Fourier transform)

$$\varphi_\nu(k) = \mathcal{F}\mathbb{P}_\nu(k) = \exp\left(\Delta \int_I (e^{2\pi i k x} - 1) d\nu\right), \ k \in \mathbb{Z}, \tag{3}$$

e.g., by the Lévy–Khintchine formula for Lévy processes in compact groups (Chapter IV.4 in [29]). Obviously $(\varphi_\nu(k) : k \in \mathbb{Z})$ identifies $\mathbb{P}_\nu$ but under the hypotheses we will employ below it will also identify ν and thus the law of the jump process $(Y_t : t \geqslant 0)$. The inverse problem is to recover ν from i.i.d. samples drawn from the probability measure $\mathbb{P}_\nu$.

We denote by $C(I)$ the space of bounded continuous functions on I equipped with the uniform norm $\|\cdot\|_\infty$, and let $M(I) = C(I)^*$ denote the (dual) space of finite signed (Borel) measures on I. For $\kappa_1, \kappa_2 \in M(I)$ their convolution is defined by

$$\kappa_1 * \kappa_2(g) = \int_I \int_I g(x+y) d\kappa_1(x) d\kappa_2(y), \quad g \in C(I),$$

and the last identity holds in fact for arbitrary $g \in L^\infty(I)$ by approximation, see Proposition 8.48 in [14]. This coincides with the usual definition of convolution of functions when the measures involved have densities with respect to the Lebesgue measure. We shall freely use standard properties of convolution integrals, see, e.g., Section 8.2 in [14].

An equivalent representation of $\mathbb{P}_\nu$ is by the infinite convolution series

$$\mathbb{P}_\nu = e^{-\Delta\nu(I)} \sum_{k=0}^{\infty} \frac{\Delta^k \nu^{*k}}{k!} \tag{4}$$

where $\nu^0 = \delta_0, \nu^{*1} = \nu, \nu^{*2} = \nu * \nu$ and ν^{*k} is the $k-1$-fold convolution of ν with itself. [To see this just check the obvious fact that the Fourier transform of the last representation coincides with φ_ν in (3), and use injectivity of the Fourier transform.]

We will denote by $\mathbb{P}_\nu^{\mathbb{N}}$ the infinite product measures describing the laws of infinite sequences of i.i.d. samples (2) arising from a compound Poisson process with Lévy measure ν, and $\mathbb{E}_\nu$ will denote the corresponding expectation operator. We denote by $L^p = L^p(I), 1 \leqslant p < \infty$, the standard spaces of functions f for which $|f|^p$ is Lebesgue-integrable on I, whereas, in slight abuse of notation, for a finite measure κ we will denote by $L^p(\kappa), 1 \leqslant p \leqslant \infty$, the corresponding spaces of κ-integrable functions on I, predominantly for the choices $\kappa = \nu, \kappa = \mathbb{P}_\nu$. The spaces $L^2(I), L^2(\kappa)$ are Hilbert spaces equipped with natural inner products $\langle \cdot, \cdot \rangle, \langle \cdot, \cdot \rangle_{L^2(\kappa)}$, respectively. The symbol $L^\infty(I)$ denotes the usual space of bounded measurable functions on I normed by $\|\cdot\|_\infty$. We also write $\lesssim, \approx$ for (in-)equalities that hold up to fixed multiplicative constants, and employ the usual o_P, O_P-notation to indicate stochastic orders of magnitude of sequences of random variables.

2.2. Likelihood, prior and posterior

We study here the problem of conducting nonparametric Bayesian inference on the parameters ν, μ, λ, assuming a regularity constraint $\nu \in C^s(I), s > 0$, where C^s is the usual Hölder space over I normed by $\|\cdot\|_{C^s}$ (when $s \in \mathbb{N}$ these are the ordinary spaces of s-times continuously differentiable functions, e.g., Section 2.2.2 in [34]). To define the likelihood function we need a common dominating measure for the statistical model $(\mathbb{P}_\nu : \nu \in \mathcal{V})$ where $\mathcal{V}$ is some family of Lévy measures possessing densities with respect to Lebesgue measure Λ with density $\Lambda = 1_{(-1/2,1/2]}$. Since Λ is idempotent – $\Lambda * \Lambda = \int_I \Lambda(\cdot - y)\Lambda(y)dy = \Lambda$ –

we can consider the resulting compound Poisson measure $\mathbb{P}_\Lambda = e^{-\Delta}\delta_0 + (1 - e^{-\Delta})\Lambda$ as a fixed reference measure on I. Then for any absolutely continuous ν on I the densities p_ν of $\mathbb{P}_\nu$ with respect to $\mathbb{P}_\Lambda$ exist. The likelihood function of the observations $X_1, \dots, X_n$ is defined as

$$L_n(\nu) = \prod_{i=1}^n p_\nu(X_i), \quad \nu \in \mathcal{V}. \tag{5}$$

We also write $\ell_n(\nu) = \log L_n(\nu)$ for the log-likelihood function. Next, if Π is a prior distribution on a σ-field $\mathcal{S}_\mathcal{V}$ of $\mathcal{V}$ such that the map $(\nu, x) \mapsto p_\nu(x)$ is jointly measurable, then standard arguments imply that the resulting posterior distribution given observations $X_1, \dots, X_n$ is

$$\Pi(B|X_1, \dots, X_n) = \frac{\int_B L_n(\nu) d\Pi(\nu)}{\int_\mathcal{V} L_n(\nu) d\Pi(\nu)}. \tag{6}$$

We shall model an s-regular function by a high-dimensional product prior expressed through a wavelet basis: Let

$$\left\{\psi_{lk} : k = 0, \dots, (2^l \vee 1) - 1,\ l = -1, \dots, J-1\right\}, J \in \mathbb{N}, \tag{7}$$

form a periodised Daubechies' type wavelet basis of $L^2 = L^2(I)$, orthogonal for the usual L^2-inner product $\langle \cdot, \cdot \rangle$ (described in Section 4.3.4 in [18]; where the constant 'scaling function' is written as the first element $\psi_{-1,0} \equiv 1$, in slight abuse of notation). Basic localisation and approximation properties of this basis are, for any $g \in C^s(I)$ and $j \in \mathbb{N}$,

$$\begin{aligned} &\sup_{x \in I} \sum_k |\psi_{jk}(x)| \lesssim 2^{j/2}, \quad |\langle g, \psi_{jk} \rangle| \lesssim \|g\|_{C^s} 2^{-j(s+1/2)}, \\ &\|P_{V_j}(g) - g\|_{L^2(I)} \lesssim \|g\|_{C^s} 2^{-js}, \end{aligned} \tag{8}$$

where P_{V_j} is the usual L^2-projector onto the linear span V_j of the ψ_{lk}'s with $l \leqslant j-1$.

Now consider the random function

$$v = \sum_{l \leqslant J-1} \sum_k a_l u_{lk} \psi_{lk}(\cdot),\ a_l = 2^{-l}(l^2+1)^{-1}, \quad J \in \mathbb{N}, \tag{9}$$

where u_{lk} are i.i.d. uniform $U(-B, B)$ random variables, and B is a fixed constant. The support of this prior is isomorphic to the hyper-ellipsoid

$$V_{B,J} := \prod_{l=-1}^{J-1} (-Ba_l, Ba_l)^{2^l \vee 1} \subseteq \mathbb{R}^{2^J}$$

of wavelet coefficients. To model an s-regular Lévy measure ν we define the random function

$$\nu = e^v, \quad \Pi = \Pi_J = \text{ the law } \mathcal{L}(\nu) \text{ of } \nu \text{ in } V_{B,J} \tag{10}$$

and shall choose $J = J_n$ such that 2^J grows as a function of n approximately as

$$2^J \approx n^{\frac{1}{2s+1}}. \tag{11}$$

We note that the weights $a_l = 2^{-l}(l^2+1)^{-1}$ ensure that the random function v has some minimal regularity, in particular is contained in a bounded subset of $C(I)$.

Throughout we shall work under the following assumption on the Lévy measure and on the prior identifying the law of the compound Poisson process generating the data.

Assumption 1. *Assume the true Lévy measure ν_0 has a Lebesgue density, still denoted by ν_0, which is contained in $C^s(I)$ for some $s > 5/2$, that ν_0 is bounded away from zero on I, and that for $v_0 = \log \nu_0$ and some $\gamma > 0$,*

$$|\langle v_0, \psi_{lk} \rangle| \leqslant (B-\gamma)a_l \quad \forall l, k, \tag{12}$$

where a_l was defined in (9). Assume moreover that B, Δ are such that $\lambda = \int_I \nu < \pi/\Delta$ for all ν in the support of the prior.

The assumption $s > 5/2$ (in place of, say, $s > 1/2$) may be an artefact of our proof methods (which localise the likelihood function by an initially suboptimal contraction rate) but, in absence of a general 'Hellinger-distance' testing theory (cf. Appendix D in [16] or Section 7.1 in [18]) for the inverse problem considered here, appears unavoidable.

The assumption (12) with $\gamma > 0$ guarantees that the true Lévy density is an 'interior' point of the parameter space $V_{B,J}$ for all J – a standard requirement if one wishes to obtain Gaussian asymptotics for posterior distributions.

Finally, the bound on λ ensures identifiability of ν, and thus of the law of the compound Poisson process, from the measure $\mathbb{P}_\nu$ generating the observations. That such an upper bound is necessary is a consequence of the fact that we are considering the periodic setting, see the discussion after Assumption 19 below. For the present parameter space $V_{B,J}$, Assumption 1 enforces a fixed upper bound on Δ – alternatively for a given value of Δ we could also renormalise ν by a large enough constant to make the intensities λ small enough, but we avoid this for conciseness of exposition.

3. Main results

3.1. Supremum norm contraction rates

Even though the standard 'Hellinger-distance' testing theory to obtain contraction rates is not directly viable in our setting, following ideas in [4] we can use the Bernstein–von Mises techniques underlying the main theorems of this paper to obtain (near-) optimal contraction rates for the Lévy density ν_0 in supremum norm loss. The idea is basically to represent the norm by a maximum over suitable collections of linear functionals, and to then treat each functional individually by semi-parametric methods. It can be shown that the minimax rate of estimation for Lévy densities in $C^s(I)$ with respect to the supremum loss is $(\log n/n)^{s/(2s+1)}$, see [9] for a discussion. The following theorem achieves this rate up to the power of the log-factor.

Theorem 2. *Suppose that $X_1,\dots,X_n$ are generated from (2) and grant Assumption 1. Let $\Pi(\cdot|X_1,\dots,X_n)$ be the posterior distribution arising from prior $\Pi = \Pi_J$ in (10) with J as in (11). Then for every $\kappa > 3$ we have as $n \to \infty$ that*

$$\Pi\Big(\nu : \|\nu-\nu_0\|_\infty > n^{-s/(2s+1)}\log^\kappa n | X_1,\dots,X_n\Big) \to^{\mathbb{P}^{\mathbb{N}}_{\nu_0}} 0.$$

Unlike in the standard i.i.d. setting in [4], we cannot rely on an initial optimal contraction rate in Hellinger distance for ν, which introduces new difficulties when dealing with 'semi-parametric bias terms'. Our proofs (via Lemma 14 below) overcome these problems at the expense of an additional $\log^\kappa n$-factor.

The only comparable posterior contraction rate result of this kind we are aware of in the literature can be found in [20], who obtain contraction rates for the Hellinger distance $h(\mathbb{P}_\nu, \mathbb{P}_{\nu_0})$ between the infinitely divisible distributions $\mathbb{P}_\nu, \mathbb{P}_{\nu_0}$ induced by the Lévy measures ν, ν_0. Without any sharp 'stability estimates' that would allow to derive optimal bounds on the distance $\|\nu-\nu_0\|_\infty$, or even just on $\|\nu-\nu_0\|_{L^2}$, in terms of $h(\mathbb{P}_\nu, \mathbb{P}_{\nu_0})$, the results in [20] do a fortiori not imply any guarantees for Bayesian inference on the statistically relevant parameters ν, μ, λ.

The above contraction rate result shows that the Bayesian method works in principle and that estimators that converge with the minimax optimal rate up to log-factors can be derived from the posterior distribution, see [15].

3.2. Information geometry of the jump process model

3.2.1. LAN-expansion of the log-likelihood ratio process

In order to formulate, and prove, Bernstein–von Mises type theorems, and to derive a notion of semi-parametric optimality of the limit distributions that will occur, we now obtain, for L_n the likelihood function defined in (5), the LAN-expansion of the log-likelihood ratio process

$$\ell_n(\nu_{h,n}) - \ell_n(\nu) = \log\frac{L_n(\nu_{h,n})}{L_n(\nu)}, \quad n\in\mathbb{N},$$

of the observation scheme considered here, in perturbation directions $\nu_{h,n}$ that are additive on the log-scale. This will induce the score operator for the model and allow us to derive the inverse Fisher information (Cramér–Rao lower bound) for a large class of semi-parametric subproblems. Some ideas of what follows are implicit in the work by Trabs (2015), although we need a finer analysis for our results, including inversion of the score operator itself.

Proposition 3 (LAN expansion)**.** *Let $\nu = e^v$ be a Lévy density that is bounded and bounded away from zero, and for $h \in L^\infty(I)$ consider a perturbation $\nu_{h,n} =$*

$e^{v+h/\sqrt{n}}$. *Then if* $X_i \sim^{i.i.d.} \mathbb{P}_\nu$ *we have*

$$\ell_n(\nu_{h,n}) - \ell_n(\nu) = \frac{1}{\sqrt{n}} \sum_{i=1}^n A_\nu(h)(X_i) - \frac{1}{2}\|A_\nu(h)\|^2_{L^2(\mathbb{P}_\nu)} + o_{\mathbb{P}_\nu^{\mathbb{N}}}(1), \tag{13}$$

where the score operator is given by the Radon–Nikodym density

$$A_\nu(h) \equiv \Delta \frac{d(h\nu - \int_I h d\nu \cdot \delta_0) * \mathbb{P}_\nu}{d\mathbb{P}_\nu}. \tag{14}$$

The operator A_ν *defines a continuous linear map from* $L^2(\nu)$ *into* $L^2_0(\mathbb{P}_\nu) := \{g \in L^2(\mathbb{P}_\nu) : \int_I g d\mathbb{P}_\nu = 0\}$.

The proposition is proved in Section 7.

In the remainder of this section we study properties of A_ν and of its adjoint A^*_ν, in particular we construct certain inverse mappings. Due to the presence of the Dirac measure in (14) some care has to be exercised when identifying the natural domain of the inverse of the 'information' operator $A^*_\nu A_\nu$. In particular we can invert $A^*_\nu A_\nu$ only along directions ψ for which $\psi(0) = 0$. An intuitive explanation is that the axiomatic property $\nu(\{0\}) = 0$ is required for ν to identify the law of the compound Poisson process (otherwise 'no jumps' and 'jumps of size zero' are indistinguishable), and as a consequence when making inference on the functional $\int_I \psi d\nu$ one should a priori restrict to $\int_I \psi 1_{\{0\}^c} d\nu$, a fact that features in the Cramér–Rao information lower bound (25) to be established below.

3.2.2. Derivation of the (right-)inverse of the score operator

To proceed we will set $\Delta = 1$ without loss of generality for the moment. If $\kappa \in M(I)$ is a finite signed measure on I and $g : I \to \mathbb{R}$ a function such that $\int_I |g| d|\kappa| < \infty$, we use the notation $g\kappa$ for the element of $M(I)$ given by $(g\kappa)(A) = \int_A g d\kappa$, A a Borel subset of I. Then, for a fixed Lévy density $\nu \in L^\infty(I)$, consider the operator

$$h \mapsto A_\nu(h) := \frac{d[(\nu h) * \mathbb{P}_\nu]}{d\mathbb{P}_\nu}(x) - \left[\int_I d(\nu h)\right], \quad x \in I, \tag{15}$$

defined on the subset of $M(I)$ given by

$$\mathcal{D} \equiv \{\kappa = \kappa_a + c\delta_0,\ \kappa_a \in M(I) \text{ has Lebesgue-density } h_a \in L^2(\nu); c \in \mathbb{R}\}.$$

This operator serves as an extension of A_ν from (14) to the larger domain $\mathcal{D}$. It still takes values in $L^2_0(\mathbb{P}_\nu)$; in fact δ_0 is in the kernel of A_ν since

$$A_\nu(\delta_0) = \frac{\nu(0) d\mathbb{P}_\nu}{d\mathbb{P}_\nu} - \int_I \nu(x) d\delta_0(x) = \nu(0) - \nu(0) = 0, \tag{16}$$

but extending A_ν formally to $\mathcal{D}$ is convenient since the inverse of A_ν to be constructed next will take values in $\mathcal{D}$. Define

$$\pi_\nu = e^{\nu(I)} \sum_{m=0}^{\infty} \frac{(-1)^m \nu^{*m}}{m!}, \tag{17}$$

a finite signed measure for which $\mathbb{P}_\nu * \pi_\nu = \delta_0$ (by checking Fourier transforms). Formally, up to a constant, π_ν equals the inverse Fourier transform $\mathcal{F}^{-1}(1/\varphi_\nu)$ of $1/\varphi_\nu$, and convolution with π_ν can be thought of as a 'deconvolution operation'.

Lemma 4. *Assume the Lévy density $\nu \in L^\infty(I)$ is bounded away from zero on I. The operator $A_\nu : \mathcal{D} \to L^2_0(\mathbb{P}_\nu)$ from (15) has inverse*

$$\widetilde{A}_\nu : L^2_0(\mathbb{P}_\nu) \to \mathcal{D}, \quad \widetilde{A}_\nu(g) := \frac{1}{\nu(\cdot)} \pi_\nu * (g\mathbb{P}_\nu)(\cdot), \tag{18}$$

in the sense that $A_\nu \widetilde{A}_\nu = Id$ on $L^2_0(\mathbb{P}_\nu)$.

Proof. For any $g \in L^2_0(\mathbb{P}_\nu)$, by the Cauchy–Schwarz inequality, $g\mathbb{P}_\nu$ defines a finite signed measure, so that $\widetilde{A}_\nu$ is well-defined and takes values in $M(I)$. Since $\mathbb{P}_\nu * \pi_\nu = \delta_0$ the Radon–Nikodym theorem (Theorem 5.5.4 in [12]) implies

$$\frac{d\left[\mathbb{P}_\nu * \pi_\nu * (g\mathbb{P}_\nu)\right]}{d\mathbb{P}_\nu} = \frac{d(g\mathbb{P}_\nu)}{d\mathbb{P}_\nu} = g, \qquad \mathbb{P}_\nu \text{ a.s..}$$

We then have

$$A_\nu(\widetilde{A}_\nu(g)) = \frac{d\left[\mathbb{P}_\nu * \pi_\nu * (g\mathbb{P}_\nu)\right]}{d\mathbb{P}_\nu} - \int_I d[\pi_\nu * (g\mathbb{P}_\nu)] = g, \tag{19}$$

where the second term vanishes since for such g, by the definition of convolution,

$$\int_I d[\pi_\nu * (g\mathbb{P}_\nu)] = \int_I g d\mathbb{P}_\nu \int_I d\pi_\nu = 0.$$

That $\widetilde{A}_\nu$ takes values in $\mathcal{D}$ is immediate from the definition of π_ν and (4). □

3.2.3. The adjoint score operator

We now calculate the adjoint operator of A_ν.

Lemma 5. *Assume the Lévy density $\nu \in L^\infty(I)$ is bounded away from zero on I. If we regard A_ν from (14) as an operator mapping the Hilbert spaces $L^2(\nu)$ into $L^2_0(\mathbb{P}_\nu)$ then its adjoint $A^*_\nu : L^2_0(\mathbb{P}_\nu) \to L^2(\nu)$ is given by $A^*_\nu(w) = \Delta\mathbb{P}_\nu(-\cdot) * w$.*

Proof. We set without loss of generality $\Delta = 1$. Let $h \in L^2(\nu)$ and $w \in C(I) \subseteq L^2(\mathbb{P}_\nu)$ such that $\int w d\mathbb{P}_\nu = 0$. Then by Fubini's theorem

$$\langle A_\nu(h), w\rangle_{L^2(\mathbb{P}_\nu)} = \int_I A_\nu(h) w d\mathbb{P}_\nu = \int_I w d(\mathbb{P}_\nu * (h\nu)) - \int h\nu \int w d\mathbb{P}_\nu$$

$$= \int_I \int_I w(x+y)h(x)\nu(x)dxd\mathbb{P}_\nu(y) = \int_I h(\mathbb{P}_\nu(-\cdot) * w)d\nu = \langle h, A_\nu^*(w)\rangle_{L^2(\nu)}$$

so that the formula for the adjoint holds on the dense subspace $C(I)$ of $L_0^2(\mathbb{P}_\nu)$. The Cauchy–Schwarz inequality implies that $\mathbb{P}_\nu(-\cdot)*w \in L^2(\nu)$ so that the case of general $w \in L_0^2(\mathbb{P}_\nu)$ follows from standard approximation arguments. □

Inspecting the formula for A_ν^* we can formally define the 'inverse' map

$$(A_\nu^*)^{-1}(g) = \pi_\nu(-\cdot) * g \text{ with } (\pi_\nu(-\cdot) * g)(x) = \int_I g(x+y)d\pi_\nu(y),\ g \in L^2(\mathbb{P}_\Lambda),$$

for $\nu \in L^\infty(I)$ and scaled by $1/\Delta$ if $\Delta \neq 1$. If $g \in L^\infty(I)$ satisfies $g(0) = 0$ then using $\mathbb{P}_\nu * \pi_\nu = \delta_0$ (cf. after (17)) we have that $(A_\nu^*)^{-1}(g) \in L_0^2(\mathbb{P}_\nu)$ since

$$\int_I (A_\nu^*)^{-1}(g)d\mathbb{P}_\nu = \int_I \pi_\nu(-\cdot) * g\ d\mathbb{P}_\nu = \int_I g\ d(\mathbb{P}_\nu * \pi_\nu) = g(0) = 0. \tag{20}$$

3.2.4. Inverse information operator and least favourable directions

Now let $\psi \in L^\infty(I)$ be arbitrary but such that $\psi(0) = 0$, for instance we can take $\psi 1_{\{0\}^c}$ for any $\psi \in C(I)$. If $\nu \in L^\infty(I)$ is bounded away from zero then $\psi/\nu \in L^2(\mathbb{P}_\Lambda)$ and by what precedes $(A_\nu^*)^{-1}(\psi/\nu) \in L_0^2(\mathbb{P}_\nu)$ and hence in view of Lemma 4 we can define, for any such ψ, the new function

$$\widetilde{\psi}_d = -\widetilde{A}_\nu\left[(A_\nu^*)^{-1}\left(\frac{\psi}{\nu}\right)\right] \tag{21}$$

as an element of $\mathcal{D}$. Concretely, in view of (4), (17), (when $\Delta = 1$, otherwise divide the right hand side in the following expression by Δ^2)

$$\widetilde{\psi}_d = -\widetilde{A}_\nu\left[\pi_\nu(-\cdot) * \frac{\psi}{\nu}\right] = -\frac{1}{\nu}\,\pi_\nu * \left(\left(\pi_\nu(-\cdot) * \frac{\psi}{\nu}\right)\mathbb{P}_\nu\right)(\cdot). \tag{22}$$

We can then write $\widetilde{\psi}_d = \widetilde{\psi} + c\delta_0$ where

$$\widetilde{\psi} = \widetilde{\psi}_d - c\delta_0 \tag{23}$$

is the part of $\widetilde{\psi}_d$ that is absolutely continuous with respect to Lebesgue measure Λ, and $c\delta_0$ is the discrete part (for some constant c).

The content of the next lemma is that $\widetilde{\psi}$ allows to represent the LAN inner product

$$\langle f, g\rangle_{LAN} \equiv \langle A_\nu(f), A_\nu(g)\rangle_{L^2(\mathbb{P}_\nu)}, \quad f, g \in L^2(\nu), \tag{24}$$

in the standard L^2-inner product $\langle\cdot,\cdot\rangle$ of $L^2(I)$.

Lemma 6. *Assume the Lévy density $\nu \in L^\infty(I)$ is bounded away from zero on I. If $\psi \in L^\infty(I)$ satisfies $\psi(0) = 0$ then for all $h \in L^2(\nu)$ and $\widetilde{\psi}_d, \widetilde{\psi}$ given as in (22), (23),*

$$\int_I A_\nu(h)A_\nu(\widetilde{\psi})d\mathbb{P}_\nu = \int_I A_\nu(h)A_\nu(\widetilde{\psi}_d)d\mathbb{P}_\nu = -\langle h, \psi\rangle.$$

Proof. From (16) and (23) we have $A_\nu(\widetilde{\psi}_d - \widetilde{\psi}) = 0$, so the first identity is immediate. By Lemma 4 and the definition of $\widetilde{\psi}_d$ we see $A_\nu(\widetilde{\psi}_d) = -\pi_\nu(-\cdot) * (\psi/\nu)$ in $L^2_0(\mathbb{P}_\nu)$ and from Lemma 5 we hence deduce

$$\int_I A_\nu(h)A_\nu(\widetilde{\psi}_d)d\mathbb{P}_\nu = -\int_I h[\mathbb{P}_\nu(-\cdot) * \pi_\nu(-\cdot) * (\psi/\nu)]\nu = -\int_I h\psi,$$

using also that $\mathbb{P}_\nu(-\cdot) * \pi_\nu(-\cdot) = \delta_0$ (cf. after (17)). □

3.2.5. *Cramér–Rao information lower bound*

Using the LAN expansion and the previous lemma we derive the Cramér–Rao lower bound for $1/\sqrt{n}$-consistently estimable functional parameters of the Lévy measure of a compound Poisson process, following the theory laid out in Chapter 25 in [35]. We recall some standard facts from efficient estimation in Banach spaces: assume for all h in some linear subspace H of a Hilbert space with Hilbert norm $\|\cdot\|_{LAN}$ that the LAN expansion

$$\log \frac{d\mathbb{P}^n_{v+h/\sqrt{n}}}{d\mathbb{P}^n_v} = \Delta_n(h) - \frac{1}{2}\|h\|^2_{LAN}, \quad v \in H,$$

holds, where $\mathbb{P}^n_v$ are laws on some measurable space $\mathcal{X}_n$ and where $\Delta_n(h) \to^d \Delta(h)$ as $n \to \infty$ with $\Delta(h) \sim N(0, \|h\|^2_{LAN}), h \in H$. Consider a map

$$K : (H, \|\cdot\|_{LAN}) \to \mathbb{R}$$

that is suitably differentiable with continuous linear derivative map $\kappa : H \to \mathbb{R}$. By Theorem 3.11.5 in [36] the Cramér–Rao information lower bound for estimating the parameter $K(\nu)$ is given by $\|\kappa^*\|^2_{LAN}$ where κ^* is the Riesz-representer of the map $\kappa : (H, \|\cdot\|_{LAN}) \to \mathbb{R}$.

We now apply this in the setting of the LAN expansion obtained from Proposition 3, with laws $\mathbb{P}^n_v$ parametrised by $v = \log\nu$, tangent space $H = L^\infty$ and LAN-norm $\|h\|_{LAN} = \|A_{\nu_0}h\|_{L^2(\mathbb{P}_{\nu_0})}$, where $A_{\nu_0} : (H, \|\cdot\|_{L^2(\nu_0)}) \to L^2_0(\mathbb{P}_{\nu_0})$ is the score operator studied above corresponding to the true absolutely continuous Lévy density ν_0 generating the data (note that the central limit theorem ensures $\Delta_n(h) \to^d \Delta(h)$ for these choices). For $\psi \in L^\infty(I)$ we consider the map

$$K : v \mapsto \int_I \psi\nu = \int_I \psi e^v,$$

which can be linearised at ν_0 with derivative

$$\kappa : h \mapsto \int_I \psi h\nu_0 = \langle\psi_{(0)}, h\rangle_{L^2(\nu_0)} = \int_I \psi 1_{\{0\}^c}\nu_0 h,$$

where by definition $\psi_{(0)} = \psi 1_{\{0\}^c}$. Using Lemma 6 we have

$$\kappa(h) = \langle\psi_{(0)}\nu_0, h\rangle = -\langle\widetilde{(\psi_{(0)}\nu_0)}_d, h\rangle_{LAN} \equiv \langle\kappa^*, h\rangle_{LAN}.$$

We conclude that the Cramér–Rao information lower bound for estimating $\int_I \psi\nu_0$ from discretely observed increments of the compound Poisson process equals

$$\begin{aligned}\|\kappa^*\|^2_{LAN} &= \|A_{\nu_0}(\widetilde{(\psi_{(0)}\nu_0)_d})\|^2_{L^2(\mathbb{P}_{\nu_0})} = \|(A^*_{\nu_0})^{-1}[\psi_{(0)}]\|^2_{L^2(\mathbb{P}_{\nu_0})}\\ &= \|\pi_\nu(-\cdot) * (\psi 1_{\{0\}^c})\|^2_{L^2(\mathbb{P}_{\nu_0})},\end{aligned} \tag{25}$$

where we used Lemma 4 in the second equality. Note that the last identity holds under the notational assumption $\Delta = 1$ employed in the preceding arguments and the far right hand side needs to be scaled by $1/\Delta^2$ when $\Delta \neq 1$.

3.3. A multi-scale Bernstein–von Mises theorem

We now formulate a Bernstein–von Mises theorem that entails a Gaussian approximation of the posterior distribution arising from prior (10) in an infinite-dimensional multi-scale space. We will show in the next subsection how one can deduce from it various Bernstein–von Mises theorems for statistically relevant aspects of ν, μ, λ. Following [7] (see also p.596f. in [18]) the idea is to study the asymptotics of the measure induced in sequence space by the action $(\langle \nu, \psi_{lk}\rangle)$ of draws $\nu \sim \Pi(\cdot|X_1, \dots, X_n)$ of the conditional posterior distribution on the wavelet basis $\{\psi_{lk}\}$ from (7). In sequence space we introduce weighted supremum norms

$$\|x\|_{\mathcal{M}(w)} = \sup_l \frac{\max_k |x_{lk}|}{w_l}, \quad \mathcal{M}(w) = \{(x_{lk}) : \|x\|_{\mathcal{M}(w)} < \infty\}, \tag{26}$$

with monotone increasing weighting sequence (w_l) to be chosen. Define further the closed separable subspace $\mathcal{M}_0(w)$ of $\mathcal{M}(w)$ consisting of sequences for which $w_l^{-1} \max_k |x_{lk}|$ converges to zero as $l \to \infty$, equipped with the same norm.

The Bernstein–von Mises theorem will be derived for the case where the posterior distribution is centred at the random element $\widehat{\nu}(J) = (\widehat{\nu}(J)_{l,k})$ of $\mathcal{M}_0(w)$ defined as follows

$$\widehat{\nu}(J)_{l,k} \equiv \int_I \psi_{lk}\nu_0 + \frac{1}{n}\sum_{i=1}^n (A^*_{\nu_0})^{-1}[\psi_{lk} 1_{\{0\}^c}](X_i), \qquad l \leqslant J-1, k, \tag{27}$$

with the convention that $\widehat{\nu}(J)_{l,k} = 0$ whenever $l \geqslant J$ (the operator $(A^*_{\nu_0})^{-1}$ was defined just after Lemma 5 above). A standard application of the central limit theorem and of (20) implies as $n \to \infty$ and under $\mathbb{P}^{\mathbb{N}}_{\nu_0}$ that, for every fixed k, l,

$$\sqrt{n}\big(\widehat{\nu}(J)_{l,k} - \int_I \psi_{lk}\nu_0\big) \to^d N(0, \|(A^*_{\nu_0})^{-1}[\psi_{lk} 1_{\{0\}^c}]\|^2_{L^2(\mathbb{P}_{\nu_0})}),$$

and hence in view of (25) the random variable $\widehat{\nu}(J)$ is a natural centring for a Bernstein–von Mises theorem. Since $\nu \in L^\infty(I)$ the law of $\sqrt{n}(\nu - \widehat{\nu}(J))$ defines a probability measure in the space $\mathcal{M}_0(\omega)$ for ω as in the next theorem. Next, denote by $\mathcal{N}_{\nu_0}$ the law $\mathcal{L}(\mathbb{X})$ of the centred Gaussian random variable $\mathbb{X}$ on $\mathcal{M}(w)$ whose coordinate process has covariances

$$E\mathbb{X}_{l,k}\mathbb{X}_{l',k'} = \langle (A^*_{\nu_0})^{-1}(\psi_{lk}1_{\{0\}^c}), (A^*_{\nu_0})^{-1}(\psi_{l'k'}1_{\{0\}^c})\rangle_{L^2(\mathbb{P}_{\nu_0})}.$$

The proof of the following theorem implies in particular that $\mathcal{N}_{\nu_0}$ is a tight Gaussian probability measure concentrated on the space $\mathcal{M}_0(w)$ where weak convergence occurs. Recall (Theorem 11.3.3 in [12]) that weak convergence of a sequence of probability measures on a separable metric space (S, d) can be metrised by the bounded Lipschitz (BL) metric

$$\beta_S(\kappa, \kappa') = \sup_{F:S\to\mathbb{R}, \|F\|_{Lip}\leqslant 1} \left| \int_S F(s) d(\kappa - \kappa')(s) \right|,$$
$$\|F\|_{Lip} = \sup_{s\in S} |F(s)| + \sup_{s\neq t, s,t\in S} \frac{|F(s) - F(t)|}{d(s,t)}.$$

Theorem 7. *Suppose that $X_1, \ldots, X_n$ are generated from (2) and grant Assumption 1. Let $\Pi(\cdot|X_1, \ldots, X_n)$ be the posterior distribution arising from prior $\Pi = \Pi_J$ in (10) with J as in (11). Let $\beta_{\mathcal{M}_0(\omega)}$ be the BL metric for weak convergence of laws in $\mathcal{M}_0(\omega)$, with $\omega = (\omega_l)$ satisfying $\omega_l/l^4 \uparrow \infty$ as $l \to \infty$. Let $\widehat{\nu}_J$ be the random variable in $\mathcal{M}_0(\omega)$ given by (27). Then for $\nu \sim \Pi(\cdot|X_1, \ldots, X_n)$ and $\mathcal{N}_{\nu_0}$ as above we have in $\mathbb{P}^{\mathbb{N}}_{\nu_0}$-probability, as $n \to \infty$,*

$$\beta_{\mathcal{M}_0(\omega)}\left(\mathcal{L}\left(\sqrt{n}(\nu - \widehat{\nu}(J))|X_1, \ldots, X_n\right), \mathcal{N}_{\nu_0}\right) \to 0.$$

Theorem 7 is proved in Section 4.4 and has various implications for posterior-based inference on the parameter ν. Arguing as in [7], Section 4.2, we could construct credible bands for the unknown Lévy density ν with L^∞-diameter shrinking at the rate as in Theorem 2 from Bayesian multi-scale credible bands. We will leave this application to the reader and instead focus on inference on functionals of the Lévy measure ν that are continuous, or differentiable, for $\|\cdot\|_{\mathcal{M}(\omega)}$ (see Section 4.1 in [7], [5]).

Theorem 7 assumes a certain growth at infinity of the weight sequence ω_l. The requirement $\omega_l/\sqrt{l} \uparrow \infty$ is necessary for the limit process to be a tight Gaussian Borel probability measure in the space $\mathcal{M}_0(\omega)$, see [7]. Similar to the presence of an additional log-factor in Theorem 2, here we need to impose the slightly more restrictive condition $\omega_l/l^4 \uparrow \infty$ for the control of semi-parametric bias terms in our proofs.

3.4. Bernstein–von Mises theorem for functionals of the Lévy measure

We now deduce from Theorem 7 Bernstein–von Mises theorems for the functionals

$$V(t) = \int_{-1/2}^t \nu(x)dx, \ t \in I,$$

which for $t = 1/2$ also includes the intensity $\lambda = \int_I d\nu = V(1/2)$ of the underlying Poisson process. From the usual 'Delta method' we can then also deduce a Bernstein–von Mises theorem for the distribution function $M(t) = \int_I 1_{(-1/2,t]} d\mu$ of the jump measure $\mu = \nu/\lambda = \nu/\int_I \nu$. The key to this is the following lemma, proved in (the proof of) Theorem 4 of [7].

Lemma 8. *Suppose the weights (ω_l) satisfy $\sum_l 2^{-l/2}\omega_l < \infty$. Then the mapping*

$$L : (\nu_{lk}) \mapsto V = \int_0^{\cdot} \sum_{l,k} \nu_{lk}\psi_{lk}$$

is linear and continuous from $\mathcal{M}_0(\omega)$ to $L^\infty(I)$ for the respective norm topologies.

For the next theorem we require some more definitions: We denote $V_0(t) = \int_{-1/2}^t \nu_0(x)dx$. Let $\mathcal{N}_{V_0}$ be the law of the tight Gaussian random variable in $L^\infty(I)$ given by $L(Z), Z \sim \mathcal{N}_{\nu_0}$. We define l_{ν_0} to be the linear mapping $L^\infty(I) \to L^\infty(I)$ with $l_{\nu_0}[h] = (hV_0(\frac{1}{2}) - V_0 h(\frac{1}{2}))/V_0^2(\frac{1}{2})$. Finally we denote by $\mathcal{N}'_{M_0}$ the law of the tight Gaussian random variable in $L^\infty(I)$ given by $l_{\nu_0}[L(Z)]$.

The measures $\mathcal{N}_{V_0}, \mathcal{N}'_{M_0}$ have separable range in the image in $L^\infty(I)$ of $\mathcal{M}_0(\omega)$ under a continuous map. The metrisation of weak convergence of laws towards $\mathcal{N}_{V_0}, \mathcal{N}'_{M_0}$ in the non-separable space L^∞ by $\beta_{L^\infty(I)}$ thus remains valid (Theorem 3.28 in [13]).

Theorem 9. *Suppose that $X_1, \dots, X_n$ are generated from (2) and grant Assumption 1. Let $\nu \sim \Pi(\cdot|X_1, \dots, X_n)$ be a draw from the posterior distribution arising from prior $\Pi = \Pi_J$ in (10) with J as in (11) and let L be the linear mapping from Lemma 8. Conditional on $X_1, \dots, X_n$ define $V = L(\nu)$ and $\widehat{V} = L(\widehat{\nu}_J)$ where $\widehat{\nu}_J$ is given in (27).*

Then we have as $n \to \infty$ and in $\mathbb{P}_{\nu_0}^{\mathbb{N}}$-probability that

$$\beta_{L^\infty(I)}\left(\mathcal{L}(\sqrt{n}(V - \widehat{V})|X_1, \dots, X_n), \mathcal{N}_{V_0}\right) \to 0.$$

In particular if N_{λ_0} is the law on $\mathbb{R}$ of $L(Z)(\frac{1}{2})$ then as $n \to \infty$,

$$\beta_{\mathbb{R}}\left(\mathcal{L}(\sqrt{n}(V(\tfrac{1}{2}) - \widehat{V}(\tfrac{1}{2}))|X_1, \dots, X_n), N_{\lambda_0}\right) \to^{\mathbb{P}_{\nu_0}^{\mathbb{N}}} 0.$$

Moreover, if $M = V/V(\frac{1}{2})$ and $\widehat{M} = \widehat{V}/\widehat{V}(\frac{1}{2})$, then as $n \to \infty$,

$$\beta_{L^\infty(I)}\left(\mathcal{L}(\sqrt{n}(M - \widehat{M})|X_1, \dots, X_n), \mathcal{N}'_{M_0}\right) \to^{\mathbb{P}_{\nu_0}^{\mathbb{N}}} 0.$$

Proof. The first two limits are immediate consequences of Theorem 7, Lemma 8 and the continuous mapping theorem. For the last limit we apply the Delta method for weak convergence ([35], Theorem 20.8) to the map $V \mapsto V/V(\frac{1}{2})$, which is Fréchet differentiable from $L^\infty(I) \to L^\infty(I)$ at any $\nu \in L^\infty(I)$ that is bounded away from zero, with derivative l_ν. $\square$

Arguing just as before (25) one shows that the above Gaussian limit distributions all attain the semi-parametric Cramér–Rao lower bounds for the problems of estimating $V, M, \lambda = V(\frac{1}{2})$, respectively. In particular they imply that 'Bayesian credible sets' are optimal asymptotic frequentist confidence sets for these parameters – the arguments are the same as in [7], Section 4.1, and hence omitted. These results are the 'Bayesian' versions of the Donsker type limit theorems obtained for frequentist estimators in [26, 10], where the same limit distributions were obtained.

3.5. Concluding remarks

Adaptive prior choices Our series prior is defined via asymptotic growth of J (see (11)) that depends on n and on knowledge of the smoothness s. A possible extension of our work would be to make the results adaptive to the choice of J, e.g., by placing a hyperprior on $J \in \mathbb{N}$ whose probability mass function is proportional to $\exp(-c2^J L(J))$ with $L(J) = J$ or $= 1$. While it seems possible to prove an upper bound for 2^J of order $(n/\log n)^{1/(2s+1)}$ with such a hyperprior, it is unclear whether a corresponding lower bound holds as well. Small values of J can entail a large bias and the control of the semi-parametric bias poses considerable difficulties in our proofs. As in [31], a self-similarity condition on ν may help to overcome such problems, but this is beyond the scope of the present paper.

Scaling of the observation distance Δ For identifiability reasons, Assumption 1 imposes an upper bound on the (fixed) distance between observations Δ. Otherwise the observation distance Δ enters the contraction rate result in Theorem 2 only via multiplicative constants. In the Bernstein–von Mises results (Theorems 7 and 9), the limiting processes scale with $1/\Delta$, as can be seen from the scaling of $(A_\nu^*)^{-1}$ before equation (20). This suggests that 'high-frequency' analogues of our Bernstein–von Mises results, comparable to those in [27], should hold true as well, with convergence rate $1/\sqrt{n\Delta}$ instead of $1/\sqrt{n}$.

Bernstein–von Mises theorems for general inverse problems This paper builds on key ideas for nonparametric Bernstein–von Mises theorems in direct models [6, 7, 4, 8, 5]. For inverse problems previous work on Bernstein–von Mises theorems treated regression-type problems where the likelihood has a more explicit Gaussian structure, see [24, 21] and also the more recent contributions [19, 25]. In our jump process setting, the log-likelihood function does not have the form of a Gaussian process, but we show how empirical process methods [18] can be used to obtain exact Gaussian posterior asymptotics in such situations as well. Our proof techniques are thus potentially relevant for other models with independent and identically distributed observations.

4. Proofs of the main theorems

4.1. Asymptotics for the localised posterior distribution

The first step will be to localise the posterior distribution near the 'true' $\nu_0 \in C^s$ by obtaining a preliminary (in itself sub-optimal) contraction rate for the prior Π from (10). Recall the notation $v = \log \nu$ and define

$$D_{n,M} := \left\{ \nu : v \in V_{B,J}, \|v - v_0\|_{L^2} \leqslant M\varepsilon_n^{L^2}, \|v - v_0\|_\infty \leqslant M\varepsilon_n^{L^\infty} \right\} \tag{28}$$

with M a constant and

$$\varepsilon_n^{L^2} = n^{-\frac{s-1/2}{2s+1}} (\log n)^{1/2+\delta}, \qquad \varepsilon_n^{L^\infty} = n^{-\frac{s-1}{2s+1}} (\log n)^{1/2+\delta}$$

for any $\delta > 1/2$. We have the following

Proposition 10. *For $D_{n,M}$ as in (28), prior Π arising from (10) with J chosen as in (11) and under Assumption 1, we have for any $s > 5/2, \delta > 1/2$ and every M large enough*

$$\Pi(D_{n,M}^c | X_1, \dots, X_n) \to^{\mathbb{P}_{\nu_0}^{\mathbb{N}}} 0 \tag{29}$$

as $n \to \infty$. In particular we can choose M in (28) large enough so that the last convergence to zero occurs also for $D_{n,M/2}$ replacing $D_{n,M}$. Moreover, on the set $D_{n,M}$ we also have the same contraction rates for $\nu - \nu_0$ in place of $v - v_0$ with a possibly larger constant M.

Proof. This is proved in Section 5 below. □

As a consequence of the previous proposition together with the notation $\Pi^{D_{n,M}} := \Pi^{D_{n,M}}(\cdot | X_1, \dots, X_n)$ for the posterior measure arising from the prior $\Pi(\cdot \cap D_{n,M})/\Pi(D_{n,M})$ instead of from Π, we can deduce the basic inequality

$$\sup_{B \in \mathcal{S}_V} |\Pi(B|X_1, \dots, X_n) - \Pi^{D_{n,M}}(B|X_1, \dots, X_n)| \leqslant 2\Pi(D_{n,M}^c | X_1, \dots, X_n) \to^{\mathbb{P}_{\nu_0}^{\mathbb{N}}} 0 \tag{30}$$

as $n \to \infty$. We now study certain Laplace-transform functionals of the localised posterior measure $\Pi^{D_{n,M}}$. We use the shorthand notation V_J for the L^2-closed linear space spanned by the wavelets up to level J and $g_J = P_{V_J}(g)$ for the wavelet projection of $g \in L^2(I)$ onto V_J. For a fixed function $\eta : I \to \mathbb{R}$, consider a perturbation of ν given by

$$\begin{aligned} \nu_t &= \nu_t^\eta := e^{v_t}, \\ v_t &= v + \delta_n \Big(\frac{t}{\delta_n \sqrt{n}} \eta + v_{0,J} - v \Big) = (1 - \delta_n) v + \delta_n \Big(\frac{t}{\delta_n \sqrt{n}} \eta + v_{0,J} \Big), \end{aligned} \tag{31}$$

where $0 < t < \infty$ and $\delta_n \to 0$ such that $\delta_n \sqrt{n} \to \infty$ is a sequence to be chosen. That the perturbation ν_t equals a convex combination of points will be useful to deal with the fact that our parameter space has a boundary (see also [23, 24]).

We have the following key proposition, giving general conditions under which a (sub-) Gaussian approximation for the Laplace transform of general functionals $F(\nu)$ of the posterior distribution holds. Its proof is given in Section 6.

Proposition 11. *Under the hypotheses of Proposition 10, suppose δ_n is chosen such that (61) is satisfied and let $\mathcal{H}_n \subseteq L^\infty(I)$ be such that (62), (63) hold uniformly for all $\eta \in \mathcal{H}_n$. If $T > 0$ and if $F : \mathcal{V} \to \mathbb{R}$ is any fixed measurable function then*

$$E^{\Pi^{D_{n,M}}}\left[e^{t\sqrt{n}F(\nu)}\middle| X_1, \dots, X_n\right]$$
$$= \exp\left\{\frac{t^2}{2}\|A_{\nu_0}(\eta)\|^2_{L^2(\mathbb{P}_{\nu_0})} - \frac{t}{\sqrt{n}}\sum_{i=1}^n A_{\nu_0}(\eta)(X_i) + r_n\right\} \times Z_n$$

where $r_n = O_{\mathbb{P}^{\mathbb{N}}_{\nu_0}}(a_n)$ as $n \to \infty$ with a nonstochastic null sequence $a_n \to 0$ that is uniform in $|t| \leqslant T$, $\eta \in \mathcal{H}_n$; and where

$$Z_n = \frac{\int_{D_{n,M}} e^{S_n(\nu)+\ell_n(\nu_t)} d\Pi(\nu)}{\int_{D_{n,M}} e^{\ell_n(\nu)} d\Pi(\nu)}, \quad \nu_t \text{ as in (31)},$$
$$S_n(\nu) = t\sqrt{n}\Big(F(\nu) + \int A_{\nu_0}(v - v_0) A_{\nu_0}(\eta) d\mathbb{P}_{\nu_0}\Big), \quad v = \log \nu,\ v_0 = \log \nu_0,$$

and $A_\nu : L^2(\nu) \to L^2_0(\mathbb{P}_\nu)$ was defined in Proposition 3.

Given a functional F of interest, we can use Proposition 11 to show Bernstein–von Mises theorems by selecting appropriate η so that $S(\nu)$ vanishes (or converges to zero). When this is the case it remains to deal with Z_n by a change of measure argument for $\nu \mapsto \nu_t$.

4.2. Change of measure in the posterior

We now study the ratio Z_n for η, δ_n satisfying certain conditions, and under the assumption that $\sup_{\nu \in D_{n,M}} |S_n(\nu)|$ is either $O(1)$ or $o(1)$. Note that by Assumption 1, $v_0 = \log \nu_0$ is an 'interior' point of the support

$$V_{B,J} = \prod_{l=-1}^{J-1} (-Ba_l, Ba_l)^{2^l \vee 1} \subseteq \mathbb{R}^{2^J}, \quad a_l = 2^{-l}(l^2+1)^{-1},$$

of the prior Π. We shall require that $(t/\delta_n\sqrt{n})\eta + v_{0,J}$ is also contained in $V_{B,J}$, implied by

$$t|\langle \eta, \psi_{lk}\rangle| \leqslant \gamma 2^{-l}(l^2+1)^{-1}\sqrt{n}\delta_n \quad \forall l < J-1, k, \qquad \langle \eta, \psi_{lk}\rangle = 0 \quad \forall l > J. \tag{32}$$

Note that under (32) the function v_t from (31) is a convex combination of elements $v, (t/\delta_n\sqrt{n})\eta + v_{0,J}$ of $V_{J,B}$ and hence itself contained in the support

$V_{J,B}$ of Π. We can thus write

$$\frac{\int_{D_{n,M}} e^{\ell_n(\nu_t)} d\Pi(\nu)}{\int_{D_{n,M}} e^{\ell_n(\nu)} d\Pi(\nu)} = \frac{\int_{D^t_{n,M}} e^{\ell_n(\nu)} \frac{d\Pi^t(\nu)}{d\Pi(\nu)} d\Pi(\nu)}{\int_{D_{n,M}} e^{\ell_n(\nu)} d\Pi(\nu)},$$

where Π^t is the law of ν_t, absolutely continuous with respect to Π, and where

$$D^t_{n,M} = \{\nu_t : \nu \in D_{n,M}\}.$$

The measure Π^t corresponds to transforming each coordinate v_{lk} of the 2^J-dimensional product integral defining the prior Π into the convex combination $v_{t,lk} = (1-\delta_n)v_{lk} + \delta_n i_{t,lk}$ where $i_{t,lk} = \langle \frac{t}{\delta_n \sqrt{n}} \eta + v_{0,J}, \psi_{lk} \rangle$ is a deterministic (under Π) point in $(-Ba_l, Ba_l) = I_{l,B}$ for every $k, l \leqslant J$. The density of the law of $v_{t,lk}$ with respect to v_{lk} is constant on a subinterval of $I_{l,B}$ of length $2B(1-\delta_n)$ and thus has constant density $(1-\delta_n)^{-1}$. The density of the product integrals is then also constant in v and equal to

$$\left(\frac{1}{1-\delta_n}\right)^{2^J} = 1 + o(1) \text{ whenever } 2^J \delta_n = o(1), \tag{33}$$

independently of ν. We conclude that if (32), (33) hold then

$$\begin{aligned} \frac{\int_{D_{n,M}} e^{\ell_n(\nu_t)} d\Pi(\nu)}{\int_{D_{n,M}} e^{\ell_n(\nu)} d\Pi(\nu)} &= (1+o(1)) \times \frac{\int_{D^t_{n,M}} e^{\ell_n(\nu)} d\Pi(\nu)}{\int_{D_{n,M}} e^{\ell_n(\nu)} d\Pi(\nu)} \\ &= (1+o(1)) \times \frac{\Pi(D^t_{n,M}|X_1,\dots,X_n)}{\Pi(D_{n,M}|X_1,\dots,X_n)}, \end{aligned} \tag{34}$$

where the last identity follows from renormalising both numerator and denominator by $\int_{\mathcal{V}} e^{\ell_n(\nu)} d\Pi(\nu)$. The numerator in the last expression is always less than or equal to one and by Proposition 10 the denominator converges to one in probability, so that we have

Lemma 12. *Suppose* $\sup_{\nu \in D_{n,M}} |S_n(\nu)| = O(1)$ *holds as* $n \to \infty$ *and assume* η, δ_n, t *are such that (32), (33) hold. Then the random variable* Z_n *in Proposition 11 is* $O_{\mathbb{P}^{\mathbb{N}}_{\nu_0}}(1)$*, uniformly in* η*, as* $n \to \infty$.

To prove the exact asymptotics in the Bernstein–von Mises theorem we need:

Lemma 13. *Suppose* η, δ_n *are such that (32), (33) hold and assume in addition that* $\|\eta\|_\infty \leqslant d$ *for some fixed constant* d.

A) Let $D_{n,M}$ *be as in (28) and define the set* $D^t_{n,M} = \{\nu_t : \nu \in D_{n,M}\}$*. Then for all* $n \geqslant n_0(t)$ *and* M *large enough we have* $D_{n,M/2} \subseteq D^t_{n,M}$ *and thus by Proposition 10 also* $\Pi(D^t_{n,M}|X_1,\dots,X_n) \to 1$ *in* $\mathbb{P}^{\mathbb{N}}_{\nu_0}$*-probability.*

B) Assume also that $\sup_{\nu \in D_{n,M}} |S_n(\nu)| = o(1)$ *then* Z_n *from Proposition 11 satisfies* $Z_n = 1 + o_{\mathbb{P}^{\mathbb{N}}_{\nu_0}}(1)$ *as* $n \to \infty$.

Proof. A) Let $\nu \in D_{n,M/2}$ be arbitrary. We need to show that there exists $\zeta = \zeta(\nu) \in D_{n,M}$ such that $\zeta_t = \nu$. For $v = \log \nu$ notice that by definition of $D_{n,M/2}$ we have $\|v - v_{0,J}\|_{L^2} \leqslant \|v - v_0\|_{L^2} \leqslant (M/2)\varepsilon_n^{L^2}$ and similarly $\|v - v_{0,J}\|_\infty \leqslant (M/2)\varepsilon_n^{L^\infty}$. Now define $\zeta = e^z$ where

$$z = z(\nu) := v_{0,J} + \frac{(v - v_{0,J}) - \frac{t}{\sqrt{n}}\eta}{1 - \delta_n},\ \nu \in D_{n,M/2}.$$

Then by definition

$$\begin{aligned} z_t &= (1-\delta_n)z + \frac{t}{\sqrt{n}}\eta + \delta_n v_{0,J} \\ &= (1-\delta_n)v_{0,J} + (v - v_{0,J}) - \frac{t}{\sqrt{n}}\eta + \frac{t}{\sqrt{n}}\eta + \delta_n v_{0,J} = v \end{aligned}$$

so $\zeta_t(\nu) = \nu$ follows. It remains to verify that also $\zeta(\nu) \in D_{n,M}$ for every $\nu \in D_{n,M/2}$. To see this we let n large enough such that in particular $\delta_n < 1/4$ and then

$$\|z(\nu) - v_0\|_{L^2} \leqslant \|v_0 - v_{0,J}\|_{L^2} + \frac{4}{3}\|v - v_{0,J}\|_{L^2} + \frac{4t}{3\sqrt{n}}\|\eta\|_{L^2} \leqslant M\varepsilon_n^{L^2} \quad (35)$$

using $\|v_0 - v_{0,J}\|_{L^2} \lesssim 2^{-Js} = o(\varepsilon_n^{L^2})$ from (8) and also $1/\sqrt{n} = o(\varepsilon_n^{L^2})$. The same arguments imply

$$\|z(\nu) - v_0\|_\infty \leqslant M\varepsilon_n^{L^\infty}.$$

Finally we need to check that $z(\nu) \in V_{J,B}$ holds true. We notice that for all $l \leqslant J$

$$|\langle z(\nu) - v_0, \psi_{lk}\rangle| \leqslant \|z(\nu) - v_0\|_{L^2} \leqslant \gamma 2^{-l}(l^2+1)^{-1} = \gamma a_l$$

is implied by

$$\varepsilon_n^{L^2} \approx n^{-\frac{s-1/2}{2s+1}}(\log n)^{1/2+\delta} = o(2^{-J}(J^2+1)^{-1}), \quad s > 5/2,$$

for n large enough, so that from Assumption 1 and (35) we deduce

$$|\langle z(\nu), \psi_{lk}\rangle| \leqslant |\langle v_0, \psi_{lk}\rangle| + |\langle z(\nu) - v_0, \psi_{lk}\rangle| \leqslant (B-\gamma)a_l + \gamma a_l, \quad l \leqslant J-1,$$

for n large enough, hence $\zeta \in V_{J,B}$. The last claim in Part A) now follows directly from Proposition 10, and Part B) also follows, from (34). □

4.3. Proof of Theorem 2

Given the results from Sections 4.1, 4.2, the proof follows ideas in [4]. By (30) it suffices to prove the theorem with the posterior $\Pi(\cdot|X_1, \dots, X_n)$ replaced by $\Pi^{D_{n,M}}(\cdot|X_1, \dots, X_n)$. Using that $\nu = e^v$ are uniformly bounded and that $v_J = P_{V_J} v = v$ for $v \sim \Pi^{D_{n,M}}(\cdot|X_1, \dots, X_n)$, we can write

$$\|\nu - \nu_0\|_\infty \lesssim \|v - v_0\|_\infty \le \|v_J - v_{0,J}\|_\infty + \|v_{0,J} - v_0\|_\infty.$$

The second term is of deterministic order $2^{-J_n s} = O(n^{-s/(2s+1)})$ by (8) and since $v_0 = \log \nu_0 \in C^s$, so it remains to deal with the first. We can write, using (8) again,

$$\begin{aligned}
\|v_J - v_{0,J}\|_\infty &= \sup_x \left| \sum_{\ell<J,m} \langle v - v_0, \psi_{\ell m}\rangle \psi_{\ell m}(x) \right| \\
&\lesssim \sum_{\ell<J} \frac{2^{\ell/2}}{\sqrt{n}} (\log n)^{1/2+\delta} \max_{m=0,\dots,2^\ell-1} \frac{\sqrt{n}}{(\log n)^{1/2+\delta}} |\langle v - v_0, \psi_{\ell m}\rangle| \\
&\lesssim \frac{2^{J/2}(J+1)}{\sqrt{n}} (\log n)^{1/2+\delta} \max_{\ell<J,m=0,\dots,2^\ell-1} \sqrt{n}\, |\langle v - v_0, c_{\ell J}\psi_{\ell m}\rangle| ,
\end{aligned} \tag{36}$$

where we have set $c_{\ell J} = \frac{2^{\ell/2}}{2^{J/2}}(\log n)^{-1/2-\delta}$, bounded by 1 since $\ell \leqslant J$.

Fix $\ell < J, m$ for the moment and let $\widetilde{\psi} \equiv (\widetilde{\psi})_{\ell m}$ be the absolutely continuous part (23) of $\widetilde{\psi}_d$ from (21) where we choose $\psi = c_{\ell J}\psi_{\ell m} 1_{I\setminus\{0\}}$. We will apply Proposition 11 to the functional $F(\nu) = \langle v - v_0, c_{\ell J}\psi_{\ell m}\rangle$ and for the choices

$$\eta = \widetilde{\psi}_J \quad \text{and} \quad \delta_n = \frac{K 2^J (J^2+1)}{\sqrt{n}}, \tag{37}$$

where $K > 0$ is a constant. To bound the term $S_n(\nu)$ in Proposition 11 we need the following approximation lemma.

Lemma 14. *For any $\psi = c_{\ell J}\psi_{\ell m} 1_{I\setminus\{0\}}$ with fixed $\ell < J, m$, let $\widetilde{\psi}_d$ be the corresponding finite measure defined in (21), let $\widetilde{\psi}$ be its absolutely continuous part from (23), and let $\widetilde{\psi}_J = P_{V_J}(\widetilde{\psi})$ be its wavelet projection onto V_J. Then we have, for some constant c_0 independent of ℓ, m, J, that*

$$\left| c_{\ell J} \int_I (v - v_0)\psi_{\ell m} + \int_I A_{\nu_0}(v - v_0) A_{\nu_0}(\widetilde{\psi}_J) d\mathbb{P}_{\nu_0} \right| \leq c_0 \frac{\|\nu - \nu_0\|_{L^2}}{2^J (\log n)^{1/2+\delta}}.$$

Proof. We notice that Lemma 6 implies

$$c_{\ell J} \int_I (v - v_0)\psi_{\ell m} = c_{\ell J} \int_I (v - v_0)\psi_{\ell m} 1_{I\setminus\{0\}} = -\int_I A_{\nu_0}(v - v_0) A_{\nu_0}(\widetilde{\psi}) d\mathbb{P}_{\nu_0},$$

so that by linearity of the operator A_{ν_0} and Lemma 5 it suffices to bound

$$\begin{aligned}
\int_I A_{\nu_0}(v - v_0) A_{\nu_0}(\widetilde{\psi}_J - \widetilde{\psi}) d\mathbb{P}_{\nu_0} &= \int_I \nu_0 A^*_{\nu_0}[A_{\nu_0}(v - v_0)](\widetilde{\psi}_J - \widetilde{\psi}) \\
&= \sum_{l>J} \sum_k \langle h(\nu, \nu_0), \psi_{lk}\rangle \langle \widetilde{\psi}, \psi_{lk}\rangle,
\end{aligned}$$

where we have used Parseval's identity, and the shorthand notation $h(\nu, \nu_0) := \nu_0 A^*_{\nu_0}[A_{\nu_0}(v - v_0)]$. Now $\widetilde{\psi}$ is the absolutely continuous part of $\widetilde{\psi}_d$ which according to (22) (with $\Delta = 1$ without loss of generality) is given by

$$\widetilde{\psi}_d = -\frac{1}{\nu_0}\, \pi_{\nu_0} * \left(\left(\pi_{\nu_0}(-\cdot) * \frac{\psi}{\nu_0} \right) \mathbb{P}_{\nu_0} \right)$$

$$= -\frac{e^{2\nu_0(I)}}{\nu_0}\Big(\sum_{\iota=0}^{\infty}\sum_{\kappa=0}^{\infty}\frac{(-1)^{\iota+\kappa}}{\iota!\kappa!}\Big(\nu_0^{*\iota} * \nu_0(-\cdot)^{*\kappa} * \frac{\psi}{\nu_0}\Big)\,\mathbb{P}_{\nu_0}\Big).$$

By standard properties of convolutions, using (4) and since ψ/ν_0 is absolutely continuous, removing the discrete part of $\widetilde{\psi}_d$ means removing Dirac measure from the series expansion of $\mathbb{P}_{\nu_0}$ – denote the resulting absolutely continuous measure by P_{ν_0}. First we consider the part $\bar{\psi}$ of $\widetilde{\psi}$ corresponding to the terms in the last series where either $\iota > 0$ or $\kappa > 0$, so that not all of the convolution factors in

$$\nu_0^{*\iota} * \nu_0(-\cdot)^{*\kappa} * \frac{\psi}{\nu_0}$$

are Dirac measures δ_0. Since $C^s(I), s > 5/2$, is imbedded into the standard periodic Sobolev space $H^\alpha(I), \alpha \leqslant 2$, we can use the basic convolution inequality $\|f * g\|_{C^\alpha(I)} \leqslant \|f\|_{H^\alpha(I)}\|g\|_{L^2}, \alpha = 0, 2$, (proved, e.g., just as Lemma 4.3.18 in [18]), the fact that $\psi/\nu_0 = c_{\ell J}\psi_{\ell m}/\nu_0$ is bounded in $L^2 = H^0$, and the multiplier property $\|fg\|_{H^2} \lesssim \|f\|_{C^2}\|g\|_{H^2}$ combined with the fact that the density of P_{ν_0} is contained in $C^s(I) \subseteq C^2(I)$, to deduce that $\bar{\psi}$ is contained in $C^2(I)$ and thus, by (8)

$$\begin{aligned}\Big|\sum_{l>J}\sum_k \langle h(\nu,\nu_0),\psi_{lk}\rangle\langle\bar{\psi},\psi_{lk}\rangle\Big| &\leqslant \sum_{l>J}\|\langle h(\nu,\nu_0),\psi_{l\cdot}\rangle\|_{L^2}\|\langle\bar{\psi},\psi_{l\cdot}\rangle\|_{L^2}\\ &\lesssim \sum_{l>J}\|\nu-\nu_0\|_{L^2}2^{-2l} \lesssim \|\nu-\nu_0\|_{L^2}2^{-2J},\end{aligned}$$

which is of the desired order.

Setting $\iota = \kappa = 0$ in the preceding representation of $\widetilde{\psi}$ and using the convolution series representation of P_{ν_0} (without discrete part) yields the 'critical' term which is given by $-\psi g$ where

$$g = c\frac{1}{\nu_0^2}\sum_{j=1}^{\infty}\frac{\nu_0^{*j}}{j!},$$

for a suitable constant $c > 0$. By arguments similar to above the function g is at least in C^2 and for x_{lk} the mid-point of the support set S_{lk} of ψ_{lk} (an interval of width $O(2^{-l})$ at most) we can write

$$\begin{aligned}\langle\psi_{\ell m}g,\psi_{lk}\rangle &= \int_I \psi_{\ell m}(g - g(x_{lk}) + g(x_{lk}))\psi_{lk}\\ &= \int_I \psi_{\ell m}\psi_{lk}(g - g(x_{lk})) + g(x_{lk})\int_I \psi_{\ell m}\psi_{lk}.\end{aligned}$$

The last term vanishes by orthogonality ($\ell \leqslant J < l$), and using the mean value theorem the absolute value of the first is bounded by

$$\|g'\|_\infty \int_{S_{lk}} |x - x_{lk}||\psi_{\ell m}(x)||\psi_{lk}(x)|dx \lesssim 2^{-l}\int_I |\psi_{\ell m}(x)||\psi_{lk}(x)|dx.$$

Then, using (8) and the standard convolution inequalities for L^2-norms,

$$\begin{aligned}\sum_{l>J} 2^{-l} \sum_k |\langle h(\nu,\nu_0), \psi_{lk}\rangle| \int_I |\psi_{\ell m}||\psi_{lk}| \\ \leqslant \sum_{l>J} 2^{-l} \|h(\nu,\nu_0)\|_{L^2} \int_I |\psi_{\ell m}(x)| \sum_k |\psi_{lk}(x)| dx \\ \lesssim \sum_{l>J} 2^{-l/2} \|h(\nu,\nu_0)\|_{L^2} \|\psi_{\ell m}\|_{L^1} \lesssim 2^{-J/2} 2^{-\ell/2} \|\nu-\nu_0\|_{L^2}\end{aligned}$$

Scaling the last estimate by a multiple of $c_{\ell J} = 2^{\ell/2-J/2}(\log n)^{-1/2-\delta}$ leads to the result. □

Conclude from Proposition 10 and our choice of J that

$$\sup_{\nu \in D_{n,M}} |S_n(\nu)| \lesssim \frac{\sqrt{n}\|\nu-\nu_0\|_{L^2}}{2^J (\log n)^{1/2+\delta}} \lesssim \sqrt{n} n^{-(s+1/2)/(2s+1)} = O(1).$$

Simple calculations (using that (22) implies that $\widetilde{\psi}_J, 2^{-J/2}\widetilde{\psi}_J$ are uniformly bounded in L^2, L^∞, respectively, proved by arguments similar to those used in Lemma 14) show that for $s > 5/2$ the three conditions (61), (62), (63) and the two conditions (32), (33) are all satisfied for such η, δ_n chosen as in (37) and K large enough. We thus deduce from Proposition 11 and Lemma 12 that for some sequence $C_n = O_{\mathbb{P}_{\nu_0}^{\mathbb{N}}}(1)$ and $|t| \leqslant T$,

$$\begin{aligned}E^{\Pi^{D_{n,M}}}\left[e^{t\sqrt{n}\int (v-v_0)c_{\ell J}\psi_{\ell m}} | X_1, \ldots, X_n\right] \\ \leqslant C_n \exp\left\{\frac{t^2}{2}\|\widetilde{\psi}_J\|_{LAN}^2 - \frac{t}{\sqrt{n}}\sum_{k=1}^n A_{\nu_0}(\widetilde{\psi}_J)(X_k)\right\}.\end{aligned}$$

If we define $\widetilde{\nu}_{\ell m} = -\frac{1}{n}\sum_{k=1}^n A_{\nu_0}(\widetilde{\psi}_J)(X_k) + c_{\ell J}\int v_0 \psi_{\ell m}$ then for $|t| \leqslant T$ this becomes the sub-Gaussian estimate

$$E^{\Pi^{D_{n,M}}}\left[e^{t\sqrt{n}(c_{\ell J}\int v\psi_{\ell m} - \widetilde{\nu}_{\ell m})} | X_1, \ldots, X_n\right] \leqslant C_n \exp\left\{\frac{t^2}{2}\|\widetilde{\psi}_J\|_{LAN}^2\right\} \tag{38}$$

for the stochastic process $Z_{\ell,m} = (c_{\ell J}\int v\psi_{\ell m} - \widetilde{\nu}_{\ell m})|X_1, \ldots, X_n$ conditional on $X_1, \ldots, X_n$, with constants η, t uniform. We can then decompose

$$\sqrt{n} c_{\ell J} |\langle v - v_0, \psi_{\ell m}\rangle| \leqslant \sqrt{n}|Z_{\ell,m}| + \left|\frac{1}{\sqrt{n}}\sum_{k=1}^n A_{\nu_0}((\widetilde{\psi}_{\ell m})_J)(X_k)\right|,$$

and the maximum over 2^J many variables in (36) can now be estimated by the sum of the maxima of each of the preceding processes. For the first process we observe that the sub-Gaussian constants are uniformly bounded through

$$\|\widetilde{\psi}_J\|_{LAN}^2 = \|A_{\nu_0}(\widetilde{\psi}_J)\|_{L^2(\mathbb{P}_{\nu_0})}^2 \lesssim \|\widetilde{\psi}_J\|_{L^2(I)}^2 \leqslant \|\widetilde{\psi}\|_{L^2(I)} \lesssim \|\psi_{\ell m}\|_{L^2(I)}^2 \lesssim 1, \tag{39}$$

using Lemma 26, that $\nu_0 \in L^\infty$ is bounded away from zero, that P_{V_J} is a L^2-projector, combined with standard convolution inequalities. Using the sub-Gaussian estimate for $|t| \leqslant T$, the display in the proof of Lemma 2.3.4 in [18] yields that this maximum has expectation of order at most $O(J)$ with $\mathbb{P}_{\nu_0}^{\mathbb{N}}$-probability as close to one as desired. To the maximum of the second (empirical) process we apply Lemma 3.5.12 in [18] (and again Lemma 26 combined with the inequality in the previous display and also that $\|g\|_\infty \lesssim 2^{J/2}\|g\|_{L^2}$ for any $g \in V_J$) to see that its $\mathbb{P}_{\nu_0}^{\mathbb{N}}$-expectation is of order $O(\sqrt{J} + J2^{J/2}/\sqrt{n}) = O(\sqrt{J})$ uniformly in $\ell \leqslant J, m$. Feeding these bounds into (36) we see that on an event of $\mathbb{P}_{\nu_0}^{\mathbb{N}}$-probability as close to one as desired,

$$E^{\Pi^{D_{n,M}}}[\|\nu - \nu_0\|_\infty | X_1, \dots, X_n] \lesssim \frac{2^{J/2} J}{\sqrt{n}} (\log n)^{1/2+\delta} J \lesssim \frac{2^{J/2}}{\sqrt{n}} (\log n)^{5/2+\delta}. \tag{40}$$

Since $\delta > 1/2$ was arbitrary an application of Markov's inequality completes the proof.

4.4. Proof of Theorem 7

Given results from Sections 4.1, 4.2, the proof follows ideas in [7]. Let $\widehat{\nu}(J)$ be the random element of $\mathcal{M}_0(w)$ from (27) with J chosen as in (11). For $D_{n,M}$ as in (28) let $\Pi^{D_{n,M}}(\cdot|X_n, \dots, X_n)$ be as before (30), and suppose $\nu \sim \Pi^{D_{n,M}}(\cdot|X_1, \dots, X_n)$. In view of (30), and since the total variation distance dominates the metric $\beta_{\mathcal{M}_0(\omega)}$, it suffices to prove the result for $\Pi^{D_{n,M}}(\cdot|X_1, \dots, X_n)$ replacing $\Pi(\cdot|X_1, \dots, X_n)$. Let $\widetilde{\Pi}_n$ denote the laws of $\sqrt{n}(\nu - \widehat{\nu}(J))$ conditionally on $X_1, \dots, X_n$ and let $\mathcal{N}_{\nu_0}$ be the Gaussian probability measure on $\mathcal{M}_0(w)$ defined (cylindrically) before Theorem 7, arising from the law of $\mathbb{X} = (\mathbb{X}_{l,k})$. The following norm estimate is the main step to establish tightness of the process Z in $\mathcal{M}_0(\omega)$.

Lemma 15. *For any monotone increasing sequence $\bar{w} = (\bar{w}_l)$, $\bar{w}_l/l^4 \geqslant 1$, if Z equals either $\mathbb{X}$ or the process $\sqrt{n}(\nu - \widehat{\nu}(J))|X_1, \dots, X_n$, then for some fixed constant $C > 0$ we have*

$$E\big[\|Z\|_{\mathcal{M}_0(\bar{w})}\big] = E\Big[\sup_l \bar{w}_l^{-1} \max_k |Z_{l,k}|\Big] \leq C, \tag{41}$$

where in case $Z = \sqrt{n}(\nu - \widehat{\nu}(J))|X_1, \dots, X_n$ the operator E denotes conditional expectation $E^{D_{n,M}}[\cdot|X_1, \dots, X_n]$ and the inequality holds with $\mathbb{P}_{\nu_0}^{\mathbb{N}}$-probability as close to one as desired.

Proof. We first consider the more difficult case where Z is the centred and scaled posterior process. We decompose, with $\nu_J = P_{V_J}(\nu)$,

$$\sqrt{n}(\nu - \widehat{\nu}(J)) = \sqrt{n}(\nu_J - \widehat{\nu}(J)) + \sqrt{n}(\nu_0 - \nu_{0,J}) + \sqrt{n}[(\nu - \nu_0) - (\nu - \nu_0)_J].$$

The second term on the right hand side has multi-scale norm $\|\nu_0 - \nu_{0,J}\|_{\mathcal{M}(w)}$ bounded by $2^{-J(s+1/2)} w_J^{-1} = o(1/\sqrt{n})$ in view of (8), $\|\psi_{lk}\|_{L^1} \lesssim 2^{-l/2}$. Similarly the expectation of the multi-scale norm of the third term is bounded by

$$\begin{aligned}
&\int \|\nu - \nu_0 - (\nu-\nu_0)_J\|_{\mathcal{M}(w)} d\Pi^{D_{n,M}}(\nu|X_1,\dots,X_n)\\
&= \int \sup_{l>J} w_l^{-1} \max_k |\langle \nu-\nu_0, \psi_{lk}\rangle| d\Pi^{D_{n,M}}(\nu|X_1,\dots,X_n)\\
&\leqslant w_J^{-1} \sup_{l>J}\max_k \|\psi_{lk}\|_{L^1} \int \|\nu-\nu_0\|_\infty d\Pi^{D_{n,M}}(\nu|X_1,\dots,X_n)\\
&\lesssim \frac{2^{-J/2}2^{J/2}}{J^4\sqrt{n}} \log^{5/2+\delta} n = o_{\mathbb{P}_{\nu_0}^{\mathbb{N}}}(1/\sqrt{n}),
\end{aligned}$$

using (40). We turn to bounding the multi-scale norm of the first term, corresponding to

$$\sqrt{n}\|\nu_J - \widehat{\nu}(J)\|_{\mathcal{M}(w)} = \sqrt{n}\sup_{l<J} w_l^{-1} \max_k \left|\int_I \nu\psi_{lk} - \widehat{\nu}(J)_{lk}\right|.$$

The first term in the decomposition

$$\int_I \nu\psi_{lk} - \widehat{\nu}(J)_{lk} = \int_I (\nu-\nu_0)\psi_{lk} - \left(\widehat{\nu}(J)_{lk} - \int_I \nu_0\psi_{lk}\right) \equiv \int_I (\nu-\nu_0)\psi_{lk} - W_{lk} \tag{42}$$

equals

$$\int_I (\nu-\nu_0)\psi_{lk} = \int_I (e^v - e^{v_0})\psi_{lk} = \int_I (v-v_0)\nu_0\psi_{lk} + O(\|\nu-\nu_0\|_\infty^2), \tag{43}$$

and the quadratic remainder is of order $o(1/\sqrt{n})$ uniformly in k, l by definition of $D_{n,M}$ and since $s > 5/2$.

Lemma 16. *Let $\psi = \nu_0\psi_{lk}1_{I\setminus\{0\}}$ for some $l < J, k$ with corresponding $\widetilde{\psi} = (\widetilde{\psi})_{lk}$ from (21), (23) and wavelet approximation $\widetilde{\psi}_J \in V_J$. We have*

$$\left|\int_I A_{\nu_0}(v-v_0)A_{\nu_0}(\widetilde{\psi}_J)d\mathbb{P}_{\nu_0} + \int_I (v-v_0)\nu_0\psi_{lk}\right| \lesssim \|\nu-\nu_0\|_\infty 2^{-J}.$$

Proof. The proof requires only notational adaptation of the proof of Lemma 14, except for the last display, where now we use Lemma 26 (and its variant for A_ν^*) in the estimate $|\langle h(\nu,\nu_0), \psi_{lk}\rangle| \le \|h(\nu,\nu_0)\|_\infty \|\psi_{lk}\|_{L^1} \lesssim 2^{-l/2}\|\nu-\nu_0\|_\infty$ so that scaling by $c_{\ell J}$ is not necessary. □

The upper bound in the display of Lemma 16 has $E^{D_{n,M}}[\cdot|X_1,\dots,X_n]$-expectation of order $o(1/\sqrt{n})$ in view of (40). We now apply Proposition 11 to the functional

$$F(\nu) \equiv F_{lk}(\nu) = -\int_I A_{\nu_0}(v-v_0)A_{\nu_0}(\widetilde{\psi}_J)d\mathbb{P}_{\nu_0}, \tag{44}$$

with choices $\delta_n = K2^J(J^2+1)/\sqrt{n}$ for $K > 0$ a large enough constant and $\eta = \widetilde{\psi}_J$. Simple calculations (using that $\widetilde{\psi}_J, 2^{-J/2}\widetilde{\psi}_J$ are uniformly bounded in

L^2, L^∞, respectively) show that for $s > 5/2$ the three conditions (61), (62), (63) and the two conditions (32), (33) are all satisfied. Conclude from Proposition 11 and Lemma 12 that

$$E^{\Pi^{D_{n,M}}}\left[e^{t\sqrt{n}F(\nu)}|X_1,\dots,X_n\right] \leqslant C_n \exp\left\{\frac{t^2}{2}\|\widetilde{\psi}_J\|^2_{LAN} - \frac{t}{\sqrt{n}}\sum_{k=1}^n A_{\nu_0}(\widetilde{\psi}_J)(X_k)\right\}$$

for $|t| \leqslant T$, or equivalently, if $V_{lk} = \frac{1}{n}\sum_{k=1}^n A_{\nu_0}(\widetilde{\psi}_J)(X_k)$, then for some $C_n' = O_{\mathbb{P}^{\mathbb{N}}_{\nu_0}}(1)$,

$$E^{\Pi^{D_{n,M}}}\left[e^{t\sqrt{n}F(\nu)+t\sqrt{n}V_{lk}}|X_1,\dots,X_n\right] \leqslant C_n' \exp\left\{\frac{t^2}{2}\|\widetilde{\psi}_J\|^2_{LAN}\right\}. \tag{45}$$

Arguing just as in (39) the sub-Gaussian constants $\|\widetilde{\psi}_J\|^2_{LAN}$ are bounded by a fixed constant. We then have, for M a fixed constant and using $w_l \geqslant l$,

$$\begin{aligned}
&E^{\Pi^{D_{n,M}}}\left[\sup_{l<J} w_l^{-1}\max_k\left|\sqrt{n}F_{lk}(\nu)+\sqrt{n}V_{lk}\right|\,\middle|\, X_1,\dots,X_n\right] \\
&\leqslant M + \int_M^\infty \Pi^{D_{n,M}}\left(\sup_{l<J} l^{-1}\max_k\left|\sqrt{n}F_{lk}(\nu)+\sqrt{n}V_{lk}\right| > u \,\middle|\, X_1,\dots,X_n\right) du
\end{aligned}$$

We bound the tail integrals using (45) as follows:

$$\begin{aligned}
&\sum_{l<J,k}\int_M^\infty \Pi^{D_{n,M}}\left(\left|\sqrt{n}F_{lk}(\nu)+\sqrt{n}V_{lk}\right| > lu|X_1,\dots,X_n\right) du \\
&\leqslant \sum_{l<J,k}\int_M^\infty \Pi^{D_{n,M}}\left(e^{T|\sqrt{n}F_{lk}(\nu)+\sqrt{n}V_{lk}|} > e^{Tlu}|X_1,\dots,X_n\right) du \\
&\leqslant \sum_{l<J,k}\int_M^\infty E^{\Pi^{D_{n,M}}}\left[e^{T|\sqrt{n}F_{lk}(\nu)+\sqrt{n}V_{lk}|}|X_1,\dots,X_n\right] e^{-Tlu} du \\
&\lesssim C_n' \sum_{l<J} 2^l \int_M^\infty e^{-Tlu} du \lesssim C_n' \sum_{l<J} 2^l e^{-TMl} = O_{\mathbb{P}^{\mathbb{N}}_{\nu_0}}(1)
\end{aligned}$$

for M large enough. Moreover, one proves $E_{\nu_0}\sup_{l<J} w_l^{-1}\max_k|V_{lk}| \lesssim 1/\sqrt{n}$ and also $E_{\nu_0}\sup_{l<J} w_l^{-1}\max_k|W_{lk}| \lesssim 1/\sqrt{n}$ just as in the proof of Theorem 1 in [7] (or Theorem 5.2.16 in [18]), using Bernstein's inequality combined with the previous bound on the sub-Gaussian constants and a uniform bound of order $2^{J/2}$ (proved just as after (39)) on the envelopes $\|A_{\nu_0}(\widetilde{\psi}_J)\|_\infty$, $\|(A^*_{\nu_0})^{-1}(\psi_{lk}1_{\{0\}^c})\|_\infty$, $l \leqslant J$, of the empirical processes involved. Combining what precedes with Lemma 16 (and the remark after it), (42), (43) proves (41) for the 'posterior' process. The Gaussian process $\mathbb{X}$ admits by definition the same (sub-) Gaussian bound as in (45) so that the result follows from the same arguments just given. □

The inequality (41) implies in particular that for any weighting sequence ω as in Theorem 7, the processes Z concentrate in the separable subspace $\mathcal{M}_0(\omega)$

of $\mathcal{M}(\omega)$, and their laws define tight (in the case of $\mathcal{N}_{\nu_0}$, Gaussian) Borel probability measures in it (by Ulam's theorem, see p.225 in [12]). Then, using the estimate (41) and arguing as in the proof of Proposition 6 in [7] (or in Theorem 7.3.20 in [18]), Theorem 7 will follow if we can establish convergence of the finite-dimensional distributions $\widetilde{\Pi}_n \circ P_{V_L}^{-1}$ towards those of $\mathcal{N}_{\nu_0} \circ P_{V_L}^{-1}$, $L \in \mathbb{N}$ fixed, as $n \to \infty$, where P_{V_L} is the projection operator onto the finite-dimensional subspace V_L of $\mathcal{M}_0(w)$ corresponding to the first 2^L coordinates $(x_{lk} : l \leqslant L, k)$. For this we proceed as in the previous lemma, combining (42), (43) with Lemma 16 and the definition of W_{lk}, to reduce the problem to showing for $\nu \sim \Pi^{D_{n,M}}(\cdot|X_1, \dots, X_n)$ weak convergence in probability of the conditional laws of

$$Y_n \equiv -\sqrt{n}\int_I A_{\nu_0}(v - v_0)A_{\nu_0}(\widetilde{\psi}_J)d\mathbb{P}_{\nu_0} - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(A_{\nu_0}^*)^{-1}(\psi_{lk}1_{\{0\}^c})(X_i),$$

to the law of $\mathcal{N}_{\nu_0}$ for every *fixed* $k, l \leqslant L \in \mathbb{N}$. Applying Proposition 11 as after (44) combined with Lemma 13 (for k, l fixed the corresponding $\widetilde{\psi}_J$'s are bounded in L^∞) gives convergence of Z_n in Proposition 11 to one and hence one has, as $n \to \infty$ and for all t,

$$E^{\Pi^{D_{n,M}}}\left[e^{tY_n}|X_1, \dots, X_n\right] = (1 + o_{\mathbb{P}_{\nu_0}^{\mathbb{N}}}(1)) \exp\left\{\frac{t^2}{2}\|A_{\nu_0}(\widetilde{\psi}_J)\|^2_{L^2(\mathbb{P}_{\nu_0})}\right\}\exp(t\rho_n)$$

where

$$\rho_n = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(A_{\nu_0}^*)^{-1}(\psi_{lk}1_{\{0\}^c})(X_i) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}A_{\nu_0}(\widetilde{\psi}_J)(X_i).$$

Using Lemma 4, (21), $A_\nu(\widetilde{\psi}_d - \widetilde{\psi}) = 0$ by (16) and (23), and then also Lemma 26 combined with $\widetilde{\psi} \in L^2$ one has

$$\begin{aligned}\|A_{\nu_0}(\widetilde{\psi}_J) + (A_{\nu_0}^*)^{-1}(\psi_{lk}1_{\{0\}^c})\|_{L^2(\mathbb{P}_{\nu_0})} &= \|A_{\nu_0}(\widetilde{\psi}_J) - A_{\nu_0}(\widetilde{\psi})\|_{L^2(\mathbb{P}_{\nu_0})} \\ &\lesssim \|\widetilde{\psi}_J - \widetilde{\psi}\|_{L^2(I)} \to 0\end{aligned}$$

as $J \to \infty$, in particular by Chebyshev's inequality $\rho_n = o_{\mathbb{P}_{\nu_0}^{\mathbb{N}}}(1)$ for every fixed $l \leqslant L, k$. Thus the Laplace-transforms of each such coordinate projection converge to the Laplace transform of the correct normal limit distribution, for all t,

$$E^{\Pi^{D_{n,M}}}\left[e^{tY_n}|X_1, \dots, X_n\right] = (1 + o_{\mathbb{P}_{\nu_0}^{\mathbb{N}}}(1)) \times \exp\left\{\frac{t^2}{2}\|(A_{\nu_0}^*)^{-1}(\psi_{lk}1_{\{0\}^c})\|^2_{L^2(\mathbb{P}_{\nu_0})}\right\},$$

and convergence in distribution now follows from standard arguments (see, e.g., Proposition 29 in [24]). This argument extends directly to all linear combinations $\sum_{l \leqslant L, k} a_{l,k}\psi_{lk}$, so that we can apply the Cramer–Wold device to obtain joint convergence in V_L for any $L \in \mathbb{N}$. The proof is complete.

5. Proof of Proposition 10

We first derive a general contraction theorem from which we will deduce Proposition 10 (after Proposition 23). We follow the usual 'testing and small ball probability approach' (as in Theorem 7.3.1 in [18], see also [16]), which in our setting gives the following starting point to prove contraction rates, where $K(\mathbb{P}_\nu, \mathbb{P}_{\nu'})$ denotes the usual Kullback–Leibler (KL-) divergence between two probability measures $\mathbb{P}_\nu, \mathbb{P}_{\nu'}$.

Proposition 17. *Consider a prior Π on a σ-field $\mathcal{S}_V$ of some set $\mathcal{V}$ of Lévy measures for which the map $(\nu, x) \mapsto p_\nu(x)$, defined before (5) is jointly measurable. Let d be some metric on $\mathcal{V}$ such that $\nu \mapsto d(\nu, \nu')$ is measurable for all $\nu' \in \mathcal{V}$. Suppose for some sequence $\varepsilon_n \to 0$ such that $\sqrt{n}\varepsilon_n \to \infty$, constant $C > 0$ and n large enough we have*

$$\Pi\left(\nu \in \mathcal{V} : K(\mathbb{P}_{\nu_0}, \mathbb{P}_\nu) \leqslant \varepsilon_n^2, \mathrm{Var}_{\mathbb{P}_{\nu_0}}\left(\log \frac{d\mathbb{P}_\nu}{d\mathbb{P}_{\nu_0}}\right) \leqslant \varepsilon_n^2\right) \geqslant e^{-Cn\varepsilon_n^2}$$

and that for $\mathcal{V}_n \subseteq \mathcal{V}$ such that $\Pi(\mathcal{V} \setminus \mathcal{V}_n) \leqslant Le^{-(C+4)n\varepsilon_n^2}$ we can find tests $\Psi_n = \Psi(X_1, \dots, X_n)$ and $\delta_n > 0, M_0 > 0$, such that

$$\mathbb{E}_{\nu_0}\Psi_n \to 0, \qquad \sup_{\nu \in \mathcal{V}_n,\, d(\nu,\nu_0) \geqslant M_0\delta_n} \mathbb{E}_\nu(1 - \Psi_n) \leqslant Le^{-(C+4)n\varepsilon_n^2}.$$

Then if $\Pi(\cdot|X_1, \dots, X_n)$ is the posterior distribution from (6) we have, for every $M \geqslant M_0$,

$$\Pi(\nu : d(\nu, \nu_0) \geqslant M\delta_n | X_1, \dots, X_n) \to 0$$

as $n \to \infty$ in $\mathbb{P}_{\nu_0}^{\mathbb{N}}$-probability.

As in previously studied 'inverse problems' settings [30, 28, 24], to apply this proposition with a metric d different from the Hellinger distance $h(\mathbb{P}_\nu, \mathbb{P}_{\nu_0})$ requires new approaches to the construction of frequentist tests, and as in these references we use tools from 'concentration of measure' theory put forward in [17], where we initially choose for d the weak (or 'robust') metric induced by the norm $\|\cdot\|_{\mathbb{H}(\delta)}$ of

$$\mathbb{H}(\delta) = \left\{f : \|f\|_{\mathbb{H}(\delta)}^2 = \sum_{l,k} 2^{-l} l^{-2\delta} \langle f, \psi_{lk}\rangle^2 < \infty\right\}, \quad \delta > 1/2, \qquad (46)$$

a negative order Sobolev space. Contraction rates in stronger norms will then be deduced from interpolation arguments. Before doing so, however, we need to calculate KL-divergences for the observation scheme relevant in our context, and show that they can be bounded in terms of the distance of their Lévy measures.

Lemma 18. *Let $D > 0$ such that $e^{-D} \leqslant d\nu/d\Lambda \leqslant e^D$ and $e^{-D} \leqslant d\nu_0/d\Lambda \leqslant e^D$ on I. Then there exists $K_D > 0$ such that*

$$K(\mathbb{P}_{\nu_0}, \mathbb{P}_\nu) = \int_I \log \frac{d\mathbb{P}_{\nu_0}}{d\mathbb{P}_\nu} d\mathbb{P}_{\nu_0} \leqslant K_D \|\nu - \nu_0\|_{L^2}^2,$$

$$\mathrm{Var}_{\mathbb{P}_{\nu_0}}\left(\log\frac{d\mathbb{P}_\nu}{d\mathbb{P}_{\nu_0}}\right)\leqslant\int_I\left(\log\frac{d\mathbb{P}_\nu}{d\mathbb{P}_{\nu_0}}\right)^2 d\mathbb{P}_{\nu_0}\leqslant K_D\|\nu-\nu_0\|_{L^2}^2.$$

Proof. We define the path $s\mapsto\exp(s(v-v_0)+v_0)=\nu^{(s)}$, $s\in[0,1]$, from ν_0 to ν and consider the function $f(s)=\int\log(d\mathbb{P}_{\nu^{(s)}}/d\mathbb{P}_{\nu_0})d\mathbb{P}_{\nu_0}$. Observing $f(0)=0$ a Taylor expansion at $s=0$ yields some $s\in[0,1]$ such that $f(1)=f'(0)+\frac{1}{2}f''(s)$. By the upper and lower bounds on the Lévy densities the differentiation may be performed under the integral and we obtain

$$\begin{aligned}\int\log\frac{d\mathbb{P}_{\nu_0}}{d\mathbb{P}_\nu}d\mathbb{P}_{\nu_0}&=-\int\frac{d\frac{d}{ds}\,\mathbb{P}_{\nu^{(s)}}}{d\,\mathbb{P}_{\nu^{(s)}}}\bigg|_{s=0}+\frac{1}{2}\frac{d\frac{d^2}{ds^2}\,\mathbb{P}_{\nu^{(s)}}}{d\,\mathbb{P}_{\nu^{(s)}}}-\frac{1}{2}\Big(\frac{d\frac{d}{ds}\,\mathbb{P}_{\nu^{(s)}}}{d\,\mathbb{P}_{\nu^{(s)}}}\Big)^2d\mathbb{P}_{\nu_0}\\
&=-\int A_{\nu_0}(v-v_0)d\mathbb{P}_{\nu_0}\\
&\quad-\frac{1}{2}\int A_{\nu^{(s)}}((v-v_0)^2)+A_{\nu^{(s)}}(v-v_0,v-v_0)-(A_{\nu^{(s)}}(v-v_0))^2d\mathbb{P}_{\nu_0}\\
&=-\frac{1}{2}\int A_{\nu^{(s)}}((v-v_0)^2)+A_{\nu^{(s)}}(v-v_0,v-v_0)-(A_{\nu^{(s)}}(v-v_0))^2d\mathbb{P}_{\nu_0}\\
&\lesssim\|A_{\nu^{(s)}}((v-v_0)^2)\|_{L^1(\mathbb{P}_{\nu^{(s)}})}+\|A_{\nu^{(s)}}(v-v_0,v-v_0)\|_{L^1(\mathbb{P}_{\nu^{(s)}})}\\
&\quad+\|A_{\nu^{(s)}}(v-v_0)\|_{L^2(\mathbb{P}_{\nu^{(s)}})}^2,\end{aligned}$$

where the last step contains a change of measure from $\mathbb{P}_{\nu_0}$ to $\mathbb{P}_{\nu^{(s)}}$ such that we may now apply Lemma 26

$$\begin{aligned}\int\log\frac{d\mathbb{P}_{\nu_0}}{d\mathbb{P}_\nu}d\mathbb{P}_{\nu_0}&\lesssim\|(v-v_0)^2\|_{L^1(\nu^{(s)})}+\|v-v_0\|_{L^1(\nu^{(s)})}^2+\|v-v_0\|_{L^2(\nu^{(s)})}^2\\
&\lesssim\|v-v_0\|_{L^2(\nu^{(s)})}^2\lesssim\|v-v_0\|_{L^2}^2\lesssim\|\nu-\nu_0\|_{L^2}^2.\end{aligned}$$

For the second inequality we consider the following function g and its derivatives

$$\begin{aligned}g(s)&=\int\left(\log\frac{d\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu_0}}\right)^2d\mathbb{P}_{\nu_0},\\
g'(s)&=\int2\Big(\log\frac{d\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu_0}}\Big)\frac{d\frac{d}{ds}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}d\mathbb{P}_{\nu_0},\\
g''(s)&=\int2\Big(\log\frac{d\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu_0}}\Big)\bigg(\Big(\frac{d\frac{d}{ds}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}\Big)^2+\frac{d\frac{d^2}{ds^2}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}-\Big(\frac{d\frac{d}{ds}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}\Big)^2\bigg)d\mathbb{P}_{\nu_0}\\
&=\int2\Big(\log\frac{d\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu_0}}\Big)\frac{d\frac{d^2}{ds^2}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}d\mathbb{P}_{\nu_0}.\end{aligned}$$

Observing $g(0)=g'(0)=0$ we obtain by a Taylor expansion $g(1)=g''(s)$ for some $s\in[0,1]$ and thus

$$\begin{aligned}&\int\left(\log\frac{d\mathbb{P}_\nu}{d\mathbb{P}_{\nu_0}}\right)^2d\mathbb{P}_{\nu_0}\\
&\quad=\int2\Big(\log\frac{d\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu_0}}\Big)\frac{d\frac{d^2}{ds^2}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}d\mathbb{P}_{\nu_0}\lesssim\int\Big|\frac{d\frac{d^2}{ds^2}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}\Big|d\mathbb{P}_{\nu^{(s)}}\end{aligned}$$

$$\lesssim \|A_{\nu^{(s)}}((v-v_0)^2)\|_{L^1(\mathbb{P}_{\nu^{(s)}})} + \|A_{\nu^{(s)}}(v-v_0, v-v_0)\|_{L^1(\mathbb{P}_{\nu^{(s)}})}$$
$$\lesssim \|(v-v_0)^2\|_{L^1(\nu^{(s)})} + \|v-v_0\|^2_{L^1(\nu^{(s)})}$$
$$\lesssim \|v-v_0\|^2_{L^2} + \|v-v_0\|^2_{L^1} \lesssim \|v-v_0\|^2_{L^2} \lesssim \|\nu-\nu_0\|^2_{L^2}. \qquad \square$$

Assumption 19. *The intensity λ of ν satisfies $\lambda < \pi/\Delta$.*

For Lévy processes on $\mathbb{R}$ the Lévy measure can be identified by taking the complex logarithm of the characteristic function of $\mathbb{P}_\nu$ in such a way that the resulting function is continuous. (This is known as the distinguished logarithm.) For Lévy processes on a circle the characteristic function is defined only on the integer lattice and a continuous version of the logarithm cannot be defined. However, this problem can be resolved by assuming $\lambda < \pi/\Delta$ since then the exponent in the Lévy-Khintchine representation always coincides with the principle branch of the logarithm of the characteristic function, ensuring identifiability. This condition is sharp as the following examples show.

Examples. By the Lévy–Khintchine representation (3) we see that $\mathbb{P}_{\nu_1}$ and $\mathbb{P}_{\nu_2}$ coincide if $\mathcal{F}\nu_1(k)$ equals $\mathcal{F}\nu_2(k)$ modulo multiples of $2\pi i/\Delta$ for all $k \in \mathbb{Z}$.

1. For $\nu_1 = (\pi/\Delta)\delta_{1/4}$ and $\nu_2 = (\pi/\Delta)\delta_{-1/4}$ we have $\mathcal{F}\nu_1(k) = \mathcal{F}\nu_2(k)$ for all even k and $\mathcal{F}\nu_1(k) = \mathcal{F}\nu_2(k) + (2\pi/\Delta)i$ or $\mathcal{F}\nu_1(k) = \mathcal{F}\nu_2(k) - (2\pi/\Delta)i$ for all odd k. This shows that the intensity bound in Assumption 19 is sharp.
2. For $\nu_1(x) = (4\pi/\Delta)(\sin(2\pi x))_+$ and $\nu_2(x) = (4\pi/\Delta)(\sin(2\pi x))_-$ we have $\mathcal{F}\nu_1(1) = \mathcal{F}\nu_2(1) + (2\pi/\Delta)i$ and $\mathcal{F}\nu_1(-1) = \mathcal{F}\nu_2(-1) - (2\pi/\Delta)i$. For all other k it can be shown that $\mathcal{F}\nu_1(k) = \mathcal{F}\nu_2(k)$. This demonstrates that there exist nonidentifiable Lévy measures which are absolutely continuous with respect to Lebesgue measure.

Lemma 20. *For any $c, x, D > 0, \delta > 1/2$, and integer $K \geqslant 2$, there exist constants $R_1(c, D, \Delta) > 0$, $R_2(c, D, \Delta) > 0$ and an estimator $\widehat{\nu} = \widehat{\nu}(X_1, \ldots, X_n)$ such that*

$$\sup_{\nu: \|\nu\|_{L^1} < \pi/\Delta, \|\nu\|_{L^2} \leqslant c} \mathbb{P}^{\mathbb{N}}_\nu \left(\|\widehat{\nu} - \nu\|_{\mathbb{H}(\delta)} > R_1 \left(\frac{\sqrt{\log K} + x}{\sqrt{n}} + \frac{1}{\sqrt{K}} \right) \right)$$
$$\leq e^{-Dx^2} + \frac{e^{-\frac{nR_2}{\log K}}}{R_2}. \qquad (47)$$

Proof. We first show the above concentration inequality with $\|\widehat{\nu} - \nu\|_{\mathbb{H}(\delta)}$ replaced by $|\widehat{\lambda} - \lambda|$, where $\lambda = \int_I \nu = (\mathcal{F}\nu)(0)$ is the intensity and $\widehat{\lambda}$ is an estimator defined as follows: Let $\varphi_n(k) = (1/n)\sum_{j=1}^n \exp\{2\pi i k X_j\}$ be the empirical characteristic function, set $\Phi_n(k) = \Delta^{-1} \log \varphi_n(k)$ for $\varphi_n(k) \neq 0$ and $\Phi_n(k) = 0$ otherwise, where we take the principal branch of the complex logarithm. For $K \geqslant 2$ consider the estimator $\widehat{\lambda} = -(1/K)\sum_{k=1}^K \operatorname{Re} \Phi_n(k)$. The Lévy–Khintchine representation (3) yields $\Phi_\nu(k) := \Delta^{-1} \log \varphi_\nu(k) = \mathcal{F}\nu(k) - \lambda$, where thanks to the restriction $\|\nu\|_{L^1} < \pi/\Delta$ the imaginary part on the r.h.s. lies in $(-\pi/\Delta, \pi/\Delta)$

and hence log is the logarithm in the principle branch. We obtain

$$\begin{aligned}\widehat{\lambda}-\lambda &= -\frac{1}{K}\sum_{k=1}^{K}\operatorname{Re}(\Phi_n(k)-\Phi_\nu(k)) - \frac{1}{K}\sum_{k=1}^{K}(\operatorname{Re}\Phi_\nu(k)+\lambda)\\ &= -\frac{1}{K}\sum_{k=1}^{K}\operatorname{Re}(\Phi_n(k)-\Phi_\nu(k)) - \frac{1}{K}\sum_{k=1}^{K}\operatorname{Re}\mathcal{F}\nu(k) \qquad (48)\end{aligned}$$

In order to linearise the first term in previous equation we define the event

$$A_n = \left\{\left\|\frac{\varphi_n-\varphi_\nu}{\varphi_\nu}\right\|_K \leqslant \frac{1}{2}\right\} \qquad \text{with } \|f\|_K = \sup_{|k|\leqslant K}|f(k)|.$$

It holds $|\log(1+z)-z| \leqslant 2|z|^2$ for $|z|\leqslant 1/2$. Thus we have on the event A_n for $|k|\leqslant K$

$$\begin{aligned}\Phi_n(k)-\Phi_\nu(k) &= \frac{1}{\Delta}\log\left(\frac{\varphi_n(k)-\varphi_\nu(k)}{\varphi_\nu(k)}+1\right)\\ &= \frac{1}{\Delta}\left\{\frac{\varphi_n(k)-\varphi_\nu(k)}{\varphi_\nu(k)} + O\left(\left|\frac{\varphi_n(k)-\varphi_\nu(k)}{\varphi_\nu(k)}\right|^2\right)\right\}.\end{aligned}$$

The first term in (48), up to linearisation, is purely stochastic and bounded by a term of the form

$$\frac{1}{\Delta K}\sum_{k=1}^{K}\frac{|\varphi_n(k)-\varphi_\nu(k)|}{|\varphi_\nu(k)|}.$$

Since $\|\nu\|_1 < \pi/\Delta$ we know that $\sup_k|1/\varphi_\nu(k)| \leqslant c'$ for some constant $c' = c'(\Delta)$. For the numerator we consider the $4K+4$ random variables

$$\begin{aligned}&\pm\operatorname{Re}(\varphi_n(-K)-\varphi_\nu(-K)),\ldots,\pm\operatorname{Re}(\varphi_n(K)-\varphi_\nu(K)),\\ &\pm\operatorname{Im}(\varphi_n(-K)-\varphi_\nu(-K)),\ldots,\pm\operatorname{Im}(\varphi_n(K)-\varphi_\nu(K))\end{aligned}$$

and denote them by Z_j with $j = 1,\ldots,4K+4$. These have bounded differences with constant $c^2 = 4/n$ which follows from using example b) before Theorem 3.3.14 in [18] and observing that $e^{2\pi ik(\cdot)}$ are uniformly bounded by 1. Applying this theorem we have $Ee^{\lambda Z_j} \leqslant e^{\lambda^2c^2/8} = e^{\lambda^2/(2n)}$. By Lemma 2.3.4 in [18] we further obtain that

$$E\left[\max_{j=1,\ldots,4K+4} Z_j\right] \leqslant \sqrt{\frac{2}{n}\log(4K+4)}$$

and denoting $Z = \max_{|k|\leqslant K}|\varphi_n(k)-\varphi_\nu(k)|$ we have

$$\begin{aligned}E[Z] &\leqslant 2E\left[\max_{|k|\leqslant K}(\operatorname{Re}(\varphi_n(k)-\varphi_\nu(k)),\operatorname{Im}(\varphi_n(k)-\varphi_\nu(k)))\right]\\ &\leqslant \sqrt{\frac{8}{n}\log(4K+4)} \lesssim \sqrt{\frac{\log K}{n}}.\end{aligned}$$

For the concentration around the mean we observe that Z itself also has bounded differences with $c^2 = 4/n$ and applying Theorem 3.3.14 in [18] yields

$$\mathbb{P}(Z \geqslant EZ + t) \leqslant e^{-2t^2/c^2} = e^{-nt^2/2}, \quad \mathbb{P}(Z \leqslant EZ - t) \leqslant e^{-nt^2/2}.$$

This shows that the linearisation of the first term in (48) is bounded by a multiple of $(\sqrt{\log K} + x)/\sqrt{n}$. On A_n we can bound the remainder in the linearisation by a multiple of the same quantity. For $n/\log K$ large enough EZ is smaller than $1/(4c')$ and we can bound $\mathbb{P}(A_n^c)$ by $\exp(-R_2 n) \leqslant \exp(-R_2 n/\log K)$ using the concentration of Z. The bound $\mathbb{P}(A_n^c) \leqslant (1/R_2)\exp(-R_2 n/\log K)$ for all n and K is obtained by choosing a possibly smaller constant R_2.

For the bias we bound, using the Cauchy–Schwarz inequality,

$$\left|\frac{1}{K}\sum_{k=1}^{K} \operatorname{Re} \mathcal{F}\nu(k)\right| \leqslant K^{-1/2}\sqrt{\sum_{k=1}^{K} |\mathcal{F}\nu(k)|^2} \leq \frac{\|\nu\|_{L^2}}{\sqrt{K}},$$

which explains the second regime in the inequality in Lemma 20.

Now to estimate ν we first estimate $\mathcal{F}\nu(k), k \neq 0$, by $\mathcal{F}\widehat{\nu}(k) = (\Phi_n(k) + \widehat{\lambda})1_{[-K,K]}(k)$, where K is a spectral cut-off parameter. By standard theory of Sobolev spaces on the unit circle, an equivalent norm on $\mathbb{H}(\delta)$ is given by

$$\|f\|'_{\mathbb{H}(\delta)} = \sum_k |\mathcal{F}f(k)|^2 k^{-1}(\log(e+k))^{-2\delta}.$$

Using that $\sum_k k^{-1}(\log(e+k))^{-2\delta}$ converges for $\delta > 1/2$ we obtain

$$\begin{aligned}
\|\widehat{\nu} - \nu\|^2_{\mathbb{H}(\delta)} &= \sum_k k^{-1}(\log(e+k))^{-2\delta}|\mathcal{F}\widehat{\nu}(k) - \mathcal{F}\nu(k)|^2 \\
&= \sum_{|k|\leqslant K} k^{-1}(\log(e+k))^{-2\delta}|\Phi_n(k) - \Phi_\nu(k) + \widehat{\lambda} - \lambda|^2 \\
&\qquad + \sum_{|k|>K} k^{-1}(\log k)^{-2\delta}|\mathcal{F}\nu(k)|^2 \\
&\lesssim (\widehat{\lambda} - \lambda)^2 + \sum_{|k|\leqslant K} k^{-1}(\log(e+k))^{-2\delta}|\Phi_n(k) - \Phi_\nu(k)|^2 \\
&\qquad + \sum_{|k|>K} k^{-1}(\log(e+k))^{-2\delta}|\mathcal{F}\nu(k)|^2 \\
&\lesssim (\widehat{\lambda} - \lambda)^2 + \max_{|k|\leqslant K}|\Phi_n(k) - \Phi_\nu(k)|^2 + \|\nu\|^2_{L^2}/K,
\end{aligned}$$

which, repeating the above, gives the same bounds as those obtained for error of the intensity $\widehat{\lambda} - \lambda$. □

The proof of the following proposition is contained in Section 8.

Proposition 21. *Denote* $\overline{\mathcal{V}} = \{\nu \in \mathcal{V} : \|\nu\|_{L^1} < \pi/\Delta$ *and* $\|\nu\|_{L^2} \leqslant c\}$ *for some* $c, \Delta > 0$. *Let* ε_n *be such that* $\sqrt{(\log n)/n} \lesssim \varepsilon_n$ *and* $\varepsilon_n = o(1/\sqrt{\log n})$. *Then for*

$\nu_0 \in \overline{\mathcal{V}}$ there exists a sequence of tests (indicator functions) $\Psi_n \equiv \Psi(X_1, \dots, X_n)$ such that for every $C > 0$, there exist $M = M(C, c, \Delta) > 0$ such that for all n large enough

$$E_{\nu_0}[\Psi_n] \to_{n\to\infty} 0, \qquad \sup_{\nu\in\overline{\mathcal{V}}:\|\nu-\nu_0\|_{\mathbb{H}(\delta)}\geqslant M\varepsilon_n} E_\nu[1-\Psi_n] \leqslant 2e^{-(C+4)n\varepsilon_n^2}.$$

Proposition 22. *Suppose we have for some constants $c, C, D > 0$, for a sequence ε_n such that $\sqrt{(\log n)/n} \lesssim \varepsilon_n$ and $\varepsilon_n = o(1/\sqrt{\log n})$, for ν_0 such that $e^{-D} \leqslant d\nu_0/d\Lambda \leqslant e^D$, for some prior Π on a set $\{\nu \in \mathcal{V} : e^{-D} \leqslant d\nu/d\Lambda \leqslant e^D\}$ of Lévy measures bounded from above and away from zero, for n large enough and with K_D from Lemma 18 that*

$$\Pi\left(\nu \in \mathcal{V} : \|\nu - \nu_0\|_{L^2} \leqslant \varepsilon_n/\sqrt{K_D}\right) \geqslant e^{-Cn\varepsilon_n^2} \tag{49}$$

and that

$$\Pi(\nu \in \mathcal{V} : \|\nu\|_{L^1} \geqslant \pi/\Delta \text{ or } \|\nu\|_{L^2} > c) \leqslant Le^{-(C+4)n\varepsilon_n^2}. \tag{50}$$

If $\Pi(\cdot|X_1, \dots, X_n)$ is the posterior distribution from (6), then there exists M_0 such that for every $M \geqslant M_0$, as $n \to \infty$ and in $\mathbb{P}_{\nu_0}^{\mathbb{N}}$-probability,

$$\Pi(\nu : \|\nu - \nu_0\|_{\mathbb{H}(\delta)} \geqslant M\varepsilon_n | X_1, \dots, X_n) \to 0.$$

Proof. Starting with Proposition 17 we replace the condition on the Kullback–Leibler neighbourhood by a condition on a L^2 neighbourhood using Lemma 18. Further we choose $\mathcal{V}_n = \{\nu \in \mathcal{V} : \|\nu\|_{L^1} < \pi/\Delta, \|\nu\|_{L^2} \leqslant c\}$, $d(\nu, \nu_0) = \|\nu - \nu_0\|_{\mathbb{H}(\delta)}$ and $\delta_n = \varepsilon_n$. The existence of tests follows by Proposition 21. ☐

Proposition 23. *Grant Assumption 1 for some $s > 5/2$, $B > 0$, and set*

$$\varepsilon_n = n^{-s/(2s+1)}(\log n)^{1/2}. \tag{51}$$

For the choice $J = J_n$ with $2^{J_n} \approx n^{1/(2s+1)}$ the prior (10) satisfies for n large enough the small ball probability condition (49).

The above proposition is proved in Section 9. We now turn to the proof of Proposition 10. When modelling an s-regular function ν, and when $\nu_0 \in C^s$ as well, Proposition 23 shows (49) for the choice $\varepsilon_n \approx n^{-s/(2s+1)}(\log n)^{1/2}$, and so we obtain the lower bound on the small ball probabilities. By Assumption 1 we have $\|\nu\|_{L^1} < \pi/\Delta$ and we also see that the prior concentrates almost surely on a fixed L^∞- (and then also L^2-) ball since $\|v\|_\infty^2 \lesssim \sum_l 2^{-l/2}$, thus (50) holds for Π too. As a consequence we obtain

$$\Pi(\nu : \|\nu - \nu_0\|_{\mathbb{H}(\delta)} \leqslant M\varepsilon_n | X_1, \dots, X_n) \to^{\mathbb{P}_{\nu_0}^{\mathbb{N}}} 1. \tag{52}$$

Restricting to this event we can further bound L^2-distances: by $v_0 = \log \nu_0 \in C^s$ and (8) and using Lemma 24 below (and the remark before it) we have on an event with posterior probability tending to one

$$\|\nu - \nu_0\|_{L^2}^2 \lesssim \|v - v_0\|_{L^2}^2 = \sum_{l<J}\sum_k \langle v - v_0, \psi_{lk}\rangle^2 + \sum_{l\geqslant J,k} \langle v_0, \psi_{lk}\rangle^2$$

$$\leqslant 2^J J^{2\delta}\|v-v_0\|^2_{\mathbb{H}(\delta)} + O(2^{-2Js}) \lesssim 2^J J^{2\delta}\|\nu-\nu_0\|^2_{\mathbb{H}(\delta)} + O(2^{-2Js}) \lesssim 2^J J^{2\delta}\varepsilon_n^2$$

so that, as $n\to\infty$,

$$\Pi(\nu : \|\nu-\nu_0\|_{L^2} \geqslant C2^{J/2}J^\delta\varepsilon_n|X_1,\dots,X_n) \to^{\mathbb{P}^{\mathbb{N}}_{\nu_0}} 0$$

and further using that with posterior probability tending to one

$$\|\nu-\nu_0\|_\infty \lesssim \|v-v_0\|_\infty = \sum_{l<J} 2^{l/2}\max_k |\langle v-v_0,\psi_{lk}\rangle| + \sum_{l\geqslant J} 2^{l/2}\max_k |\langle v_0,\psi_{lk}\rangle|$$
$$\lesssim 2^{J/2}\|v-v_0\|_{L^2} + O(2^{-Js}) \lesssim 2^J J^\delta \varepsilon_n$$

which also implies that

$$\Pi(\nu : \|\nu-\nu_0\|_\infty \geqslant C2^{J}J^\delta\varepsilon_n|X_1,\dots,X_n) \to^{\mathbb{P}^{\mathbb{N}}_{\nu_0}} 0.$$

For $\delta > 1/2$ we have posterior contraction with rates $\varepsilon_n^{L^2}$ and $\varepsilon_n^{L^\infty}$ in L^2 and L^∞, respectively, where

$$\varepsilon_n^{L^2} = n^{-\frac{s-1/2}{2s+1}}(\log n)^{1/2+\delta} \qquad \text{and} \qquad \varepsilon_n^{L^\infty} = n^{-\frac{s-1}{2s+1}}(\log n)^{1/2+\delta}.$$

Estimating $\|v-v_0\|_{L^p} \lesssim \|\nu-\nu_0\|_p$ for $p=2,\infty$ implies Proposition 10. Moreover, using $(\varepsilon_n^{L^p})^p \leqslant (\varepsilon_n^{L^\infty})^{p-2}(\varepsilon_n^{L^2})^2$ we obtain for contraction in L^p the rate

$$\varepsilon_n^{L^p} = n^{-\frac{s+1/p-1}{2s+1}}(\log n)^{1/2+\delta}. \tag{53}$$

It remains to prove Lemma 24. Let us introduce the spaces

$$\mathbb{B}(\delta) = \left\{ f : \|f\|^2_{\mathbb{B}(\delta)} = \sum_{l,k} 2^l l^{2\delta}\langle f,\psi_{lk}\rangle^2 < \infty \right\}, \quad \delta > 1/2,$$

which are equal to the (logarithmically refined) Sobolev spaces $H^{1/2,\delta}(I)$. As in Proposition 4.3.12 in [18] one shows that $\mathbb{H}(\delta)$ is the topological dual space of $\mathbb{B}(\delta)$. We further see directly from the definition of the prior that $v = \log\nu$ satisfies

$$\|v\|^2_{\mathbb{B}(\delta')} = \sum_{l,k} 2^l l^{2\delta'} a_l^2 u_{lk}^2 \leqslant \sum_{l\leqslant J} l^{2\delta'-4} \leqslant c, \quad \text{any } \delta' < 3/2,$$

and one further shows that also $\|\nu\|_{\mathbb{B}(\delta')} = \|e^v\|_{\mathbb{B}(\delta')}$ is bounded by a fixed constant Π-almost surely (e.g., using the modulus of continuity characterisation of the $\mathbb{B}(\delta)$-norm, proved as in Section 4.3.5 in [18]). This justifies the application of the following lemma with $1/2 < \delta < \delta' < 3/2$ in the above estimate. The lemma is proved in Section 10.

Lemma 24. *a) For any $\nu,\nu_0 \in \mathbb{B}(\delta), \delta > 1/2$, such that ν,ν_0 are bounded away from zero on I and such that $\|\nu-\nu_0\|_{\mathbb{B}(\delta)} \to 0$, we have $\|\log\nu - \log\nu_0\|_{\mathbb{H}(\delta)} \lesssim \|\nu-\nu_0\|_{\mathbb{H}(\delta)}$.*

b) If $\|\nu-\nu_0\|_{\mathbb{H}(\delta)} \to 0$ and ν,ν_0 are uniformly bounded in $\mathbb{B}(\delta')$, then for any $\delta < \delta'$ we have $\|\nu-\nu_0\|_{\mathbb{B}(\delta)} \to 0$.

6. Proof of Proposition 11

Using the definition of $S_n(\nu)$ and the formula for the posterior distribution we obtain

$$\begin{aligned} &E^{\Pi^{D_{n,M}}}\left[e^{t\sqrt{n}F(\nu)}\Big|X_1,\dots,X_n\right] \\ &\quad= E^{\Pi^{D_{n,M}}}\left[e^{S_n(\nu)-t\sqrt{n}\int A_{\nu_0}(v-v_0)A_{\nu_0}(\eta)d\mathbb{P}_{\nu_0}}\Big|X_1,\dots,X_n\right] \\ &\quad= \frac{\int_{D_{n,M}} e^{S_n(\nu)-t\sqrt{n}\int A_{\nu_0}(v-v_0)A_{\nu_0}(\eta)d\mathbb{P}_{\nu_0}+\ell_n(\nu)}d\Pi(\nu)}{\int_{D_{n,M}} e^{\ell_n(\nu)}d\Pi(\nu)}. \end{aligned} \tag{54}$$

By Assumption 1 we have $s > 5/2$ so that by Remark 28 condition (63) implies condition (64) and we conclude that the entire Assumption 27 is satisfied. By Lemma 29, the choice of J as in (11), Assumption 27 and the L^p-contraction rates (53) derived from Proposition 10 we have that Assumption 25 is satisfied. In Section 6.2 we prove that under Assumption 25

$$\begin{aligned} &-t\sqrt{n}\int A_{\nu_0}(v-v_0)A_{\nu_0}(\eta)d\mathbb{P}_{\nu_0}+\ell_n(\nu) \\ &\quad= \frac{t^2}{2}\|A_{\nu_0}(\eta)\|^2_{L^2(\mathbb{P}_{\nu_0})} - \frac{t}{\sqrt{n}}\sum_{k=1}^n A_{\nu_0}(\eta)(X_k) + \ell_n(\nu_t) + r_n'(\nu), \end{aligned} \tag{55}$$

where $\sup_{\nu\in D_{n,M}}|r_n'(\nu)| = o_{\mathbb{P}^{\mathbb{N}}_{\nu_0}}(1)$ with the nonstochastic null sequence implicit in the $o_{\mathbb{P}^{\mathbb{N}}_{\nu_0}}$ notation uniform in $\eta\in\mathcal{H}_n$. Since the first two terms on the right hand side do not depend on ν they can be taken outside the posterior integral in (54) so that

$$\begin{aligned} &E^{\Pi^{D_{n,M}}}\left[e^{t\sqrt{n}F(\nu)}\Big|X_1,\dots,X_n\right] \\ &= \exp\left\{\frac{t^2}{2}\|A_{\nu_0}(\eta)\|^2_{L^2(\mathbb{P}_{\nu_0})} - \frac{t}{\sqrt{n}}\sum_{k=1}^n A_{\nu_0}(\eta)(X_k)\right\} \\ &\qquad\times\frac{\int_{D_{n,M}} e^{S_n(\nu)+\ell_n(\nu_t)+r_n'(\nu)}d\Pi(\nu)}{\int_{D_{n,M}} e^{\ell_n(\nu)}d\Pi(\nu)}. \end{aligned}$$

By the mean value theorem for integrals $r_n'(\nu)$ can be replaced by r_n not depending on ν with $|r_n| \leqslant \sup_{\nu\in D_{n,M}}|r_n'(\nu)| = o_{\mathbb{P}^{\mathbb{N}}_{\nu_0}}(1)$ in the above display finishing the proof of the proposition.

In order to prove the crucial perturbation approximation (55), we first need to obtain formulas for the directional derivatives of the likelihood function, which is done in the next section.

6.1. Directional derivatives of the likelihood function

We fix a positive and absolutely continuous Lévy measure $\nu_0 = \lambda_0\mu_0$ with corresponding infinitely divisible distribution $\mathbb{P}_{\nu_0}$. We set $v_0 = \log\nu_0$ so that

$\nu_0 = \exp v_0$ and parametrise a path away from ν_0 as

$$\nu^{(s)} = \exp(s(v - v_0) + v_0),\ \ s \in [0, 1].$$

The resulting compound Poisson measure can be identified in the Fourier domain as

$$\begin{aligned}
&\mathcal{F}\mathbb{P}_{\nu^{(s+h)}}(k) = \exp\left(\Delta \int (e^{2\pi ikx} - 1) d\nu^{(s+h)}(x)\right) \\
&= \exp\left(\Delta \int (e^{2\pi ikx} - 1)\nu^{(s)}(x) e^{h(v-v_0)(x)} dx\right) \\
&= \exp\left(\Delta \int (e^{2\pi ikx} - 1)\nu^{(s),h}(x) dx + \Delta \int (e^{2\pi ikx} - 1)\nu^{(s)}(x) dx\right) \\
&= \mathcal{F}\mathbb{P}_{\nu^{(s)}}(k) \times \exp\left(\Delta \int (e^{2\pi ikx} - 1)\nu^{(s),h}(x) dx\right),
\end{aligned}$$

where $\nu^{(s),h}(x) := \nu^{(s)}(x)\left(e^{h(v-v_0)(x)} - 1\right)$ is a finite signed measure on I. One checks by the usual properties of convolution and definition of e^z that the second factor in the last product is the Fourier transform of the finite signed measure

$$e^{-\Delta\nu^{(s),h}(I)} \sum_{k=0}^{\infty} \frac{\Delta^k (\nu^{(s),h})^{*k}}{k!}$$

and so we conclude by injectivity of $\mathcal{F}$ that

$$\mathbb{P}_{\nu^{(s+h)}} = e^{-\Delta\nu^{(s),h}(I)} \sum_{k=0}^{\infty} \frac{\Delta^k (\nu^{(s),h})^{*k}}{k!} * \mathbb{P}_{\nu^{(s)}}. \tag{56}$$

Let Λ denote the Lebesgue (probability) measure on I. We observe that the resulting compound Poisson measure is of the form $\mathbb{P}_\Lambda = e^{-\Delta}\delta_0 + (1 - e^{-\Delta})\Lambda$. Both $\mathbb{P}_{\nu^{(s)}}$ and $\mathbb{P}_{\nu^{(s+h)}}$ are absolutely continuous with respect to $\mathbb{P}_\Lambda$. We will now determine the first five derivatives of $d\,\mathbb{P}_{\nu^{(s)}}/d\,\mathbb{P}_\Lambda$. To this end we expand (56) in terms of h. We start with the factor in front of the sum and expand

$$\begin{aligned}
&e^{-\Delta\nu^{(s),h}(I)} = \exp\left(-\Delta \int (e^{h(v-v_0)(x)} - 1) d\nu^{(s)}\right) \\
&= \exp\left(-\Delta \int h(v - v_0)(x) + \frac{h^2}{2}(v - v_0)^2(x) + \frac{h^3}{6}(v - v_0)^3(x) + O(h^4) d\nu^{(s)}\right) \\
&= 1 - \Delta \int h(v - v_0)(x) + \frac{h^2}{2}(v - v_0)^2(x) + \frac{h^3}{6}(v - v_0)^3(x) d\nu^{(s)} \\
&\quad + \frac{\Delta^2}{2}\left(\int h(v - v_0)(x) + \frac{h^2}{2}(v - v_0)^2(x) + \frac{h^3}{6}(v - v_0)^3(x) d\nu^{(s)}\right)^2 \\
&\quad - \frac{\Delta^3}{6}\left(\int h(v - v_0)(x) + \frac{h^2}{2}(v - v_0)^2(x) + \frac{h^3}{6}(v - v_0)^3(x) d\nu^{(s)}\right)^3 + O(h^4)
\end{aligned}$$

$$
\begin{aligned}
&= 1 - \Delta h \int v - v_0 d\nu^{(s)} - \Delta \frac{h^2}{2} \int (v - v_0)^2 d\nu^{(s)} - \Delta \frac{h^3}{6} \int (v - v_0)^3 d\nu^{(s)} \\
&\quad + \frac{\Delta^2}{2} h^2 \left(\int v - v_0 d\nu^{(s)} \right)^2 + \frac{\Delta^2}{2} h^3 \int v - v_0 d\nu^{(s)} \int (v - v_0)^2 d\nu^{(s)} \\
&\quad - \frac{\Delta^3}{6} h^3 \left(\int v - v_0 d\nu^{(s)} \right)^3 + O(h^4).
\end{aligned}
$$

From the definition of $\nu^{(s),h}$ we observe that $(\nu^{(s),h})^{*k} = O(h^k)$. Using (56) we obtain

$$
\begin{aligned}
&\frac{d\,\mathbb{P}_{\nu^{(s+h)}}}{d\,\mathbb{P}_\Lambda} - \frac{d\,\mathbb{P}_{\nu^{(s)}}}{d\,\mathbb{P}_\Lambda} = \frac{d}{d\,\mathbb{P}_\Lambda} \left\{ e^{-\Delta \nu^{(s),h}(I)} \sum_{k=0}^{\infty} \frac{\Delta^k (\nu^{(s),h})^{*k}}{k!} * \mathbb{P}_{\nu^{(s)}} - \mathbb{P}_{\nu^{(s)}} \right\} \\
&= \frac{d}{d\,\mathbb{P}_\Lambda} \left\{ \left(1 - \Delta h \int v - v_0 d\nu^{(s)} - \Delta \frac{h^2}{2} \int (v - v_0)^2 d\nu^{(s)} \right. \right. \\
&\quad - \Delta \frac{h^3}{6} \int (v - v_0)^3 d\nu^{(s)} + \frac{\Delta^2}{2} h^2 \left(\int v - v_0 d\nu^{(s)} \right)^2 \\
&\quad \left. + \frac{\Delta^2}{2} h^3 \int v - v_0 d\nu^{(s)} \int (v - v_0)^2 d\nu^{(s)} - \frac{\Delta^3}{6} h^3 \left(\int v - v_0 d\nu^{(s)} \right)^3 + O(h^4) \right) \\
&\left(\delta_0 + \Delta \nu^{(s)} (e^{h(v - v_0)(x)} - 1) + \frac{\Delta^2}{2} (\nu^{(s)} (e^{h(v - v_0)(x)} - 1))^{*2} \right. \\
&\left. \left. + \frac{\Delta^3}{6} (\nu^{(s)} (e^{h(v - v_0)(x)} - 1))^{*3} + O(h^4) \right) * \mathbb{P}_{\nu^{(s)}} - \mathbb{P}_{\nu^{(s)}} \right\}.
\end{aligned}
$$

To find the first derivative we gather all terms that are linear in h and obtain

$$
\begin{aligned}
&\frac{d}{d\,\mathbb{P}_\Lambda} \left\{ \left(\Delta \nu^{(s)} h (v - v_0) - \Delta h \int v - v_0 d\nu^{(s)} \delta_0 \right) * \mathbb{P}_{\nu^{(s)}} \right\} \\
&\qquad = h\Delta \frac{d((\nu^{(s)}(v - v_0)) * \mathbb{P}_{\nu^{(s)}} - \int v - v_0 d\nu^{(s)}\, \mathbb{P}_{\nu^{(s)}})}{d\,\mathbb{P}_\Lambda}.
\end{aligned}
$$

This gives the first derivative

$$
\frac{d}{ds} \frac{d\,\mathbb{P}_{\nu^{(s)}}}{d\,\mathbb{P}_\Lambda} = \Delta \frac{d((\nu^{(s)}(v - v_0)) * \mathbb{P}_{\nu^{(s)}} - \int v - v_0 d\nu^{(s)}\, \mathbb{P}_{\nu^{(s)}})}{d\,\mathbb{P}_\Lambda}.
$$

Gathering all terms quadratic in h we find

$$
\begin{aligned}
&\frac{d}{d\,\mathbb{P}_\Lambda} \left\{ \left(\Delta \nu^{(s)} \frac{h^2}{2} (v - v_0)^2 + \frac{\Delta^2}{2} (\nu^{(s)} h (v - v_0))^{*2} \right. \right. \\
&\quad - \Delta^2 h \int v - v_0 d\nu^{(s)} \nu^{(s)} h (v - v_0) - \frac{\Delta h^2}{2} \int (v - v_0)^2 d\nu^{(s)} \delta_0 \\
&\quad \left. \left. + \frac{\Delta^2 h^2}{2} \left(\int v - v_0 d\nu^{(s)} \right)^2 \delta_0 \right) * \mathbb{P}_{\nu^{(s)}} \right\}
\end{aligned}
$$

$$= \frac{h^2}{2}\frac{d}{d\,\mathbb{P}_\Lambda}\Bigg\{\Bigg(\Delta\nu^{(s)}(v-v_0)^2 - \Delta\int(v-v_0)^2 d\nu^{(s)}\delta_0 + \Delta^2(\nu^{(s)}(v-v_0))^{*2}$$
$$- 2\Delta^2\int v - v_0 d\nu^{(s)}(v-v_0)\nu^{(s)} + \Delta^2\Big(\int v-v_0 d\nu^{(s)}\Big)^2\delta_0\Bigg) * \mathbb{P}_{\nu^{(s)}}\Bigg\}.$$

And this gives the second derivative

$$\frac{d^2}{ds^2}\frac{d\,\mathbb{P}_{\nu^{(s)}}}{d\,\mathbb{P}_\Lambda} = \frac{d}{d\,\mathbb{P}_\Lambda}\Bigg\{\Delta((v-v_0)^2\nu^{(s)}) * \mathbb{P}_{\nu^{(s)}} - \Delta\int(v-v_0)^2 d\nu^{(s)}\,\mathbb{P}_{\nu^{(s)}}$$
$$+ \Delta^2\left(((v-v_0)\nu^{(s)}) - \delta_0\int(v-v_0)d\nu^{(s)}\right)^{*2} * \mathbb{P}_{\nu^{(s)}}\Bigg\}.$$

Finally we gather all terms which are cubic in h. This yields

$$\frac{d}{d\,\mathbb{P}_\Lambda}\Bigg\{\Bigg(\Delta\nu^{(s)}\frac{h^3}{6}(v-v_0)^3 + \Delta^2\Big((\nu^{(s)}h(v-v_0)) * \Big(\nu^{(s)}\frac{h^2}{2}(v-v_0)^2\Big)\Big)$$
$$+ \frac{\Delta^3}{6}(\nu^{(s)}h(v-v_0))^{*3}$$
$$- \Delta^2 h\int v-v_0 d\nu^{(s)}\nu^{(s)}\frac{h^2}{2}(v-v_0)^2 - \Delta^3 h\int v-v_0 d\nu^{(s)}\frac{1}{2}(\nu^{(s)}h(v-v_0))^{*2}$$
$$+ \frac{h^2}{2}\Big(\Delta^3\Big(\int v-v_0 d\nu^{(s)}\Big)^2 - \Delta^2\int(v-v_0)^2 d\nu^{(s)}\Big)\nu^{(s)}h(v-v_0)$$
$$- \frac{h^3\Delta^3}{6}\Big(\int v-v_0 d\nu^{(s)}\Big)^3\delta_0$$
$$- \frac{h^3\Delta}{6}\int(v-v_0)^3 d\nu^{(s)}\delta_0 + \frac{h^3\Delta^2}{2}\int v-v_0 d\nu^{(s)}\int(v-v_0)^2 d\nu^{(s)}\delta_0\Bigg) * \mathbb{P}_{\nu^{(s)}}\Bigg\}.$$

In this way we obtain the third derivative

$$\frac{d^3}{ds^3}\frac{d\,\mathbb{P}_{\nu^{(s)}}}{d\,\mathbb{P}_\Lambda} = \frac{d}{d\,\mathbb{P}_\Lambda}\Bigg\{\Delta((v-v_0)^3\nu^{(s)}) * \mathbb{P}_{\nu^{(s)}} - \Delta\int(v-v_0)^3 d\nu^{(s)}\,\mathbb{P}_{\nu^{(s)}}$$
$$+ 3\Delta^2\left(((v-v_0)\nu^{(s)}) - \delta_0\int(v-v_0)d\nu^{(s)}\right)$$
$$* \left(((v-v_0)^2\nu^{(s)}) - \delta_0\int(v-v_0)^2 d\nu^{(s)}\right) * \mathbb{P}_{\nu^{(s)}}$$
$$+ \Delta^3\left(((v-v_0)\nu^{(s)}) - \delta_0\int(v-v_0)d\nu^{(s)}\right)^{*3} * \mathbb{P}_{\nu^{(s)}}\Bigg\}.$$

In a similar way we obtain for the fourth and fifth derivative

$$\frac{d^4}{ds^4}\frac{d\,\mathbb{P}_{\nu^{(s)}}}{d\,\mathbb{P}_\Lambda} = \frac{d}{d\,\mathbb{P}_\Lambda}\Bigg\{\Delta((v-v_0)^4\nu^{(s)}) * \mathbb{P}_{\nu^{(s)}} - \Delta\int(v-v_0)^4 d\nu^{(s)}\,\mathbb{P}_{\nu^{(s)}}$$
$$+ 3\Delta^2\left(((v-v_0)^2\nu^{(s)}) - \delta_0\int(v-v_0)^2 d\nu^{(s)}\right)^{*2} * \mathbb{P}_{\nu^{(s)}}$$

$$
\begin{aligned}
&+4\Delta^2\left(((v-v_0)\nu^{(s)})-\delta_0\int(v-v_0)d\nu^{(s)}\right)\\
&\qquad\qquad *\left(((v-v_0)^3\nu^{(s)})-\delta_0\int(v-v_0)^3d\nu^{(s)}\right)*\mathbb{P}_{\nu^{(s)}}\\
&+6\Delta^3\left(((v-v_0)\nu^{(s)})-\delta_0\int(v-v_0)d\nu^{(s)}\right)^{*2}\\
&\qquad\qquad *\left(((v-v_0)^2\nu^{(s)})-\delta_0\int(v-v_0)^2d\nu^{(s)}\right)*\mathbb{P}_{\nu^{(s)}}\\
&+\Delta^4\left(((v-v_0)\nu^{(s)})-\delta_0\int(v-v_0)d\nu^{(s)}\right)^{*4}*\mathbb{P}_{\nu^{(s)}}\Bigg\},\\
&\frac{d^5}{ds^5}\frac{d\,\mathbb{P}_{\nu^{(s)}}}{d\,\mathbb{P}_\Lambda}=\frac{d}{d\,\mathbb{P}_\Lambda}\Bigg\{\Delta((v-v_0)^5\nu^{(s)})*\mathbb{P}_{\nu^{(s)}}-\Delta\int(v-v_0)^5d\nu^{(s)}\,\mathbb{P}_{\nu^{(s)}}\\
&+10\Delta^2\left(((v-v_0)^2\nu^{(s)})-\delta_0\int(v-v_0)^2d\nu^{(s)}\right)\\
&\qquad\qquad *\left(((v-v_0)^3\nu^{(s)})-\delta_0\int(v-v_0)^3d\nu^{(s)}\right)*\mathbb{P}_{\nu^{(s)}}\\
&+5\Delta^2\left(((v-v_0)\nu^{(s)})-\delta_0\int(v-v_0)d\nu^{(s)}\right)\\
&\qquad\qquad *\left(((v-v_0)^4\nu^{(s)})-\delta_0\int(v-v_0)^4d\nu^{(s)}\right)*\mathbb{P}_{\nu^{(s)}}\\
&+10\Delta^3\left(((v-v_0)\nu^{(s)})-\delta_0\int(v-v_0)d\nu^{(s)}\right)^{*2}\\
&\qquad\qquad *\left(((v-v_0)^3\nu^{(s)})-\delta_0\int(v-v_0)^3d\nu^{(s)}\right)*\mathbb{P}_{\nu^{(s)}}\\
&+15\Delta^3\left(((v-v_0)^2\nu^{(s)})-\delta_0\int(v-v_0)^2d\nu^{(s)}\right)^{*2}\\
&\qquad\qquad *\left(((v-v_0)\nu^{(s)})-\delta_0\int(v-v_0)d\nu^{(s)}\right)*\mathbb{P}_{\nu^{(s)}}\\
&+10\Delta^4\left(((v-v_0)\nu^{(s)})-\delta_0\int(v-v_0)d\nu^{(s)}\right)^{*3}\\
&\qquad\qquad *\left(((v-v_0)^2\nu^{(s)})-\delta_0\int(v-v_0)^2d\nu^{(s)}\right)*\mathbb{P}_{\nu^{(s)}}\\
&+\Delta^5\left(((v-v_0)\nu^{(s)})-\delta_0\int(v-v_0)d\nu^{(s)}\right)^{*5}*\mathbb{P}_{\nu^{(s)}}\Bigg\}.
\end{aligned}
$$

Let $L^2_0(\mathbb{P}_\nu):=\{g\in L^2(\mathbb{P}_\nu):\int g d\,\mathbb{P}_\nu=0\}$. Motivated by the structure of the derivatives we define the multilinear form

$$A_\nu|_{L^2(\nu)^{\otimes k}}:L^2(\nu)^{\otimes k}\to L^2_0(\mathbb{P}_\nu), \tag{57}$$

$$(w_1,\dots,w_k)\mapsto \Delta^k\frac{d((w_1\nu-\delta_0\int w_1 d\nu)*\cdots*(w_k\nu-\delta_0\int w_k d\nu)*\mathbb{P}_\nu)}{d\,\mathbb{P}_\nu}.$$

In view of the derivatives of the log-likelihood we divide the derivatives by $d\,\mathbb{P}_{\nu^{(s)}}/d\,\mathbb{P}_\Lambda$. Then the dominating measure $\mathbb{P}_\Lambda$ cancels and we suppress it in the notation. We obtain the following expressions

$$\begin{aligned}
\frac{d\frac{d}{ds}\,\mathbb{P}_{\nu^{(s)}}}{d\,\mathbb{P}_{\nu^{(s)}}} &= A_{\nu^{(s)}}(v-v_0),\\
\frac{d\frac{d^2}{ds^2}\,\mathbb{P}_{\nu^{(s)}}}{d\,\mathbb{P}_{\nu^{(s)}}} &= A_{\nu^{(s)}}(v-v_0)^2+A_{\nu^{(s)}}(v-v_0,v-v_0),\\
\frac{d\frac{d^3}{ds^3}\,\mathbb{P}_{\nu^{(s)}}}{d\,\mathbb{P}_{\nu^{(s)}}} &= A_{\nu^{(s)}}(v-v_0)^3+3A_{\nu^{(s)}}(v-v_0,(v-v_0)^2)\\
&\quad + A_{\nu^{(s)}}(v-v_0,v-v_0,v-v_0),\\
\frac{d\frac{d^4}{ds^4}\,\mathbb{P}_{\nu^{(s)}}}{d\,\mathbb{P}_{\nu^{(s)}}} &= A_{\nu^{(s)}}(v-v_0)^4+4A_{\nu^{(s)}}(v-v_0,(v-v_0)^3)\\
&\quad +3A_{\nu^{(s)}}((v-v_0)^2,(v-v_0)^2)+6A_{\nu^{(s)}}(v-v_0,v-v_0,(v-v_0)^2)\\
&\quad + A_{\nu^{(s)}}(v-v_0,v-v_0,v-v_0,v-v_0)\\
\frac{d\frac{d^5}{ds^5}\,\mathbb{P}_{\nu^{(s)}}}{d\,\mathbb{P}_{\nu^{(s)}}} &= A_{\nu^{(s)}}(v-v_0)^5+5A_{\nu^{(s)}}(v-v_0,(v-v_0)^4)\\
&\quad +10A_{\nu^{(s)}}((v-v_0)^2,(v-v_0)^3)+10A_{\nu^{(s)}}(v-v_0,v-v_0,(v-v_0)^3)\\
&\quad +15A_{\nu^{(s)}}((v-v_0)^2,(v-v_0)^2,v-v_0)\\
&\quad +10A_{\nu^{(s)}}(v-v_0,v-v_0,v-v_0,(v-v_0)^2)\\
&\quad + A_{\nu^{(s)}}(v-v_0,v-v_0,v-v_0,v-v_0,v-v_0).
\end{aligned}$$

With the densities at hand we can determine the derivatives of the empirical log-likelihood

$$\begin{aligned}
&D\ell_n(\nu_0)[v-v_0]=\sum_{j=1}^n\frac{d\frac{d}{ds}\,\mathbb{P}_{\nu^{(s)}}}{d\,\mathbb{P}_{\nu^{(s)}}}\bigg|_{s=0}(X_j),\\
&D^2\ell_n(\nu_0)[v-v_0,v-v_0]=\sum_{j=1}^n\frac{d\frac{d^2}{ds^2}\,\mathbb{P}_{\nu^{(s)}}}{d\,\mathbb{P}_{\nu^{(s)}}}\bigg|_{s=0}(X_j)-\sum_{j=1}^n\left(\frac{d\frac{d}{ds}\,\mathbb{P}_{\nu^{(s)}}}{d\,\mathbb{P}_{\nu^{(s)}}}\right)^2\bigg|_{s=0}(X_j)\\
&D^3\ell_n(\nu^{(s)})[v-v_0,v-v_0,v-v_0]\\
&=\sum_{j=1}^n\frac{d\frac{d^3}{ds^3}\,\mathbb{P}_{\nu^{(s)}}}{d\,\mathbb{P}_{\nu^{(s)}}}(X_j)-3\sum_{j=1}^n\frac{d\frac{d^2}{ds^2}\,\mathbb{P}_{\nu^{(s)}}}{d\,\mathbb{P}_{\nu^{(s)}}}(X_j)\frac{d\frac{d}{ds}\,\mathbb{P}_{\nu^{(s)}}}{d\,\mathbb{P}_{\nu^{(s)}}}(X_j)\\
&\quad+2\sum_{j=1}^n\left(\frac{d\frac{d}{ds}\,\mathbb{P}_{\nu^{(s)}}}{d\,\mathbb{P}_{\nu^{(s)}}}(X_j)\right)^3\\
&D^4\ell_n(\nu^{(s)})[v-v_0,v-v_0,v-v_0,v-v_0]
\end{aligned}$$

$$= \sum_{j=1}^n \frac{d\frac{d^4}{ds^4}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}(X_j) - 4\sum_{j=1}^n \frac{d\frac{d^3}{ds^3}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}(X_j)\frac{d\frac{d}{ds}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}(X_j)$$
$$- 3\sum_{j=1}^n \left(\frac{d\frac{d^2}{ds^2}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}(X_j)\right)^2 + 12\sum_{j=1}^n \frac{d\frac{d^2}{ds^2}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}(X_j)\left(\frac{d\frac{d}{ds}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}(X_j)\right)^2$$
$$- 6\sum_{j=1}^n \left(\frac{d\frac{d}{ds}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}(X_j)\right)^4$$

$$D^5\ell_n(\nu^{(s)})[v-v_0, v-v_0, v-v_0, v-v_0, v-v_0]$$
$$= \sum_{j=1}^n \frac{d\frac{d^5}{ds^5}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}(X_j) - 5\sum_{j=1}^n \frac{d\frac{d^4}{ds^4}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}(X_j)\frac{d\frac{d}{ds}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}(X_j)$$
$$+ 20\sum_{j=1}^n \frac{d\frac{d^3}{ds^3}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}(X_j)\left(\frac{d\frac{d}{ds}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}(X_j)\right)^2$$
$$- 10\sum_{j=1}^n \frac{d\frac{d^3}{ds^3}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}(X_j)\frac{d\frac{d^2}{ds^2}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}(X_j)$$
$$- 60\sum_{j=1}^n \frac{d\frac{d^2}{ds^2}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}(X_j)\left(\frac{d\frac{d}{ds}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}(X_j)\right)^3$$
$$+ 30\sum_{j=1}^n \left(\frac{d\frac{d^2}{ds^2}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}(X_j)\right)^2\frac{d\frac{d}{ds}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}(X_j) + 24\sum_{j=1}^n \left(\frac{d\frac{d}{ds}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}(X_j)\right)^5.$$

The previous quantities simply denote one-dimensional derivatives of the empirical log-likelihood along the curve $\nu^{(s)}$. These derivatives can be viewed as values on the diagonal of symmetric multilinear forms and by means of polarization we extend the derivatives to symmetric multilinear forms.

6.2. Likelihood expansion

In this section we will use a likelihood expansion to show the statement used in Section 6 that

$$-t\sqrt{n}\int A_{\nu_0}(v-v_0)A_{\nu_0}(\eta)d\mathbb{P}_{\nu_0} + \ell_n(\nu)$$
$$= \frac{t^2}{2}\|A_{\nu_0}(\eta)\|^2_{L^2(\mathbb{P}_{\nu_0})} - \frac{t}{\sqrt{n}}\sum_{k=1}^n A_{\nu_0}(\eta)(X_k) + \ell_n(\nu_t) + r_n'(\nu),$$

where $\sup_{\nu\in D_{n,M}}|r_n'(\nu)| = o_{\mathbb{P}_{\nu_0}^{\mathbb{N}}}(1)$. Let $\varepsilon_n^{L^p}$ with $2<p<\infty$ be rates such that for

$$D_{n,p} = D_{n,p,M} := \left\{\nu : v \in V_{B,J}, \|v-v_0\|_{L^p} \leqslant M\varepsilon_n^{L^p}\right\}$$

we have

$$\Pi(D_{n,p}^c|X_1,\dots,X_n) \to^{\mathbb{P}_{\nu_0}^{\mathbb{N}}} 0.$$

For example we can take $(\varepsilon_n^{L^p})^p = (\varepsilon_n^{L^\infty})^{p-2}(\varepsilon_n^{L^2})^2$. Setting $\omega_n^{L^p} = tn^{-1/2}\|\eta\|_{L^p} + \delta_n\varepsilon_n^{L^p}$ we work under the following conditions.

Assumption 25. *Let $\mathcal{H}_n \subseteq L^\infty(I)$. Assume J, δ_n, $\varepsilon_n^{L^p}$ and $\omega_n^{L^p}$ satisfy uniformly over $\eta \in \mathcal{H}_n$*

$$2^{-Js} = o(\varepsilon_n^{L^2}), \quad 2^{-Js} = o(\varepsilon_n^{L^\infty}), \quad \textit{(bias conditions)}$$

$$\sqrt{n}\delta_n\varepsilon_n^{L^2}2^{J/2}\sqrt{\log\frac{c}{\varepsilon_n^{L^2}}} = o(1), \quad \textit{(for term II)}$$

$$\frac{2^{J/2}}{\sqrt{n}}\sqrt{\log\frac{c}{\varepsilon_n^{L^2}}} \lesssim \varepsilon_n^{L^2}, \quad \textit{(first term dominates in II)}$$

$$n\delta_n\big(\varepsilon_n^{L^2}\big)^2 = o(1), \quad \textit{(for centring of III(ii))}$$

$$t\|\eta\|_\infty\varepsilon_n^{L^2}2^{J/2}\sqrt{\log\frac{c}{\varepsilon_n^{L^2}}} = o(1), \quad \textit{(for term III(i))}$$

$$\frac{t^2}{\sqrt{n}}\|\eta\|_{L^4}^2 = o(1), \quad \textit{(for deviation from mean of IV(i))}$$

$$t\delta_n\sqrt{n}\|\eta\|_{L^2}\varepsilon_n^{L^2} = o(1), \quad \textit{(for centring of IV(iii))}$$

$$n\,\omega_n^{L^3}\left(\varepsilon_n^{L^3}\right)^2 = o(1), \qquad n\left(\omega_n^{L^3}\right)^3 = o(1), \quad \textit{(for centring of third derivative)}$$

$$n\,\omega_n^{L^4}\left(\varepsilon_n^{L^4}\right)^3 = o(1), \qquad n\left(\omega_n^{L^4}\right)^4 = o(1), \quad \textit{(for centring of fourth derivative)}$$

$$n\left(\varepsilon_n^{L^5}\right)^5 = o(1), \qquad n\left(\omega_n^{L^5}\right)^5 = o(1), \quad \textit{(for centring of fifth derivative)}$$

$$\sqrt{n}\Big(\varepsilon_n^{L^\infty} + \omega_n^{L^\infty}\Big)^2\Big(\varepsilon_n^{L^2} + \omega_n^{L^2}\Big)2^{J/2}\left(\log\frac{c}{\varepsilon_n^{L^2} + \omega_n^{L^2}}\right)^{1/2} = o(1), \quad \textit{(for } R_n\textit{)}$$

$$\frac{1}{\sqrt{n}}2^{J/2}\left(\log\frac{c}{\varepsilon_n^{L^2} + \omega_n^{L^2}}\right)^{1/2} \lesssim \varepsilon_n^{L^2} + \omega_n^{L^2}. \quad \textit{(first term dominates in } R_n\textit{)}$$

We consider the following path from ν_0 to ν, $s \mapsto \exp(s(v - v_0) + v_0) = \nu^{(s)}$. A Taylor expansion of the log-likelihood ℓ_n along this path gives

$$\begin{aligned}\ell_n(\nu) - \ell_n(\nu_0) = D\ell_n(\nu_0)[v - v_0] + \tfrac{1}{2}D^2\ell_n(\nu_0)[v - v_0, v - v_0]\\ + \tfrac{1}{6}D^3\ell_n(\nu^{(s)})[v - v_0, v - v_0, v - v_0],\end{aligned}$$

where the first two terms denote first and second derivative at zero and the last term denotes the third derivative at some intermediate point $s \in [0,1]$. We will see later that the derivatives depend linearly on the directions. Thus it is possible to extend them to symmetric multilinear forms. The corresponding path from ν_0 to $\nu_t = \exp(v_t)$ is $u \mapsto \exp(u(v_t - v_0) + v_0) = \nu_t^{(u)}$.

We recall the perturbation (31) and define $\widetilde{\delta}_n(v)$ by

$$v_t = v + \delta_n\Big(\frac{t}{\delta_n\sqrt{n}}\eta + v_{0,J} - v\Big) = v + \widetilde{\delta}_n(v).$$

With this definition we calculate

$$\begin{aligned}
&\ell_n(\nu) - \ell_n(\nu_0) - (\ell_n(\nu_t) - \ell_n(\nu_0))\\
&\quad = D\ell_n(\nu_0)[v - v_0] - D\ell_n(\nu_0)[v_t - v_0] + \tfrac{1}{2}D^2\ell_n(\nu_0)[v - v_0, v - v_0]\\
&\qquad - \tfrac{1}{2}D^2\ell_n(\nu_0)[v_t - v_0, v_t - v_0] + R_n\\
&\quad = D\ell_n(\nu_0)[v - v_t] + \tfrac{1}{2}D^2\ell_n(\nu_0)[v - v_0, v - v_0]\\
&\qquad - \tfrac{1}{2}D^2\ell_n(\nu_0)[v - v_0 + \widetilde{\delta}_n(v), v - v_0 + \widetilde{\delta}_n(v)] + R_n\\
&\quad = -D\ell_n(\nu_0)[(t/\sqrt{n})\eta] - \delta_n D\ell_n(\nu_0)[v_{0,J} - v]\\
&\qquad - D^2\ell_n(\nu_0)[v - v_0, \widetilde{\delta}_n(v)] - \tfrac{1}{2}D^2\ell_n(\nu_0)[\widetilde{\delta}_n(v), \widetilde{\delta}_n(v)] + R_n\\
&\quad = I + II + III + IV + R_n,
\end{aligned}$$

where

$$R_n = \tfrac{1}{6}D^3\ell_n(\nu^{(s)})[v - v_0, v - v_0, v - v_0] - \tfrac{1}{6}D^3\ell_n(\nu_t^{(u)})[v_t - v_0, v_t - v_0, v_t - v_0]$$

with intermediate points $s, u \in [0,1]$.

We need to show that

$$\begin{aligned}
I + II + III + IV + R_n = t\sqrt{n}\int A_{\nu_0}(v - v_0)A_{\nu_0}(\eta)d\mathbb{P}_{\nu_0}&\\
+ \frac{t^2}{2}\|A_{\nu_0}(\eta)\|^2_{L^2(\mathbb{P}_{\nu_0})} - \frac{t}{\sqrt{n}}\sum_{k=1}^n A_{\nu_0}(\eta)(X_k) + r'_n(\nu).&
\end{aligned} \tag{58}$$

The first term is given by $I = -\frac{t}{\sqrt{n}}D\ell_n(\nu_0)[\eta] = -\frac{t}{\sqrt{n}}\sum_{k=1}^n A_{\nu_0}[\eta](X_k)$. For the second term we have

$$II = -\delta_n D\ell_n(\nu_0)[v_{0,J} - v] = \sqrt{n}\delta_n \frac{1}{\sqrt{n}}\sum_{k=1}^n A_{\nu_0}(v - v_{0,J})(X_k) = \sqrt{n}\delta_n \mathbb{G}_n f_v,$$

where $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_{\nu_0,n} - \mathbb{P}_{\nu_0})$ is the empirical process and $f_v = A_{\nu_0}(v - v_{0,J})$.

On $D_{n,M}$ we have $\|v - v_0\|_{L^2} \leqslant M\varepsilon_n^{L^2}$ and $\|v - v_0\|_\infty \leqslant M\varepsilon_n^{L^\infty}$. Using the usual bias bounds $\|v_{0,J} - v_0\|_{L^2} \lesssim 2^{-Js}$, $\|v_{0,J} - v_0\|_\infty \lesssim 2^{-Js}$ and the bias condition in Assumption 25 we obtain $\|v - v_{0,J}\|_{L^2} \leqslant M\varepsilon_n^{L^2}$ and $\|v - v_{0,J}\|_\infty \leqslant M\varepsilon_n^{L^\infty}$ with a possibly larger constant M. We recall $f_v = A_{\nu_0}(v - v_{0,J})$ and consider the finite dimensional class of functions

$$\mathcal{F} := \left\{ f_v : v \in V_{B,J}, \|v - v_{0,J}\|_{L^2} \leqslant M\varepsilon_n^{L^2}, \|v - v_{0,J}\|_\infty \leqslant M\varepsilon_n^{L^\infty} \right\}. \tag{59}$$

We observe that there is $D > 0$ such that $\|v_0\|_\infty \leqslant D$ and $\|v\|_\infty \leqslant D$ for all $v \in V_{B,J}$. We will bound the norms of functions in $\mathcal{F}$ using the following lemma.

Lemma 26. *Let $\|v\|_\infty \leqslant D$ and $\nu = \exp(v)$. Then for A_ν defined in (57) and for $1 \leqslant p \leqslant \infty$*

$$\|A_\nu(w_1, \dots, w_k)\|_{L^p(\mathbb{P}_\nu)} \lesssim \|w_1\|_{L^p(\nu)} \dots \|w_k\|_{L^p(\nu)}.$$

The constants only depends on k, D and Δ.

Proof. We write ν for both the Lévy measure and its density. The measure $\mathbb{P}_\nu$ can be written as a convolution exponential $\mathbb{P}_\nu = e^{-\Delta\lambda}\sum_{k=0}^\infty \frac{\Delta^k}{k!}\nu^{*k}$ with intensity $\lambda = \nu((-1/2,1/2])$. The function v is bounded such that the corresponding Lévy density $\nu = \exp(v)$ is bounded from above and bounded away from zero. Likewise the intensity λ is bounded from above and bounded away from zero. We denote by Λ the Lebesgue measure on $[-1/2,1/2]$. Then $\frac{d\Lambda}{d\mathbb{P}_\nu}$ is in $L^\infty(\mathbb{P}_\nu)$ with norm bounded by a constant depending on D and Δ only. Defining by $\mathbb{P}_\nu^a = e^{-\Delta\lambda}\sum_{k=1}^\infty \frac{\Delta^k}{k!}\nu^{*k}$ the absolutely continuous part with respect to the Lebesgue measure Λ we see likewise that the density $\frac{d\mathbb{P}_\nu^a}{d\Lambda}$ is bounded in $L^\infty(\Lambda)$ from above depending on D and Δ only. By definition we have

$$\begin{aligned}&\|A_\nu(w_1,\dots,w_k)\|_{L^p(\mathbb{P}_\nu)}\\&\leqslant \Delta^k\left\|\frac{d((w_1\nu-\delta_0\int w_1 d\nu)*\cdots*(w_k\nu-\delta_0\int w_k d\nu)*\mathbb{P}_\nu)}{d\,\mathbb{P}_\nu}\right\|_{L^p(\mathbb{P}_\nu)}.\end{aligned}$$

The nominator consists of 2^k terms and a typical term is of the from

$$\int w_1 d\nu \dots \int w_j d\nu \cdot (w_{j+1}\nu)*\cdots*(w_k\nu)*\mathbb{P}_\nu$$

and up to permutation and choice of j between 0 and k all terms are of this form. So it suffices to bound

$$\begin{aligned}&\left\|\frac{d(\int w_1 d\nu\dots\int w_j d\nu\cdot(w_{j+1}\nu)*\cdots*(w_k\nu)*\mathbb{P}_\nu)}{d\,\mathbb{P}_\nu}\right\|_{L^p(\mathbb{P}_\nu)}\\&\lesssim \|w_1\|_{L^1(\nu)}\dots\|w_j\|_{L^1(\nu)}\left\|\frac{d((w_{j+1}\nu)*\cdots*(w_k\nu)*\mathbb{P}_\nu)}{d\,\mathbb{P}_\nu}\right\|_{L^p(\mathbb{P}_\nu)}\\&\lesssim \|w_1\|_{L^p(\nu)}\dots\|w_j\|_{L^p(\nu)}\left\|\frac{d((w_{j+1}\nu)*\cdots*(w_k\nu)*\mathbb{P}_\nu)}{d\,\mathbb{P}_\nu}\right\|_{L^p(\mathbb{P}_\nu)}.\end{aligned}$$

For $j=k$ this gives the desired bound and for $j<k$ the previous line can be bounded by

$$\begin{aligned}&\|w_1\|_{L^p(\nu)}\dots\|w_j\|_{L^p(\nu)}\left\|\frac{d((w_{j+1}\nu)*\cdots*(w_k\nu)*\mathbb{P}_\nu)}{d\Lambda}\right\|_{L^p(\mathbb{P}_\nu)}\left\|\frac{d\Lambda}{d\,\mathbb{P}_\nu}\right\|_{L^\infty(\mathbb{P}_\nu)}\\&\lesssim \|w_1\|_{L^p(\nu)}\dots\|w_j\|_{L^p(\nu)}\left\|\frac{d((w_{j+1}\nu)*\cdots*(w_k\nu)*\mathbb{P}_\nu)}{d\Lambda}\right\|_{L^p(\Lambda)},\end{aligned}$$

where we have used boundedness of $\frac{d\Lambda}{d\mathbb{P}_\nu}$ and $\frac{d\mathbb{P}_\nu^a}{d\Lambda}$. Young's inequality for convolutions yields the bound

$$\begin{aligned}&\|w_1\|_{L^p(\nu)}\dots\|w_j\|_{L^p(\nu)}\|w_{j+1}\nu\|_{L^1(\Lambda)}\dots\|w_{k-1}\nu\|_{L^1(\Lambda)}\|w_k\nu\|_{L^p(\Lambda)}\\&\lesssim \|w_1\|_{L^p(\nu)}\dots\|w_k\|_{L^p(\nu)}\end{aligned}$$

and the lemma follows by treating all 2^k terms in this way. □

We define $v(u) = \sum_{l\leqslant J-1}\sum_k a_l u_{lk}\psi_{lk}$ with $a_l = 2^{-l}(l^2+1)^{-1}$. For $u, u' \in \mathbb{R}^{2^J}$ we denote $v = v(u), v' = v(u')$. Applying Lemma 26 with $w_1 = v - v'$ yields $\|f_v - f_{v'}\|_\infty \lesssim \|v - v'\|_\infty \lesssim \|u - u'\|_\infty$, where the constant only depends on D and Δ. It follows that $\sup_{\mathbb{Q}} \|f_v - f_{v'}\|_{L^2(\mathbb{Q})} \lesssim \|u - u'\|_\infty$, where the supremum is over all Borel probability measures $\mathbb{Q}$. Consequently we have $\sup_{\mathbb{Q}} N(\mathcal{F}, L^2(\mathbb{Q}), \varepsilon\|F\|_{L^2(\mathbb{Q})}) \leqslant (A/\varepsilon)^{2^J}$, for some $A \geqslant 2$ and for $0 < \varepsilon < A$ and where the envelope can be taken as a constant function F with constant only depending on D and Δ.

Let $\sigma^2 = \sup_{f\in\mathcal{F}} \mathbb{P}_{\nu_0} f^2$. Lemma 26 yields

$$\begin{aligned}\sigma &\leqslant \sup_{\|v-v_{0,J}\|\leqslant M\varepsilon_n^{L^2}} \|A_{\nu_0}(v - v_{0,J})\|_{L^2(\mathbb{P}_{\nu_0})} \\ &\lesssim \sup_{\|v-v_{0,J}\|\leqslant M\varepsilon_n^{L^2}} \|v - v_{0,J}\|_{L^2(\nu_0)} \lesssim \varepsilon_n^{L^2}.\end{aligned}$$

Then we have by Corollary 3.5.8 in [18] for some $c > 0$

$$E\|\mathbb{G}_n\|_{\mathcal{F}} \lesssim \varepsilon_n^{L^2} 2^{J/2}\sqrt{\log\frac{c}{\varepsilon_n^{L^2}}} + \frac{1}{\sqrt{n}}2^J \log\frac{c}{\varepsilon_n^{L^2}}.$$

We obtain $II = o_{\mathbb{P}}(1)$ using the conditions

$$\sqrt{n}\delta_n\varepsilon_n^{L^2}2^{J/2}\sqrt{\log\frac{c}{\varepsilon_n^{L^2}}} = o(1) \quad \text{and} \quad \frac{2^{J/2}}{\sqrt{n}}\sqrt{\log\frac{c}{\varepsilon_n^{L^2}}} \lesssim \varepsilon_n^{L^2}.$$

Next we consider the term III. It equals

$$\begin{aligned}&- D^2\ell_n(\nu_0)[v - v_0, \widetilde{\delta}_n(v)] \\ &= \underbrace{-n^{-1/2} t D^2\ell_n(\nu_0)[v - v_0, \eta]}_{(i)} \underbrace{+\delta_n D^2\ell_n(\nu_0)[v - v_0, v - v_{0,J}]}_{(ii)} \\ &= \underbrace{-\frac{t}{\sqrt{n}}\sum_{j=1}^n \frac{d\frac{d^2}{ds^2}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}\bigg|_{s=0}[v - v_0, \eta](X_j)}_{(i)(a)} \\ &\quad \underbrace{+\frac{t}{\sqrt{n}}\sum_{j=1}^n \left(\frac{d\frac{d}{ds}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}\right)^2\bigg|_{s=0}[v - v_0, \eta](X_j)}_{(i)(b)} \\ &\quad \underbrace{+\delta_n\sum_{j=1}^n \frac{d\frac{d^2}{ds^2}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}\bigg|_{s=0}[v - v_0, v - v_{0,J}](X_j)}_{(ii)(a)} \\ &\quad \underbrace{-\delta_n\sum_{j=1}^n \left(\frac{d\frac{d}{ds}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}\right)^2\bigg|_{s=0}[v - v_0, v - v_{0,J}](X_j),}_{(ii)(b)}\end{aligned}$$

where we understand the bilinear forms through polarization and by abuse of notation $\nu^{(s)}$ denotes a generic path.

The terms $(i)(a)$ and $(ii)(a)$ are both centred. The term $(i)(b)$ is centred after subtracting

$$\sqrt{n}\, t \int A_{\nu_0}(v - v_0) A_{\nu_0}(\eta) d\, \mathbb{P}_{\nu_0}$$

yielding the corresponding term in (58). The centring of the term $(ii)(b)$ is of order

$$\begin{aligned}
&\delta_n n \left| \int A_{\nu_0}(v - v_0) A_{\nu_0}(v - v_{0,J}) d\, \mathbb{P}_{\nu_0} \right| \\
&\lesssim \delta_n n \left(E_{\nu_0}\left[(A_{\nu_0}(v - v_0))^2 \right] \right)^{1/2} \left(E_{\nu_0}\left[(A_{\nu_0}(v - v_{0,J}))^2 \right] \right)^{1/2} \\
&\lesssim \delta_n n \|v - v_0\|_{L^2(\nu_0)} \|v - v_{0,J}\|_{L^2(\nu_0)} \lesssim \delta_n n (\varepsilon_n^{L^2})^2 = o(1).
\end{aligned}$$

We start with the term $(i)(a)$. We define functions

$$f_v = A_{\nu_0}((v - v_0)\eta) + A_{\nu_0}(v - v_0, \eta)$$

and consider the corresponding class of functions as in (59). For $u, u' \in \mathbb{R}^{2^J}$ we denote again $v = v(u), v' = v(u')$ and apply Lemma 26 to the function $f_v - f_{v'}$. This yields

$$\|f_v - f_{v'}\|_\infty \lesssim \|\eta\|_\infty \|v - v'\|_\infty \lesssim \|\eta\|_\infty \|u - u'\|_\infty,$$

where the constant only depends on D and Δ. We choose the envelope F of the class $\mathcal{F}$ as a constant function $C\|\eta\|_\infty$, where the constant C depends only on D and Δ. Then the bound $\|f_v - f_{v'}\|_\infty \lesssim \|\eta\|_\infty \|u - u'\|_\infty$ shows that we have $\sup_{\mathbb{Q}} N(\mathcal{F}, L^2(\mathbb{Q}), \varepsilon \|F\|_{L^2(\mathbb{Q})}) \leqslant (A/\varepsilon)^{2^J}$ for some $A \geqslant 2$ and for all $0 < \varepsilon < A$.

The next step is to bound $\sigma^2 = \sup_{f \in \mathcal{F}} \mathbb{P}_{\nu_0} f^2$. By Lemma 26 we have

$$\sigma = \sup_{f \in \mathcal{F}} \|f\|_{L^2(\mathbb{P}_{\nu_0})} \lesssim \|\eta\|_\infty \varepsilon_n^{L^2}.$$

Corollary 3.5.8 in [18] allows to bound the empirical process appearing in term $(i)(a)$. For some $c > 0$ we obtain

$$E\|\mathbb{G}_n\|_{\mathcal{F}} \lesssim \|\eta\|_\infty\, \varepsilon_n^{L^2} 2^{J/2} \sqrt{\log \frac{c}{\varepsilon_n^{L^2}}} + \frac{1}{\sqrt{n}} \|\eta\|_\infty 2^J \log \frac{c}{\varepsilon_n^{L^2}}.$$

The conditions for the first term dominating the second term is the same as for the term II. To bound the term $(i)(a)$ we use

$$t\|\eta\|_\infty\, \varepsilon_n^{L^2} 2^{J/2} \sqrt{\log \frac{c}{\varepsilon_n^{L^2}}} = o(1).$$

Next we treat term $(i)(b)$, which is given by

$$\frac{t}{\sqrt{n}}\sum_{j=1}^{n} A_{\nu_0}(v-v_0)(X_j)A_{\nu_0}(\eta)(X_j).$$

We define $g_v = A_{\nu_0}(v-v_0)A_{\nu_0}(\eta)$ and $f_v = g_v - E_{\nu_0}[g_v]$. So after centring the term is given by $t\mathbb{G}_n f_v$. We have by Lemma 26

$$\begin{aligned}\|g_v - g_{v'}\|_\infty = \|A_{\nu_0}(v-v')A_{\nu_0}(\eta)\|_\infty &\leqslant \|A_{\nu_0}(v-v')\|_\infty\|A_{\nu_0}(\eta)\|_\infty \\ &\lesssim \|v-v'\|_\infty\|\eta\|_\infty\end{aligned}$$

and thus also $\|f_v - f_{v'}\|_\infty \lesssim \|v-v'\|_\infty\|\eta\|_\infty$. We consider the class of functions $\mathcal{F}$ as in (59) corresponding to the functions of the form f_v here and bound

$$\begin{aligned}\sigma = \sup_{f\in\mathcal{F}}\|f\|_{L^2(\mathbb{P}_{\nu_0})} &\leqslant \sup_{\|v-v_0\|_{L^2}\leqslant 2M\varepsilon_n^{L^2}}\|g_v\|_{L^2(\mathbb{P}_{\nu_0})} \\ &\leqslant \sup_{\|v-v_0\|_{L^2}\leqslant 2M\varepsilon_n^{L^2}}\|A_{\nu_0}(\eta)\|_\infty\|A_{\nu_0}(v-v_0)\|_{L^2(\mathbb{P}_{\nu_0})} \lesssim \|\eta\|_\infty\varepsilon_n^{L^2}.\end{aligned}$$

Just as for term $(i)(a)$ we apply now Corollary 3.5.8 in [18] with envelop proportional to $\|\eta\|_\infty$. So the conditions for term $(ii)(b)$ are the same as for the term $(i)(a)$.

We move on to the term $(ii)(a)$. We define

$$f_{vv'} = A_{\nu_0}((v-v_0)(v'-v_{0,J})) + A_{\nu_0}(v-v_0, v'-v_{0,J})$$

and $f_v = f_{vv}$. We now consider the class of functions $\mathcal{F}$ with this definition of f_v. Then we have

$$\|f_v - f_{v'}\|_\infty \lesssim \|f_{vv} - f_{vv'}\|_\infty + \|f_{vv'} - f_{v'v'}\|_\infty \lesssim \varepsilon_n^{L^\infty}\|v-v'\|_\infty.$$

Choosing the envelope as a constant function proportional to $\varepsilon_n^{L^\infty}$ we obtain for the covering numbers $\sup_{\mathbb{Q}} N(\mathcal{F}, L^2(\mathbb{Q}), \varepsilon\|F\|_{L^2(\mathbb{Q})}) \leqslant (A/\varepsilon)^{2^J}$. Turning to σ we see

$$\sigma = \sup_{f\in\mathcal{F}}\|f\|_{L^2(\mathbb{P}_{\nu_0})} \lesssim \varepsilon_n^{L^\infty}\varepsilon_n^{L^2}.$$

Again we apply Corollary 3.5.8 in [18], which gives the following bound for term $(ii)(a)$

$$\delta_n\sqrt{n}E\|\mathbb{G}_n\| \lesssim \delta_n\sqrt{n}\varepsilon_n^{L^\infty}\varepsilon_n^{L^2}2^{J/2}\sqrt{\log\frac{c}{\varepsilon_n^{L^2}}} + \delta_n\varepsilon_n^{L^\infty}2^J\log\frac{c}{\varepsilon_n^{L^2}}.$$

This tends to zero by the assumption for the term II.

The only remaining term of III is $(ii)(b)$. This term takes the from

$$-\delta_n\sum_{j=1}^{n} A_{\nu_0}(v-v_0)(X_j)A_{\nu_0}(v-v_{0,J})(X_j).$$

With the definitions $g_{vv'} = A_{\nu_0}(v - v_0)A_{\nu_0}(v' - v_{0,J})$ and $f_v = g_{vv} - E_{\nu_0}[g_{vv}]$ the term $(ii)(b)$ can be written after centring as $-\delta_n\sqrt{n}\mathbb{G}_n f_v$ and we bound

$$\begin{aligned}\|g_{vv} - g_{v'v'}\|_\infty &\leqslant \|g_{vv} - g_{vv'}\|_\infty + \|g_{vv'} - g_{v'v'}\|_\infty \\ &\lesssim \|v - v_0\|_\infty \|v - v'\|_\infty + \|v - v'\|_\infty \|v' - v_{0,J}\|_\infty \lesssim \varepsilon_n^{L^\infty} \|v - v'\|_\infty.\end{aligned}$$

Consequently we also have $\|f_v - f_{v'}\|_\infty \lesssim \varepsilon_n^{L^\infty}\|v - v'\|_\infty$. We denote by $\mathcal{F}$ the class of functions corresponding to f_v as in (59) and further bound

$$\begin{aligned}\sigma &= \sup_{f\in\mathcal{F}} \|f\|_{L^2(\mathbb{P}_{\nu_0})} \\ &\leqslant \sup\{\|g_{vv}\|_{L^2(\mathbb{P}_{\nu_0})} : \|v - v_0\|_{L^2} \leqslant M\varepsilon_n^{L^2}, \|v - v_0\|_{L^\infty} \leqslant M\varepsilon_n^{L^\infty}\} \\ &\leqslant \sup\{\|A_{\nu_0}(v - v_0)\|_\infty \|A_{\nu_0}(v - v_{0,J})\|_{L^2(\mathbb{P}_{\nu_0})} \\ &\qquad\qquad : \|v - v_0\|_{L^2} \leqslant M\varepsilon_n^{L^2}, \|v - v_0\|_{L^\infty} \leqslant M\varepsilon_n^{L^\infty}\} \\ &\lesssim \varepsilon_n^{L^\infty}\varepsilon_n^{L^2}.\end{aligned}$$

We see that $(ii)(b)$ leads to the same condition as the term $(ii)(a)$.

The term IV equals

$$\underbrace{-\frac{t^2}{2n}D^2\ell_n(\nu_0)[\eta,\eta]}_{(i)} \underbrace{-\frac{\delta_n^2}{2}D^2\ell_n(\nu_0)[v - v_{0,J}, v - v_{0,J}]}_{(ii)} \underbrace{+\frac{t\delta_n}{\sqrt{n}}D^2\ell_n(\nu_0)[\eta, v - v_{0,J}]}_{(iii)}$$

$$\begin{aligned}= &\underbrace{-\frac{t^2}{2n}\sum_{j=1}^n \frac{d\frac{d^2}{ds^2}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}\bigg|_{s=0}[\eta,\eta](X_j)}_{(i)(a)} \underbrace{+\frac{t^2}{2n}\sum_{j=1}^n \left(\frac{d\frac{d}{ds}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}\bigg|_{s=0}[\eta](X_j)\right)^2}_{(i)(b)} \\ &\underbrace{-\frac{\delta_n^2}{2}\sum_{j=1}^n \frac{d\frac{d^2}{ds^2}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}\bigg|_{s=0}[v - v_{0,J}, v - v_{0,J}](X_j)}_{(ii)(a)} \\ &\underbrace{+\frac{\delta_n^2}{2}\sum_{j=1}^n \left(\frac{d\frac{d}{ds}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}\bigg|_{s=0}[v - v_{0,J}](X_j)\right)^2}_{(ii)(b)} \\ &\underbrace{+\frac{t\delta_n}{\sqrt{n}}\sum_{j=1}^n \frac{d\frac{d^2}{ds^2}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}\bigg|_{s=0}[\eta, v - v_{0,J}](X_j)}_{(iii)(a)} \\ &\underbrace{-\frac{t\delta_n}{\sqrt{n}}\sum_{j=1}^n \left(\frac{d\frac{d}{ds}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}\right)^2\bigg|_{s=0}[\eta, v - v_{0,J}](X_j)}_{(iii)(b)}.\end{aligned}$$

The terms $(i)(a)$, $(ii)(a)$ and $(iii)(a)$ are centred. The term $(i)(b)$ can be centred by subtracting

$$\frac{t^2}{2}\|A_{\nu_0}(\eta)\|^2_{L^2(\mathbb{P}_{\nu_0})}$$

and gives the corresponding expression in (58). For the centring of term $(ii)(b)$ we subtract

$$\frac{\delta_n^2 n}{2}\|A_{\nu_0}(v-v_{0,J})\|^2_{L^2(\mathbb{P}_{\nu_0})} \lesssim \frac{\delta_n^2 n}{2}\|v-v_{0,J}\|^2_{L^2(\nu_0)} \lesssim \delta_n^2 n(\varepsilon_n^{L^2})^2 = o(1).$$

To centre the term $(iii)(b)$ we add $t\delta_n\sqrt{n}E_{\nu_0}[A_{\nu_0}(\eta)A_{\nu_0}(v-v_{0,J})]$ and this is bounded in absolute value by

$$\begin{aligned}
&|t\delta_n\sqrt{n}E_{\nu_0}[A_{\nu_0}(\eta)A_{\nu_0}(v-v_{0,J})]| \\
&\quad\lesssim t\delta_n\sqrt{n}\|A_{\nu_0}(\eta)\|_{L^2(\mathbb{P}_{\nu_0})}\|A_{\nu_0}(v-v_{0,J})\|_{L^2(\mathbb{P}_{\nu_0})} \\
&\quad\lesssim t\delta_n\sqrt{n}\|\eta\|_{L^2(\nu_0)}\|v-v_{0,J}\|_{L^2(\nu_0)} \lesssim t\delta_n\sqrt{n}\|\eta\|_{L^2}\varepsilon_n^{L^2} = o(1).
\end{aligned}$$

For term $(i)(a)$ we bound using Lemma 26

$$\begin{aligned}
E_{\nu_0}\left[\left(\frac{d\frac{d^2}{ds^2}\mathbb{P}_{\nu^{(s)}}}{d\mathbb{P}_{\nu^{(s)}}}\Bigg|_{s=0}[\eta,\eta]\right)^2\right] &\lesssim \|A_{\nu_0}\eta^2\|^2_{L^2(\mathbb{P}_{\nu_0})} + \|A_{\nu_0}(\eta,\eta)\|^2_{L^2(\mathbb{P}_{\nu_0})} \\
&\lesssim \|\eta^2\|^2_{L^2(\nu_0)} + \|\eta\|^4_{L^2(\nu_0)} \lesssim \|\eta\|^4_{L^4}
\end{aligned}$$

and for term $(i)(b)$ we bound using Lemma 26

$$E_{\nu_0}\left[(A_{\nu_0}(\eta))^4\right] = \|A_{\nu_0}(\eta)\|^4_{L^4(\mathbb{P}_{\nu_0})} \lesssim \|\eta\|^4_{L^4}.$$

We conclude that after centring term (i) is of order $O_{\mathbb{P}}(t^2 n^{-1/2}\|\eta\|^2_{L^4})$ and we use $t^2 n^{-1/2}\|\eta\|^2_{L^4} = o(1)$.

The terms $IV(ii)$ and $IV(iii)$ are treated in the same way as the terms $III(ii)$ and $III(i)$, respectively. Since the terms $IV(ii)$ and $IV(iii)$ both have an additional factor δ_n, no extra condition is needed.

The remainder term can be expressed as

$$\begin{aligned}
R_n &= \tfrac{1}{3!}D^3\ell_n(\nu_0)[v-v_0,v-v_0,v-v_0] - \tfrac{1}{3!}D^3\ell_n(\nu_0)[v_t-v_0,v_t-v_0,v_t-v_0] \\
&\quad + \tfrac{1}{4!}D^4\ell_n(\nu_0)[v-v_0,v-v_0,v-v_0,v-v_0] \\
&\quad - \tfrac{1}{4!}D^4\ell_n(\nu_0)[v_t-v_0,v_t-v_0,v_t-v_0,v_t-v_0] \\
&\quad + \tfrac{1}{5!}D^5\ell_n(\nu^{(s)})[v-v_0,v-v_0,v-v_0,v-v_0,v-v_0] \\
&\quad - \tfrac{1}{5!}D^5\ell_n(\nu_t^{(u)})[v_t-v_0,v_t-v_0,v_t-v_0,v_t-v_0,v_t-v_0] \\
&= -\tfrac{3}{3!}D^3\ell_n(\nu_0)[\widetilde{\delta}_n(v),v-v_0,v-v_0] - \tfrac{3}{3!}D^3\ell_n(\nu_0)[\widetilde{\delta}_n(v),\widetilde{\delta}_n(v),v-v_0] \\
&\quad - \tfrac{1}{3!}D^3\ell_n(\nu_0)[\widetilde{\delta}_n(v),\widetilde{\delta}_n(v),\widetilde{\delta}_n(v)] \\
&\quad - \tfrac{4}{4!}D^4\ell_n(\nu_0)[\widetilde{\delta}_n(v),v-v_0,v-v_0,v-v_0]
\end{aligned}$$

$$
\begin{aligned}
&- \tfrac{6}{4!} D^4\ell_n(\nu_0)[\widetilde{\delta}_n(v), \widetilde{\delta}_n(v), v - v_0, v - v_0] \\
&- \tfrac{4}{4!} D^4\ell_n(\nu_0)[\widetilde{\delta}_n(v), \widetilde{\delta}_n(v), \widetilde{\delta}_n(v), v - v_0] \\
&- \tfrac{1}{4!} D^4\ell_n(\nu_0)[\widetilde{\delta}_n(v), \widetilde{\delta}_n(v), \widetilde{\delta}_n(v), \widetilde{\delta}_n(v)] \\
&+ \tfrac{1}{5!} D^5\ell_n(\nu^{(s)})[v - v_0, v - v_0, v - v_0, v - v_0, v - v_0] \\
&- \tfrac{1}{5!} D^5\ell_n(\nu_t^{(u)})[v_t - v_0, v_t - v_0, v_t - v_0, v_t - v_0, v_t - v_0].
\end{aligned}
$$

We start with the centring of the third derivatives. So the aim is to bound $E_{\nu_0}[|D^3\ell_n(\nu_0)[w_1, w_2, w_3]|]$.

$$
\begin{aligned}
&D^3\ell_n(\nu_0)[w, w, w] \\
&= \underbrace{\sum_{j=1}^n \frac{d\frac{d^3}{dr^3}\mathbb{P}_{\nu^{(r)}}}{d\mathbb{P}_{\nu^{(r)}}}\bigg|_{r=0}(X_j)}_{(a)} \underbrace{-3\sum_{j=1}^n \frac{d\frac{d^2}{dr^2}\mathbb{P}_{\nu^{(r)}}}{d\mathbb{P}_{\nu^{(r)}}}\bigg|_{r=0}(X_j)\frac{d\frac{d}{dr}\mathbb{P}_{\nu^{(r)}}}{d\mathbb{P}_{\nu^{(r)}}}\bigg|_{r=0}(X_j)}_{(b)} \\
&\quad \underbrace{+2\sum_{j=1}^n \left(\frac{d\frac{d}{dr}\mathbb{P}_{\nu^{(r)}}}{d\mathbb{P}_{\nu^{(r)}}}\bigg|_{r=0}(X_j)\right)^3}_{(c)}.
\end{aligned}
$$

The term (a) is centred. For term (b) we calculate using Hölder's inequality

$$
\begin{aligned}
&E_{\nu_0}[|(A_{\nu_0}(w_1w_2) + A_{\nu_0}(w_1, w_2))A_{\nu_0}(w_3)|] \\
&\leqslant \|A_{\nu_0}(w_1w_2) + A_{\nu_0}(w_1, w_2)\|_{L^{3/2}(\mathbb{P}_{\nu_0})}\|A_{\nu_0}(w_3)\|_{L^3(\mathbb{P}_{\nu_0})} \\
&\lesssim (\|w_1w_2\|_{L^{3/2}(\nu_0)} + \|w_1\|_{L^{3/2}(\nu_0)}\|w_2\|_{L^{3/2}(\nu_0)})\|w_3\|_{L^3(\nu_0)} \\
&\lesssim \|w_1\|_{L^3}\|w_2\|_{L^3}\|w_3\|_{L^3}
\end{aligned}
$$

and for term (c) we likewise obtain

$$
E_{\nu_0}[|A_{\nu_0}(w_1)A_{\nu_0}(w_2)A_{\nu_0}(w_3)|] \lesssim \|w_1\|_{L^3}\|w_2\|_{L^3}\|w_3\|_{L^3}.
$$

We conclude

$$
E_{\nu_0}\left[|D^3\ell_n(\nu_0)[w_1, w_2, w_3]|\right] \lesssim \|w_1\|_{L^3}\|w_2\|_{L^3}\|w_3\|_{L^3}.
$$

Using Lemma 26 and the generalization of Hölder's inequality $\|\prod_{j=1}^k f_j\|_{L^1(\mu)} \leqslant \prod_{j=1}^k \|f_j\|_{L^{p_j}(\mu)}$ for $\sum_{j=1}^k \frac{1}{p_j} = 1$ and some measure μ, it follows in the same way that

$$
E_{\nu_0}\left[|D^4\ell_n(\nu_0)[w_1, w_2, w_3, w_4]|\right] \lesssim \|w_1\|_{L^4}\|w_2\|_{L^4}\|w_3\|_{L^4}\|w_4\|_{L^4}.
$$

For the fifth derivative we let $\widetilde{\nu}$ be either $\nu^{(s)}$ or $\nu_t^{(u)}$ and first apply a measure change

$$
E_{\nu_0}\left[|D^5\ell_n(\widetilde{\nu})[w_1, w_2, w_3, w_4]|\right] \lesssim E_{\widetilde{\nu}}\left[|D^5\ell_n(\widetilde{\nu})[w_1, w_2, w_3, w_4]|\right]
$$

$$\lesssim \|w_1\|_{L^5}\|w_2\|_{L^5}\|w_3\|_{L^5}\|w_4\|_{L^5}\|w_5\|_{L^5}.$$

We observe that

$$\omega_n^{L^p} = \frac{t}{\sqrt{n}}\|\eta\|_{L^p} + \delta_n \varepsilon_n^{L^p}$$

is the rate at which $\widetilde{\delta}_n(v)$ converges to zero in L^p. For the centring of the third, fourth and fifth derivative we use the following conditions

$$\begin{aligned} n\,\omega_n^{L^3}\left(\varepsilon_n^{L^3}\right)^2 &= o(1), \qquad & n\left(\omega_n^{L^3}\right)^3 &= o(1),\\ n\,\omega_n^{L^4}\left(\varepsilon_n^{L^4}\right)^3 &= o(1), \qquad & n\left(\omega_n^{L^4}\right)^4 &= o(1),\\ n\left(\varepsilon_n^{L^5}\right)^5 &= o(1), \qquad & n\left(\omega_n^{L^5}\right)^5 &= o(1). \end{aligned}$$

For the empirical process part we develop the remainder term R_n only to the third derivative so that it takes the form

$$\underbrace{\tfrac{1}{6}D^3\ell_n(\nu^{(s')})[v-v_0, v-v_0, v-v_0]}_{(i)}\underbrace{-\tfrac{1}{6}D^3\ell_n(\nu_t^{(u')})[v_t-v_0, v_t-v_0, v_t-v_0]}_{(ii)}.$$

We have $\|v-v_0\|_{L^p} \lesssim \varepsilon_n^{L^p}$ and $\|v_t-v_0\|_{L^p} \lesssim \varepsilon_n^{L^p} + \omega_n^{L^p}$. Both (i) and (ii) can be treated jointly by bounding a term of the form $D^3\ell_n(\widetilde{\nu}_n)[w,w,w]$ with $\widetilde{\nu}_n = \exp(\widetilde{v}_n)$, $\|\widetilde{v}_n\|_\infty \leqslant D$, and either $w = v - v_0$ or $w = v + \widetilde{\delta}_n - v_0$.

Let $\nu^{(r)} = \widetilde{\nu}_n \exp(rw)$ so that

$$\begin{aligned} &D^3\ell_n(\widetilde{\nu}_n)[w,w,w]\\ &= \underbrace{\sum_{j=1}^n \frac{d\frac{d^3}{dr^3}\mathbb{P}_{\nu^{(r)}}}{d\mathbb{P}_{\nu^{(r)}}}\bigg|_{r=0}(X_j)}_{(a)}\underbrace{-3\sum_{j=1}^n \frac{d\frac{d^2}{dr^2}\mathbb{P}_{\nu^{(r)}}}{d\mathbb{P}_{\nu^{(r)}}}\bigg|_{r=0}(X_j)\frac{d\frac{d}{dr}\mathbb{P}_{\nu^{(r)}}}{d\mathbb{P}_{\nu^{(r)}}}\bigg|_{r=0}(X_j)}_{(b)}\\ &\quad\underbrace{+2\sum_{j=1}^n\left(\frac{d\frac{d}{dr}\mathbb{P}_{\nu^{(r)}}}{d\mathbb{P}_{\nu^{(r)}}}\bigg|_{r=0}(X_j)\right)^3}_{(c)}. \end{aligned}$$

For term (a) we define the functions

$$g_v = A_{\widetilde{\nu}_n}w^3 + 3A_{\widetilde{\nu}_n}(w,w^2) + A_{\widetilde{\nu}_n}(w,w,w).$$

We denote $f_v = g_v - E_{\nu_0}[g_v]$. After centring the term (a) is given by $\sqrt{n}\mathbb{G}_n f_v$ with f_v varying in the class of functions corresponding to (59), where the functions f_v are defined as here. We bound using Lemma 26

$$\begin{aligned} &\|g_v - g_{v'}\|_\infty \lesssim (\varepsilon_n^{L^\infty} + \omega_n^{L^\infty})^2\|v-v'\|_\infty \text{ so that}\\ &\|f_v - f_{v'}\|_\infty \lesssim (\varepsilon_n^{L^\infty} + \omega_n^{L^\infty})^2\|v-v'\|_\infty. \end{aligned}$$

With $v = v(u)$ and $v' = v(u')$ from the definition of the prior we further bound $\|v - v'\|_\infty \lesssim \|u - u'\|_\infty$. We take the envelope F to be a constant function proportional to $\left(\varepsilon_n^{L^\infty} + \omega_n^{L^\infty}\right)^2$ and obtain $\sup_{\mathbb{Q}} N(\mathcal{F}, L^2(\mathbb{Q}), \varepsilon\|F\|_{L^2(\mathbb{Q})}) \leqslant (A/\varepsilon)^{2^J}$ for some $A \geqslant 2$ and for all $0 < \varepsilon < A$.

We bound σ by

$$\begin{aligned}\sigma &= \sup_{f\in\mathcal{F}} \|f\|_{L^2(\mathbb{P}_{\nu_0})} \leqslant \sup_{\|v-v_0\|_{L^2}\leqslant 2M\varepsilon_n^{L^2}} \|g_v\|_{L^2(\mathbb{P}_{\nu_0})} \lesssim \sup_{\|v-v_0\|_{L^2}\leqslant 2M\varepsilon_n^{L^2}} \|g_v\|_{L^2(\mathbb{P}_{\tilde\nu_n})} \\ &\lesssim \|w^3\|_{L^2(\tilde\nu_n)} + \|w^2\|_{L^2(\tilde\nu_n)}\|w\|_{L^2(\tilde\nu_n)} + \|w\|^3_{L^2(\tilde\nu_n)} \lesssim \|w\|^3_{L^6(\tilde\nu_n)} \\ &\lesssim \left(\varepsilon_n^{L^6} + \omega_n^{L^6}\right)^3 \lesssim \left(\varepsilon_n^{L^6}\right)^3 + \left(\omega_n^{L^6}\right)^3 \lesssim \left(\varepsilon_n^{L^\infty} + \omega_n^{L^\infty}\right)^2\left(\varepsilon_n^{L^2} + \omega_n^{L^2}\right).\end{aligned}$$

Using Corollary 3.5.8 in [18] this yields some $c > 0$ such that

$$\begin{aligned}E\|\mathbb{G}_n\|_{\mathcal{F}} \lesssim{}& \left(\varepsilon_n^{L^\infty} + \omega_n^{L^\infty}\right)^2\left(\varepsilon_n^{L^2} + \omega_n^{L^2}\right)2^{J/2}\left(\log\frac{c}{\varepsilon_n^{L^2} + \omega_n^{L^2}}\right)^{1/2} \\ &+ \frac{1}{\sqrt{n}}\left(\varepsilon_n^{L^\infty} + \omega_n^{L^\infty}\right)^2 2^J \log\frac{c}{\varepsilon_n^{L^2} + \omega_n^{L^2}}. \qquad (60)\end{aligned}$$

For the term (b) and (c) we obtain the same bounds for the uniform covering numbers and for σ as for term (a). So the bound (60) applies likewise to terms (b) and (c).

6.3. Simplification of Assumption 25

In this section we simplify Assumption 25 and reduce it to a condition involving η and δ_n only. To this end we recall ε_n from(51) and the L^p-contraction rates $\varepsilon_n^{L^p}$ from (53) both in Section 5. We set $2^J \approx n^{1/(2s+1)}$.

Assumption 27. *Suppose $t = O(1)$, $s > 11/6$ and $\mathcal{H}_n \subseteq L^\infty(I)$. Furthermore, assume for δ_n and uniformly for all $\eta \in \mathcal{H}_n$*

$$\delta_n n^{2/(2s+1)}(\log n)^{1+2\delta} = o(1), \qquad (61)$$

$$\|\eta\|_{L^2} = O(1), \qquad (62)$$

$$\|\eta\|_\infty n^{(-s+1)/(2s+1)}(\log n)^{1+\delta} = o(1), \qquad (63)$$

$$\|\eta\|_\infty n^{(-3s+11/2)/(2s+1)}(\log n)^{3+6\delta} = o(1). \qquad (64)$$

Remark 28. For $s > 9/4$ (and so in particular for $s > 10/4 = 5/2$) condition (63) implies condition (64).

Lemma 29. *Let $2^J \approx n^{1/(2s+1)}$ and grant Assumption 27. Then t, δ_n, $\mathcal{H}_n$ and $\varepsilon_n^{L^p}$ from (53) satisfy Assumption 25.*

Proof. The bias conditions are satisfied for this choice of 2^J. Further we have

$$\sqrt{n}\delta_n\varepsilon_n^{L^2}2^{J/2}\sqrt{\log\frac{c}{\varepsilon_n^{L^2}}} \lesssim \sqrt{n}\delta_n n^{-\frac{s-1/2}{2s+1}}(\log n)^{1/2+\delta} n^{\frac{1/2}{2s+1}}\sqrt{\log n}$$

$$= \delta_n n^{\frac{3/2}{2s+1}} (\log n)^{1+\delta} = o(1)$$

by (61). Next we verify

$$\frac{2^{J/2}}{\sqrt{n}} \sqrt{\log \frac{c}{\varepsilon_n^{L^2}}} \lesssim n^{-1/2} n^{\frac{1/2}{2s+1}} \sqrt{\log n} = n^{-s/(2s+1)} (\log n)^{1/2} \lesssim \varepsilon_n^{L^2}$$

and

$$n\delta_n (\varepsilon_n^{L^2})^2 = \delta_n n^{2/(2s+1)} (\log n)^{1+2\delta} = o(1)$$

using (61). For term III(i) we bound

$$t\|\eta\|_\infty \varepsilon_n^{L^2} 2^{J/2} \sqrt{\log \frac{c}{\varepsilon_n^{L^2}}} \lesssim \|\eta\|_\infty n^{(-s+1)/(2s+1)} (\log n)^{1+\delta} = o(1)$$

by (63). We check that

$$\frac{t^2}{\sqrt{n}} \|\eta\|_{L^4}^2 \lesssim n^{-1/2} \|\eta\|_\infty = o(1)$$

by (63) and that

$$t\delta_n \sqrt{n} \|\eta\|_{L^2} \varepsilon_n^{L^2} \lesssim \delta_n n^{1/(2s+1)} (\log n)^{1/2+\delta} = o(1)$$

by (61). For the centring of the third derivatives we bound

$$\begin{aligned} &n\omega_n^{L^3} \left(\varepsilon_n^{L^3}\right)^2 \lesssim n^{1/2} \|\eta\|_\infty^{1/3} \|\eta\|_{L^2}^{2/3} \left(\varepsilon_n^{L^3}\right)^2 + n\delta_n \left(\varepsilon_n^{L^3}\right)^3 \\ &\lesssim \|\eta\|_\infty^{1/3} n^{(-s+11/6)/(2s+1)} (\log n)^{1+2\delta} + \delta_n n^{(-s+3)/(2s+1)} (\log n)^{3/2+3\delta} = o(1), \end{aligned}$$

where we used (64) for the first term and (61) for the second term. Further we have

$$n\left(\omega_n^{L^3}\right)^3 \lesssim n\frac{t^3}{n^{3/2}} \|\eta\|_{L^3}^3 + n\delta_n^3 \left(\varepsilon_n^{L^3}\right)^3 \lesssim n^{-1/2} \|\eta\|_\infty + o(1) = o(1)$$

using (63) for the first term and $n\delta_n(\varepsilon_n^{L^3})^3 = o(1)$ from the next to last display for the second term. The terms for the centering of the fourth derivates are treated by

$$\begin{aligned} &n\omega_n^{L^4} (\varepsilon_n^{L^4})^3 \lesssim n\frac{t}{n^{1/2}} \|\eta\|_{L^4} (\varepsilon_n^{L^4})^3 + n\delta_n (\varepsilon_n^{L^4})^4 \\ &\lesssim n^{(-2s+11/4)/(2s+1)} (\log n)^{3/2+3\delta} \|\eta\|_\infty^{1/2} + n^{(-2s+4)/(2s+1)} (\log n)^{2+4\delta} \delta_n = o(1), \end{aligned}$$

where we used (64) for the first term and (61) for the second term, and by

$$n(\omega_n^{L^4})^4 \lesssim n\frac{t^4}{n^2} \|\eta\|_{L^4}^4 + n\delta_n^4 (\varepsilon_n^{L^4})^4$$

$$\lesssim n^{-1}\|\eta\|_\infty^2 + o(1) = o(1),$$

where we used (63) for the first term and the next to last display for the second term. Turning to the centring of the fifth derivatives we observe

$$n(\varepsilon_n^{L^5})^5 = n^{(-3s+5)/(2s+1)}(\log n)^{5/2+5\delta} = o(1)$$

and

$$n(\omega_n^{L^5})^5 \lesssim n\frac{t^5}{n^{5/2}}\|\eta\|_{L^5}^5 + n\delta_n^5(\varepsilon_n^{L^5})^5 \lesssim n^{-3/2}\|\eta\|_\infty^3 + o(1) = o(1)$$

using (63) for the first term and the next to last display for the second term. For the remainder term R_n we bound

$$\begin{aligned}
&\sqrt{n}\Big(\varepsilon_n^{L^\infty} + \omega_n^{L^\infty}\Big)^2\Big(\varepsilon_n^{L^2} + \omega_n^{L^2}\Big)2^{J/2}\bigg(\log\frac{c}{\varepsilon_n^{L^2} + \omega_n^{L^2}}\bigg)^{1/2}\\
&\lesssim \sqrt{n}\Big(\varepsilon_n^{L^\infty} + \frac{t}{\sqrt{n}}\|\eta\|_\infty\Big)^2\Big(\varepsilon_n^{L^2} + \frac{t}{\sqrt{n}}\|\eta\|_{L^2}\Big)2^{J/2}(\log n)^{1/2}\\
&\lesssim \sqrt{n}\Big(\big(\varepsilon_n^{L^\infty}\big)^2 + \frac{\|\eta\|_\infty^2}{n}\Big)\Big(\varepsilon_n^{L^2} + n^{-1/2}\Big)n^{(1/2)/(2s+1)}(\log n)^{1/2}\\
&\lesssim \Big(\big(\varepsilon_n^{L^\infty}\big)^2\varepsilon_n^{L^2} + \big(\varepsilon_n^{L^\infty}\big)^2 n^{-1/2} + \frac{\|\eta\|_\infty^2}{n}\varepsilon_n^{L^2} + \frac{\|\eta\|_\infty^2}{n^{3/2}}\Big)n^{(s+1)/(2s+1)}(\log n)^{1/2}\\
&\lesssim n^{(-2s+7/2)/(2s+1)}(\log n)^{2+3\delta} + n^{(-2s+5/2)/(2s+1)}(\log n)^{3/2+2\delta}\\
&\quad + \|\eta\|_\infty^2 n^{(-2s+1/2)/(2s+1)}(\log n)^{1+\delta} + \|\eta\|_\infty^2 n^{(-2s-1/2)/(2s+1)}(\log n)^{1/2} = o(1)
\end{aligned}$$

using that $s > 11/6$ for the first and the second term and (63) for the third and the fourth term. Finally for the condition that the first term dominates in R_n we verify

$$\begin{aligned}
&\frac{1}{\sqrt{n}}2^{J/2}\frac{1}{\varepsilon_n^{L^2} + \omega_n^{L^2}}\sqrt{\log\frac{c}{\varepsilon_n^{L^2} + \omega_n^{L^2}}}\\
&\qquad\lesssim n^{(-s-1/2)/(2s+1)}n^{(1/2)/(2s+1)}\frac{1}{\varepsilon_n^{L^2}}\sqrt{\log\frac{c}{\varepsilon_n^{L^2}}}\\
&\qquad\lesssim n^{(-1/2)/(2s+1)}(\log n)^{-\delta} = O(1).
\end{aligned}$$

□

7. Proof of Proposition 3

The Radon–Nikodym density in (14) is well defined in view of the convolution series representation of $\mathbb{P}_\nu$ in (4). That A_ν maps $L^2(\nu)$ into $L^2(\mathbb{P}_\nu)$ is proved in Lemma 26, and an application of Fubini's theorem gives $\int_I A_\nu(h)d\mathbb{P}_\nu = 0$ for all $h \in L^2(\nu)$. The expansion (13) follows by the same arguments used for the proof in Section 6.2 but is in fact easier and no empirical process tools are needed here. In the case $v \in V_J$ for some J the expansion follows directly from setting $v_0 = v$ and $\eta = h$ in (58). For the general case we consider the path $s \mapsto \exp(v + sh/\sqrt{n}) = \nu^{(s)}$ and obtain by a Taylor expansion for some $s \in [0,1]$

$$\begin{aligned}
&\ell_n(\nu_{h,n}) - \ell_n(\nu)\\
&= D\ell_n(\nu_0)\Big[\frac{h}{\sqrt{n}}\Big] + \tfrac{1}{2}D^2\ell_n(\nu_0)\Big[\frac{h}{\sqrt{n}}, \frac{h}{\sqrt{n}}\Big] + \tfrac{1}{6}D^3\ell_n(\nu^{(s)})\Big[\frac{h}{\sqrt{n}}, \frac{h}{\sqrt{n}}, \frac{h}{\sqrt{n}}\Big]\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n A_\nu(h)(X_i) - \frac{1}{2}\|A_\nu(h)\|^2_{L^2(\mathbb{P}_\nu)} + \sum_{j=1}^n \frac{d\frac{d^2}{ds^2}\mathbb{P}_{\nu^{(s)}}}{d\,\mathbb{P}_{\nu^{(s)}}}\bigg|_{s=0}[h,h](X_j)\\
&\quad + \bigg(- \sum_{j=1}^n \bigg(\frac{d\frac{d}{ds}\mathbb{P}_{\nu^{(s)}}}{d\,\mathbb{P}_{\nu^{(s)}}}[h](X_j)\bigg)^2\bigg|_{s=0} + \frac{1}{2}\|A_\nu(h)\|^2_{L^2(\mathbb{P}_\nu)}\bigg)\\
&\quad + \frac{1}{6n^{3/2}}D^3\ell_n(\nu^{(s)})[h,h,h]\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n A_\nu(h)(X_i) - \frac{1}{2}\|A_\nu(h)\|^2_{L^2(\mathbb{P}_\nu)} + I + II + III.
\end{aligned}$$

The terms I and II are both centred and are treated exactly as the term $IV(i)(a)$ and the centred version of $IV(i)(b)$ in Section 6.2. This yields $I + II = O_{\mathbb{P}_\nu^{\mathbb{N}}}(n^{-1/2}\|h\|^2_{L^4})$. The centring of term III is shown to be $O_{\mathbb{P}_\nu^{\mathbb{N}}}(n^{-3/2}\|h\|^3_{L^3})$, which is proved along the same lines as the centring of the third derivatives of the term R_n in Section 6.2 combined with the measure change there applied to the fifth derivatives. After centring the term III is shown to be of order $O_{\mathbb{P}_\nu^{\mathbb{N}}}(n^{-1}\|h\|^3_{L^6})$ with the same bounds as used for bounding σ when treating the empirical process part of R_n except that here h is fixed and so a simple variance bound suffices instead of the empirical process inequality used for R_n. We conclude $I + II + III = o_{\mathbb{P}_\nu^{\mathbb{N}}}(1)$.

8. Proof of Proposition 21

We define, for $L' > 0$ to be chosen

$$\Psi_n = \begin{cases} 0 & \text{if } \|\widehat{\nu} - \nu_0\|_{\mathbb{H}(\delta)} < L'\varepsilon_n \\ 1 & \text{if } \|\widehat{\nu} - \nu_0\|_{\mathbb{H}(\delta)} \geqslant L'\varepsilon_n. \end{cases}$$

Applying Lemma 20 with $K = n$ and $x = \sqrt{n}\varepsilon_n$ yields, for L' large enough, $E_{\nu_0}[\Psi_n] \to 0$ as $n \to \infty$. For the error of second type we obtain, for M large enough depending on L', C that, again by Lemma 20,

$$\begin{aligned}
&\sup_{\nu \in \overline{\mathcal{V}}: \|\nu - \nu_0\|_{\mathbb{H}(\delta)} \geqslant M\varepsilon_n} E_\nu[1 - \Psi_n]\\
&= \sup_{\nu \in \overline{\mathcal{V}}: \|\nu - \nu_0\|_{\mathbb{H}(\delta)} \geqslant M\varepsilon_n} \mathbb{P}_\nu^{\mathbb{N}}\big(\|\widehat{\nu} - \nu_0\|_{\mathbb{H}(\delta)} < L'\varepsilon_n\big)\\
&\leqslant \sup_{\nu \in \overline{\mathcal{V}}: \|\nu - \nu_0\|_{\mathbb{H}(\delta)} \geqslant M\varepsilon_n} \mathbb{P}_\nu^{\mathbb{N}}\big(\|\nu_0 - \nu\|_{\mathbb{H}(\delta)} - \|\nu - \widehat{\nu}\|_{\mathbb{H}(\delta)} < L'\varepsilon_n\big)
\end{aligned}$$

$$\leqslant \sup_{\nu\in\overline{\mathcal{V}}} \mathbb{P}_\nu^{\mathbb{N}} \left(\|\nu-\widehat{\nu}\|_{\mathbb{H}(\delta)} > (M/2)\varepsilon_n\right)$$
$$\leqslant e^{-(C+4)n\varepsilon_n^2} + \frac{1}{R_2} e^{-nR_2/\log n} \leqslant 2e^{-(C+4)n\varepsilon_n^2},$$

where we used $\varepsilon_n = o(1/\sqrt{\log n})$ and n large enough in the last inequality.

9. Proof of Proposition 23

Since v, v_0 are bounded and thus exp is Lipschitz on the range of v, v_0 we have

$$\mathbb{P}\left(\|\nu-\nu_0\|_{L^2} \leqslant \frac{\varepsilon_n}{\sqrt{K_D}}\right) \geqslant \mathbb{P}\left(\|v-v_0\|_\infty \leqslant c\varepsilon_n\right)$$
$$\geqslant \mathbb{P}\left(\sum_l 2^{l/2} \max_k |\beta_{lk} - 2^{-l}(l^2+1)^{-1}u_{lk}| < c'\varepsilon_n\right),$$

where $u_{lk}=0$ for $l \geqslant J$ and $\beta_{lk} = \langle v_0, \psi_{lk}\rangle$. We define $b_{lk} = 2^l(l^2+1)\beta_{lk}$ such that $|b_{lk}| \leqslant B$, and $M(J) = \sum_{l=-1}^{J-1}\sum_{k=0}^{(2^l-1)\vee 0} 1 = 2^J$. We can bound the last probability from below by

$$\mathbb{P}\left(\sum_{l\leqslant J-1} 2^{-l/2}(l^2+1)^{-1}\max_k |b_{lk}-u_{lk}| < c'\varepsilon_n - \bar{c}2^{-J_n s}/(J_n^2+1)\right)$$
$$\geqslant \mathbb{P}\left(\max_{l\leqslant J-1}\max_k |b_{lk}-u_{lk}| < c''\varepsilon_n\right) = \prod_{l\leqslant J-1}\prod_k \mathbb{P}\left(|b_{lk}-u_{lk}| < c''\varepsilon_n\right)$$
$$\geqslant \left(\frac{c''\varepsilon_n}{2B}\right)^{M(J)} \geqslant e^{-Cn\varepsilon_n^2}$$

for n large enough and for some constant $C>0$.

10. Proof of Lemma 24

a) Write B for the unit ball of the space $\mathbb{B} = \mathbb{B}(\delta)$ which can be shown to be closed under pointwise multiplication in the sense that $\|fg\|_{\mathbb{B}} \leqslant c_0\|f\|_{\mathbb{B}}\|g\|_{\mathbb{B}}$. Since $\nu_0^{-1} \in \mathbb{B}$, $\|\nu-\nu_0\|_{\mathbb{B}} \to 0$ we also have $\|(\nu-\nu_0)/\nu_0\|_\infty \lesssim \|(\nu-\nu_0)/\nu_0\|_{\mathbb{B}} \to 0$ and thus $\|[(\nu-\nu_0)/\nu_0]^k\|_{\mathbb{B}} \leqslant c_0^k \|(\nu-\nu_0)/\nu_0\|_{\mathbb{B}}^k$. Since eventually $\|(\nu-\nu_0)/\nu_0\|_{\mathbb{B}} < 1/(2c_0)$ we deduce that the series

$$g = \sum_k \frac{(-1)^k}{k}\left(\frac{\nu-\nu_0}{\nu_0}\right)^{k-1}$$

converges absolutely uniformly and in $\mathbb{B}$ and has $\|\cdot\|_{\mathbb{B}}$-norm less than a constant multiple of $\|\nu-\nu_0\|_{\mathbb{B}}$. Thus, using again the multiplication property of the norm

$$\|\log\nu - \log\nu_0\|_{\mathbb{H}(\delta)} = \sup_{f\in B}\left|\int f \log\left(1+\frac{\nu-\nu_0}{\nu_0}\right)\right|$$

$$= \sup_{f\in B}\left|\int(\nu-\nu_0)\sum_{k=1}^{\infty}\frac{(-1)^k}{k}\frac{(\nu-\nu_0)^{k-1}}{\nu_0^{k-1}}\frac{f}{\nu_0}\right|$$
$$= \sup_{f\in B}\left|\int(\nu-\nu_0)g\frac{f}{\nu_0}\right| \leqslant \sup_{h\in c_1 B}\left|\int h(\nu-\nu_0)\right| = c_1\|\nu-\nu_0\|_{\mathbb{H}(\delta)}.$$

b) For any j we have, using the Cauchy–Schwarz inequality,

$$\begin{aligned}
&\|\nu-\nu_0\|^2_{\mathbb{B}(\delta)}\\
&\lesssim \sum_{l\leqslant j}2^l l^{2\delta}\sum_k|\langle\nu-\nu_0,\psi_{lk}\rangle|^2 + j^{2\delta-2\delta'}\sum_{l>j}2^l l^{2\delta'}\sum_k|\langle\nu-\nu_0,\psi_{lk}\rangle|^2\\
&\leq 2^{2j}j^{4\delta}\sum_{l\leqslant j}2^{-l}l^{-2\delta}\sum_k|\langle\nu-\nu_0,\psi_{lk}\rangle|^2 + j^{2\delta-2\delta'}\|\nu-\nu_0\|_{B_{22}^{1/2,\delta'}}\\
&\lesssim 2^{2j}j^{4\delta}\|\nu-\nu_0\|_{\mathbb{H}(\delta)} + j^{-2(\delta'-\delta)}.
\end{aligned}$$

Using $\|\nu-\nu_0\|_{\mathbb{H}(\delta)} = o(1)$ and letting $j\to\infty$ slowly enough we deduce $\|\nu-\nu_0\|_{\mathbb{B}(\delta)}\to 0$.

References

[1] Denis Belomestny, Fabienne Comte, Valentine Genon-Catalot, Hiroki Masuda, and Markus Reiß. *Lévy matters IV*. Lecture Notes in Mathematics. Springer, 2015. MR3364253

[2] Denis Belomestny and Markus Reiß. Spectral calibration of exponential Lévy models. *Finance Stoch.*, 10(4):449–474, 2006. MR2276314

[3] Boris Buchmann and Rudolf Grübel. Decompounding: an estimation problem for Poisson random sums. *Ann. Statist.*, 31(4):1054–1074, 2003. MR2001642

[4] Ismaël Castillo. On Bayesian supremum norm contraction rates. *Ann. Statist.*, 42(5):2058–2091, 2014. MR3262477

[5] Ismaël Castillo. Pólya tree posterior distributions on densities. *Ann. Inst. Henri Poincaré Probab. Stat.*, 53(4):2074–2102, 2017. MR3729648

[6] Ismaël Castillo and Richard Nickl. Nonparametric Bernstein–von Mises theorems in Gaussian white noise. *Ann. Statist.*, 41(4):1999–2028, 2013. MR3127856

[7] Ismaël Castillo and Richard Nickl. On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures. *Ann. Statist.*, 42(5):1941–1969, 2014. MR3262473

[8] Ismaël Castillo and Judith Rousseau. A Bernstein–von Mises theorem for smooth functionals in semiparametric models. *Ann. Statist.*, 43(6):2353–2383, 2015. MR3405597

[9] Alberto J. Coca. Adaptive nonparametric estimation for compound Poisson processes robust to the discrete-observation scheme. arXiv:1803.09849, 2018.

[10] Alberto J. Coca. Efficient nonparametric inference for discretely observed compound Poisson processes. *Probab. Theory Related Fields*, 170(1-2):475–523, 2018. MR3748329

[11] Masoumeh Dashti and Andrew Stuart. The Bayesian approach to inverse problems. In: *Handbook of Uncertainty Quantification, Eds. Ghanem et al., Springer*, 2016. MR3839555

[12] Richard M. Dudley. *Real analysis and probability*. Cambridge Univ. Press, 2002. MR1932358

[13] Richard M. Dudley. *Uniform central limit theorems*. Cambridge Univ.Press, 2014. MR3445285

[14] Gerald B. Folland. *Real analysis*. Wiley, second edition, 1999. MR1681462

[15] Subhashis Ghosal, Jayanta K. Ghosh, and Aad W. van der Vaart. Convergence rates of posterior distributions. *Ann. Statist.*, 28(2):500–531, 2000. MR1790007

[16] Subhashis Ghosal and Aad W. van der Vaart. *Fundamentals of nonparametric Bayesian inference*. Cambridge University Press, New York, 2017. MR3587782

[17] Evarist Giné and Richard Nickl. Rates of contraction for posterior distributions in L^r-metrics, $1 \leq r \leq \infty$. *Ann. Statist.*, 39(6):2883–2911, 2011. MR3012395

[18] Evarist Giné and Richard Nickl. *Mathematical foundations of infinite-dimensional statistical models*. Cambridge University Press, 2016. MR3588285

[19] Matteo Giordano and Hanne Kekkonen. Bernstein-von Mises theorems and uncertainty quantification for linear inverse problems. *arXiv preprint arXiv:1811.04058*, 2018.

[20] Shota Gugushvili, Frank van der Meulen, and Peter Spreij. Nonparametric Bayesian inference for multidimensional compound Poisson processes. *Mod. Stoch. Theory Appl.*, 2(1):1–15, 2015. MR3356921

[21] François Monard, Richard Nickl, and Gabriel P. Paternain. Efficient nonparametric Bayesian inference for X-ray transforms. *Ann. Statist.*, 47(2):1113–1147, 2019. MR3909962

[22] Michael H. Neumann and Markus Reiß. Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli*, 15(1):223–248, 2009. MR2546805

[23] Richard Nickl. Donsker-type theorems for nonparametric maximum likelihood estimators. *Probab. Theory Related Fields*, 138(3-4):411–449, 2007. MR2299714

[24] Richard Nickl. Bernstein–von Mises theorems for statistical inverse problems I: Schrödinger equation. *J. Eur. Math. Soc. (JEMS)*, to appear, 2018.

[25] Richard Nickl and Kolyan Ray. Nonparametric statistical inference for drift vector fields of multi-dimensional diffusions. *Ann. Statist.*, to appear.

[26] Richard Nickl and Markus Reiß. A Donsker theorem for Lévy measures. *J. Funct. Anal.*, 263(10):3306–3332, 2012. MR2973342

[27] Richard Nickl, Markus Reiß, Jakob Söhl, and Mathias Trabs. High-frequency Donsker theorems for Lévy measures. *Probab. Th. Rel. Fields*,

164:61–108, 2016. MR3449386

[28] Richard Nickl and Jakob Söhl. Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions. *Ann. Statist.*, 45(4):1664–1693, 2017. MR3670192

[29] Kalyanapuram R. Parthasarathy. *Probability measures on metric spaces*. Probability and Mathematical Statistics, No. 3. Academic Press, Inc., New York-London, 1967. MR0226684

[30] Kolyan Ray. Bayesian inverse problems with non-conjugate priors. *Electron. J. Stat.*, 7:2516–2549, 2013. MR3117105

[31] Kolyan Ray. Adaptive Bernstein–von Mises theorems in Gaussian white noise. *Ann. Statist.*, 45(6):2511–2536, 2017. MR3737900

[32] Andrew M. Stuart. Inverse problems: a Bayesian perspective. *Acta Numer.*, 19:451–559, 2010. MR2652785

[33] Mathias Trabs. Information bounds for inverse problems with application to deconvolution and Lévy models. *Ann. Inst. H. Poincaré*, 51(4):1620–1650, 2015. MR3414460

[34] Hans Triebel. *Theory of function spaces*. Birkhäuser Verlag, Basel, 1983. MR0781540

[35] Aad W. van der Vaart. *Asymptotic statistics*. Cambridge University Press, 1998. MR1652247

[36] Aad W. van der Vaart and Jon A. Wellner. *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996. MR1385671

[37] Bert van Es, Shota Gugushvili, and Peter Spreij. A kernel type nonparametric density estimator for decompounding. *Bernoulli*, 13(3):672–694, 2007. MR2348746