On the Bernstein-von Mises theorem for the Dirichlet process[*]

Kolyan Ray[1] and Aad van der Vaart[2]

[1]*Department of Mathematics,*
Imperial College London,
United Kingdom
e-mail: kolyan.ray@imperial.ac.uk

[2]*Mathematical Institute,*
Leiden University,
Netherlands
e-mail: avdvaart@math.leidenuniv.nl

Abstract: We establish that Laplace transforms of the posterior Dirichlet process converge to those of the limiting Brownian bridge process in a neighbourhood about zero, uniformly over Glivenko-Cantelli function classes. For real-valued random variables and functions of bounded variation, we strengthen this result to hold for all real numbers. This last result is proved via an explicit strong approximation coupling inequality.

1. Results

Let $\mathbb{P}_n = n^{-1} \sum_{i=1}^n \delta_{Z_i}$ be the empirical distribution of an i.i.d. sample $Z_1, \ldots, Z_n$ from a distribution P_0 on some measurable space $(\mathcal{X}, \mathcal{A})$, and given $Z_1, \ldots, Z_n$ let P_n be a draw from the Dirichlet process with base measure $\nu + n\mathbb{P}_n$. Thus ν is a finite measure on the sample space and $P_n | Z_1, \ldots, Z_n \sim \mathrm{DP}(\nu + n\mathbb{P}_n)$ for all n, which is the posterior distribution obtained when equipping the distribution of the observations $Z_1, Z_2, \ldots, Z_n$ with a Dirichlet process prior with base measure ν. The case $\nu = 0$ is allowed; the process P_n is then known as the Bayesian bootstrap. For full definitions and properties, see the review in Chapter 4 of [15].

The Dirichlet process is the standard "nonparametric prior" on the set of probability distributions on a (Polish) sample space and was first made popular in Bayesian nonparametrics by Ferguson [13] and has subsequently been used in numerous statistical applications. The purpose of this note is to prove the

---

following result concerning the Bernstein-von Mises theorem for the Dirichlet process posterior.

Theorem 1. *Suppose $\mathcal{G}$ is a P_0-Glivenko-Cantelli class of measurable functions $g : \mathcal{X} \to \mathbb{R}$ with measurable envelope function G, such that $\nu e^{tG} \leq C e^{ct^2}$ for every $t > 0$ and some $c, C > 0$, and $P_0 G^{2+\delta} < \infty$ for some $\delta > 0$. Then there exists a neighbourhood of 0, such that for every t in the neighbourhood,*

$$\sup_{g \in \mathcal{G}} \left| \mathrm{E}\left[e^{t\sqrt{n}(P_n g - \mathbb{P}_n g)} \,\middle|\, Z_1, \ldots, Z_n \right] - e^{t^2 P_0 (g - P_0 g)^2 / 2} \right| \overset{as*}{\to} 0. \qquad (1.1)$$

Here we write $S_n \overset{as*}{\to} 0$ if there exist measurable random variables Δ_n with $|S_n| \leq \Delta_n$ and $\Delta_n \to 0$ for P_0^∞-almost every sequence $Z_1, Z_2, \ldots$. This is because the supremum in the theorem is not necessarily measurable in $Z_1, Z_2, \ldots$, and, when it is not, convergence on a set of non-measurability one may be less informative than it appears, as pointed out by a referee. (Cf. Chapter 1.9 in [35] for further discussion.)

The function $t \mapsto e^{t^2 \sigma^2 / 2}$ is the Laplace transform of the normal distribution with mean 0 and variance σ^2. The theorem thus says that the Laplace transform of the posterior Dirichlet process centered at the empirical measure tends to the Laplace transform of a centered normal distribution with variance $P_0(g - P_0 g)^2$ in a neighbourhood of 0. This implies that the posterior Dirichlet process tends in distribution to a normal distribution (see Section 2.4), which is a version of the Bernstein-von Mises theorem for the Dirichlet process prior (a weak version, as the usual theorem gives the approximation in the total variation distance; see Section 12.2 of [15] for discussion). The convergence of the Laplace transform is useful for handling for instance moments of the posterior distribution.

The main contribution of the theorem is, however, to provide uniformity in a class of functions g. This uniformity refers to the *marginal* posterior distributions of the process $\left(\sqrt{n}(P_n g - \mathbb{P}_n g) : g \in \mathcal{G} \right)$. The stronger sense of uniformity of distributional convergence of this process as a random element in the set $\ell^\infty(\mathcal{G})$ is known to be true if $\mathcal{G}$ is a Donsker class, as shown in [18] (also see [21, 22]). This is a much stronger property than Glivenko-Cantelli as assumed here.

Remark 1. *The assumption in Theorem 1 that the class $\mathcal{G}$ possesses an envelope function G with sub-Gaussian tails under the base measure of the Dirichlet process prior $Q \sim \mathrm{DP}(\nu)$ ensures that the distribution of QG possesses a finite moment generating function. If $G \in \mathcal{G}$, this is clearly necessary (see also Lemma 6(iii) below). It results from the fact that although vanishing in the limit, the prior remains present in the posterior for every n. In the case of the Bayesian bootstrap, which formally corresponds to taking $\nu = 0$, this condition can be omitted.*

Remark 2. *Theorem 1 can be extended to the assertion (1.1) for a sequence $\mathcal{G}_n$ of classes of measurable functions. Inspection of the proofs below shows that if the total mass of the base measures remains bounded, then it suffices that these*

classes satisfy

$$\sup_{g \in \mathcal{G}_n \cup \mathcal{G}_n^2} |\mathbb{P}_n g - P_0 g| \overset{as*}{\to} 0, \qquad \sup_{g \in \mathcal{G}_n} P_0 g^2 = O(1),$$

and possess envelope functions G_n such that $\max_{1 \leq i \leq n} G_n(Z_i) = o(\sqrt{n}/\log n)$, almost surely, and $\mathrm{E}e^{tQG_n/\sqrt{n}} \leq e^{ct}$ for $Q \sim \mathrm{DP}(\nu)$ and every $0 < t < \sqrt{n}$. The last condition is implied by uniform sub-Gaussianity of the envelopes G_n, but it is of course also satisfied if $\|G_n\|_\infty \lesssim \sqrt{n}$ for every n. (If the total mass of the base measures increases to infinity, then this last condition must be replaced by a condition that ensures the assumption of Lemma 6(iii).) For convergence in probability in (1.1), it suffices that these conditions hold in probability. If the classes $\mathcal{G}_n$ are separable, then uniformity over $\mathcal{G}_n^2$ is implied by uniformity over $\mathcal{G}_n$, as shown by Lemma 11 of [30].

Major applications of studying posterior Laplace transforms of functionals as in (1.1) include establishing semiparametric and nonparametric Bernstein-von Mises theorems [4, 5, 28, 31], especially for inverse problems [24, 25, 27], posterior contraction rates in the supremum norm [2, 6, 26] and convergence rates for Tikhonov-type penalised least squares estimators [24, 26]. Such proofs typically require uniformity over function classes as established in (1.1) and use likelihood expansions based on local asymptotic normality (LAN) of the model. Because the Dirichlet process prior does not give probability one to a dominated set of measures, the resulting posterior distribution cannot be derived using Bayes' formula; one cannot thus use the LAN approach of the aforementioned papers to prove (1.1).

Our result is applicable when a Dirichlet process prior is assigned to some distributional component of the model, such as the covariate distribution in regression models with random design. For example, Theorem 1 has recently been applied to establish semiparametric Bernstein-von Mises results for estimating average treatment effects in causal inference problems [29, 30]. Indeed, results there suggest that for estimating functionals, using a Dirichlet process prior on the covariate distribution can yield better performance than other common priors choices, especially in high-dimensional covariate settings.

Bernstein-von Mises and Donsker-type results have also been obtained for generalizations of the Dirichlet process, such as Pólya trees [3] and the Pitman-Yor process [18, 14], and similar style results to Theorem 1 can be expected to hold for such priors. While the main outline of our proof should be applicable, it relies on the explicit posterior representation (2.1) of the Dirichlet process for several technical computations, which would have to be extended to the explicit posterior representations available for these other priors.

The case $\mathcal{X} = \mathbb{R}$

The proof of Theorem 1 requires uniformly bounded exponential moments of the process $(\sqrt{n}(P_n g - \mathbb{P}_n g) : g \in \mathcal{G})$, which only holds for small $|t|$ under the moment condition $P_0 G^{2+\delta} < \infty$ of the theorem (see Lemmas 3 and 4). When

$\mathcal{X} = \mathbb{R}$, we can strengthen Theorem 1 to hold for all $t \in \mathbb{R}$ under significantly stronger conditions on $\mathcal{G}$.

We now assume $Z_1, Z_2, \ldots$ are i.i.d. random variables taking values in $\mathcal{X} = \mathbb{R}$. Recall that the total variation of a function $f : \mathbb{R} \to \mathbb{R}$ on an interval $[a, b]$ is

$$V_a^b(f) = \sup_{\Pi \in \mathcal{P}_a^b} \sum_{i=1}^{n_\Pi} |f(x_i) - f(x_{i-1})|,$$

where $\mathcal{P}_a^b = \{\Pi = (x_0, \ldots, x_{n_P}) : a = x_0 \leq x_1 \leq \cdots \leq x_{n_\Pi}, n_\Pi \in \mathbb{N}\}$ is the set of all partitions of $[a, b]$, and define $|f|_{BV} = \sup_{a,b} V_a^b(f)$.

Proposition 1. *Suppose $\mathcal{G}$ is a class of right-continuous functions $g : \mathbb{R} \to \mathbb{R}$ such that $\sup_{g \in \mathcal{G}} |g|_{BV} < \infty$. Then for every $t \in \mathbb{R}$, for P_0^∞-almost every sequence $Z_1, Z_2, \ldots$,*

$$\sup_{g \in \mathcal{G}} \left| \mathrm{E}\left[e^{t\sqrt{n}(P_n g - \mathbb{P}_n g)} \mid Z_1, \ldots, Z_n \right] - e^{t^2 P_0 (g - P_0 g)^2 / 2} \right| \to 0.$$

Since bounded variation balls are universal Donsker classes, this is a significantly stronger requirement than $\mathcal{G}$ being P_0-Glivenko-Cantelli in Theorem 1. We prove this result by exploiting a strong approximation, which establishes a rate of convergence for representations of these random variables defined on a common probability space and has various applications in probability and statistics, for instance studying distributional approximations of transformed random variables $\psi_n(\sqrt{n}(P_n - \mathbb{P}_n))$, where the functions ψ_n depend on n. For an overview of the theory of strong approximations and a survey of their applications in probability and statistics, see Csörgő and Révész [8] and Csörgő and Hall [9], respectively.

Let $F_n(t) = P_n 1_{(-\infty, t]}$ and $\mathbb{F}_n(t) = \mathbb{P}_n 1_{(-\infty, t]}$ denote the distribution function of the posterior Dirichlet process draw P_n and empirical distribution function, respectively. In a slight abuse of notation, we shall write $F \sim \mathrm{DP}(\nu)$ to mean $F = P 1_{(-\infty, \cdot]}$ for $P \sim DP(\nu)$. We write $|\nu| = \nu(\mathbb{R})$. Recall that a *Brownian bridge* $\{B(s) : s \in [0, 1]\}$ is a mean-zero Gaussian process with covariance function $\mathrm{E}B(s_1)B(s_2) = s_1 \wedge s_2 - s_1 s_2$. A *Kiefer* process $\{K(s, t) : s \in [0, 1], t \geq 0\}$ is a two-parameter mean-zero Gaussian process with covariance function $\mathrm{E}K(s_1, t_1)K(s_2, t_2) = (s_1 \wedge s_2 - s_1 s_2)(t_1 \wedge t_2)$. For each $t > 0$, $\{t^{-1/2}K(s, t) : s \in [0, 1]\}$ is a Brownian bridge, while $\{K(\cdot, n + 1) - K(\cdot, n) : n \geq 1\}$ is a sequence of independent Brownian bridges.

An almost sure strong approximation of the posterior Dirichlet process was established by Lo [23]. He showed that on a suitable probability space, there exist random elements F, K and $Z_1, Z_2, \cdots \sim^{iid} F_0$ such that $F | Z_1, \ldots, Z_n \sim DP(\nu + n\mathbb{P}_n)$ for every n, K is a Kiefer process independent of $Z_1, Z_2, \ldots$ and such that

$$\sup_{z \in \mathbb{R}} \left| \sqrt{n}(F - \mathbb{F}_n)(z) - \frac{K(F_0(z), n)}{n^{1/2}} \right| = O\left(\frac{(\log n)^{1/2}(\log \log n)^{1/4}}{n^{1/4}} \right) \quad \text{a.s.} \quad (1.2)$$

Applications of (1.2) include studying the large sample behaviour of the Bayesian bootstrap and smoothed Dirichlet process posterior [23], as well as receiver operating characteristic (ROC) curves [17]. We revisit this result by establishing an explicit coupling inequality in order to make uniform the constants in (1.2). This for instance allows control of exponential moments, which is needed to prove Proposition 1.

We henceforth assume that the underlying probability space is rich enough that all random variables and processes subsequently introduced may be defined on it. Since the posterior distribution is conditional on the observations $Z_1, \ldots, Z_n$, it is natural for a Bayesian to index the Gaussian process in (1.2) by the empirical distribution function $\mathbb{F}_n$ to obtain a conditional Gaussian approximation. The following is the explicit coupling inequality analogue of Lemma 6.3 of [23].

Theorem 2. *On a suitable probability space, there exist i.i.d. random variables* $Z_1, Z_2, \cdots \sim^{iid} P_0$ *and* F_n, *with* $F_n | Z_1, \ldots, Z_n \sim \mathrm{DP}(\nu + n\mathbb{P}_n)$, *and a sequence of Brownian bridges* (B_n) *independent of* $Z_1, Z_2, \ldots$, *such that*

$$
P\left(\sup_{z \in \mathbb{R}} \left| \sqrt{n}(F_n - \mathbb{F}_n)(z) - B_n(\mathbb{F}_n(z)) \right| \geq \frac{C_1(\log n + |\nu|) + x}{\sqrt{n}} \,\middle|\, Z_1, \ldots, Z_n \right)
$$
$$
\leq C_2 e^{-C_3 x},
$$

for all $x > 0$ *and* $n \geq 2$, *where* $C_1 - C_3$ *are universal constants.*

This result says one can couple the Dirichlet process posteriors to a sequence of Brownian bridges independent of the underlying data. The theorem could also be rephrased with the random variables $Z_1, Z_2, \ldots$ replaced by any real numbers $z_1, z_2, \ldots$ to emphasize this independence.

For $x = x_n$ taken equal to a constant times $\log n$, the right side sums finite over n, and hence the complement of the events at x_n are valid for every sufficiently large n, by the Borel-Cantelli lemma, for almost every sequence $Z_1, Z_2, \ldots$. Provided that $|\nu| = O(\log n)$, this yields that, for almost every sequence $Z_1, Z_2, \ldots$,

$$
\sup_{z \in \mathbb{R}} \left| \sqrt{n}(F - \mathbb{F}_n)(z) - B_n(\mathbb{F}_n(z)) \right| = O(n^{-1/2} \log n),
$$

which improves on the rate $n^{-1/4}(\log n)^{1/2}(\log \log n)^{1/4}$ in Lemma 6.3 of [23]. This is because we replace the KMT coupling used in [23], which involves a Kiefer process, with a direct quantile coupling due to [7] involving dependent Brownian bridges. The following is the analogous result when the Brownian bridges are related amongst themselves by tying them to a Kiefer process $K(\cdot, n) = \sum_{i=1}^{n} B_i$ for (B_i) independent Brownian bridges.

Theorem 3. *On a suitable probability space, there exist i.i.d. random variables* $Z_1, Z_2, \cdots \sim^{iid} P_0$ *and* F_n, *with* $F_n | Z_1, \ldots, Z_n \sim \mathrm{DP}(\nu + n\mathbb{P}_n)$, *and a Kiefer*

process K independent of $Z_1, Z_2, \ldots$, such that

$$P\left(\sup_{z \in \mathbb{R}} \left| \sqrt{n}(F_n - \mathbb{F}_n)(z) - n^{-1/2} K(\mathbb{F}_n(z), n) \right| \right.$$

$$\left. \geq \frac{C_1 |\nu| + x \log n}{n^{1/2}} + \frac{C_2 \sqrt{\log n}\, x^{3/4}}{n^{1/4}} \right| Z_1, \ldots, Z_n \right) \leq C_3 e^{-C_4 x},$$

for all $x > 0$ and $n \geq 2$, where $C_1 - C_4$ are universal constants.

Theorem 2 does not say anything about the joint distribution in n of the corresponding Brownian bridges and thus only "in probability" or "in distribution" limit results can be proved. On the other hand, despite the slower convergence rate, Theorem 3 can be used to establish the almost sure limiting behaviour of statistics of interest based upon $\sqrt{n}(F_n - \mathbb{F}_n)(z)$, for instance a law of the iterated logarithm.

If $|\nu| = O(n^{1/4}(\log n)^{5/4})$, the above yields $P(\cdot | Z_1, Z_2, \ldots)$-almost sure order $n^{-1/4}(\log n)^{5/4}$, significantly slower than the rate in Theorem 2. In Theorem 3 we follow the approach of [23] of using the KMT coupling rather than a quantile coupling as in Theorem 2. Indeed, up to logarithmic factors, a better rate is not obtainable for coupling a quantile process with a Kiefer process [10], as opposed to dependent Brownian bridges. We obtain a slightly slower almost sure rate than the $n^{-1/4}(\log n)^{1/2}(\log \log n)^{1/4}$ achieved in Lemma 6.3 of [23] due to technical arguments used to make the coupling non-asymptotic.

We may also index the Brownian bridges by the true distribution function F_0 at the expense of a slower rate. The following is the coupling inequality analogue of Theorem 2.1 of [23].

Corollary 1. *On a suitable probability space, there exist random variables $Z_1, Z_2, \cdots \sim^{iid} F_0$ and F_n, with $F_n | Z_1, \ldots, Z_n \sim \mathrm{DP}(\nu + n\mathbb{P}_n)$, and a sequence of Brownian bridges (B_n) independent of $Z_1, Z_2, \ldots$, such that for any $y > 0$, the event*

$$A_{n,y} = \{\sqrt{n}\|\mathbb{F}_n - F_0\|_\infty \leq y\}$$

satisfies $P(A_{n,y}) \geq 1 - 2e^{-2y^2}$ and

$$P\left(\sup_{z \in \mathbb{R}} \left| \sqrt{n}(F_n - \mathbb{F}_n)(z) - B_n(F_0(z)) \right| \geq \frac{C_1(\log n + |\nu|) + x}{n^{1/2}} \right.$$

$$\left. + \frac{C_2 \sqrt{y}(\sqrt{\log n} + \sqrt{x})}{n^{1/4}} \right| Z_1, \ldots, Z_n \right) 1_{A_{n,y}} \leq C_3 e^{-C_4 x},$$

for all $x > 0$ and $n \geq 2$, where $C_1 - C_4$ are universal constants.

The Bayesian interpretation is that there are events $(A_{n,y})$ of high P_0^n-probability depending only on the observations $Z_1, \ldots, Z_n$ on which one can approximate the posterior Dirichlet process with a sequence of Brownian bridges independent of the underlying data. If $|\nu| = O(n^{-1/4}(\log n)^{3/4})$, setting $y =$

$\sqrt{\delta \log n}$ with $\delta > 1/2$ gives that for P_0^∞-almost every sequence $Z_1, Z_2, \ldots$, we have approximation rate $n^{-1/4}(\log n)^{3/4}$, $P(\cdot | Z_1, Z_2, \ldots)$-almost surely. A similar, if more complicated, expression can be proved with the Brownian bridges (B_n) replaced by the Kiefer process K, in particular yielding an almost sure rate $O(n^{-1/4}(\log n)^{5/4})$, for P_0^∞-almost every sequence $Z_1, Z_2, \ldots$.

2. Proofs

2.1. Proof of Theorem 1

For given $Z_1, Z_2, \ldots$, the Dirichlet process posterior distribution can be represented in law as the convex combination

$$P_n g = V_n \, Qg + (1 - V_n) \frac{\sum_{i=1}^n W_i g(Z_i)}{\sum_{i=1}^n W_i}, \tag{2.1}$$

where $V_n \sim \mathrm{Beta}(|\nu|, n)$, $Q \sim \mathrm{DP}(\nu)$, and $W_1, W_2, \ldots$ are i.i.d. exponential variables with mean 1, all variables $V_n, Q, , W_1, W_2, \ldots$ independent. This follows for instance from Theorem 14.37 in [15] (with $\sigma = 0$), or representation properties of the Dirichlet distribution as in Proposition G.10 in the same reference. For $\nu = 0$, the variables Q and V_n should be interpreted as 0, and the first term vanishes. With the notation $\bar{W}_n = n^{-1} \sum_{i=1}^n W_i$, some algebra gives

$$\sqrt{n}(P_n g - \mathbb{P}_n g) = \sqrt{n} V_n \Big(Qg - \frac{\sum_{i=1}^n W_i g(Z_i)}{\sum_{i=1}^n W_i} \Big)$$
$$+ \frac{1}{\bar{W}_n} \frac{1}{\sqrt{n}} \sum_{i=1}^n (W_i - 1)(g(Z_i) - \mathbb{P}_n g). \tag{2.2}$$

We show that the first term is negligible and the second tends to a normal distribution.

The variable V_n is of the order $1/n$ and the first term in brackets on the right side of (2.2) is bounded above by $Q|g| + \max_{1 \le i \le n} |g(Z_i)|$. Because $\mathrm{E}Q|g| = \nu|g|/|\nu| < \infty$, and by Lemma 5 (below), since $P_0 g^2 < \infty$,

$$\frac{1}{\sqrt{n}} \max_{1 \le i \le n} |g(Z_i)| \to 0, \qquad \text{a.s.}, \tag{2.3}$$

the first term on the right tends to zero, in distribution conditionally given almost every sequence $Z_1, Z_2, \ldots$, as $n \to \infty$.

The leading factor $1/\bar{W}_n$ of the second term on the right of (2.2) tends to 1 almost surely, by the strong law of large numbers. If $P_0 g^2 < \infty$, then $\mathbb{P}_n g \to P_0 g$ and $\mathbb{P}_n g^2 \to P_0 g^2$, almost surely. This may be used together with (2.3) to show that, for every $\varepsilon > 0$,

$$\frac{1}{n} \sum_{i=1}^n \mathrm{E}\Big[(W_i - 1)^2 (g(Z_i) - \mathbb{P}_n g)^2 \, | \, Z_1, Z_2, \ldots \Big] \to P_0 (g - P_0 g)^2 =: \sigma_g^2, \qquad \text{a.s.}$$

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\Big[(W_i-1)^2(g(Z_i)-\mathbb{P}_ng)^2 1_{|W_i-1||g(Z_i)-\mathbb{P}_ng|>\varepsilon\sqrt{n}}\big|\, Z_1,Z_2,\ldots\Big]\to 0,\qquad\text{a.s.}$$

The Lindeberg central limit theorem then gives that, given almost every sequence $Z_1, Z_2, \ldots$, the sequence $n^{-1/2}\sum_{i=1}^{n}(W_i-1)(g(Z_i)-\mathbb{P}_ng)$ tends in distribution to a $N(0,\sigma_g^2)$-distribution.

The preceding together with Slutsky's lemma give that

$$\sqrt{n}(P_ng-\mathbb{P}_ng)|\,Z_1,Z_2,\ldots\rightsquigarrow N(0,\sigma_g^2),\qquad\text{a.s.}\tag{2.4}$$

If we would know that the moment generating function of the variables on the left were bounded, then this would imply convergence of exponential moments, and the proposition would be proved for $\mathcal{G}=\{g\}$. The approach to proving the proposition will be to strengthen first the preceding display to uniformity in g, and next show that exponential moments of the variables on the left are suitably bounded.

For the uniformity, we use the assumption that $\mathcal{G}$ is Glivenko-Cantelli. This is not overly strong, and it may be not far off from necessary. Indeed, if the variables W_i-1 were standard normal instead of exponential, and the leading factor $1/\bar{W}_n$ were not present and $\nu=0$, then the conditional distribution of the left side of the preceding display would be $N\big(0,\mathbb{P}_n(g-\mathbb{P}_ng)^2\big)$, and convergence of these normal distributions to $N(0,\sigma_g^2)$ would imply the convergence $\mathbb{P}_n(g-\mathbb{P}_ng)^2\to\sigma_g^2$, uniformly in g if the convergence in distribution were uniform. This is close to the Glivenko-Cantelli property.

To take account of possible non-measurability of the supremum in Theorem 1, we use an explicit bound on the distance between the distributions in (2.4).

Lemma 1. *For independent mean-zero random variables $X_1,\ldots,X_n$, the bounded Lipschitz distance between the law of $n^{-1/2}\sum_{i=1}^{n}X_i$ and the mean-zero normal distribution with the same variance satisfies, for any $0<\varepsilon<1$, and $H(u)=u+u^{1/3}$ and $H_0(u)=u^{1/4}\big(1+|\log u|^{1/2}\big)$,*

$$d\Big(\mathcal{L}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i\Big),N\Big(0,\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}X_i^2\Big)\Big)$$
$$\lesssim \varepsilon+\frac{1}{\varepsilon^2}H\Big(\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}X_i^2 1_{|X_i|>\varepsilon\sqrt{n}}\Big)+H_0\Big(\frac{\varepsilon}{n}\sum_{i=1}^{n}\mathrm{E}X_i^2\Big)\tag{2.5}$$

Proof. We combine results by [36] and [11], as outlined in the proof of Proposition A.5.2 in [35], which gives the lemma for i.i.d. variables. The first step is to note that, for $Z_i=X_i 1_{|X_i|\le\varepsilon\sqrt{n}}-\mathrm{E}X_i 1_{|X_i|\le\varepsilon\sqrt{n}}$, by Chebychev's inequality,

$$P\Big(\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-Z_i)\Big|>\varepsilon\Big)\le\frac{1}{\varepsilon^2 n}\sum_{i=1}^{n}\mathrm{E}X_i^2 1_{|X_i|>\varepsilon\sqrt{n}}.$$

We can then use Strassen's theorem (see e.g. [12], Theorem 11.6.2) to see that the Prohorov distance between the laws of $n^{-1/2}\sum_{i=1}^{n}X_i$ and $n^{-1/2}\sum_{i=1}^{n}Z_i$

is bounded above by the maximum of ε and the right side of the last display. Next, we apply Theorem B in [11] to bound the Prohorov distance between the law of $n^{-1/2}\sum_{i=1}^{n} Z_i$ and the mean-zero normal distribution with the same variance by a multiple of $H_0(\varepsilon n^{-1}\sum_{i=1}^{n} \mathrm{E}X_i^2 1_{X_i \le \varepsilon\sqrt{n}})$, where we have used $\mathrm{E}|Z_i|^3 \le 2\varepsilon\sqrt{n}\mathrm{E}X_i^2$. Since $\sup_{0<v\le u} H_0(v) \le CH_0(u)$ for an absolute constant $C > 0$, this is bounded by a multiple of the third term of (2.5). Finally, the distance between this normal distribution and the normal distribution as in the lemma is bounded by $(n^{-1}\sum_{i=1}^{n}\mathrm{E}X_i^2 1_{|X_i|>\varepsilon\sqrt{n}})^{1/3}$ using Lemma 2.1 of [11]

These estimates use the Prohorov distance, but are also valid for the bounded Lipschitz distance, which is bounded above by twice the Prohorov distance (see [12], Corollary 11.6.5). (Then we may also replace $u^{1/3}$ by $u^{1/2}$ in H.) $\qquad\square$

The idea of the lemma is that the first and third terms on the right can be made arbitrarily small by choice of ε if $n^{-1}\sum_{i=1}^{n}\mathrm{E}X_i^2$ remains bounded (as $H_0(u) \to 0$ as $u \downarrow 0$), while for fixed $\varepsilon > 0$ the middle term tends to zero as $n \to \infty$ if the Lindeberg condition holds.

Lemma 2. *Suppose $\mathcal{G}$ is a P_0-Glivenko-Cantelli class of measurable functions $g : \mathcal{X} \to \mathbb{R}$ with envelope function G such that $\nu G < \infty$ and $P_0 G^2 < \infty$. Then the convergence in (2.4) is uniform in $g \in \mathcal{G}$. More precisely, for d the bounded Lipschitz distance,*

$$\sup_{g\in\mathcal{G}} d\Big(\mathcal{L}\big(\sqrt{n}(P_n g - \mathbb{P}_n g)\big| Z_1, Z_2, \ldots\big), N(0, \sigma_g^2)\Big) \overset{as*}{\to} 0.$$

Proof. By the square-integrability of G and conservation of the Glivenko-Cantelli property under continuous transformations (see [34]), the set $\{g^2 : g \in \mathcal{G}\}$ is also Glivenko-Cantelli. Thus, for $\sigma_{n,g}^2 = \mathbb{P}_n(g - \mathbb{P}_n g)^2$,

$$\sup_{g\in G} |\sigma_{n,g}^2 - \sigma_g^2| \overset{as*}{\to} 0.$$

We now apply (2.5) to the variables $X_i = (W_i - 1)\big(g(Z_i) - \mathbb{P}_n g\big)$ conditionally given $Z_1, Z_2, \ldots$, and that $\sup_{0<v\le u} H_0(v) \le CH_0(u)$, to see that

$$d\Big(\mathcal{L}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(W_i - 1)\big(g(Z_i) - \mathbb{P}_n g\big)\big| Z_1, Z_2, \ldots, \Big), N(0, \sigma_{n,g}^2)\Big)$$

$$\lesssim \varepsilon + \frac{1}{\varepsilon^2} H\Big(\sup_g \sigma_{n,g}^2 \mathrm{E}(W_1 - 1)^2 1_{|W_1-1|2\max_{1\le i\le n} G(Z_i)>\varepsilon\sqrt{n}}\Big)$$

$$+ H_0\Big(\varepsilon \sup_g \sigma_{n,g}^2\Big).$$

Here, $\sup_g \sigma_{n,g}^2$ can be further bounded by $\mathbb{P}_n G^2$, resulting in a uniform upper on the distance that is a measurable function of $Z_1, Z_2, \ldots$. The resulting measurable upper bound tends to zero almost surely, as $n \to \infty$ followed by $\varepsilon \to 0$, because $\max_{1\le i\le n} G(Z_i) = o(\sqrt{n})$ almost surely by Lemma 5 and $P_0 G^2 < \infty$. Because $d\big(N(0, \sigma^2), N(0, \tau^2)\big) \le \sqrt{|s^2 - \tau^2|}$ for any $\sigma, \tau > 0$, the variances $\sigma_{n,g}^2$ in the left side can be replaced by their limits σ_g^2 by the preceding display.

This shows the convergence of the second term in (2.2), without the denominator $\bar{W}_n$, to the claimed Gaussian limits. Here $\bar{W}_n \to 1$, almost surely, and the first term is bounded above by $\sqrt{n} V_n \big(QG + \max_{1 \le i \le n} G(Z_i) \big)$, which tends to zero almost surely, by the argument given before. Since $d\big(\mathcal{L}(\mu + \sigma X), \mathcal{L}(X)\big) \le |\mu| + |\sigma - 1| \, \mathrm{E}|X|$, for any μ, σ and random variable X, the scaling by $\bar{W}_n$ and shifting by the first term does not change the Gaussian limits. $\qquad\square$

Lemma 3. *Suppose that the conclusion of Lemma 2 holds and for some $T > 0$, there exists a measurable random variable U such that*

$$\limsup_{n \to \infty} \sup_{g \in \mathcal{G}} \mathrm{E}\big[e^{T\sqrt{n}(P_n g - \mathbb{P}_n g)} \,\big|\, Z_1, \ldots, Z_n \big] \le U < \infty, \qquad a.s. \qquad (2.6)$$

Then (1.1) holds for $0 \le t < T$. Furthermore, if (2.6) holds for some $T < 0$, then (1.1) holds for $T < t \le 0$.

Proof. For a fixed $t > 0$ and $M > 0$, the function $h_M(x) = e^{tx} \wedge M$ is bounded and Lipschitz. Lemma 2 thus gives that, with E_Z denoting the conditional expectation given $Z_1, Z_2, \ldots$,

$$\sup_{g \in \mathcal{G}} \Big| \mathrm{E}_Z \big[e^{t\sqrt{n}(P_n g - \mathbb{P}_n g)} \wedge M \big] - \int h_M \, dN(0, \sigma_g^2) \Big| \overset{as*}{\to} 0.$$

Since $\sup_{g \in \mathcal{G}} \sigma_g^2 \le P_0 G^2 < \infty$, we can choose M such that $\big| \int h_M \, dN(0, \sigma_g^2) - e^{t^2 \sigma_g^2 / 2} \big|$ is arbitrary small, uniformly in $g \in \mathcal{G}$. Furthermore,

$$\Big| \mathrm{E}_Z \big[e^{t\sqrt{n}(P_n g - \mathbb{P}_n g)} \wedge M - e^{t\sqrt{n}(P_n g - \mathbb{P}_n g)} \big] \Big| \le \mathrm{E}_Z \big[e^{t\sqrt{n}(P_n g - \mathbb{P}_n g)} 1_{e^{t\sqrt{n}(P_n g - \mathbb{P}_n g)} \ge M} \big]$$

$$\le \frac{1}{M^{(T-t)/t}} \, \mathrm{E}_Z \big[e^{T\sqrt{n}(P_n g - \mathbb{P}_n g)} \big].$$

For $t < T$ and sufficiently large M and n, this is arbitrarily small, uniformly in $g \in \mathcal{G}$, by assumption (2.6).

The proof of the assertion with $T < 0$ is similar (or replace $P_n - \mathbb{P}_n$ by $\mathbb{P}_n - P_n$ in the argument). $\qquad\square$

Lemma 4. *If $\mathcal{G}$ has envelope function G such that $\int e^{tG} \, d\nu \le C e^{ct^2}$ for every $t > 0$ and some $C, c > 0$, and $P_0 G^{2+\delta} < \infty$ for some $\delta > 0$, then (2.6) holds for every T in a sufficiently small neighbourhood of 0.*

Proof. By the Cauchy-Schwarz inequality, $\mathrm{E} e^{T(Y_1 + Y_2)} < \infty$ if $\mathrm{E} e^{2TY_i} < \infty$, for $i = 1, 2$. Thus it suffices to prove that the two terms on the right side of (2.2) both possess finite exponential moments that are bounded in n.

Since $P_0 G^2 < \infty$, we have that $\varepsilon_n := \max_{1 \le i \le n} G(Z_i)/\sqrt{n} \to 0$, almost surely by Lemma 5. The absolute value of the first term of (2.2) is bounded above by $n V_n (n^{-1/2} QG + \varepsilon_n)$, where ε_n tends to zero almost surely. Thus this term has bounded exponential moments by Lemma 6.

Next consider the second term on the right side of (2.2), or equivalently, assume that $\nu = 0$. The absolute value satisfies

$$\sqrt{n} |P_n g - \mathbb{P}_n g| \le \frac{1}{\bar{W}_n} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_i |g(Z_i) - \mathbb{P}_n g| \le 2\sqrt{n} \max_{1 \le i \le n} G(Z_i) = 2n\varepsilon_n.$$

Since $\mathrm{E}e^X = 1 + \int_0^\infty P(X \geq x)e^x dx$, it follows that for $T > 0$,

$$\mathrm{E}_Z\left[e^{T\sqrt{n}(P_n g - \mathbb{P}_n g)}\right]$$

$$= 1 + \int_0^{2T\varepsilon_n n} \mathrm{P}_Z\left(\frac{T}{\sqrt{n}}\sum_{i=1}^n (W_i - 1)(g(Z_i) - \mathbb{P}_n g) > x\bar{W}_n\right)e^x \, dx$$

$$\leq 1 + \int_0^\infty \mathrm{P}\left(\bar{W}_n < 1 - \sqrt{3x/n}\right)e^x \, dx$$

$$+ \int_0^{2T\varepsilon_n n} \mathrm{P}_Z\left(\frac{T}{\sqrt{n}}\sum_{i=1}^n (W_i - 1)(g(Z_i) - \mathbb{P}_n g) > x(1 - \sqrt{3x/n})\right)e^x \, dx.$$

The probability in the first integral on the far right is bounded above by $e^{-3x/2}$, for every $x > 0$, using (2.8). Thus the integral involving this term is bounded above by $\int_0^\infty e^{-x/2} \, dx = 2$. For x in the integration range of the second integral on the far right, the number $1 - \sqrt{3x/n}$ is at least $1/2$ for large enough n. Hence the preceding display is bounded above by

$$3 + \int_0^\infty \mathrm{P}_Z\left(\frac{T}{\sqrt{n}}\sum_{i=1}^n (W_i - 1)(g(Z_i) - \mathbb{P}_n g) > \frac{x}{2}\right)e^x \, dx$$

$$= 2 + \mathrm{E}_Z\left[e^{2Tn^{-1/2}\sum_{i=1}^n (W_i - 1)(g(Z_i) - \mathbb{P}_n g)}\right].$$

It suffices to show that the last expectation is finite and bounded in $g \in \mathcal{G}$ for some $T > 0$.

Let $\psi_1(x) = e^x - 1$, and let $\|\cdot\|_{\psi_1}$ be the corresponding Orlicz norm. Then, by Proposition A.1.6 and Lemma 2.2.2 in [35], with the norms interpreted conditionally given Z,

$$\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n (W_i - 1)(g(Z_i) - \mathbb{P}_n g)\right\|_{\psi_1}$$

$$\lesssim \left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n (W_i - 1)(g(Z_i) - \mathbb{P}_n g)\right\|_2 + \frac{1}{\sqrt{n}}\left\|\max_{1 \leq i \leq n}|W_i - 1||g(Z_i) - \mathbb{P}_n g|\right\|_{\psi_1}$$

$$\lesssim \sqrt{\mathbb{P}_n(g - \mathbb{P}_n g)^2} + \frac{\log n}{\sqrt{n}}\|W_1 - 1\|_{\psi_1}\max_{1 \leq i \leq n}|g(Z_i)|$$

$$\lesssim \sqrt{\mathbb{P}_n G^2} + \frac{\log n}{\sqrt{n}}\max_{1 \leq i \leq n} G(Z_i).$$

Under the condition $P_0 G^{2+\delta} < \infty$, the first term is bounded almost surely by the law of large numbers, while the second term tends to zero almost surely by Lemma 5.

By the definition of the Orlicz norm, $\mathrm{E}e^{|Y|/C} \leq 2$ for $C \geq \|Y\|_{\psi_1}$ and any random variable Y. This concludes the proof for $T > 0$. For $T < 0$, we copy the preceding argument, but replace $W_i - 1$ by $1 - W_i$ and T by $|T|$. $\qquad\square$

Lemma 5. *If $Y_1, Y_2, \ldots$ are i.i.d. random variables with $\mathrm{E}|Y_i|^r < \infty$ for some $r > 0$, then $\max_{1 \leq i \leq n} |Y_i|/n^{1/r} \to 0$, almost surely.*

Proof. For any $y > 0$ we have

$$\max_{1 \le i \le n} \frac{|Y_i|^r}{n} \le \frac{y^r}{n} + \frac{1}{n} \sum_{i=1}^n |Y_i|^r 1_{|Y_i| > y}.$$

As $n \to \infty$ the first term tends to zero, for fixed y, while the second tends to $\mathrm{E}|Y_1|^r 1_{|Y_1| > y}$ by the law of large numbers, and can be made arbitrarily small by choice of y. $\quad\square$

Lemma 6. *Suppose that $Q \sim \mathrm{DP}(\nu)$ and $V_n \sim \mathrm{Beta}(|\nu|, n)$ are independent, and let G be a nonnegative, measurable function.*

   (i) *If $t_n \to t \in [0, 1)$, then $\mathrm{E}e^{nt_n V_n} \to (1 - t)^{-|\nu|}$. In particular, $\mathrm{E}e^{nt_n V_n} \to 1$ when $t_n \to 0$.*
   (ii) *If $\int e^{tG} d\nu \le Ce^{ct^2}$ for every $t > 0$ and some $C, c > 0$, then $\mathrm{E}e^{t\sqrt{n}V_n QG} \to 1$, for every $t \in [0, 1/\sqrt{c})$.*
   (iii) *If $\mathrm{E}e^{t\sqrt{n}V_n QG} = O(1)$ for some $t > 0$, then there exist $C, c > 0$ such that $\int e^{tG} d\nu \le Ce^{ct^2}$, for every $t > 0$.*

Proof. We have that

$$\int_0^1 e^{ntv} v^{|\nu|-1} (1 - v)^{n-1} dv = n^{-|\nu|} \int_0^n e^{tu} u^{|\nu|-1} (1 - u/n)^{n-1} du.$$

The integrand is dominated by $e^{tu} u^{|\nu|-1} e^{-u(1-1/n)}$, which is uniformly integrable for sufficiently large n and $t < 1$. Therefore, for fixed $t < 1$, the integral times $n^{|\nu|}$ is asymptotic to $\int_0^\infty u^{|\nu|-1} e^{-u(1-t)} du = (1 - t)^{-|\nu|} \Gamma(|\nu|)$ by the dominated convergence theorem. By the definition of the beta distribution, the expectation $\mathrm{E}e^{nt_n V_n}$ is the quotient of two of the integrals as in the display, with $t = t_n$ and with $t = 0$, respectively. This concludes the proof of (i).

For (ii), we first note that $\mathrm{E}e^{tQG} \le \mathrm{E}Qe^{tG} = \int e^{tG} d\nu/|\nu|$ by Jensen's inequality, whence $\mathrm{E}e^{tQG} \lesssim e^{ct^2}$ by the assumption. By the independence of Q and V_n and a coordinate substitution as under (i),

$$\mathrm{E}e^{t\sqrt{n}V_n QG} = \frac{\int_0^n \mathrm{E}e^{(tu/\sqrt{n})QG} u^{|\nu|-1} (1 - u/n)^{n-1} du}{\int_0^n u^{|\nu|-1} (1 - u/n)^{n-1} du}.$$

The integrand in the numerator tends pointwise to $u^{|\nu|-1} e^{-u}$ and is dominated by a multiple of $e^{ct^2 u^2/n} u^{|\nu|-1} e^{-u(1-1/n)} \le u^{|\nu|-1} e^{-du}$ on $[0, n]$, for a constant $d < 1 - ct^2$ and sufficiently large n. The denominator is as before. The integrals thus have the same limit.

For (iii), note that by the stick-breaking representation of the Dirichlet process, the variable QG is stochastically larger than $WG(\theta)$, for $W \sim \mathrm{Beta}(1, |\nu|)$ independent of $\theta \sim \nu/|\nu|$. It follows that for any $t \ge 0$,

$$\mathrm{E}e^{tQG} \ge \mathrm{E}e^{tWG(\theta)} = \int_0^1 \int e^{twG} d\nu \, (1 - w)^{|\nu|-1} dw \ge \psi(t/2) \frac{1}{|\nu| 2^{|\nu|}}$$

for $\psi(t) = \int e^{tG} \, d\nu$. Then by the preceding calculations, for sufficiently large n,

$$
\begin{aligned}
\mathrm{E}e^{t\sqrt{n}V_n QG} &\geq \frac{1}{|\nu|2^{|\nu|+1}\Gamma(|\nu|)} \int_0^n \psi\big(tu/(2\sqrt{n})\big)u^{|\nu|-1}(1-u/n)^{n-1} \, du \\
&\geq \frac{\psi(t\sqrt{n}/4)}{|\nu|2^{|\nu|+1}\Gamma(|\nu|)} \int_{n/2}^n u^{|\nu|-1}(1-u/n)^{n-1} \, du.
\end{aligned}
$$

The integral is bounded below by $n^{-1}2^{-n}$, whence $\psi(t\sqrt{n}/4) \leq C(\nu)e^n$ for large enough n, if the left side of the display remains bounded in n. For sufficiently large s, there exists large enough n such that $t\sqrt{n-1}/4 < s \leq t\sqrt{n}/4$, and $\psi(s) \leq C(\nu)e^n \leq C(\nu)e^{32s^2/t^2}$. □

2.2. Proof of Proposition 1

Proof of Proposition 1. Because by assumption the variation of a function $g \in \mathcal{G}$ is bounded uniformly over all intervals $[a, b]$, the limits of $g(x)$ as $x \to \pm\infty$ exist and are finite. (Indeed the values $|g(a)|$, for $a < 0$, are bounded by $|g(0)| + V_a^0(g) \leq |g(0)| + V$ and hence every sequence $g(x_n)$ with $x_n \to -\infty$ has a converging subsequence. If there were two subsequences x_n and y_n with different limits, then these could without loss of generality be chosen alternating: $x_1 \geq y_1 \geq x_2 \geq y_2 \geq \cdots$ and the variation over the partitions containing $y_N, x_N, \ldots y_1, x_1$ would tend to infinity with N.) It can be seen that the variation of the extended function g over $[-\infty, \infty]$ is the supremum of the variations over all intervals $[a, b]$, and hence is also finite. In particular, the functions $g - g(-\infty)$ are uniformly bounded. As shifting the functions by a constant does not change the claim of the proposition, we can assume without loss of generality that $g(-\infty) = 0$, and that the class $\mathcal{G}$ has a uniformly bounded envelope function G. We can then decompose g as $g = g^+ - g^-$, for right-continuous, nondecreasing functions $g^+, g^- : [-\infty, \infty] \to \mathbb{R}$, uniformly bounded by $2V$ (e.g. Section 6.3 of [32]). Let $dg = dg^+ - dg^-$ be the corresponding signed (Stieltjes) measure, and $|dg| = dg^+ + dg^-$ its total variation.

We work on the probability space from Theorem 2. For (B_n) the Brownian bridges in that theorem and $g \in \mathcal{G}$, set

$$
\mathbb{W}_n g = -\int B_n \circ \mathbb{F}_n \, dg.
$$

It can be seen that given $Z_1, Z_2, \ldots$, the variable $\mathbb{W}_n g$ possesses a $N(0, \|g - \mathbb{P}_n g\|_{L^2(\mathbb{P}_n)}^2)$-distribution, whence $\mathbb{W}_n$ is a $\mathbb{P}_n$-Brownian bridge process, indexed by $\mathcal{G}$.

The process $F = F_n$ in Theorem 2 is the distribution function of the posterior Dirichlet process P_n. By partial integration, we have

$$
(P_n - P)g = \int g \, d(F_n - \mathbb{F}_n) = -\int (F_n - \mathbb{F}_n) \, dg.
$$

Writing $\Delta_n = \|\sqrt{n}(F_n - \mathbb{F}_n) - B_n \circ \mathbb{F}_n\|_\infty$, we thus find that, for every sequence $Z_1, Z_2, \ldots,$

$$|\sqrt{n}(P_n - \mathbb{P}_n)g - \mathbb{W}_n(g)| = \left| \int \left(\sqrt{n}(F_n - \mathbb{F}_n) - B_n \circ \mathbb{F}_n \right) dg \right| \leq 4V\Delta_n.$$

Since bounded variation balls are uniform Donsker classes, $\mathcal{G}$ is P_0-Glivenko-Cantelli. Thus to prove the proposition, by Lemmas 2 and 3, we need show only that (2.6) holds for all $T \in \mathbb{R}$. Using the last display and Cauchy-Schwarz,

$$\sup_{g \in \mathcal{G}} \mathrm{E}_Z \left[e^{T\sqrt{n}(P_n - \mathbb{P}_n)g} \right] \leq \left(\mathrm{E}_Z \left[e^{8TV\Delta_n} \right] \right)^{1/2} \times \sup_{g \in \mathcal{G}} \left(\mathrm{E}_Z \left[e^{2T\mathbb{W}_n(g)} \right] \right)^{1/2}.$$

The first term converges to 1 as $n \to \infty$ for every sequence $Z_1, Z_2, \ldots$ by Lemma 8 below. The second term equals $\sup_{g \in \mathcal{G}} e^{T^2 \mathbb{P}_n (g - \mathbb{P}_n g)^2} \leq e^{T^2 \mathbb{P}_n G^2} \to e^{T^2 P_0 G^2} < \infty$, P_0^∞-a.s. This establishes (2.6) and completes the proof. $\square$

2.3. Proofs of strong approximation results

We recall some useful facts. For a centered Gaussian process $(G_t)_{t \in T}$ with countable index set T satisfying $\sup_{t \in T} |G_t| < \infty$ (Borell's inequality – Theorem 7.1 of [20]):

$$P \left(\sup_{t \in T} |G_t| \geq \mathrm{E} \sup_{t \in T} |G_t| + x \right) \leq e^{-\frac{x^2}{2\sigma^2}}, \qquad (2.7)$$

for every $x > 0$, where $\sigma^2 = \sup_{t \in T} \mathrm{E}G_t^2 < \infty$. Note that if G has continuous sample paths and $T \subset \mathbb{R}$ is uncountable, (2.7) still holds, since we may restrict the supremum to a countable skeleton of T.

For $X_\theta \sim \mathrm{Gamma}(\theta, 1)$, we have the pair of exponential inequalities

$$P(X_\theta > \theta + \sqrt{2\theta x} + x) \leq e^{-x}, \qquad P(X_\theta < \theta - \sqrt{2\theta x}) \leq e^{-x}, \qquad (2.8)$$

for every $x > 0$, see p. 28-29 of [1]. We also denote by P_Z the conditional probability given $Z_1, \ldots, Z_n$.

Proof of Theorem 2. Recall that $F|Z_1, \ldots, Z_n = P_n 1_{(-\infty, \cdot]}$ and let $\bar{F}|Z_1, \ldots, Z_n = \bar{P}_n 1_{(-\infty, \cdot]}$ for $\bar{P}_n \sim \mathrm{DP}(n\mathbb{P}_n)$. Using the representation (2.1), conditionally on $Z_1, \ldots, Z_n$,

$$\|\sqrt{n}(F - \mathbb{F}_n) - \sqrt{n}(\bar{F} - \mathbb{F}_n)\|_\infty = \sup_{t \in \mathbb{R}} \sqrt{n} \left| (P_n - \bar{P}_n) 1_{(-\infty, t]} \right| \leq \sqrt{n} V_n,$$

where $V_n \sim \mathrm{Beta}(|\nu|, n)$ is independent of $\bar{F}$. The random variable V_n is equal in distribution to $X/(X + Y_n)$, where $X \sim \mathrm{Gamma}(|\nu|, 1)$ and $Y_n \sim \mathrm{Gamma}(n, 1)$ are independent. Applying (2.8) gives $P(Y_n < Cn) \leq e^{-x}$ for $C = 1 - \sqrt{2/3} > 1/6$ and any $0 < x \leq n/3$, and then that $P(X/Y_n > 6n^{-1}(|\nu| + \sqrt{2|\nu|x} + x)) \leq 2e^{-x}$ for all $0 < x \leq n/3$. For $x \geq n/3$, we have the trivial probability bound

$$P(X/(X + Y_n) \geq 6n^{-1}(|\nu| + \sqrt{2|\nu|x} + x)) \leq P(X/(X + Y_n) \geq 2) = 0.$$

Combining the above and using $2\sqrt{|\nu|x} \le |\nu| + x$,

$$P_Z(\|\sqrt{n}(F - \mathbb{F}_n) - \sqrt{n}(\bar{F} - \mathbb{F}_n)\|_\infty \ge 12n^{-1/2}(|\nu| + x)) \le 2e^{-x}, \qquad (2.9)$$

for all $x > 0$. It therefore remains to show the desired exponential inequality with $\sqrt{n}(\bar{F} - \mathbb{F}_n)$ instead of $\sqrt{n}(F - \mathbb{F}_n)$.

Let $U_1, \ldots, U_{n-1} \sim U(0,1)$ be i.i.d. and independent of $(Z_i)_{i \ge 1}$ and denote the corresponding order statistics by $0 = U_{(0)} < U_{(1)} < \cdots < U_{(n-1)} < U_{(n)} = 1$. For given $Z_1, \ldots, Z_n$, the Bayesian bootstrap posterior distribution can be represented in law as $\bar{P}_n = \sum_{i=1}^{n}(U_{(i)} - U_{(i-1)})\delta_{Z_{(i)}}$, where $Z_{(1)} \le \cdots \le Z_{(n)}$ are the order statistics of the sample and we have used the exchangeability of $(U_{(i)} - U_{(i-1)} : 1 \le i \le n-1)$. This gives

$$\bar{F}(z) = \sum_{i=1}^{n}(U_{(i)} - U_{(i-1)})1_{\{Z_{(i)} \le z\}}. \qquad (2.10)$$

Define the empirical quantile function $Q_{n-1}(t)$ of the U_i's by

$$Q_{n-1}(t) = U_{(i)} \qquad \text{if} \quad \frac{i-1}{n-1} < t \le \frac{i}{n-1}, \quad i = 1, 2, \ldots, n-1$$

and set $q_{n-1}(t) = \sqrt{n-1}(Q_{n-1}(t) - t)$ to be the uniform quantile process. By Theorem 1 of Csörgő and Révész [7], one can define for each n a Brownian bridge $\{\tilde{B}_n(t) : 0 \le t \le 1\}$ on the same probability space such that for all $x \ge 0$,

$$P\left(\sup_{0 \le t \le 1}\left|q_n(t) - \tilde{B}_n(t)\right| \ge \frac{c_1 \log n + x}{\sqrt{n}}\right) \le c_2 e^{-c_3 x}, \qquad (2.11)$$

where c_1, c_2, c_3 are universal constants. Since these Brownian bridges are constructed based on $(U_i)_{i \ge 1}$, which are independent of $(Z_i)_{i \ge 1}$, they may also be taken to be independent of $(Z_i)_{i \ge 1}$. Setting $B_n = \tilde{B}_{n-1}$ and following [23],

$$\begin{aligned}
\|\sqrt{n}(\bar{F} - \mathbb{F}_n) - B_n(\mathbb{F}_n)\|_\infty &= \max_{1 \le i < n}\left|\sqrt{n}\left(U_{(i)} - \frac{i}{n}\right) - B_n(i/n)\right| \\
&\le \max_{1 \le i < n}\left|\sqrt{n}\left(U_{(i)} - \frac{i}{n-1}\right) - B_n(i/n)\right| + n^{-1/2} \\
&\le \sqrt{\frac{n}{n-1}}\max_{1 \le i < n}\left|\sqrt{n-1}\left(U_{(i)} - \frac{i}{n-1}\right) - B_n(i/(n-1))\right| \\
&\quad + \sqrt{\frac{n}{n-1}}\max_{1 \le i < n}|B_n(i/(n-1)) - B_n(i/n)| \\
&\quad + \left(\sqrt{\frac{n}{n-1}} - 1\right)\max_{1 \le i < n}|B_n(i/n)| + n^{-1/2} \\
&=: I_B + II_B + III_B + n^{-1/2}.
\end{aligned} \qquad (2.12)$$

We prove separate exponential inequalities for $I_B - III_B$. For $n \ge 2$,

$$I_B \le \sqrt{2}\max_{1 \le i < n}\left|q_{n-1}(i/(n-1)) - \tilde{B}_{n-1}(i/(n-1))\right|$$

$$\leq \sqrt{2} \sup_{0 \leq t \leq 1} \left| q_{n-1}(t) - \tilde{B}_{n-1}(t) \right|,$$

and the required inequality follows from (2.11).

Setting $V_i = V_i^{(n)} = B_n(i/(n-1)) - B_n(i/n)$, we have $II_B \leq \sqrt{2} \max_{1 \leq i < n} |V_i|$ for $n \geq 2$. Since B_n is a Brownian bridge, $V_1, \ldots, V_{n-1}$ are Gaussian random variables with $V_i \sim N(0, \frac{i}{n(n-1)}(1 - \frac{i}{n(n-1)}))$. Thus $\mathrm{Var}(V_i) \leq 1/n$ for all i, and so the standard Gaussian maximal inequality Lemma 2.3.4 of [16] yields $\mathrm{E} \max_{1 \leq i < n} |V_i| \leq C\sqrt{\log n/n}$ for an absolute constant $C > 0$. Applying Borell's inequality (2.7), for $x > 0$,

$$P\left(II_B \geq \frac{C\sqrt{\log n} + x}{\sqrt{n}} \right) \leq e^{-cx^2}. \tag{2.13}$$

For III_B, recall that for a Brownian bridge B_n, $P(\|B_n\|_\infty > x) \leq 2e^{-2x^2}$ for $x > 0$ (Proposition 12.3.4 of [12]). Using the mean value theorem with $h(x) = (1-x)^{1/2}$, for some $\xi \in [0, 1/n]$,

$$\sqrt{\frac{n}{n-1}} - 1 = \sqrt{\frac{n}{n-1}}\left(1 - \sqrt{1 - 1/n} \right) = \sqrt{\frac{n}{n-1}} \frac{1}{2n\sqrt{1-\xi}} \leq \frac{1}{n}.$$

Therefore, $P(III_B \geq n^{-1}x) \leq 2e^{-2x^2}$ for $x > 0$.

Combining the exponential inequalities for $I_B - III_B$ via a union bound and comparing the dominating terms,

$$P_Z\left(\|\sqrt{n}(\bar{F} - \mathbb{F}_n) - B_n(\mathbb{F}_n)\|_\infty \geq \frac{C_1 \log n + x}{\sqrt{n}} \right) \leq C_2 e^{-C_3 x},$$

for all $x > 0$ and universal constants $C_1, C_2, C_3 > 0$. Together with (2.9) this gives the result. □

Proof of Theorem 3. Using the exponential inequality (2.9), we need show only the result with $\sqrt{n}(\bar{F} - \mathbb{F}_n)$ instead of $\sqrt{n}(F - \mathbb{F}_n)$, where $\bar{F}$ is defined in (2.10). Let $H_n(s) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{U_i \leq s\}}$ be the empirical distribution function of the i.i.d. random variables $U_1, U_2, \cdots \sim U(0, 1)$ in (2.10). With $\alpha_n(s) = \sqrt{n}(H_n(s) - s)$, $s \in [0, 1]$, the uniform empirical process, the KMT inequality (Theorem 4 of [19]) implies that there exists a Kiefer process $\tilde{K}(s, t)$ with

$$P\left(\sup_{0 \leq s \leq 1} \left| \alpha_n(s) - n^{-1/2}\tilde{K}(s, n) \right| \geq \frac{C(\log n)^2 + x \log n}{\sqrt{n}} \right) \leq De^{-cx} \tag{2.14}$$

for all $n \geq 1$, $x > 0$ and universal constants $C, c, D > 0$. We take as Kiefer process $K(s, t) = -\tilde{K}(s, t)$.

Arguing as in (2.12),

$$\|\sqrt{n}(\bar{F} - \mathbb{F}_n) - n^{-1/2}K(\mathbb{F}_n, n)\|_\infty$$
$$\leq \sqrt{\frac{n}{n-1}} \max_{1 \leq i < n} \left| \sqrt{n-1}\left(U_{(i)} - \frac{i}{n-1} \right) - (n-1)^{-1/2}K\left(\frac{i}{n-1}, n-1 \right) \right|$$

$$+ \max_{1 \le i < n} \left| \frac{\sqrt{n}}{n-1} K\left(\frac{i}{n-1}, n-1\right) - n^{-1/2} K(i/n, n) \right| + n^{-1/2}$$

$$=: I_K + II_K + n^{-1/2}.$$

We again establish separate exponential inequalities for I_K and II_K. For $n \ge 2$,

$$I_K \le \sqrt{2} \max_{1 \le i < n} \left| -\alpha_{n-1}(U_{(i)}) - (n-1)^{-1/2} K(U_{(i)}, n-1) \right|$$

$$+ \sqrt{\frac{2}{n-1}} \max_{1 \le i < n} \left| K(U_{(i)}, n-1) - K\left(\frac{i}{n-1}, n-1\right) \right|$$

$$\le \sqrt{2} \sup_{0 \le s \le 1} |\alpha_{n-1}(s) - (n-1)^{-1/2} \widetilde{K}(s, n-1)|$$

$$+ \sqrt{2}(n-1)^{-1/2} \sup_{0 \le s \le 1} |K(s, n-1) - K(H_{n-1}(s), n-1)|.$$

For the first term, we use the KMT inequality (2.14). Since $\{(n-1)^{-1/2} K(s, n-1) : s \in [0,1]\}$ is a Brownian bridge for each $n \ge 2$, we use the first inequality in Lemma 7 to deal with the second term. Together these yield

$$P_Z\left(I_K \ge \frac{C_1 (\log n)^2 + x \log n}{n^{1/2}} + \frac{C_2 \sqrt{\log n} x^{3/4}}{n^{1/4}} \right) \le C_3 e^{-C_4 x}$$

for all $x > 0$ and universal constants $C_1 - C_4 > 0$.

For $n \ge 2$,

$$II_K \le \frac{\sqrt{n}}{n-1} \max_{1 \le i < n} \left| K\left(\frac{i}{n-1}, n-1\right) - K\left(\frac{i}{n}, n-1\right) \right|$$

$$+ \frac{\sqrt{n}}{n-1} \max_{1 \le i < n} |K(i/n, n-1) - K(i/n, n)|$$

$$+ \left(\frac{\sqrt{n}}{n-1} - \frac{1}{\sqrt{n}} \right) \max_{1 \le i < n} |K(i/n, n)|$$

$$= II_K^{(1)} + II_K^{(2)} + II_K^{(3)}.$$

Using again that $\{(n-1)^{-1/2} K(s, n-1) : s \in [0,1]\}$ is a Brownian bridge, $II_K^{(1)}$ is equal in distribution to II_B, for which we use the inequality (2.13). Similarly, $II_K^{(3)} = (n-1)^{-1} \max_{1 \le i < n} |n^{-1/2} K(i/n, n)|$, which is equal in distribution to III_B up to a universal constant factor, and hence satisfies $P(II_K^{(3)} \ge Cn^{-1} x) \le 2e^{-2x^2}$ for all $x > 0$. For $n \ge 2$, we have $II_K^{(2)} \le Cn^{-1/2} \max_{1 \le i < n} |X_i|$, where $X_i = K(i/n, n-1) - K(i/n, n) \sim N(0, \frac{i}{n}(1 - \frac{i}{n}))$ satisfies $\text{Var}(X_i) \le 1/4$. By Lemma 2.3.4 of [16], $\mathbb{E} \max_{1 \le i < n} |X_i| \le C\sqrt{\log n}$ and so by Borell's inequality (2.7), $P(II_K^{(2)} \ge n^{-1/2}(C\sqrt{\log n} + x)) \le e^{-2x^2}$ for a universal constant $C > 0$ and all $x > 0$. Together these give for all $x > 0$,

$$P\left(II_K \ge \frac{C_1 \sqrt{\log n} + \sqrt{x}}{n^{1/2}} \right) \le C_2 e^{-C_3 x}.$$

Using the exponential inequalities for I_K and II_K, a union bound and that $x^{1/2} \lesssim \log n + x$,

$$P_Z \left(\| \sqrt{n}(\bar{F} - \mathbb{F}_n) - n^{-1/2} K(\mathbb{F}_n, n) \|_\infty \geq \frac{C_1 (\log n)^2}{n^{1/2}} \right.$$
$$\left. + \frac{x \log n}{n^{1/2}} + \frac{C_2 \sqrt{\log n} x^{3/4}}{n^{1/4}} \right) \leq C_3 e^{-C_4 x},$$

for all $x > 0$ and universal constants $C_1 - C_4 > 0$. The first term on the right-hand side dominates if and only if $x \leq D(\log n)^2 / n^{1/3}$ for a universal constant $D > 0$. For such x, the upper bound in the last display is bounded by $C_3 \exp(-C_4 (\log 2)^2 / 2^{1/3})$ for all $n \geq 2$, which can be made larger than 1 by taking C_3 universal and large enough. The last display is thus trivially satisfied for such x, which implies

$$P \left(\| \sqrt{n}(\bar{F} - \mathbb{F}_n) - n^{-1/2} K(\mathbb{F}_n, n) \|_\infty \geq \frac{x \log n}{n^{1/2}} + \frac{C_2 \sqrt{\log n} x^{3/4}}{n^{1/4}} \right) \leq C_3 e^{-C_4 x},$$

for all $x > 0$ and (different) universal constants $C_2 - C_4 > 0$. Together with (2.9) this yields the result. □

Proof of Corollary 1. That $P(A_{n,y}) \geq 1 - 2e^{-2y^2}$ follows from the Dvoretzky-Kiefer-Wolfowitz-Massart inequality. Let $\{B_n : n \geq 1\}$ be the Brownian bridges from Theorem 2. By the triangle inequality,

$$\left| \sqrt{n}(F - \mathbb{F}_n)(z) - B_n(F_0(z)) \right| \leq \left| \sqrt{n}(F - \mathbb{F}_n)(z) - B_n(\mathbb{F}_n(z)) \right|$$
$$+ \left| B_n(\mathbb{F}_n(z)) - B_n(F_0(z)) \right|,$$

and the exponential inequality for the first term follows from Theorem 2. Since $\{B_n : n \geq 1\}$ are independent of $(Z_i)_{i \geq 1}$ by Theorem 2, applying the second inequality in Lemma 7 gives

$$P \left(\sup_{z \in \mathbb{R}} \left| B_{\mathbb{F}_n(z)} - B_{F_0(z)} \right| 1 \geq K \frac{\sqrt{y} \left(\sqrt{\log n} + \sqrt{x} \right)}{n^{1/4}} \, \middle| \, Z_1, \ldots, Z_n \right) 1_{A_{n,y}} \leq e^{-x},$$

for all $x > 0$ and a universal constant $K > 0$. The result follows by a union bound. □

Lemma 7. *Let $B = \{B_t : t \in [0,1]\}$ be a Brownian bridge and $\mathbb{F}_n$ be the empirical distribution function of $Z_1, \ldots, Z_n \sim^{iid} F_0$. Then there exists a universal constant $K > 0$ such that, for $n \geq 2$ and every $x > 0$,*

$$P \left(\sup_{z \in \mathbb{R}} \left| B_{\mathbb{F}_n(z)} - B_{F_0(z)} \right| \geq K \frac{\sqrt{\log n}}{n^{1/4}} \, x^{3/4} \right) \leq 2 e^{-x}.$$

If B is independent of $Z_1, \ldots, Z_n$, then there also exists $K > 0$ such that, for $n \geq 2$ and every $x > 0$,

$$P \left(\sup_{z \in \mathbb{R}} \left| B_{\mathbb{F}_n(z)} - B_{F_0(z)} \right| \geq K \| \mathbb{F}_n - F_0 \|_\infty^{1/2} \left(\sqrt{\log n} + \sqrt{x} \right) \, \middle| \, Z_1, \ldots, Z_n \right) \leq e^{-x}.$$

Proof. The intrinsic metric of the Brownian bridge is bounded above by the square root of the Euclidean distance, whence its metric entropy integral is a multiple of $\delta \mapsto \delta \max(\sqrt{\log(1/\delta)}, 1)$. Therefore, by Dudley's theorem (see [16], Theorem 2.3.8). $\mathrm{E} \sup_{s,t}\big[|B_s - B_t|/J(|s - t|)\big] < \infty$, for $J(\delta) = \sqrt{\delta} \max(\sqrt{\log(1/\delta)}, 1)$. Because the process $(s,t) \mapsto (B_s - B_t)/J(|s - t|)$ is centered Gaussian with uniformly bounded variance, we can apply Borell's inequality (2.7) to see that there exist constants $D, E > 0$ such that, for $y > 0$,

$$\mathrm{P}\Big(\sup_{0 < s,t < 1} \frac{|B_s - B_t|}{J(|s - t|)} > E + y\Big) \le 2e^{-Dy^2}.$$

There exists a constant $C > 0$ such that $Cy^2 \le D(y - E)^2$, for $y > 2E$. Then, for $y > 2E$,

$$\mathrm{P}\Big(\sup_{0 < s,t < 1} \frac{|B_s - B_t|}{J(|s - t|)} > y\Big) \le 2e^{-Cy^2}.$$

By making C if necessary still smaller, we can ensure that the right side is bigger than 1 for $y \le 2E$, and then the preceding inequality is valid for every $y > 0$.

By the Dvoretsky-Kiefer-Wolfowitz-Massart inequality, we also have, for $y > 0$,

$$\mathrm{P}\Big(\sup_{z \in \mathbb{R}} |\mathbb{F}_n(z) - F_0(z)| > y\Big) \le 2e^{-2ny^2}.$$

Combining these two inequalities, we see that, for every $y_1, y_2 > 0$,

$$\mathrm{P}\Big(\sup_{z \in \mathbb{R}} |B_{\mathbb{F}_n(z)} - B_{F_0(z)}| \ge y_1 J(y_2)\Big) \le 2e^{-Cy_1^2} + 2e^{-2ny_2^2}.$$

We choose $y_1 = \sqrt{2x/C}$ and $y_2 = \sqrt{x/n}$ to reduce the right side to $4e^{-2x}$, and then have $y_1 J(y_2) \ge K_1 x^{3/4} \max(\sqrt{\log(n/x)}, 1)/n^{1/4}$, for some constant $K_1 > 0$. For $x < \log 2$, we have that $2e^{-x} > 1$ and hence the first inequality of the lemma is trivially satisfied. For $x \ge \log 2$, we have $4e^{-2x} \le 2e^{-x}$ and $\max(\sqrt{\log(n/x)}, 1) \ge K_2\sqrt{\log n}$, for some constant $K_2 > 0$ and $n \ge 2$. The first inequality of the lemma follows.

For the second inequality of the lemma, note that

$$\mathrm{E}_B[\sup_z |B_{\mathbb{F}_n(z)} - B_{F_0(z)}||Z_1, \ldots, Z_n] \lesssim J\big(\|\mathbb{F}_n - F_0\|_\infty\big) = \sqrt{\|\mathbb{F}_n - F_0\|_\infty}\, \eta_n,$$

for $\eta_n^2 = \max(\log(1/\|\mathbb{F}_n - F_0\|_\infty), 1)$, by Dudley's bound, while $\sup_z \mathrm{Var}_B(B_{\mathbb{F}_n(z)} - B_{F_0(z)}|Z_1, \ldots, Z_n) \lesssim \|\mathbb{F}_n - F_0\|_\infty$. Therefore, by Borell's inequality (2.7), there exists $K > 0$ such that

$$\mathrm{P}\Big(\sup_{z \in \mathbb{R}} \frac{|B_{\mathbb{F}_n(z)} - B_{F_0(z)}|}{\sqrt{\|\mathbb{F}_n - F_0\|_\infty}} \ge K(\eta_n + y)\Big|Z_1, \ldots, Z_n\Big) \le 2e^{-y^2}.$$

We conclude by noting that $\liminf_n \sqrt{2n \log \log n}\|\mathbb{F}_n - F_0\|_\infty = \pi/2 > 0$ a.s. by Mogulskii's law (p. 526 of [33]), so that $\eta_n \lesssim \sqrt{\log n}$, a.s.   □

Lemma 8. *Consider the setting of Theorem 2 and let $\Delta_n = \|\sqrt{n}(F - \mathbb{F}_n) - B_n(\mathbb{F}_n)\|_\infty$. Then for any $t \in \mathbb{R}$ and every sequence $Z_1, Z_2, \ldots$, as $n \to \infty$,*

$$\mathrm{E}[e^{t\Delta_n} \,|\, Z_1, \ldots, Z_n] \to 1.$$

Proof. Suppose $t > 0$. For $\alpha_n = C_1(\log n + |\nu|)/\sqrt{n}$, with C_1 the universal constant from Theorem 2, and using the change of variable $u = e^{t\alpha_n + tx/\sqrt{n}}$,

$$\mathrm{E}[e^{t\Delta_n} \,|\, Z_1, \ldots, Z_n] \le e^{t\alpha_n} + \int_{e^{t\alpha_n}}^\infty P(e^{t\Delta_n} \ge u | Z_1, \ldots, Z_n) du$$

$$= e^{t\alpha_n} + \frac{te^{t\alpha_n}}{\sqrt{n}} \int_0^\infty P\left(\Delta_n \ge \frac{C_1(\log n + |\nu|) + x}{\sqrt{n}} \,\middle|\, Z_1, \ldots, Z_n\right) e^{tx/\sqrt{n}} dx.$$

Using Theorem 2 and that $C_3 - t/\sqrt{n} > 0$ for n large enough,

$$\mathrm{E}[e^{t\Delta_n} | Z_1, \ldots, Z_n] \le e^{t\alpha_n} + \frac{te^{t\alpha_n}}{\sqrt{n}} \int_0^\infty C_2 e^{-(C_3 - t/\sqrt{n})x} dx$$

$$= e^{t\alpha_n} \left(1 + \frac{C_2 t}{C_3\sqrt{n} - t}\right) \to 1$$

as $n \to \infty$. Since $\Delta_n \ge 0$ the lower bound $\mathrm{E}[e^{t\Delta_n} | Z_1, \ldots, Z_n] \ge 1$ holds trivially, which completes the proof for $t > 0$. The case $t < 0$ follows by a similar argument. $\square$

2.4. Some weak convergence facts

For completeness, we include the proof that (1.1) implies $\sqrt{n}(P_n - \mathbb{P}_n)g \rightsquigarrow N(0, P_0(g - P_0 g)^2)$ for every $g \in \mathcal{G}$ and P_0^∞-almost every sequence $Z_1, Z_2, \ldots$.

Lemma 9. *If Y_n are random variables with $\mathrm{E}e^{tY_n} \to e^{t^2\sigma^2/2}$, for every t in a subset of $\mathbb{R}$ that contains both a strictly increasing sequence with limit 0 and a strictly decreasing sequence with limit 0, then $Y_n \rightsquigarrow N(0, \sigma^2)$.*

Proof. Let $\mathcal{T}$ be the set of points and let $a < 0$ and $b > 0$ be contained in $\mathcal{T}$. Because $\mathrm{E}e^{tY_n}$ is bounded in n, for both $t = a$ and $t = b$, the sequence Y_n is tight, by Markov's inequality. For every $t \in \mathcal{T}$ strictly between a and b, some power larger than 1 of the variable e^{tY_n} is bounded in L_1, and hence the sequence e^{tY_n} is uniformly integrable. Consequently, if Y is a weak limit point of Y_n, then $\mathrm{E}e^{tY_n}$ tends to $\mathrm{E}e^{tY}$ along the same subsequence for every $t \in (a, b) \cap \mathcal{T}$. In view of the assumption of the lemma, it follows that $\mathrm{E}e^{tY} = e^{t^2\sigma^2/2}$. The set $t \in (a, b) \cap \mathcal{T}$ is infinite by assumption. Finiteness of $\mathrm{E}e^{tY}$ on this set implies that the function $z \mapsto \mathrm{E}e^{zY}$ is analytic in an open strip containing the real axis. By analytic continuation it is equal to $e^{z^2\sigma^2/2}$, whence $\mathrm{E}e^{isY} = e^{-s^2\sigma^2/2}$, for every $s \in \mathbb{R}$. $\square$

Corollary 2. *If (Y_n, Z_n) are random elements with $\mathrm{E}(e^{tY_n} | Z_n) \to e^{t^2\sigma^2/2}$, in probability, for every t in a set that contains both a strictly increasing sequence*

with limit 0 and a strictly decreasing sequence with limit 0, then $Y_n | Z_n \rightsquigarrow N(0, \sigma^2)$, in probability. If the convergence in the assumption is in the almost sure sense, then the conclusion is also true in the almost sure sense.

Proof. For the conclusion in probability it suffices to show that every subsequence of $\{n\}$ has a further subsequence with $d\big(\mathcal{L}(Y_n | Z_n), N(0, \sigma^2)\big) \to 0$, almost surely, where d is a metric defining weak convergence. From the assumption we know that every subsequence has a further subsequence with $\mathrm{E}(e^{tY_n} | Z_n) \to e^{t^2\sigma^2/2}$, almost surely. For a countable set of t, we can construct a single subsequence with this property for every t, by a diagonalization scheme. The preceding lemma gives that $d\big(\mathcal{L}(Y_n | Z_n), N(0, \sigma^2)\big) \to 0$, almost surely, along this subsequence. $\qquad\square$

Acknowledgments

References

[1] BOUCHERON, S., LUGOSI, G., AND MASSART, P. *Concentration inequalities*. Oxford University Press, Oxford, 2013. A nonasymptotic theory of independence, With a foreword by Michel Ledoux. MR3185193

[2] CASTILLO, I. On Bayesian supremum norm contraction rates. *Ann. Statist. 42*, 5 (2014), 2058–2091. MR3262477

[3] CASTILLO, I. Pólya tree posterior distributions on densities. *Ann. Inst. Henri Poincaré Probab. Stat. 53*, 4 (2017), 2074–2102. MR3729648

[4] CASTILLO, I., AND NICKL, R. On the Bernstein-von Mises phenomenon for nonparametric Bayes procedures. *Ann. Statist. 42*, 5 (2014), 1941–1969. MR3262473

[5] CASTILLO, I., AND ROUSSEAU, J. A Bernstein–von Mises theorem for smooth functionals in semiparametric models. *Ann. Statist. 43*, 6 (2015), 2353–2383. MR3405597

[6] CASTILLO, I., AND VAN DER PAS, S. Multiscale Bayesian survival analysis, 2020.

[7] CSÖRGŐ, M., AND RÉVÉSZ, P. Strong approximations of the quantile process. *Ann. Statist. 6*, 4 (1978), 882–894. MR0501290

[8] CSÖRGŐ, M., AND RÉVÉSZ, P. *Strong approximations in probability and statistics*. Probability and Mathematical Statistics. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1981. MR0666546

[9] CSÖRGŐ, S., AND HALL, P. The Komlós-Major-Tusnády approximations and their applications. *Austral. J. Statist. 26*, 2 (1984), 189–218. MR0766619

[10] DEHEUVELS, P. On the approximation of quantile processes by Kiefer processes. *J. Theoret. Probab. 11*, 4 (1998), 997–1018. MR1660907

[11] DEHLING, H. Limit theorems for sums of weakly dependent Banach space valued random variables. *Z. Wahrsch. Verw. Gebiete 63*, 3 (1983), 393–432. MR0705631

[12] DUDLEY, R. M. *Real analysis and probability*, vol. 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original. MR1932358

[13] FERGUSON, T. Prior distributions on spaces of probability measures. *Ann. Statist. 2* (1974), 615–629. MR0438568

[14] FRANSSEN, S. E. M. P., AND VAN DER VAART, A. W. The Bernstein-von Mises theorem for the Pitman-Yor process of nonnegative type, 2021.

[15] GHOSAL, S., AND VAN DER VAART, A. W. *Fundamentals of Nonparametric Bayesian Inference*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2017. MR3587782

[16] GINÉ, E., AND NICKL, R. *Mathematical foundations of infinite-dimensional statistical models*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, New York, 2016. MR3588285

[17] GU, J., AND GHOSAL, S. Strong approximations for resample quantile processes and application to ROC methodology. *J. Nonparametr. Stat. 20*, 3 (2008), 229–240. MR2421767

[18] JAMES, L. Large sample asymptotics for the two-parameter Poisson-Dirichlet process. In *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*, vol. 3 of *Inst. Math. Stat. Collect.* Inst. Math. Statist., Beachwood, OH, 2008, pp. 187–199. MR2459225

[19] KOMLÓS, J., MAJOR, P., AND TUSNÁDY, G. An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 32* (1975), 111–131. MR0375412

[20] LEDOUX, M. *The concentration of measure phenomenon*, vol. 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001. MR1849347

[21] LO, A. Y. Weak convergence for Dirichlet processes. *Sankhyā Ser. A 45*, 1 (1983), 105–111. MR0749358

[22] LO, A. Y. A remark on the limiting posterior distribution of the multiparameter Dirichlet process. *Sankhyā Ser. A 48*, 2 (1986), 247–249. MR0905464

[23] LO, A. Y. A large sample study of the Bayesian bootstrap. *Ann. Statist. 15*, 1 (1987), 360–375. MR0885742

[24] MONARD, F., NICKL, R., AND PATERNAIN, G. P. Efficient nonparametric Bayesian inference for X-ray transforms. *Ann. Statist. 47*, 2 (2019), 1113–1147. MR3909962

[25] NICKL, R. Bernstein–von Mises theorems for statistical inverse problems I: Schrödinger equation. *J. Eur. Math. Soc. (JEMS) 22*, 8 (2020), 2697–2750. MR4118619

[26] NICKL, R., AND RAY, K. Nonparametric statistical inference for drift vector fields of multi-dimensional diffusions. *Ann. Statist. 48*, 3 (2020), 1383–1408. MR4124327

[27] Nickl, R., and Söhl, J. Bernstein-von Mises theorems for statistical inverse problems II: compound Poisson processes. *Electron. J. Stat. 13*, 2 (2019), 3513–3571. MR4013745

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

[29] Ray, K., and Szabó, B. Debiased Bayesian inference for average treatment effects. In *Advances in Neural Information Processing Systems 33*. 2019. MR4051359

[30] Ray, K., and van der Vaart, A. W. Semiparametric Bayesian causal inference. *Ann. Statist. 48* (2020), 2999–3020. MR4152632

[31] Rivoirard, V., and Rousseau, J. Bernstein-von Mises theorem for linear functionals of the density. *Ann. Statist. 40*, 3 (2012), 1489–1523. MR3015033

[32] Royden, H., and Fitzpatrick, P. *Real Analysis*. Prentice Hall, 2010.

[33] Shorack, G. R., and Wellner, J. A. *Empirical processes with applications to statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986. MR0838963

[34] van der Vaart, A., and Wellner, J. A. Preservation theorems for Glivenko-Cantelli and uniform Glivenko-Cantelli classes. In *High dimensional probability, II (Seattle, WA, 1999)*, vol. 47 of *Progr. Probab.* Birkhäuser Boston, Boston, MA, 2000, pp. 115–133. MR1857319

[35] van der Vaart, A. W., and Wellner, J. A. *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996. With applications to statistics. MR1385671

[36] Yurinskii, V. V. On the error of the gaussian approximation for convolutions. *Theory of Probability & Its Applications 22*, 2 (1978), 236–247. MR0517490