

# Uniform rates of the Glivenko–Cantelli convergence and their use in approximating Bayesian inferences

EMANUELE DOLERA\* and EUGENIO REGAZZINI\*\*

*Dipartimento di Matematica, via Ferrata 5, 27100 Pavia (PV), Italy. E-mail: \*emanuele.dolera@unipv.it; \*\*eugenio.regazzini@unipv.it*

This paper deals with suitable quantifications in approximating a probability measure by an “empirical” random probability measure  $\hat{p}_n$ , depending on the first  $n$  terms of a sequence  $\{\tilde{\xi}_i\}_{i \geq 1}$  of random elements. Section 2 studies the range of oscillation near zero of the Wasserstein distance  $d_{[\mathbb{S}]}^{(p)}$  between  $p_0$  and  $\hat{p}_n$ , assuming the  $\tilde{\xi}_i$ ’s i.i.d. from  $p_0$ . In Theorem 2.1  $p_0$  can be fixed in the space of all probability measures on  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  and  $\hat{p}_n$  coincides with the empirical measure  $\tilde{e}_n := \frac{1}{n} \sum_{i=1}^n \delta_{\tilde{\xi}_i}$ . In Theorem 2.2 (Theorem 2.3, respectively),  $p_0$  is a  $d$ -dimensional Gaussian distribution (an element of a distinguished statistical exponential family, respectively) and  $\hat{p}_n$  is another  $d$ -dimensional Gaussian distribution with estimated mean and covariance matrix (another element of the same family with an estimated parameter, respectively). These new results improve on allied recent works by providing also uniform bounds with respect to  $n$ , meaning the finiteness of the  $p$ -moment of  $\sup_{n \geq 1} b_n d_{[\mathbb{S}]}^{(p)}(p_0, \hat{p}_n)$  is proved for some diverging sequence  $b_n$  of positive numbers. In Section 3, assuming the  $\tilde{\xi}_i$ ’s exchangeable, one studies the range of oscillation near zero of the Wasserstein distance between the conditional distribution – also called posterior – of the directing measure of the sequence, given  $\tilde{\xi}_1, \dots, \tilde{\xi}_n$ , and the point mass at  $\hat{p}_n$ . Similarly, a bound for the approximation of predictive distributions is given. Finally, Theorems from 3.3 to 3.5 reconsider Theorems from 2.1 to 2.3, respectively, according to a Bayesian perspective.

*Keywords:* dominated ergodic theorem; empirical measure; exchangeability; Glivenko–Cantelli theorem; law of the iterated logarithm; posterior distribution; predictive distribution; Wasserstein distance

## 1. Introduction and description of the results

A recurrent problem in different branches of science concerns the approximation of a probability measure (p.m., for short) – either fixed or random – by means of another random p.m., say  $\hat{p}_n$ , of an empirical nature, as function of a random sample  $\tilde{\xi}^{(n)} := (\tilde{\xi}_1, \dots, \tilde{\xi}_n)$  from the original measure. The accuracy of such approximation is usually measured in terms of a distinguished form of distance between p.m.’s, and the final objective is to study the stochastic convergence to zero of that distance between the original p.m. and  $\hat{p}_n$ , as  $n$  goes to infinity.

This problem arose in connection with the issue, of a statistical nature, of estimating an unknown probability distribution (p.d., for short) on  $\mathbb{R}$  by means of the empirical frequency distribution. Thus, the original formulation considered  $\tilde{\xi}^{(n)}$  as the initial  $n$ -segment of a sequence  $\{\tilde{\xi}_i\}_{i \geq 1}$  of i.i.d. real random variables with common p.d.  $p_0$ , and  $\hat{p}_n$  equal to  $\tilde{e}_n := \frac{1}{n} \sum_{i=1}^n \delta_{\tilde{\xi}_i}$ .

In the same year and journal, Cantelli [17], Glivenko [38] and Kolmogorov [45] published their respective solutions concerning the almost sure convergence to zero of the so-called uniform (or Kolmogorov) distance between  $p_0$  and  $\tilde{\epsilon}_n$ , while de Finetti [24] gave a solution in terms of the Lévy distance. Since this achievement was considered by some authors so important to earn the title of “fundamental theorem of mathematical statistics”, a flourishing line of research developed from it with a view to providing refinements and extensions in various directions, such as: (i) replacing  $\tilde{\epsilon}_n$  with smoothed versions of it; (ii) providing central limit theorems for the above-mentioned distances; (iii) quantifying the almost sure convergence; (iv) relaxing the i.i.d. assumption on the  $\tilde{\xi}_i$ ’s. See, for example, the books [32,57,62] for a comprehensive treatment of both the original problem and its developments. In particular, the very fruitful line of research – also pursued in the present paper – which aims at providing *quantitative rates* of the Glivenko–Cantelli convergence never ceased to be investigated from the end of the Sixties. See, for example, [11–13,31,37,40,43,47,59,65] for an approach based on the empirical processes theory, and [1,4,18,27,56,67] for the perspective of the so-called optimal matching problem. Possible applications to various areas of pure and applied mathematics can be found in the references of the quoted papers, with a particular mention to: (i) statistics (see [31] and also [32,35,57,62]); (ii) particle physics and PDE’s (see [4,14,18,37]); (iii) numerical analysis, with particular attention to quantization (see [27,56,65,67]); (iv) machine learning (see also [28,63]).

The present paper tackles the above quantification problem for various forms of  $\hat{p}_n$  by studying rates with respect to both  $L^p$  and almost sure convergence. In Section 2,  $p_0$  is fixed and the  $\tilde{\xi}_i$ ’s are i.i.d. from  $p_0$ , while in Section 3 the  $\tilde{\xi}_i$ ’s are assumed to be exchangeable. In both sections, as in many other papers, the discrepancy between two p.m.’s is measured in terms of the distance  $d_{[\mathbb{S}]}^{(p)}$ , usually called  $p$ -Wasserstein distance,<sup>1</sup> defined as follows: given a separable metric space  $(\mathbb{S}, d_{\mathbb{S}})$ ,  $p \geq 1$ , and  $\mu_1, \mu_2$  in  $[\mathbb{S}]_p := \{\mu \text{ p.m. on } (\mathbb{S}, \mathcal{B}(\mathbb{S})) \mid \int_{\mathbb{S}} [d_{\mathbb{S}}(x, x_0)]^p \mu(dx) < +\infty \text{ for some } x_0 \in \mathbb{S}\}$ , put

$$d_{[\mathbb{S}]}^{(p)}(\mu_1, \mu_2) := \inf_{\gamma \in \mathcal{F}(\mu_1, \mu_2)} \left( \int_{\mathbb{S}^2} [d_{\mathbb{S}}(x, y)]^p \gamma(dx dy) \right)^{1/p}, \tag{1.1}$$

where  $\mathcal{F}(\mu_1, \mu_2)$  stands for the class of all p.m.’s on  $(\mathbb{S}^2, \mathcal{B}(\mathbb{S}^2))$  with  $i$ th marginal equal to  $\mu_i$ ,  $i = 1, 2$ . Here and throughout, given any topological space  $\mathbb{X}$ ,  $\mathcal{B}(\mathbb{X})$  stands for the Borel  $\sigma$ -algebra on  $\mathbb{X}$ , and the product  $\mathbb{X}^n$  is thought of as endowed with the product topology, for any  $n \in \mathbb{N} \cup \{\infty\}$ . The couple  $([\mathbb{S}]_p, d_{[\mathbb{S}]}^{(p)})$  proves to be a separable metric space, which is complete if  $(\mathbb{S}, d_{\mathbb{S}})$  is also complete, provided that  $(\mathbb{S}, d_{\mathbb{S}})$  satisfies the additional *Radon property*, that is, every p.m. on  $(\mathbb{S}, \mathcal{B}(\mathbb{S}))$  has the compact inner approximation property. See Definition 5.1.4 and Proposition 7.1.5 in [3]. Finally, recall that a metric subspace  $(\mathbb{S}, d_{\mathbb{S}})$  of another complete and separable metric space, say  $(\hat{\mathbb{S}}, d_{\hat{\mathbb{S}}})$ , has the Radon property if  $\mathbb{S} \in \mathcal{B}(\hat{\mathbb{S}})$ . See Theorem 7.1.4 in [33].

Throughout the paper, one puts  $(\Omega, \mathcal{F}) := (\mathbb{S}^{\infty}, \mathcal{B}(\mathbb{S}^{\infty}))$  and defines the sequence  $\{\tilde{\xi}_i\}_{i \geq 1}$  to be the coordinate random elements of  $\mathbb{S}^{\infty}$ , that is,  $\tilde{\xi}_i(\omega) := \omega_i$  for all  $\omega = (\omega_1, \omega_2, \dots) \in \mathbb{S}^{\infty}$ , providing a formal definition of the sample evoked at the beginning of the section. In addition,

<sup>1</sup> See however the Bibliographical Notes at the end of Chapter 6 of [64] for a correct attribution.

one confines oneself to considering forms of  $\hat{p}_n$  which, like  $\tilde{\epsilon}_n$ , are presentable as measurable functions of  $\tilde{\xi}^{(n)}$ . More formally, given some  $p \geq 1$  and defined  $\mathcal{F}_n$  to be the sub- $\sigma$ -algebra of  $\mathcal{F}$  generated by  $\tilde{\xi}^{(n)}$ , the mapping  $\hat{p}_n : (\Omega, \mathcal{F}_n) \rightarrow ([\mathbb{S}]_p, \mathcal{B}([\mathbb{S}]_p))$  is required to be measurable for every  $n \in \mathbb{N}$ .

In Section 2, one fixes  $p \geq 1$ ,  $p_0$  in  $[\mathbb{S}]_p$  and endows  $(\Omega, \mathcal{F})$  with the law  $p_0^\infty$  which makes the coordinates i.i.d. random elements with common p.d.  $p_0$ , i.e.  $p_0^\infty[\tilde{\xi}_1 \in A_1, \dots, \tilde{\xi}_n \in A_n] = \prod_{i=1}^n p_0^\infty[\tilde{\xi}_i \in A_i] = \prod_{i=1}^n p_0(A_i)$ , for every  $n \in \mathbb{N}$  and  $A_1, \dots, A_n \in \mathcal{B}(\mathbb{S})$ . The main objective is to determine an increasing sequence  $b_n$  of positive numbers, with  $\lim_{n \rightarrow +\infty} b_n = +\infty$ , and positive constants  $C_p(p_0)$  and  $Y_p(p_0)$  for which

$$p_0^\infty \left[ \left( \sup_{n \geq 1} b_n d_{[\mathbb{S}]}^{(p)}(p_0, \hat{p}_n) \right)^p \right] \leq C_p(p_0) \tag{1.2}$$

and

$$p_0^\infty \left( \left\{ \limsup_{n \rightarrow +\infty} b_n d_{[\mathbb{S}]}^{(p)}(p_0, \hat{p}_n) \leq Y_p(p_0) \right\} \right) = 1. \tag{1.3}$$

In (1.2), as well as throughout the paper, expectation of a real random variable  $X$  with respect to a p.m.  $\mu$  on  $(\Omega, \mathcal{F})$  is denoted by  $\mu(X)$ . Moreover, in the sequel, the expression “ $p_0$  satisfies (1.2)” (“ $p_0$  satisfies (1.3)”, respectively) will be used as a shorthand to mean that, given  $p_0$ , (1.2) ((1.3), respectively) is in force with a suitable constant  $C_p(p_0)$  ( $Y_p(p_0)$ , respectively).

**Remark 1.1.** It is worth mentioning how to extend the validity of (1.2) to lower exponents  $s \in [1, p]$ , once (1.2) has been established. In fact, one has

$$\begin{aligned} p_0^\infty \left[ \left( \sup_{n \geq 1} b_n d_{[\mathbb{S}]}^{(s)}(p_0, \hat{p}_n) \right)^s \right] &\leq \left\{ p_0^\infty \left[ \left( \sup_{n \geq 1} b_n d_{[\mathbb{S}]}^{(s)}(p_0, \hat{p}_n) \right)^p \right] \right\}^{\frac{s}{p}} \\ &\leq \left\{ p_0^\infty \left[ \left( \sup_{n \geq 1} b_n d_{[\mathbb{S}]}^{(p)}(p_0, \hat{p}_n) \right)^p \right] \right\}^{\frac{s}{p}}, \end{aligned}$$

where the former inequality follows from the Lyapunov inequality for moments, while the latter comes from the monotonicity of the Wasserstein distance with respect to the order. Once (1.3) has been established, the same monotonicity entails  $p_0^\infty(\{\limsup_{n \rightarrow +\infty} b_n d_{[\mathbb{S}]}^{(s)}(p_0, \hat{p}_n) \leq Y_p(p_0)\}) = 1$  for any  $s \in [1, p]$ .

The first result in Section 2 (Theorem 2.1) states the validity of (1.2)–(1.3) when  $\hat{p}_n = \tilde{\epsilon}_n$ ,  $\mathbb{S} = \mathbb{R}^d$ ,  $d_{\mathbb{S}}$  coincides with the standard Euclidean distance,  $p \in [1, +\infty) \cap (d/2, +\infty)$  and  $b_n \sim (n/\log \log n)^{1/2p}$ , provided that  $\int_{\mathbb{R}^d} |\mathbf{x}|^{2p+\delta} p_0(d\mathbf{x}) < +\infty$  obtains for some  $\delta > 0$ . To introduce the second topic of the same section, one should notice that  $b_n$  increases slower and slower as  $d$  gets large. As noted in [1,37,47,56,67], this drawback is an intrinsic feature of the problem since  $\lim_{n \rightarrow +\infty} n^{1/d} p_0^\infty[d_{[\mathbb{R}^d]}^{(1)}(p_0, \tilde{\epsilon}_n)] > 0$  when  $p_0$  coincides with the uniform measure on the cube  $[-1, 1]^d$ . This phenomenon is ascribable to the fact that the use of  $\tilde{\epsilon}_n$  in approximating  $p_0$  is thoroughly justified when there is no significant *a priori* restriction on the subclass of  $[\mathbb{S}]_p$  in

which  $\mathfrak{p}_0$  can be fixed, so that the slowdown of the divergence of  $b_n$  for large dimensions seems to be the price to pay for this generality. In point of fact, many problems in applied mathematics and statistics provide enough information to restrict the aforesaid admissible class to a family  $\mathcal{M}$  of distinguished p.m.'s  $\mu_\theta$  on  $(\mathbb{S}, \mathcal{B}(\mathbb{S}))$ , determined up to some finite-dimensional parameter  $\theta$  in  $\Theta \subseteq \mathbb{R}^k$  so that  $\theta \mapsto \mu_\theta$  is injective. In this framework, the approximation of  $\mathfrak{p}_0 = \mu_{\theta_0}$  proves to be more natural within the elements of  $\mathcal{M}$ , so that one first approximates  $\theta_0$  by a suitable random element  $\hat{\theta}_n \in \Theta$ , depending on  $\tilde{\xi}^{(n)}$  according to well-known principles of *statistical estimation*, and then approximates  $\mathfrak{p}_0$  by  $\hat{\mathfrak{p}}_n = \mu_{\hat{\theta}_n}$ . The second main result supports this last method of approximation, in contrast to the rougher one based on the choice of  $\tilde{\epsilon}_n$ , by showing that, if  $\hat{\mathfrak{p}}_n = \mu_{\hat{\theta}_n}$ , one can put  $p = 2$  and  $b_n \sim (n/\log \log n)^{1/2}$  in (1.2)–(1.3) independently of the dimension of the  $\tilde{\xi}_i$ 's and of  $\Theta$ , at least when  $\mathcal{M}$  coincides with the class of all non-singular multidimensional Gaussian distributions (Theorem 2.2) and, more in general, with a distinguished type of statistical exponential family (Theorem 2.3).

Before explaining Section 3, it is worth commenting on the value of the problems associated with (1.2)–(1.3) and their solutions contained in Theorems 2.1–2.3. First, the use of the distance  $d_{[\mathbb{S}]}^{(p)}$  connects the present problems with the so-called Vapnik–Chervonenkis theory. See [32,62,63]. However, the actual novelty of the present study consists in finding rates according to a concept of *strong* (i.e., *uniform* with respect to  $n$ ) convergence, obtained by suitable applications of classical inequalities concerning the *dominated ergodic theorem*, due to Siegmund and Teicher. In fact, the current literature has investigated, until now, only non-uniform bounds like  $\mathfrak{p}_0^\infty[(d_{[\mathbb{S}]}^{(p)}(\mathfrak{p}_0, \hat{\mathfrak{p}}_n))^p] \leq C'(\mathfrak{p}_0)\alpha(n)$ , valid for all  $n \in \mathbb{N}$  and suitable positive constants  $C'(\mathfrak{p}_0)$ , whilst (1.2) entails, of course,  $\mathfrak{p}_0^\infty[(\sup_{n \geq N} d_{[\mathbb{S}]}^{(p)}(\mathfrak{p}_0, \hat{\mathfrak{p}}_n))^p] \leq C_p(\mathfrak{p}_0)/b_N^p$ , for all  $N \in \mathbb{N}$ . In spite of the strengthening expressed by the latter inequality, the rate  $\alpha(n)$  given in the literature turns out to be comparable with our  $1/b_n^p$ , in the sense that  $\alpha(n)$  goes to zero slightly faster than  $1/b_n^p$  under the same assumption  $\int_{\mathbb{R}^d} |\mathbf{x}|^{2p+\delta} \mathfrak{p}_0(d\mathbf{x}) < +\infty$ , at least for small  $\delta > 0$ . Cf. Theorem 1 in [37] and Theorem 13 in [40]. Moreover, apart from the obvious formal improvement, uniform bounds, which are strictly connected with the concept of *uniform integrability*, prove to be of crucial importance in those applications where one is interested in showing that the union of a collection of “bad events” – and not only the “bad events” taken individually – has small probability. Indeed, one can notice that (1.2) entails  $\mathfrak{p}_0^\infty(\bigcup_{n \geq 1} \{d_{[\mathbb{S}]}^{(p)}(\mathfrak{p}_0, \hat{\mathfrak{p}}_n) > M/b_n\}) = \mathfrak{p}_0^\infty(\sup_{n \geq 1} b_n d_{[\mathbb{S}]}^{(p)}(\mathfrak{p}_0, \hat{\mathfrak{p}}_n) > M) \leq C_p(\mathfrak{p}_0)/M^p$  by means of Markov's inequality, so that, given any  $\eta > 0$ , one can choose  $M_\eta$  large enough to obtain  $\mathfrak{p}_0^\infty(\bigcap_{n \geq 1} \{d_{[\mathbb{S}]}^{(p)}(\mathfrak{p}_0, \hat{\mathfrak{p}}_n) \leq M_\eta/b_n\}) \geq 1 - \eta$ . On the other hand, certain applications to mathematical statistics and physics would prove completely justified only in the presence of such uniform bounds, as suggested, for example, by Wiener in relation to the problem of connecting von Neumann's ergodic theorem with Birkhoff's. See Section 1 of [66]. As far as our personal motivations to study the uniform convergence displayed in (1.2), we mention the paper [20], which indicates a line of research – pursued in the next Section 3 – focused on some approximations of posterior and predictive distributions in Bayesian statistics. Indeed, we realized that it is just a condition of the same type as (1.2) that allows the application of a suitable martingale argument due to Blackwell and Dubins, so that we seize the opportunity to highlight this connection.

In Section 3, the coordinate random elements  $\tilde{\xi}_1, \tilde{\xi}_2, \dots$  are assumed to be exchangeable, i.e. the joint distribution of every finite subset of  $n$  of these elements depends only on  $n$  and not on

the particular subset, for every  $n \in \mathbb{N}$ . This condition is obviously satisfied if the  $\tilde{\xi}_i$ 's are i.i.d. Actually, exchangeability is the suitable assumption to be made to translate the usual empirical condition of analogy (not necessarily identity) of the observations into precise probabilistic terms. It is well known that, given any exchangeable p.m.  $\rho$  on  $(\Omega, \mathcal{F}) = (\mathbb{S}^\infty, \mathcal{B}(\mathbb{S}^\infty))$ , there is an extension of the Glivenko–Cantelli theorem, due to de Finetti [23,25,26], which states the existence of a *random* p.m., say  $\tilde{p} : (\Omega, \mathcal{F}) \rightarrow ([\mathbb{S}], \mathcal{B}([\mathbb{S}]))$ , such that  $\rho(\{\tilde{c}_n \Rightarrow \tilde{p}, \text{ as } n \rightarrow \infty\}) = 1$ ,  $\Rightarrow$  denoting the limit in the sense of weak convergence of p.m.'s, and  $[\mathbb{S}]$  the class of all p.m.'s on  $(\mathbb{S}, \mathcal{B}(\mathbb{S}))$  endowed with the topology generated by  $\Rightarrow$ . Moreover, the  $\tilde{\xi}_i$ 's turn out to be conditionally i.i.d. given  $\tilde{p}$ , with common p.d. equal to  $\tilde{p}$ . These two facts are encapsulated in the well-known *de Finetti representation theorem*, equivalently reformulated as follows (see [2]): for any exchangeable p.d.  $\rho$  on  $(\Omega, \mathcal{F})$ , there exists one and only one p.m.  $\pi$  on  $([\mathbb{S}], \mathcal{B}([\mathbb{S}]))$ , equal to the p.d. of  $\tilde{p}$ , such that  $\rho(A) = \int_{[\mathbb{S}]} p^\infty(A) \pi(dp)$  holds for any  $A \in \mathcal{F}$ ,  $p^\infty$  denoting the p.m. on  $(\Omega, \mathcal{F})$  that makes the coordinate random variables i.i.d. with common distribution  $p$ . In Bayesian statistics,  $\pi$  is usually called *prior distribution*. One encounters therein two new questions concerning *posterior* and *predictive* distributions, i.e. regular conditional p.d.'s  $\pi(\tilde{\xi}^{(n)} | \tilde{\xi}^{(n)}) := \pi(\tilde{\xi}^{(n)}, \cdot)$  and  $p(\tilde{\xi}^{(n)} | \tilde{\xi}^{(n)}) := p(\tilde{\xi}^{(n)}, \cdot)$  of  $\tilde{p}$  and  $\{\tilde{\xi}_{n+i}\}_{i \geq 1}$ , respectively, given  $\tilde{\xi}^{(n)}$ . In particular,  $\pi(\tilde{\xi}^{(n)}, B)$  is a version of  $\rho(\tilde{p} \in B | \tilde{\xi}^{(n)})$  and  $p(\tilde{\xi}^{(n)}, A)$  a version of  $\rho(\{\tilde{\xi}_{i+n}\}_{i \geq 1} \in A | \tilde{\xi}^{(n)})$ , for  $B \in \mathcal{B}([\mathbb{S}])$  and  $A \in \mathcal{F}$ . Apropos of these questions, Section 3 completes and enriches the previous work [20] by extending some of its statements and by providing a Bayesian interpretation of the main results (1.2)–(1.3) formulated in Section 2. As to the approximation of  $\pi(\tilde{\xi}^{(n)})$ , the aim is to show that, if  $\rho[C_p(\tilde{p})] < +\infty$  with the same  $C_p$  as in (1.2), then

$$\rho\left(\left\{\limsup_{n \rightarrow +\infty} b_n d_{[[\mathbb{S}]_p]}^{(p)}(\pi(\tilde{\xi}^{(n)}), \delta_{\hat{p}_n}) \leq Y_p(\tilde{p})\right\}\right) = 1 \tag{1.4}$$

holds with the same  $Y_p$  as in (1.3), where  $d_{[[\mathbb{S}]_p]}^{(p)}$  is defined according to (1.1) with the proviso that  $(\mathbb{S}, d_{\mathbb{S}})$  therein is replaced by  $([\mathbb{S}]_p, d_{[\mathbb{S}]}^{(p)})$ . See Theorem 3.1 in Section 3. Notice that  $d_{[[\mathbb{S}]_p]}^{(p)}(\pi(\tilde{\xi}^{(n)}), \delta_{\hat{p}_n}) = (\int_{[\mathbb{S}]} [d_{[\mathbb{S}]}^{(p)}(p, \hat{p}_n)]^p \pi(\tilde{\xi}^{(n)}, dp))^{1/p}$ . It is also worth remarking that the choice of  $d_{[[\mathbb{S}]_p]}^{(p)}$  as reference distance establishes an interesting link between (1.3) and (1.4), which shows how the latter can be viewed as an extension of the former: when the  $\tilde{\xi}_i$ 's are i.i.d. with common p.d.  $p_0$ ,  $\rho$  coincides with  $p_0^\infty$ ,  $\pi(\tilde{\xi}^{(n)})$  is equal to  $\delta_{p_0}$  with  $\rho$ -probability 1, and  $d_{[[\mathbb{S}]_p]}^{(p)}(\pi(\tilde{\xi}^{(n)}), \delta_{\hat{p}_n}) = d_{[\mathbb{S}]}^{(p)}(p_0, \hat{p}_n)$ . The approximation of  $p(\tilde{\xi}^{(n)})$  is a more delicate problem, here solved, consistently with the general lines and notation given in [20,21], by considering conditional p.d.'s, say  $q_m(\tilde{\xi}^{(n)} | \tilde{\xi}^{(n)}) := q_m(\tilde{\xi}^{(n)}, \cdot)$ , of  $\tilde{c}_{n,m} := \frac{1}{m} \sum_{i=1}^m \delta_{\tilde{\xi}_{i+n}}$  given  $\tilde{\xi}^{(n)}$ , for every  $m \in \mathbb{N}$ . In particular,  $q_m(\tilde{\xi}^{(n)}, B)$  is a version of  $\rho(\tilde{c}_{n,m} \in B | \tilde{\xi}^{(n)})$  for  $B \in \mathcal{B}([\mathbb{S}])$ . Therefore, the main result concerning the approximation of the predictive distributions reads

$$\rho\left(\left\{\limsup_{n \rightarrow +\infty} b_n d_{[[\mathbb{S}]_p]}^{(p)}(q_m(\tilde{\xi}^{(n)}), \hat{p}_n^\infty \circ \tilde{c}_{n,m}^{-1}) \leq Y_p(\tilde{p})\right\}\right) = 1 \tag{1.5}$$

for every  $m \in \mathbb{N}$ , while  $\hat{p}_n^\infty \circ \tilde{c}_{n,m}^{-1}(B)$  stands for  $\hat{p}_n^\infty(\{\frac{1}{m} \sum_{i=1}^m \delta_{\tilde{\xi}_{i+n}} \in B\})$ , for  $B \in \mathcal{B}([\mathbb{S}])$ . See Theorem 3.2 in Section 3. After establishing these general facts, the next new results, that is, Theorems from 3.3 to 3.5, reconsider Theorems from 2.1 to 2.3, respectively, according to the present Bayesian perspective. Thus, explicit expressions for  $\hat{p}_n$ ,  $b_n$ ,  $C_p(\tilde{p})$  and  $Y_p(\tilde{p})$  are provided in each of the Theorems from 3.3 to 3.5.

**Remark 1.2.** In the same vein of Remark 1.1, once (1.4) has been established, the identities

$$\begin{aligned} & \rho\left(\left\{\limsup_{n \rightarrow +\infty} b_n d_{[[\mathbb{S}]_s]}^{(s)}(\pi(\tilde{\xi}^{(n)}), \delta_{\hat{p}_n}) \leq Y_p(\tilde{p})\right\}\right) \\ &= \rho\left(\left\{\limsup_{n \rightarrow +\infty} b_n d_{[[\mathbb{S}]_p]}^{(p)}(\pi(\tilde{\xi}^{(n)}), \delta_{\hat{p}_n}) \leq Y_p(\tilde{p})\right\}\right) = 1 \end{aligned}$$

are valid for any  $s \in [1, p]$ , since  $d_{[[\mathbb{S}]_s]}^{(s)}(\pi(\tilde{\xi}^{(n)}), \delta_{\hat{p}_n}) \leq d_{[[\mathbb{S}]_p]}^{(s)}(\pi(\tilde{\xi}^{(n)}), \delta_{\hat{p}_n}) \leq d_{[[\mathbb{S}]_p]}^{(p)}(\pi(\tilde{\xi}^{(n)}), \delta_{\hat{p}_n})$ . In addition, for any  $s \in [1, p]$ , (1.5) entails  $\rho(\{\limsup b_n d_{[[\mathbb{S}]_s]}^{(s)}(q_m(\tilde{\xi}^{(n)}), \hat{p}_n^\infty \circ \tilde{c}_{n,m}^{-1}) \leq Y_p(\tilde{p})\}) = 1$ .

The achievement of (1.4)–(1.5) represents a specific asymptotic analysis of both posterior and predictive distributions, which is relevant with a view to a better understanding of different kinds of empirical approximation to orthodox Bayesian methods. In fact, the impressive growth of Bayesian statistics in the last decades has produced new complex models – in particular of non-parametric type – which, although appreciated for their predictive features, generate very often serious hurdles to clear from the point of view of direct computations. These difficulties are very often circumvented by the use of unorthodox, but more manageable, Bayesian methods, such as: empirical Bayes (see [35,46,48,52]); partial and profile likelihood (see [22,55]); numerical techniques based on simulation of random quantities, like the bootstrap (see [34,35]). In particular, contrasting  $\hat{p}_n^\infty \circ \tilde{c}_{n,m}^{-1}$  to  $q_m(\tilde{\xi}^{(n)})$  as in (1.5) is a cornerstone in bootstrap techniques. Very recently, the question of merging of orthodox and empirical Bayes procedures has been studied in [50,51,54], but with a significant difference: (1.4)–(1.5) are reformulated therein by replacing  $\rho$  with some hypothetical distribution  $p_\star^\infty$ , in the spirit of the *what if method* described, for example, in [29]. Other results which are more comparable to ours, though confined to the merging of predictive distributions with  $m = 1$ , can be found in [7–9].

In conclusion, the ultimate aim of producing bounds like (1.4)–(1.5) is to quantify the degree of accuracy in the approximation ensuing from one of the aforesaid empirical techniques. In fact, for any  $\epsilon, \eta > 0$ , one can find  $n_0 = n_0(\epsilon, \eta)$  and  $L = L(\eta)$  such that  $\rho[\bigcap_{n \geq n_0} \{d_{[[\mathbb{S}]_p]}^{(p)}(\pi(\tilde{\xi}^{(n)}), \delta_{\hat{p}_n}) \leq (L + \epsilon)/b_n\}] \geq 1 - \eta$  and  $\rho[\bigcap_{n \geq n_0} \{d_{[[\mathbb{S}]_p]}^{(p)}(q_m(\tilde{\xi}^{(n)}), \hat{p}_n^\infty \circ \tilde{c}_{n,m}^{-1}) \leq (L + \epsilon)/b_n\}] \geq 1 - \eta$  are in force. Thus, one can observe that the determination of  $L$  follows from the specific form of  $Y_p(\tilde{p})$ , while that of  $n_0$  represents an interesting open problem to be tackled in future works.

## 2. Results for the i.i.d. case

This section contains four propositions. The first one, of a general character, is a re-statement of an inequality for normed sums of i.i.d. real random variables, originally due to Siegmund [58] and

Teicher [61], sometimes referred to as *dominated ergodic theorem*. See also Section 10.3 in [19]. In the following version, this inequality is reinforced, with respect to the original statements, by the explicit characterization of the upper bound, which plays a crucial role in the proofs of the remaining theorems exhibited in the present section. These last results, in turn, provide affirmative answers to the achievement of (1.2)–(1.3) in three situations of interest.

Hence, one starts with the reformulation of the Siegmund-Teicher inequality.

**Proposition 2.1.** *Let  $\{X_i\}_{i \geq 1}$  be a sequence of i.i.d. real random variables on  $(\Omega, \mathcal{F}, \mathbb{P})$  with  $E[X_1] = 0$  and  $E[|X_1|^r] < +\infty$  for some  $r > 2$ . Then, after putting  $\text{Log}_2(x) := \mathbb{1}\{x < e^e\} + (\log \log x)\mathbb{1}\{x \geq e^e\}$ , one has*

$$E \left[ \sup_{n \geq 1} \frac{|\sum_{i=1}^n X_i|^r}{(n \text{Log}_2(n))^{r/2}} \right] \leq \sigma^r \left[ \alpha_0(r) + \alpha_1(r) \left( \frac{E[|X_1|^r]}{\sigma^r} \right)^{\lceil r \rceil} \right] \tag{2.1}$$

with suitable constants  $\alpha_0(r), \alpha_1(r), \sigma^2 := E[X_1^2]$  and  $\lceil r \rceil := \inf\{n \in \mathbb{N} | n \geq r\}$ .

See Section 4.1 for the proof.

**Remark 2.1.** With a view to successive applications of this proposition, it is crucial to underline that the constants  $\alpha_0(r)$  and  $\alpha_1(r)$  appearing in the upper bound do not depend on the law of the  $X_i$ 's. Precise expressions for these constants can be drawn from the combination of specific passages of the proof, culminating in inequality (4.8).

Combination of (2.1) with certain inequalities concerning  $d_{[\mathbb{S}]}^{(p)}$ , originally proved in [27] and reformulated more recently in [37], yields the following theorem.

**Theorem 2.1.** *If  $\mathbb{S} = \mathbb{R}^d$  and  $d_{\mathbb{S}}$  coincides with the Euclidean distance, assume that, for some  $\delta > 0$  and  $p \in [1, +\infty) \cap (d/2, +\infty)$ ,  $\int_{\mathbb{R}^d} |\mathbf{x}|^{2p+\delta} p_0(d\mathbf{x}) < +\infty$  obtains. Then, (1.2)–(1.3) hold true with  $\hat{p}_n = \tilde{\epsilon}_n, b_n = (n/\text{Log}_2 n)^{1/2p}$ . Moreover, for any choice of  $r \in (2, 3/M)$ , with  $M := \max\{2 - p/d; 1 + 1/(2 + \delta/p)\}$ , one can put*

$$C_p(p_0) = \frac{K(p, d)\overline{C}(r)}{1 - 2^{-\lambda}} \left[ 1 + \frac{2^p}{1 - 2^{-\sigma}} \left( \int_{\mathbb{R}^d} |\mathbf{x}|^{2p+\delta} p_0(d\mathbf{x}) \right)^{\frac{3}{r}-1} \right], \tag{2.2}$$

$$Y_p(p_0) = \left\{ \frac{\sqrt{2}K(p, d)}{1 - 2^{-(p-d/2)}} \left[ 1 + \frac{2^p}{1 - 2^{-\delta/2}} \left( \int_{\mathbb{R}^d} |\mathbf{x}|^{2p+\delta} p_0(d\mathbf{x}) \right)^{1/2} \right] \right\}^{1/p}, \tag{2.3}$$

where  $K(p, d), \overline{C}(r), \lambda$  and  $\sigma$  are suitable positive constants independent of  $p_0$ .

See Section 4.2 for the proof, as well as for the exact evaluation of  $K(p, d), \overline{C}(r), \lambda$  and  $\sigma$ . It is also worth recalling Remark 1.1 in order to indicate how to enlarge the applicability of this theorem, since one could be interested in bounding the Wasserstein distance of low order (e.g., 1 or 2) even in the presence of a high dimension  $d$ .

Before introducing the other results, it is worth noticing once again the troublesome dependence, in the previous theorem, of  $b_n$  and  $p$  on  $d$ , which appears to be unavoidable when  $\mathfrak{p}_0$  is viewed as any element of the whole class  $[\mathbb{R}^d]_{2p+\delta}$ . See [1,37,47,56,65,67]. Then, one is naturally led to search for situations in which the aforesaid drawback does not occur. As already remarked in Section 1, there are many significant problems in which it is natural to constrain both  $\mathfrak{p}_0$  and  $\hat{\mathfrak{p}}_n$  within a distinguished class  $\mathcal{M}$  strictly contained in  $[\mathbb{S}]$ . By the way of illustration, maintaining the choice  $\mathbb{S} = \mathbb{R}^d$ , one now considers the noteworthy case in which  $\mathcal{M} = \{\mu_\theta\}_{\theta \in \Theta}$  coincides with the class of the non-degenerate Gaussian p.d.'s, namely

$$\mu_\theta(\mathbf{d}\mathbf{x}) = \left(\frac{1}{2\pi}\right)^{d/2} (\det(V))^{-1/2} \exp\left\{-\frac{1}{2} {}^t(\mathbf{x} - \mathbf{m})V^{-1}(\mathbf{x} - \mathbf{m})\right\} \mathbf{d}\mathbf{x} \quad (\mathbf{x} \in \mathbb{R}^d), \quad (2.4)$$

where  $\theta = (\mathbf{m}, V)$ ,  $\Theta = \mathbb{R}^d \times \mathbb{GL}_{\text{sym}}^+(d)$  and  $\mathbb{GL}_{\text{sym}}^+(d)$  denotes the class of symmetric and positive-definite  $d \times d$  matrices with real entries. As for the approximating p.d.  $\hat{\mathfrak{p}}_n$ , in agreement with the so-called *plug-in* method of statistical point estimation, put  $\hat{\mathfrak{p}}_n = \mu_{\hat{\theta}_n}$ ,  $\hat{\theta}_n = (\hat{\mathbf{m}}_n, \hat{V}_n)$  and

$$\hat{\mathbf{m}}_n = \frac{1}{n} \sum_{i=1}^n \tilde{\xi}_i, \quad \hat{V}_n = \frac{1}{n} \sum_{i=1}^n (\tilde{\xi}_i - \hat{\mathbf{m}}_n) {}^t(\tilde{\xi}_i - \hat{\mathbf{m}}_n), \quad (2.5)$$

with the proviso that  $\mu_{\hat{\theta}_1} = \delta_{\tilde{\xi}_1}$ , to obtain the following theorem.

**Theorem 2.2.** *Let  $\mathbb{S} = \mathbb{R}^d$  and  $d_{\mathbb{S}}$  coincide with the Euclidean distance. Assume that  $\mathfrak{p}_0 = \mu_{\theta_0}$ ,  $\mu_{\theta_0}$  being defined by (2.4) with  $\theta_0 = (\mathbf{m}_0, V_0) \in \mathbb{R}^d \times \mathbb{GL}_{\text{sym}}^+(d)$ . Then,  $\mathfrak{p}_0$  satisfies (1.2)–(1.3) with  $p = 2$ ,  $b_n = (n/\text{Log}_2 n)^{1/2}$ ,  $\hat{\mathfrak{p}}_n = \mu_{\hat{\theta}_n}$  and  $\hat{\theta}_n = (\hat{\mathbf{m}}_n, \hat{V}_n)$  as in (2.5). Moreover, for any  $\varepsilon$  such that  $0 < \varepsilon \leq [2(d+1)]^{-d}$ , there exist two positive constants  $K(\varepsilon, d)$  and  $H(\varepsilon, d)$  for which one can put*

$$C_2(\mathfrak{p}_0) = [c_\star + b_d^2(d+1)] \text{tr}[V_0] + c_\star K(\varepsilon, d) \|V_0\|_F + H(\varepsilon, d), \quad (2.6)$$

$$Y_2(\mathfrak{p}_0) = \sqrt{2} \{ \sigma_{\max}[V_0] + d \sqrt{K(\varepsilon, d)} \sqrt{\text{tr}[V_0]} \}, \quad (2.7)$$

where  $\text{tr}[\cdot]$ ,  $\|\cdot\|_F$  and  $\sigma_{\max}^2[V_0]$  stand for the trace, the Frobenius norm and the maximum eigenvalue of a matrix, respectively, while  $c_\star$  and  $c_\star$  are two numerical constants, even independent of the dimension  $d$ .

See Section 4.3 for the proof. In particular, the proof of Lemma 4.1 contains the precise evaluation of  $K(\varepsilon, d)$ , while  $c_\star$  is displayed in (4.12),  $c_\star$  in (4.17), and  $H(\varepsilon, d)$  in (4.21). Lastly, definition and properties of  $\|\cdot\|_F$  can be found in Section 5.6 of [42].

A remarkably interesting fact is that an analogous conclusion holds true for a rather general type of *exponential family* which, besides including the previous class of Gaussian distributions as a distinguished, but particular, case, enjoys important, global properties from the point of view of mathematical statistics. Two of them are worth noticing: (i) identifiability, in the sense of a parametric family of p.m.'s; (ii) existence and uniqueness of the maximum likelihood estimator,

expressed as arithmetic mean of values of a given function  $\mathbf{t}$ . In turn, (i)–(ii) are implied by certain weak conditions of a technical character, which are explained, for example, in [6,16]. A brief summary is given here, referring the reader to the quoted monographs for the details. One can start introducing a  $\sigma$ -finite measure  $\mu$  on the metric space  $(\mathbb{S}, d_{\mathbb{S}})$ , to dominate the whole family  $\mathcal{M}$  (to be defined just below), together with the aforesaid measurable map  $\mathbf{t} : (\mathbb{S}, \mathcal{B}(\mathbb{S})) \rightarrow (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ , in such a way that:

- (A) the interior  $\Theta$  of the convex hull of the support of  $\mu \circ \mathbf{t}^{-1}$  is non-void,
- (B) the set  $\Lambda := \{\mathbf{y} \in \mathbb{R}^k \mid \int_{\mathbb{S}} e^{\mathbf{y} \cdot \mathbf{t}(x)} \mu(dx) < +\infty\}$  is a non void open subset of  $\mathbb{R}^k$ , which proves to be convex.

Thus, a parametrization of the exponential family  $\mathcal{M}$ , conducive to (i)–(ii), is given by

$$\mu_{\theta}(dx) := \exp\{V^{-1}(\theta) \cdot \mathbf{t}(x) - M(V^{-1}(\theta))\} \mu(dx) \quad (\theta \in \Theta), \tag{2.8}$$

where  $M(\mathbf{y}) := \log(\int_{\mathbb{S}} e^{\mathbf{y} \cdot \mathbf{t}(x)} \mu(dx)) : \mathbb{R}^k \rightarrow (-\infty, +\infty]$  and  $V(\mathbf{y}) := \nabla M(\mathbf{y}) = \int_{\mathbb{S}} \mathbf{t}(x) \exp\{\mathbf{y} \cdot \mathbf{t}(x) - M(\mathbf{y})\} \mu(dx)$ . These definitions make sense in view of the following facts: first, Theorems 1.13, 2.2 and 2.7 in [16] state that  $M$  is strictly convex on  $\Lambda$ , lower semi-continuous on  $\mathbb{R}^k$ , of class  $C^\infty(\Lambda)$  and analytic. Second, Corollary 5.3 in [6] implies that  $M$  is steep (essentially smooth), according to Definition 3.2 of [16]. Third, from Theorem 3.6 in [16],  $\mathbf{y} \mapsto V(\mathbf{y})$  defines a smooth homeomorphism of  $\Lambda$  and  $\Theta$ . Thus, putting  $\gamma_{\mathbf{y}}(A) := \int_A \exp\{\mathbf{y} \cdot \mathbf{t}(x) - M(\mathbf{y})\} \mu(dx)$ ,  $A \in \mathcal{B}(\mathbb{S})$ , Corollary 2.5 in [16] entails the equivalence of  $\gamma_{\mathbf{y}_1} = \gamma_{\mathbf{y}_2}$ ,  $\mathbf{y}_1 = \mathbf{y}_2$  and  $\int \mathbf{t} d\gamma_{\mathbf{y}_1} = \int \mathbf{t} d\gamma_{\mathbf{y}_2}$ , making precise the requisite of identifiability. Therefore, (2.8) defines the family  $\mathcal{M}$  which is the subject of the next theorem, in such a way that  $\mathbf{p}_0$  will be a fixed element  $\mu_{\theta_0}$  of  $\mathcal{M}$ , and  $\hat{\mathbf{p}}_n$  the element  $\mu_{\hat{\theta}_n}$  corresponding to

$$\hat{\theta}_n := \frac{1}{n} \sum_{i=1}^n \mathbf{t}(\tilde{\xi}_i) \in \Theta \tag{2.9}$$

which represents, in this case, the *maximum likelihood estimator* of the parameter  $\theta$ .

The particular form of  $\hat{\theta}_n$  is conducive to an application of Proposition 2.1, but with the significant variant that the analog of the supremum appearing in (2.1) is multiplied by a function, say  $\Phi$ , of  $\hat{\theta}_n$ , which diverges as either  $|\hat{\theta}_n|$  goes to infinity or  $\hat{\theta}_n$  approaches the boundary of  $\Theta$ . Hence, to obtain any quantitative bound, the following concepts will come in useful: first, the cumulant generating function, say  $\Psi$ , of  $\mathbf{t}(\tilde{\xi}_1)$ , namely  $\Psi(\mathbf{y}) := \log(\int_{\mathbb{S}} e^{\mathbf{y} \cdot \mathbf{t}(x)} \mu_{\theta_0}(dx))$ . Second, the *Legendre transformation*, say  $I_{\mu}$ , of  $M$ , that is

$$I_{\mu}(\theta) := \sup_{\mathbf{y} \in \Lambda} \{\theta \cdot \mathbf{y} - M(\mathbf{y})\} = V^{-1}(\theta) \cdot \theta - M(V^{-1}(\theta)). \tag{2.10}$$

Third, the Legendre transformation, say  $I_{\theta_0}$ , of  $\Psi$ , given by

$$I_{\theta_0}(\theta) := \sup_{\mathbf{y} \in \Lambda} \{\theta \cdot \mathbf{y} - \Psi(\mathbf{y})\} = I_{\mu}(\theta) - \theta \cdot \mathbf{y}_0 - M(\mathbf{y}_0), \tag{2.11}$$

where  $\mathbf{y}_0 = V^{-1}(\theta_0)$ . See Section 26 of [53] (especially Theorems 26.4 and 26.6) for well-definiteness and regularity of  $I_{\mu}$ , as well as Theorem 6.9 in [16] for steepness of  $I_{\mu}$ . The way is

now paved to specify  $\Phi$ , that is,

$$\Phi(\boldsymbol{\eta}) := \sqrt{\int_0^1 (1-s) \|\text{Hess}[I_\mu](\boldsymbol{\theta}_0 + s(\boldsymbol{\eta} - \boldsymbol{\theta}_0))\|_F ds} \quad (\boldsymbol{\eta} \in \Theta), \tag{2.12}$$

where  $\|\cdot\|_F$  stands for the Frobenius norm. As a consequence of the properties of  $I_\mu$ , in order that  $\Phi(\boldsymbol{\eta})$  might diverge it is necessary that either  $|\boldsymbol{\eta}|$  goes to infinity or  $\boldsymbol{\eta}$  approaches  $\partial\Theta$ . Hence, with a view to quantifying the rapidity of the divergence of  $\Phi$ , one observes that there always exist two functions  $\Phi_1 : [0, +\infty) \rightarrow [0, +\infty)$  and  $\Phi_2 : \Theta \rightarrow [0, +\infty)$  satisfying

$$\begin{cases} \Phi(\boldsymbol{\eta}) \leq \Phi_1(|\boldsymbol{\eta}|) + \Phi_2(\boldsymbol{\eta}) & \text{for all } \boldsymbol{\eta} \in \Theta; \\ \Phi_1 \text{ is increasing and diverges at infinity;} \\ \forall \{\boldsymbol{\eta}_n\}_{n \geq 1} \subset \Theta, \Phi_2(\boldsymbol{\eta}_n) \rightarrow +\infty \text{ entails } d(\boldsymbol{\eta}_n, \partial\Theta) \rightarrow 0, & \text{as } n \rightarrow +\infty. \end{cases} \tag{2.13}$$

Then, along with  $\Phi_2$ , one introduces the non-negative integer (possibly equal to  $+\infty$ )  $N_{\Phi_2}(\rho, \sigma)$  which represents the minimum number of convex closed sets that form a covering of  $R_1 := \{\boldsymbol{\theta} \in \Theta \mid \Phi_2(\boldsymbol{\theta}) \geq \frac{\sigma}{2}, |\boldsymbol{\theta} - \boldsymbol{\theta}_0| \leq \rho\}$  with the constrain that such a covering must be included in  $R_2 := \{\boldsymbol{\theta} \in \Theta \mid \Phi_2(\boldsymbol{\theta}) \geq \frac{\sigma}{4}, |\boldsymbol{\theta} - \boldsymbol{\theta}_0| \leq \rho\}$ .

Finally, it should be noticed that the involvement of  $d_{[\mathbb{S}]}^{(p)}$  in (1.2)–(1.3) forces to assume that  $\mathcal{M}$  is included in  $[\mathbb{S}]_p$  and, with a view to the obtainment of explicit upper bounds, induces to focus attention on the so-called *Talagrand inequality*. With reference to the elements of  $\mathcal{M}$  with  $p = 2$ , this inequality reads

$$d_{[\mathbb{S}]}^{(2)}(\mu_{\boldsymbol{\theta}_0}, \mu_{\boldsymbol{\theta}})^2 \leq C_T(\boldsymbol{\theta}_0) \mathbf{K}(\mu_{\boldsymbol{\theta}} \mid \mu_{\boldsymbol{\theta}_0}) \quad (\boldsymbol{\theta}_0, \boldsymbol{\theta} \in \Theta), \tag{2.14}$$

where  $\mathbf{K}(\mu_{\boldsymbol{\theta}} \mid \mu_{\boldsymbol{\theta}_0}) := \int_{\mathbb{S}} \log\left(\frac{d\mu_{\boldsymbol{\theta}}}{d\mu_{\boldsymbol{\theta}_0}}(x)\right) \mu_{\boldsymbol{\theta}}(dx)$  is the *Kullback–Leibler information* of  $\mu_{\boldsymbol{\theta}_0}$  at  $\mu_{\boldsymbol{\theta}}$ . See, for example, [39,41] for further information.

**Theorem 2.3.** *Suppose that (A)–(B) are fulfilled and that the elements of the exponential family (2.8) belong to  $[\mathbb{S}]_2$  and meet (2.14). Then,  $\mathbf{p}_0 = \mu_{\boldsymbol{\theta}_0}$  satisfies (1.3) with  $p = 2$ ,  $b_n = (n/\text{Log}_2 n)^{1/2}$ ,  $\hat{\mathbf{p}}_n = \mu_{\hat{\boldsymbol{\theta}}_n}$ ,  $\hat{\boldsymbol{\theta}}_n$  as in (2.9), and*

$$Y_2(\mathbf{p}_0) = \sqrt{2C_T(\boldsymbol{\theta}_0)} \Phi(\boldsymbol{\theta}_0) \sigma_{\max}(\boldsymbol{\theta}_0), \tag{2.15}$$

where  $\sigma_{\max}^2(\boldsymbol{\theta}_0)$  is the largest eigenvalue of  $\text{Hess}[M](V^{-1}(\boldsymbol{\theta}_0))$ . Moreover,  $\mathbf{p}_0$  satisfies (1.2) with the same  $p$ ,  $b_n$ ,  $\hat{\mathbf{p}}_n$  and a suitable  $C_2(\mathbf{p}_0)$  specified in the proof if, in addition, there exist a positive constant  $\tau(\boldsymbol{\theta}_0)$  and functions  $\rho, \sigma : [\tau(\boldsymbol{\theta}_0), +\infty) \rightarrow [0, +\infty)$  satisfying

$$\int_{\tau(\boldsymbol{\theta}_0)}^{+\infty} m(t)^{-2} e^{-m(t)} dt + \int_{\tau(\boldsymbol{\theta}_0)}^{+\infty} N_{\Phi_2}(\rho(t), \sigma(t)) h(t)^{-2} e^{-h(t)} dt < +\infty, \tag{2.16}$$

where, with reference to (2.13), one puts  $m(t) := \min\{\rho(t), \frac{t}{\sigma(t)}, \Phi_1^{-1}(\frac{\sigma(t)}{2})\}$  and  $h(t) := \inf\{I_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}) \mid \Phi_2(\boldsymbol{\theta}) \geq \frac{\sigma(t)}{4}, |\boldsymbol{\theta} - \boldsymbol{\theta}_0| \leq \rho(t)\}$ .

See Section 4.4 for the proof, which also includes the elements to specify the constant  $C_2(p_0)$ .

Before concluding the section, it is worth remarking that the abundance of assumptions in the last theorem counterbalances the generality of the family (2.8). In any case, checking of (A)–(B) is a standard task, as shown in Chapter 1 of [16]. To check the validity of (2.14) see [39,41], which also highlight the interesting connection with the *logarithmic Sobolev inequality* as means to estimate the constant  $C_T(\theta_0)$ . Lastly, the finding of  $\tau(\theta_0)$ ,  $\rho(t)$ ,  $\sigma(t)$  and  $N(t)$  is less standard and could prove more labored, even because of lack of background literature.

### 3. Exchangeable random variables

As announced in Section 1, the random coordinates  $\tilde{\xi}_1, \tilde{\xi}_2, \dots$  are now thought of as exchangeable random elements distributed according to the p.m.  $\rho$  on  $(\Omega, \mathcal{F}) = (\mathbb{S}^\infty, \mathcal{B}(\mathbb{S}^\infty))$ . The reader is recommended to resort to the representation theorem recalled therein and, in particular, to the meaning of the random p.m.  $\tilde{p}$  as limit of the sequence of the empirical laws  $\tilde{c}_n$ 's, as well as of the symbols  $\pi(\tilde{\xi}^{(n)})$  and  $p(\tilde{\xi}^{(n)})$  to denote the posterior and the predictive distribution, respectively.

The first theorem establishes the validity of (1.4). Its novelty is that it provides information about the *range* of the sequence  $\{d_{[[\mathbb{S}]_p]}^{(p)}(\pi(\tilde{\xi}^{(n)}), \delta_{\hat{p}_n})\}_{n \geq 1}$ , when  $(\mathbb{S}, d_{\mathbb{S}})$  is a general separable metric space meeting the Radon property.

**Theorem 3.1.** *Assume that, for some  $p \geq 1$ ,  $\hat{p}_n \in [\mathbb{S}]_p$  for all  $n \in \mathbb{N}$  and  $\tilde{p} \in [\mathbb{S}]_p$  with  $\rho$ -probability 1. Moreover, let  $\{b_n\}_{n \geq 1}$  be an increasing and diverging sequence of positive numbers with respect to which  $\tilde{p}$  satisfies (1.2)–(1.3) with  $\rho$ -probability 1. Then,*

$$\rho \left[ \limsup_{n \rightarrow \infty} b_n d_{[[\mathbb{S}]_p]}^{(p)}(\pi(\tilde{\xi}^{(n)}), \delta_{\hat{p}_n}) \leq Y_p(\tilde{p}) \right] = 1 \tag{3.1}$$

*provided that  $C_p(\tilde{p})$  and  $Y_p(\tilde{p})$  are real random variables and  $\rho[C_p(\tilde{p})] < +\infty$ .*

The proof is deferred to Section 4.5. Apropos of the assumption that  $C_p(\tilde{p})$  and  $Y_p(\tilde{p})$  are real random variables, notice that their checking boils down to an analysis of their explicit expressions. Cf., for example, (2.2)–(2.3) and (2.6)–(2.7). Lastly, recall Remark 1.2 for an answer to the problem of bounding  $\limsup b_n d_{[[\mathbb{S}]_s]}^{(s)}(\pi(\tilde{\xi}^{(n)}), \delta_{\hat{p}_n})$  for  $s \in [1, p]$ .

In the same vein, one now deals with the approximation of the distributions  $q_m(\tilde{\xi}^{(n)})$  as in (1.5). To connect  $q_m(\tilde{\xi}^{(n)})$  with  $\pi(\tilde{\xi}^{(n)})$ , suffice it to notice that  $p(\tilde{\xi}^{(n)}, A) = \int_{[\mathbb{S}]} p^\infty(A) \pi(\tilde{\xi}^{(n)}, dp)$  and  $q_m(\tilde{\xi}^{(n)}, B) = p(\tilde{\xi}^{(n)}, \tilde{c}_{n,m}^{-1}(B))$  hold true for every  $A \in \mathcal{B}(\mathbb{S}^\infty)$  and  $B \in \mathcal{B}([\mathbb{S}])$ . From a technical point of view, knowledge of the latter distribution may help to make the properties of the former more explicit. This is what happens when, by following the usual frequentistic interpretation of analogy of the observations, one approximates  $q_m(\tilde{\xi}^{(n)})$  by means of the distinguished distribution of  $\frac{1}{m} \sum_{i=1}^m \delta_{\tilde{\xi}_{i+n}}$  when the  $\tilde{\xi}_{i+n}$ 's are viewed as i.i.d. with common p.d.  $\hat{p}_n$ , for every  $n$ . Apropos of this, one states the second main result of this section as the following theorem.

**Theorem 3.2.** *If, for some  $p \geq 1$ ,  $\hat{\mathfrak{p}}_n \in [\mathbb{S}]_p$  for all  $n \in \mathbb{N}$  and  $\tilde{\mathfrak{p}} \in [\mathbb{S}]_p$  with  $\rho$ -probability 1, then*

$$\rho\left[d_{[[\mathbb{S}]_p]}^{(p)}(q_m(\tilde{\xi}^{(n)}), \hat{\mathfrak{p}}_n^\infty \circ \tilde{\mathfrak{c}}_{n,m}^{-1}) \leq d_{[[\mathbb{S}]_p]}^{(p)}(\pi(\tilde{\xi}^{(n)}), \delta_{\tilde{\mathfrak{p}}_n})\right] = 1 \tag{3.2}$$

for every  $m \in \mathbb{N}$ . In particular, under the same assumptions of Theorem 3.1, one has

$$\rho\left[\limsup_{n \rightarrow \infty} b_n d_{[[\mathbb{S}]_p]}^{(p)}(q_m(\tilde{\xi}^{(n)}), \hat{\mathfrak{p}}_n^\infty \circ \tilde{\mathfrak{c}}_{n,m}^{-1}) \leq Y_p(\tilde{\mathfrak{p}})\right] = 1 \tag{3.3}$$

for every  $m \in \mathbb{N}$ .

See Section 4.6 for the proof, and resort to Remark 1.2 for an answer to the problem of bounding Wasserstein distances of lower orders  $s \in [1, p]$ .

Theorems 3.1 and 3.2 are deliberately stated in a very general form, as far as the nature of the space  $\mathbb{S}$  is concerned, but contain the strong assumption that (1.2)–(1.3) are in force. With a view to their real use, one could do a good turn by providing possibly simple sufficient conditions for the validity of (1.2)–(1.3), even if one has to give up the generality of  $\mathbb{S}$ . It is with this spirit that we state the following three theorems which, from a merely mathematical side, are immediate consequences of the combination of Theorems 3.1 and 3.2 with Theorems 2.1, 2.2 and 2.3, respectively. The first statement deals with the nonparametric model, and can be viewed as an extension of Theorem 4 in [20] from  $\mathbb{R}$  to  $\mathbb{R}^d$ .

**Theorem 3.3.** *Let  $\mathbb{S} = \mathbb{R}^d$  and  $d_{\mathbb{S}}$  be the Euclidean distance. Assume that, for some  $\delta > 0$  and  $p \in [1, +\infty) \cap (d/2, +\infty)$ ,*

$$\rho\left[\left(\int_{\mathbb{R}^d} |\mathbf{x}|^{2p+\delta} \tilde{\mathfrak{p}}(d\mathbf{x})\right)^{\frac{3}{r}-1}\right] = \int_{[\mathbb{R}^d]} \left(\int_{\mathbb{R}^d} |\mathbf{x}|^{2p+\delta} \mathfrak{p}(d\mathbf{x})\right)^{\frac{3}{r}-1} \pi(d\mathfrak{p}) < +\infty$$

is valid with a suitable  $r \in (2, 3)$  for which  $p/d > 2 - \frac{3}{r}$  and  $(2p + \delta)(\frac{3}{r} - 1) > p$ . Then, (1.4)–(1.5) hold true with  $\hat{\mathfrak{p}}_n = \tilde{\mathfrak{c}}_n$ ,  $b_n = (n/\text{Log}_2 n)^{1/2p}$  and the same constant  $Y_p$  as in Theorem 2.1, with  $\tilde{\mathfrak{p}}$  in place of  $\mathfrak{p}_0$ .

The next theorem provides a new result concerning the  $d$ -dimensional Gaussian model.

**Theorem 3.4.** *If  $\mathbb{S} = \mathbb{R}^d$  and  $d_{\mathbb{S}}$  coincides with the Euclidean distance, suppose that  $\tilde{\mathfrak{p}} = \mu_{\tilde{\theta}}$  with  $\rho$ -probability 1, where  $\tilde{\theta} = (\tilde{\mathbf{m}}, \tilde{V})$  is a random element taking values in  $\Theta = \mathbb{R}^d \times \mathbb{GL}_{\text{sym}}^+(d)$  and  $\mu_{\theta}$  is referred to (2.4). If  $\rho[\|\tilde{V}\|_F] = \int_0^{+\infty} x \pi_{\|\tilde{V}\|_F}(dx) < +\infty$  is valid, with  $\pi_{\|\tilde{V}\|_F}$  standing for the  $p.d.$  of  $\|\tilde{V}\|_F$ , then (1.4)–(1.5) hold true with  $\hat{\mathfrak{p}}_n = \mu_{\hat{\theta}_n}$ ,  $b_n = (n/\text{Log}_2 n)^{1/2}$  and the same constant  $Y_p$  as in Theorem 2.2, with  $\tilde{V}$  in place of  $V_0$ .*

Finally, the third result concerns the exponential family considered in Theorem 2.3.

**Theorem 3.5.** *Let  $\tilde{\mathfrak{p}}$  equal  $\mu_{\tilde{\theta}}$  with  $\rho$ -probability 1, where  $\mu_{\theta}$  is referred to (2.8) and  $\tilde{\theta}$  is a random element with values in  $\Theta$ . Suppose that (A)–(B) are fulfilled and that, with  $\rho$ -probability*

1,  $\mu_{\tilde{\theta}} \in [\mathbb{S}]_2$ , (2.14) holds with  $\tilde{\theta}$  in place of  $\theta_0$  and  $\hat{\theta}_n$  in place of  $\theta$ , and (2.16) is in force. Then, if the same constant  $C_2(p_0)$  mentioned in Theorem 2.3 is such that  $C_2(\mu_{\tilde{\theta}})$  is a real random variable satisfying  $\rho[C_2(\mu_{\tilde{\theta}})] < +\infty$ , properties (1.4)–(1.5) hold true with  $\hat{p}_n = \mu_{\hat{\theta}_n}$ ,  $b_n = (n/\text{Log}_2 n)^{1/2}$  and  $Y_2(\tilde{p}) = \sqrt{2C_T(\tilde{\theta})}\Phi(\tilde{\theta})\sigma_{\max}(\tilde{\theta})$ .

**Remark 3.1.** Theorems 3.1 and 3.2 face the challenging task of comparing probability laws of p.d.’s. Nevertheless, there are situations in which it suffices to have a partial knowledge of a p.d., such as mean values or appropriate synthetic measures of interesting aspects of that p.d., presentable as functionals on classes of p.d.’s. Thus, Theorems 3.1 and 3.2 are ideally accompanied by results concerning the comparison between  $\rho[g(\tilde{p}) \mid \tilde{\xi}^{(n)}]$  and  $g(\hat{p}_n)$ , and between  $\rho[g(\frac{1}{m} \sum_{i=1}^m \delta_{\tilde{\xi}_{i+n}}) \mid \tilde{\xi}^{(n)}]$  and  $\hat{p}_n^\infty[g(\frac{1}{m} \sum_{i=1}^m \delta_{\tilde{\xi}_{i+n}})]$ , respectively, where  $g$  stands for some specific functional. In particular, in view of the Kantorovich–Rubinstein representation of  $d^{(1)}$  (see, e.g., Theorem 11.8.2 of [33]), if  $g$  is Lipschitz-continuous, one has

$$|\rho[g(\tilde{p}) \mid \tilde{\xi}^{(n)}] - g(\hat{p}_n)| \leq \text{Lip}(g)d_{[[\mathbb{S}]_1]}^{(1)}(\pi(\tilde{\xi}^{(n)}), \delta_{\hat{p}_n})$$

and

$$\begin{aligned} & \left| \rho \left[ g \left( \frac{1}{m} \sum_{i=1}^m \delta_{\tilde{\xi}_{i+n}} \right) \mid \tilde{\xi}^{(n)} \right] - \hat{p}_n^\infty \left[ g \left( \frac{1}{m} \sum_{i=1}^m \delta_{\tilde{\xi}_{i+n}} \right) \right] \right| \\ & \leq \text{Lip}(g)d_{[[\mathbb{S}]_1]}^{(1)}(q_m(\tilde{\xi}^{(n)}), \hat{p}_n^\infty \circ \tilde{e}_{n,m}^{-1}), \end{aligned}$$

where  $\text{Lip}(g)$  stands for the Lipschitz constant of  $g$ . Due to its relevance, proven by a rich literature, the empirical approximation of functionals will be studied in a forthcoming paper.

## 4. Proofs

Gathered here are the proofs of the main results.

### 4.1. Proof of Proposition 2.1

This subsection complements the arguments already developed to prove Lemma 1 and Theorem 4 in Section 10.3 of [19], in order to justify the bound (2.1). Then, notation is substantially the same as therein, and numbering relates to that very same book.

To start, consider a sequence  $\{Y_n\}_{n \geq 1}$  of independent, non-negative random variables, and set  $\mathfrak{S} := \sum_{n=1}^{+\infty} Y_n$  and  $\mathfrak{M}_k := \sum_{n=1}^{+\infty} E[Y_n^k]$ , for  $k > 0$ . An application of the inequality  $(a + b)^m \leq 2^{m-1}(a^m + b^m)$ , valid for  $a, b > 0$  and  $m \geq 1$ , yields  $E[\mathfrak{S}^q] \leq 2^{q-2}(\mathfrak{M}_q + \mathfrak{M}_1 E[\mathfrak{S}^{q-1}])$  for  $q \in \{2, 3, \dots\}$ . Thus, by induction, one gets  $E[\mathfrak{S}^q] \leq \lambda_{q,q-1} \mathfrak{M}_1^q + \sum_{k=0}^{q-2} \lambda_{q,k} \mathfrak{M}_{q-k} \mathfrak{M}_1^k$  for  $q \in \{2, 3, \dots\}$ , where  $\lambda_{2,1} := 1$ ,  $\lambda_{q,0} := 2^{q-2}$  and  $\lambda_{q+1,h} := 2^{q-1} \lambda_{q,h-1}$  if  $h = 1, \dots, q$ . If  $q > 2$

and  $q \notin \mathbb{N}$ , one can write

$$\begin{aligned} \mathbb{E}[\mathfrak{S}^q] &\leq 2^{[q]-1}(\mathfrak{M}_q + \mathfrak{M}_{q-[q]}\mathbb{E}[\mathfrak{S}^{[q]}]) \\ &\leq 2^{[q]-1}\left(\mathfrak{M}_q + \lambda_{[q],[q]-1}\mathfrak{M}_{q-[q]}\mathfrak{M}_1^{[q]} + \mathfrak{M}_{q-[q]}\sum_{k=0}^{[q]-2}\lambda_{[q],k}\mathfrak{M}_{[q]-k}\mathfrak{M}_1^k\right). \end{aligned}$$

Now, the reasoning continues paralleling the method used to prove Theorem 4. First, assume that the distribution of the  $X_n$ 's is symmetric with  $\mathbb{E}[X_1^2] = 1$ . Second, define  $c_n := (n \text{Log}_2(n))^{-1/2}$ ,  $X'_n := X_n \mathbb{1}\{|X_n| \leq n^{1/r}\}$ ,  $X''_n := X_n \mathbb{1}\{|X_n| > n^{1/r}\}$  and decompose  $S_n := \sum_{i=1}^n X_i$  as  $S_n = S'_n + S''_n$ , with  $S'_n := \sum_{i=1}^n X'_i$  and  $S''_n := \sum_{i=1}^n X''_i$ . Apropos of  $S''_n$ , exploit the monotonicity of the sequence  $\{c_n\}_{n \geq 1}$ , as in the last inequality on page 389, to obtain

$$\mathbb{E}\left[\sup_{n \geq 1} \frac{|S''_n|^r}{(n \text{Log}_2(n))^{r/2}}\right] \leq \mathbb{E}\left[\left(\sup_{n \geq 1} c_n \sum_{i=1}^n |X''_i|\right)^r\right] \leq \mathbb{E}\left[\left(\sum_{n \geq 1} c_n |X''_n|\right)^r\right].$$

Therefore, if  $r \in \{3, 4, \dots\}$ , one gets

$$\begin{aligned} &\mathbb{E}\left[\sup_{n \geq 1} \frac{|S''_n|^r}{(n \text{Log}_2(n))^{r/2}}\right] \\ &\leq \lambda_{r,r-1}\left(\sum_{n=1}^{+\infty} c_n \mathbb{E}[|X''_n|]\right)^r + \sum_{k=0}^{r-2} \lambda_{r,k}\left(\sum_{n=1}^{+\infty} c_n^{r-k} \mathbb{E}[|X''_n|^{r-k}]\right) \cdot \left(\sum_{n=1}^{+\infty} c_n \mathbb{E}[|X''_n|]\right)^k \end{aligned} \tag{4.1}$$

whilst, if  $r > 2$  and  $r \notin \mathbb{N}$ , the following bound

$$\begin{aligned} &\mathbb{E}\left[\sup_{n \geq 1} \frac{|S''_n|^r}{(n \text{Log}_2(n))^{r/2}}\right] \\ &\leq 2^{[r]-1}\left\{\sum_{n=1}^{+\infty} c_n^r \mathbb{E}[|X''_n|^r] + \lambda_{[r],[r]-1}\left(\sum_{n=1}^{+\infty} c_n^{r-[r]} \mathbb{E}[|X''_n|^{r-[r]}]\right) \cdot \left(\sum_{n=1}^{+\infty} c_n \mathbb{E}[|X''_n|]\right)^{[r]} \right. \\ &\quad \left. + \left(\sum_{n=1}^{+\infty} c_n^{r-[r]} \mathbb{E}[|X''_n|^{r-[r]}]\right) \cdot \sum_{k=0}^{[r]-2} \lambda_{[r],k}\left(\sum_{n=1}^{+\infty} c_n^{[r]-k} \mathbb{E}[|X''_n|^{[r]-k}]\right) \right. \\ &\quad \left. \cdot \left(\sum_{n=1}^{+\infty} c_n \mathbb{E}[|X''_n|]\right)^k\right\} \end{aligned} \tag{4.2}$$

is in force. Now, combining the inequalities of Hölder and Markov, one gets  $\mathbb{E}[|X''_n|^h] \leq (1/n)^{\frac{r-h}{r}} \mathbb{E}[|X_1|^r]$  for every  $n \in \mathbb{N}$  and  $h \in (0, r]$  which, in turn, yields  $\sum_{n=1}^{+\infty} c_n^h \mathbb{E}[|X''_n|^h] \leq$

$K(h)\mathbb{E}[|X_1|^r]$ , with  $K(h) := \sum_{n=1}^{+\infty} (1/n)^{\frac{r-h}{r}} c_n^h < +\infty$ . Thus, one concludes that

$$\mathbb{E}\left[\sup_{n \geq 1} \frac{|S'_n|^r}{(n \operatorname{Log}_2(n))^{r/2}}\right] \leq \beta_1(r) (\mathbb{E}[|X_1|^r])^{[r]} \tag{4.3}$$

is valid with a suitable numerical constant  $\beta_1(r)$ , determined as follows. First, recall that  $\mathbb{E}[X_1^2] = 1$  entails  $\mathbb{E}[|X_1|^r] \geq 1$ . Then, if  $r \in \{3, 4, \dots\}$ , invoke (4.1) to get  $\beta_1(r) = \sum_{k=0}^{r-1} \lambda_{r,k} K(r-k)(K(1))^k$ , whilst, if  $r > 2$  and  $r \notin \mathbb{N}$ , use (4.2) to obtain

$$\beta_1(r) = 2^{[r]-1} \left\{ K(r) + \lambda_{[r],[r]-1} K(r-[r])(K(1))^{[r]} + K(r-[r]) \sum_{k=0}^{[r]-2} \lambda_{[r],k} K([r]-k)(K(1))^k \right\}.$$

Apropos of the term involving  $S'_n$ , one starts by observing that

$$\mathbb{E}\left[\left(\sup_{n \geq 1} c_n |S'_n|\right)^r\right] \leq u_0^r + r\lambda^r \int_{u_0/\lambda}^{+\infty} u^{r-1} \mathbb{P}\left[\sup_{n \geq 1} c_n |S'_n| > \lambda u\right] du \tag{4.4}$$

holds for every  $\lambda, u_0 > 0$ . Then, after putting  $n_k := [e^k]$  for  $k \in \mathbb{N}_0$ , formula (15) on page 390 can be invoked to write

$$\mathbb{P}\left[\sup_{n \geq 1} c_n |S'_n| > \lambda u\right] \leq 4 \sum_{k=0}^{+\infty} \mathbb{P}[c_{n_k} S'_{n_{k+1}} > \lambda u]. \tag{4.5}$$

Moreover, at the same page, it is proved that  $\mathbb{P}[S'_n > x] \leq \exp\{-tx + nt^2\}$  is valid for every  $n \in \mathbb{N}$ ,  $x > 0$  and  $t \in [0, n^{-1/r}]$ . Then, setting  $n = n_{k+1}$ ,  $\gamma_r := \sup_{k \in \mathbb{N}} \left(\frac{\operatorname{Log}_2(n_{k+1})}{n_{k+1}}\right)^{1/2} \times (n_{k+1})^{1/r} < +\infty$ ,  $t = \gamma_r^{-1} \left(\frac{\operatorname{Log}_2(n_{k+1})}{n_{k+1}}\right)^{1/2}$  and  $x = \frac{u}{\gamma_r c_{n_k}}$  yields

$$\mathbb{P}\left[c_{n_k} S'_{n_{k+1}} > \frac{u}{\gamma_r}\right] \leq \exp\left\{-\frac{(\alpha u - 1)}{\gamma_r^2} \operatorname{Log}_2(n_{k+1})\right\}, \tag{4.6}$$

where  $\alpha := \inf_{k \in \mathbb{N}_0} \left(\frac{n_k \operatorname{Log}_2(n_k)}{n_{k+1} \operatorname{Log}_2(n_{k+1})}\right)^{1/2} > 0$ . Putting  $\lambda = \gamma_r^{-1}$ , the combination of (4.4), (4.5) and (4.6) leads to

$$\mathbb{E}\left[\left(\sup_{n \geq 1} c_n |S'_n|\right)^r\right] \leq u_0^r + 4r\gamma_r^{-r} \sum_{k=0}^{+\infty} \int_{u_0\gamma_r}^{+\infty} u^{r-1} \exp\left\{-\frac{(\alpha u - 1)}{\gamma_r^2} \operatorname{Log}_2(n_{k+1})\right\} du.$$

To conclude, it is enough to choose  $u_0$  large enough so that  $\alpha\gamma_r u_0 > 4\gamma_r^2 + 1$  is in force. In fact, this choice entails, for every  $u \geq u_0\gamma_r$ ,  $\alpha u > 4\gamma_r^2 + 1$  and  $\frac{(\alpha u - 1)}{\gamma_r^2} \operatorname{Log}_2(n_{k+1}) \geq 2 \operatorname{Log}_2(n_{k+1}) +$

$\frac{2\alpha}{4\gamma_r^2+1}u$  which, in turn, give

$$\begin{aligned} \mathbb{E}\left[\left(\sup_{n \geq 1} c_n |S'_n|\right)^r\right] &\leq u_0^r + 4r\gamma_r^{-r} \left(\sum_{k=0}^{+\infty} \exp\{-2\text{Log}_2(n_{k+1})\}\right) \cdot \left(\frac{4\gamma_r^2+1}{2\alpha}\right)^r \Gamma(r) \\ &=: \beta_0(r). \end{aligned} \tag{4.7}$$

The proof can be now carried out by means of the following steps. First, combine (4.3) with (4.7) to get

$$\begin{aligned} \mathbb{E}\left[\left(\sup_{n \geq 1} c_n |S_n|\right)^r\right] &\leq 2^{r-1} \mathbb{E}\left[\left(\sup_{n \geq 1} c_n |S'_n|\right)^r\right] + 2^{r-1} \mathbb{E}\left[\left(\sup_{n \geq 1} c_n |S''_n|\right)^r\right] \\ &\leq 2^{r-1} (\beta_0(r) + \beta_1(r) (\mathbb{E}[|X_1|^r])^{[r]}), \end{aligned} \tag{4.8}$$

which is valid for symmetric  $X_n$ 's with  $\mathbb{E}[X_1^2] = 1$ . Then, invoke the symmetrization argument at the end of page 390 to show that this last very same bound holds also for non-symmetric  $X_n$ 's, provided that  $\mathbb{E}[X_1] = 0$  and  $\mathbb{E}[X_1^2] = 1$ . Finally, remove the assumption  $\mathbb{E}[X_1^2] = 1$ , by replacing every  $X_i$  with  $X_i/\sigma$ .

### 4.2. Proof of Theorem 2.1

Fix  $d \in \mathbb{N}$ ,  $p \in [1, +\infty) \cap (d/2, +\infty)$  and  $\delta > 0$ . After setting  $\beta := 2p + \delta$ , observe that it is possible to choose  $r \in (2, 3)$  so that  $\lambda = \lambda(p, d, r) := p - d(2 - \frac{3}{r})$  and  $\sigma = \sigma(p, d, r) := \beta(\frac{3}{r} - 1) - p$  are both strictly positive constants. In particular, the bound on  $\sigma$  is equivalent to  $r < \frac{3\beta}{\beta+p} \in (2, 3)$ . As for the bound on  $\lambda$ , notice that it is trivially satisfied for any choice of  $r \in (2, 3)$  if  $p \geq d$  whilst, if  $p \in [1, +\infty) \cap (d/2, d)$ , one gets  $r < \frac{3d}{2d-p} \in (2, 3)$ . Therefore, these arguments show that  $r$  can be chosen as any element of  $(2, 3/M)$ , where  $M := \max\{2 - p/d; 1 + 1/(2 + \delta/p)\}$ .

Now, consider the LHS of (1.2) with  $b_n^p = \sqrt{n/\text{Log}_2 n}$  and invoke Lemmata 5 and 6 in [37], whose main contents are now briefly recalled. For  $l \in \mathbb{N}_0 := \{0, 1, 2, \dots\}$ , denote by  $\mathcal{P}_l$  the natural partition of  $B_0 := (-1, 1]^d$  into  $2^{dl}$  translations of  $(-2^{-l}, 2^{-l}]^d$  and, for  $s \in \mathbb{N}$ , put  $B_s := (-2^s, 2^s]^d \setminus (-2^{s-1}, 2^{s-1}]^d$  and, for any  $F \subseteq \mathbb{R}^d$ ,  $2^s F := \{2^s \mathbf{x} \mid \mathbf{x} \in F\}$ . In this notation, for any  $p \geq 1$  and  $\mu, \nu \in [\mathbb{R}^d]_p$ ,

$$d_{[\mathbb{R}^d]}^{(p)}(\mu, \nu)^p \leq K(p, d) \sum_{s=0}^{+\infty} 2^{ps} \sum_{l=0}^{+\infty} 2^{-lp} \sum_{F \in \mathcal{P}_l} |\mu(2^s F \cap B_s) - \nu(2^s F \cap B_s)| \tag{4.9}$$

holds with  $K(p, d) := 2^{2p} d^{p/2} (2^p + 1)(2^p - 1)^{-2}$ . Whence,

$$\begin{aligned} &\mathfrak{p}_0^\infty \left[ \sup_{n \geq 1} b_n^p d_{[\mathbb{R}^d]}^{(p)}(\mathfrak{p}_0, \tilde{\mathfrak{e}}_n)^p \right] \\ &\leq K(p, d) \sum_{s=0}^{+\infty} 2^{ps} \sum_{l=0}^{+\infty} 2^{-lp} \sum_{F \in \mathcal{P}_l} \mathfrak{p}_0^\infty \left[ \sup_{n \geq 1} b_n^p \left| \mathfrak{p}_0(2^s F \cap B_s) - \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{\tilde{\xi}_i \in 2^s F \cap B_s\} \right| \right]. \end{aligned}$$

The objective is now to bound  $\mathfrak{p}_0^\infty[\sup_{n \geq 1} b_n^p | \mathfrak{p}_0(2^s F \cap B_s) - \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{\tilde{\xi}_i \in 2^s F \cap B_s\} |]$  by the following steps: first, choose  $r \in (2, 3)$  satisfying the restrictions imposed in the first lines of this proof. Second, apply Lyapunov’s inequality to get

$$\begin{aligned} & \mathfrak{p}_0^\infty \left[ \sup_{n \geq 1} b_n^p \left| \mathfrak{p}_0(2^s F \cap B_s) - \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{\tilde{\xi}_i \in 2^s F \cap B_s\} \right| \right] \\ & \leq \left\{ \mathfrak{p}_0^\infty \left[ \left( \sup_{n \geq 1} b_n^p \left| \mathfrak{p}_0(2^s F \cap B_s) - \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{\tilde{\xi}_i \in 2^s F \cap B_s\} \right| \right)^r \right] \right\}^{1/r}. \end{aligned}$$

At this stage, the identities  $\mathfrak{p}_0^\infty[\mathbb{1}\{\tilde{\xi}_i \in 2^s F \cap B_s\}] = \mathfrak{p}_0(2^s F \cap B_s)$  and

$$b_n^p \left| \mathfrak{p}_0(2^s F \cap B_s) - \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{\tilde{\xi}_i \in 2^s F \cap B_s\} \right| = \frac{|\sum_{i=1}^n [\mathbb{1}\{\tilde{\xi}_i \in 2^s F \cap B_s\} - \mathfrak{p}_0(2^s F \cap B_s)]|}{\sqrt{n \text{Log}_2 n}},$$

valid for any  $i$  and  $n$  in  $\mathbb{N}$ , respectively, pave the way for the application of Proposition 2.1 with  $X_i = \mathbb{1}\{\tilde{\xi}_i \in 2^s F \cap B_s\} - \mathfrak{p}_0(2^s F \cap B_s)$ . Notice that  $\sigma^2$  and  $\mathbb{E}[|X_i|^r]$  appearing in the RHS of (2.1) are equal to  $\mathfrak{p}_0(2^s F \cap B_s)(1 - \mathfrak{p}_0(2^s F \cap B_s))$  and  $\mathfrak{p}_0(2^s F \cap B_s)(1 - \mathfrak{p}_0(2^s F \cap B_s))^r + \mathfrak{p}_0(2^s F \cap B_s)^r(1 - \mathfrak{p}_0(2^s F \cap B_s))$ , respectively. In view of  $(A + B)^{1/r} \leq A^{1/r} + B^{1/r}$ , one can bound from above the  $(1/r)$ -power of the RHS of (2.1) by means of

$$\begin{aligned} & \alpha_0(r)^{1/r} \sqrt{\mathfrak{p}_0(2^s F \cap B_s)(1 - \mathfrak{p}_0(2^s F \cap B_s))} \\ & + \alpha_1(r)^{1/r} [\mathfrak{p}_0(2^s F \cap B_s)(1 - \mathfrak{p}_0(2^s F \cap B_s))]^{\frac{3}{r}-1} \\ & \cdot [(1 - \mathfrak{p}_0(2^s F \cap B_s))^{r-1} + \mathfrak{p}_0(2^s F \cap B_s)^{r-1}]^{\frac{3}{r}-1} \end{aligned}$$

so that, taking account of  $(1 - \mathfrak{p}_0(2^s F \cap B_s))^{r-1} + \mathfrak{p}_0(2^s F \cap B_s)^{r-1} \leq 2$ ,  $1 - \mathfrak{p}_0(2^s F \cap B_s) \leq 1$  and  $\frac{1}{2} > \frac{3}{r} - 1$ , one can write

$$\mathfrak{p}_0^\infty \left[ \sup_{n \geq 1} b_n^p \left| \mathfrak{p}_0(2^s F \cap B_s) - \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{\tilde{\xi}_i \in 2^s F \cap B_s\} \right| \right] \leq \bar{C}(r) [\mathfrak{p}_0(2^s F \cap B_s)]^{\frac{3}{r}-1},$$

with  $\bar{C}(r) := \alpha_0(r)^{1/r} + 2^{3/r} \alpha_1(r)^{1/r}$ . At this stage, after recalling that  $\#\mathcal{P}_l = 2^{dl}$  and that  $\frac{3}{r} - 1 \in (0, 1)$ , another application of Lyapunov’s inequality yields

$$\begin{aligned} & \sum_{F \in \mathcal{P}_l} \mathfrak{p}_0^\infty \left[ \sup_{n \geq 1} b_n^p \left| \mathfrak{p}_0(2^s F \cap B_s) - \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{\tilde{\xi}_i \in 2^s F \cap B_s\} \right| \right] \\ & \leq \bar{C}(r) 2^{dl(2-\frac{3}{r})} \left( \sum_{F \in \mathcal{P}_l} \mathfrak{p}_0(2^s F \cap B_s) \right)^{\frac{3}{r}-1} = \bar{C}(r) 2^{dl(2-\frac{3}{r})} [\mathfrak{p}_0(B_s)]^{\frac{3}{r}-1}. \end{aligned}$$

Whence,

$$\begin{aligned} & \sum_{l=0}^{+\infty} 2^{-lp} \sum_{F \in \mathcal{P}_l} \mathfrak{p}_0^\infty \left[ \sup_{n \geq 1} b_n^p \left| \mathfrak{p}_0(2^s F \cap B_s) - \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{\tilde{\xi}_i \in 2^s F \cap B_s\} \right| \right] \\ & \leq \overline{C}(r) [\mathfrak{p}_0(B_s)]^{\frac{3}{r}-1} \sum_{l=0}^{+\infty} 2^{dl(2-\frac{3}{r})-lp} = \frac{\overline{C}(r)}{1-2^{-\lambda}} [\mathfrak{p}_0(B_s)]^{\frac{3}{r}-1}. \end{aligned}$$

Now, observe that, for any  $s \in \mathbb{N}$ , the definition of  $B_s$  entails  $\mathfrak{p}_0(B_s) \leq \mathfrak{p}_0^\infty[\{|\tilde{\xi}_1| > 2^{s-1}\}]$  so that, from Markov’s inequality,  $\mathfrak{p}_0(B_s) \leq 2^{-\beta(s-1)} \mathfrak{p}_0^\infty[|\tilde{\xi}_1|^\beta]$  holds with the same  $\beta$  as defined at the beginning of this proof. To conclude the proof of (1.2), gather the above inequalities together to obtain

$$\mathfrak{p}_0^\infty \left[ \sup_{n \geq 1} b_n^p d_{[\mathbb{R}^d]}^{(p)}(\mathfrak{p}_0, \tilde{\epsilon}_n)^p \right] \leq \frac{K(p, d) \overline{C}(r)}{1-2^{-\lambda}} \left[ 1 + \frac{2^p}{1-2^{-\sigma}} (\mathfrak{p}_0^\infty[|\tilde{\xi}_1|^\beta])^{\frac{3}{r}-1} \right]$$

which provides the value of  $C_p(\mathfrak{p}_0)$  displayed in (2.2).

To prove (1.3), use again (4.9) to write

$$\begin{aligned} & \limsup_{n \rightarrow +\infty} b_n^p d_{[\mathbb{R}^d]}^{(p)}(\mathfrak{p}_0, \tilde{\epsilon}_n)^p \\ & \leq K(p, d) \sum_{s=0}^{+\infty} 2^{ps} \sum_{l=0}^{+\infty} 2^{-lp} \sum_{F \in \mathcal{P}_l} \limsup_{n \rightarrow +\infty} b_n^p \left| \mathfrak{p}_0(2^s F \cap B_s) - \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{\tilde{\xi}_i \in 2^s F \cap B_s\} \right|. \end{aligned}$$

Since, from Hartman–Wintner’s law of the iterated logarithm,

$$\limsup_{n \rightarrow +\infty} b_n^p \left| \mathfrak{p}_0(2^s F \cap B_s) - \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{\tilde{\xi}_i \in 2^s F \cap B_s\} \right| = \sqrt{2\mathfrak{p}_0(2^s F \cap B_s)(1 - \mathfrak{p}_0(2^s F \cap B_s))}$$

one can argue exactly as above to obtain

$$\limsup_{n \rightarrow +\infty} b_n^p d_{[\mathbb{R}^d]}^{(p)}(\mathfrak{p}_0, \tilde{\epsilon}_n)^p \leq \frac{\sqrt{2}K(p, d)}{1-2^{-(p-d/2)}} \left[ 1 + \frac{2^p}{1-2^{-\delta/2}} (\mathfrak{p}_0^\infty[|\tilde{\xi}_1|^\beta])^{1/2} \right]$$

which provides the value of  $Y_p(\mathfrak{p}_0)$  displayed in (2.3).

### 4.3. Proof of Theorem 2.2

To facilitate the proof, one premises the following technical result.

**Lemma 4.1.** *Let  $\lambda_1, \dots, \lambda_d$  be the eigenvalues (counted with their multiplicity) of a symmetric, positive definite  $d \times d$  matrix  $M$  with real entries. If  $\det(M) \geq \varepsilon$  for some  $\varepsilon \in (0, 1)$ , then there*

exists a positive constant  $K(\varepsilon, d)$  such that

$$\sum_{j=1}^d (\lambda_j - 1 - \log \lambda_j) \leq K(\varepsilon, d) \sum_{j=1}^d (\lambda_j - 1)^2 = K(\varepsilon, d) \|M - \text{Id}_d\|_F^2, \tag{4.10}$$

where  $\text{Id}_d$  is the  $d \times d$  identity matrix and  $\|\cdot\|_F$  stands for the Frobenius norm.

**Proof of Lemma 4.1.** Fix  $\delta \in (0, 1)$  and define  $d_1(M, \delta)$  to be the number of those  $\lambda_j$ 's belonging to  $[1 - \delta, 1 + \delta]$ . Since  $K_1(\delta) := \sup_{x \in [1-\delta, 1+\delta] \setminus \{1\}} \frac{(x-1)^2}{x - \log x - 1} > 0$ , if  $d_1(M, \delta) > 0$  one gets

$$\sum_{\substack{j \in \{1, \dots, d\} \\ \lambda_j \in [1-\delta, 1+\delta]}} (\lambda_j - 1 - \log \lambda_j) \leq K_1(\delta) \sum_{\substack{j \in \{1, \dots, d\} \\ \lambda_j \in [1-\delta, 1+\delta]}} (\lambda_j - 1)^2.$$

If  $d_1(M, \delta) = d$ , the lemma is proved upon putting  $K(\varepsilon, d) = K_1(\delta)$ , with  $(1 - \delta)^d = \varepsilon$ . Thus, assume that  $d_2 = d_2(M, \delta) := d - d_1(M, \delta) > 0$ , and notice that  $\sum_{(*)} (\lambda_j - 1 - \log \lambda_j) \leq \sum_{(*)} \lambda_j - \log \prod_{(*)} \lambda_j$ , where  $(*)$  is a shorthand to indicate that sums or products run over those  $j$ 's in  $\{1, \dots, d\}$  for which  $\lambda_j \notin [1 - \delta, 1 + \delta]$ . Moreover, by assumption, there holds

$$\prod_{(*)} \lambda_j = \prod_{j=1}^d \lambda_j \cdot \left( \prod_{\substack{j \in \{1, \dots, d\} \\ \lambda_j \in [1-\delta, 1+\delta]}} \lambda_j \right)^{-1} \geq \frac{\varepsilon}{(1 + \delta)^d} =: p_0,$$

where  $p_0 \in (0, 1)$ . Lyapunov's inequality entails  $\frac{1}{d_2} \sum_{(*)} \lambda_j \leq (\frac{1}{d_2} \sum_{(*)} \lambda_j^2)^{1/2}$  and the GAM inequality gives  $\frac{1}{d_2} \sum_{(*)} \lambda_j \geq (\prod_{(*)} \lambda_j)^{1/d_2}$ , so that the combination of these inequalities yields  $p_0^{1/d_2} \leq (\frac{1}{d_2} \sum_{(*)} \lambda_j^2)^{1/2}$ . Hence, recalling that  $\log p_0 < 0$ , there holds

$$\begin{aligned} \sum_{(*)} (\lambda_j - 1 - \log \lambda_j) &\leq d_2 \left( \frac{1}{d_2} \sum_{(*)} \lambda_j^2 \right)^{1/2} - \log p_0 \\ &\leq \left[ \left( \frac{1}{p_0} \right)^{1/d_2} - \frac{1}{d_2} \left( \frac{1}{p_0} \right)^{2/d_2} \log p_0 \right] \sum_{(*)} \lambda_j^2. \end{aligned}$$

Since  $K_2(\delta) := \sup_{x \notin [1-\delta, 1+\delta] \setminus \{1\}} \frac{x^2}{(x-1)^2} > 0$  for any  $\delta > 0$ , one deduces the former inequality in (4.10). Finally, the latter inequality follows from the properties of the Frobenius norm, as shown, for example, Section 5.6 of [42]. □

The way is now paved for the proof of Theorem 2.2.

From a well-known expression of the distance  $d_{[\mathbb{R}^d]}^{(2)}$  between two multivariate Gaussian distributions (see, e.g., [30,49]), one gets

$$[d_{[\mathbb{R}^d]}^{(2)}(\mu_{\theta_0}, \mu_{\hat{\theta}_n})]^2 = |\mathbf{m}_0 - \hat{\mathbf{m}}_n|^2 + \text{tr}[V_0 + \hat{V}_n - 2(V_0 \hat{V}_n)^{1/2}], \tag{4.11}$$

where the symbol  $\mu_\theta$  is referred to (2.4), while  $\hat{\mathbf{m}}_n$  and  $\hat{V}_n$  are referred to (2.5). Then, to prove (1.2), one starts by analyzing the first summand on the above RHS. The combination of Lyapunov's inequality with (2.1) yields, for any  $r > 2$ ,

$$\begin{aligned} \mu_{\theta_0}^\infty \left[ \left( \sup_{n \geq 1} b_n |\mathbf{m}_0 - \hat{\mathbf{m}}_n| \right)^2 \right] &\leq \sum_{j=1}^d \mu_{\theta_0}^\infty \left[ \left( \sup_{n \geq 1} \frac{|\sum_{i=1}^n (\tilde{\xi}_i^{(j)} - m_0^{(j)})|}{\sqrt{n \text{Log}_2 n}} \right)^2 \right] \\ &\leq \sum_{j=1}^d \left\{ \mu_{\theta_0}^\infty \left[ \left( \sup_{n \geq 1} \frac{|\sum_{i=1}^n (\tilde{\xi}_i^{(j)} - m_0^{(j)})|}{\sqrt{n \text{Log}_2 n}} \right)^r \right] \right\}^{2/r} \\ &\leq \sum_{j=1}^d \sigma_j^2 \left\{ \left[ \alpha_0(r) + \alpha_1(r) \left( \frac{\mu_{\theta_0}^\infty [|\tilde{\xi}_1^{(j)} - m_0^{(j)}|^r]}{\sigma_j^r} \right)^{r-1} \right] \right\}^{2/r} \\ &= \left[ \alpha_0(r) + \alpha_1(r) \left( \frac{2^{r/2}}{\sqrt{\pi}} \Gamma \left( \frac{r+1}{2} \right) \right)^{r-1} \right]^{2/r} \text{tr}[V_0]. \end{aligned} \quad (4.12)$$

As to the latter summand on the RHS of (4.11), one can start by writing

$$\mu_{\theta_0}^\infty \left[ \sup_{n \in \{1, \dots, d\}} b_n^2 \text{tr}[V_0 + \hat{V}_n - 2(V_0 \hat{V}_n)^{1/2}] \right] \leq b_d^2 (d+1) \text{tr}[V_0]$$

since, for every  $n \in \mathbb{N}$ ,  $\mu_{\theta_0}^\infty[\text{tr}[\hat{V}_n]] \leq \text{tr}[V_0]$ . Consequently, one can confine oneself to studying  $\mu_{\theta_0}^\infty[\sup_{n \geq d+1} b_n^2 \text{tr}[V_0 + \hat{V}_n - 2(V_0 \hat{V}_n)^{1/2}]]$ . Now, it is worth observing that the term  $\text{tr}[V_0 + \hat{V}_n - 2(V_0 \hat{V}_n)^{1/2}]$  coincides with the squared distance  $[d_{[\mathbb{R}^d]}^{(2)}]^2$  between two  $d$ -dimensional Gaussian distributions with zero means and a covariance matrix equal to  $V_0$  or  $\hat{V}_n$ , respectively. This fact paves the way for the application of the Talagrand inequality (2.14), to get

$$\begin{aligned} \text{tr}[V_0 + \hat{V}_n - 2(V_0 \hat{V}_n)^{1/2}] &\leq C_T(\theta_0) [\text{tr}(V_0^{-1} \hat{V}_n - \text{Id}_d) - \log \det(V_0^{-1} \hat{V}_n)] \\ &= C_T(\theta_0) \sum_{j=1}^d [\lambda_j^{(n)} - \log \lambda_j^{(n)} - 1], \end{aligned} \quad (4.13)$$

where  $\lambda_1^{(n)}, \dots, \lambda_d^{(n)}$  denote the strictly positive (with  $\mu_{\theta_0}$ -probability 1) eigenvalues of  $V_0^{-1} \hat{V}_n$ . Note that, as a consequence of the original formulation of the Talagrand inequality (see [60]),  $C_T(\theta_0)$  can be chosen equal to  $2\|V_0\|_F$ . At this stage, after fixing  $\varepsilon \in (0, 1)$ , an application of Lemma 4.1 gives

$$\begin{aligned} \mu_{\theta_0}^\infty \left[ \sup_{n \geq d+1} b_n^2 \text{tr}[V_0 + \hat{V}_n - 2(V_0 \hat{V}_n)^{1/2}] \mathbb{1}\{\det(V_0^{-1} \hat{V}_n) \geq \varepsilon\} \right] \\ \leq 2K(\varepsilon, d) \|V_0^{-1}\|_F \cdot \sum_{l,h=1}^d \mu_{\theta_0}^\infty \left[ \left( \sup_{n \geq d+1} \frac{|\sum_{i=1}^n (\tilde{\eta}_i^{(l,h)} - (V_0)_{l,h})|}{\sqrt{n \text{Log}_2 n}} \right)^2 \right], \end{aligned} \quad (4.14)$$

where  $\tilde{\eta}_i^{(l,h)} := (\tilde{\xi}_i^{(l)} - \hat{m}_n^{(l)})(\tilde{\xi}_i^{(h)} - \hat{m}_n^{(h)})$ . Elementary algebra shows that  $\sum_{i=1}^n (\tilde{\eta}_i^{(l,l)} - (V_0)_{l,l}) = \sum_{i=1}^n [(\tilde{\xi}_i^{(l)} - m_0^{(l)})^2 - (V_0)_{l,l}] - \frac{1}{n} (\sum_{i=1}^n [\tilde{\xi}_i^{(l)} - m_0^{(l)}])^2$  and, for  $l \neq h$ ,  $\sum_{i=1}^n (\tilde{\eta}_i^{(l,h)} - (V_0)_{l,h}) = \sum_{i=1}^n [(\tilde{\xi}_i^{(l)} - m_0^{(l)})(\tilde{\xi}_i^{(h)} - m_0^{(h)}) - (V_0)_{l,h}] - \frac{1}{n} (\sum_{i=1}^n [\tilde{\xi}_i^{(l)} - m_0^{(l)}]) (\sum_{i=1}^n [\tilde{\xi}_i^{(h)} - m_0^{(h)}])$ . Whence,

$$\begin{aligned} & \mu_{\theta_0}^\infty \left[ \left( \sup_{n \geq d+1} \frac{|\sum_{i=1}^n (\tilde{\eta}_i^{(l,l)} - (V_0)_{l,l})|}{\sqrt{n \text{Log}_2(n)}} \right)^2 \right] \\ & \leq 2\mu_{\theta_0}^\infty \left[ \left( \sup_{n \geq d+1} \frac{|\sum_{i=1}^n [(\tilde{\xi}_i^{(l)} - m_0^{(l)})^2 - (V_0)_{l,l}]|}{\sqrt{n \text{Log}_2(n)}} \right)^2 \right] \\ & \quad + 2\mu_{\theta_0}^\infty \left[ \sup_{n \geq d+1} \frac{(\sum_{i=1}^n [\tilde{\xi}_i^{(l)} - m_0^{(l)}])^4}{n^3 \text{Log}_2(n)} \right] \end{aligned} \tag{4.15}$$

and, for  $l \neq h$ ,

$$\begin{aligned} & \mu_{\theta_0}^\infty \left[ \left( \sup_{n \geq d+1} \frac{|\sum_{i=1}^n (\tilde{\eta}_i^{(l,h)} - (V_0)_{l,h})|}{\sqrt{n \text{Log}_2(n)}} \right)^2 \right] \\ & \leq 2\mu_{\theta_0}^\infty \left[ \left( \sup_{n \geq d+1} \frac{|\sum_{i=1}^n [(\tilde{\xi}_i^{(l)} - m_0^{(l)})(\tilde{\xi}_i^{(h)} - m_0^{(h)}) - (V_0)_{l,h}]|}{\sqrt{n \text{Log}_2(n)}} \right)^2 \right] \\ & \quad + 2\mu_{\theta_0}^\infty \left[ \sup_{n \geq d+1} \frac{(\sum_{i=1}^n [\tilde{\xi}_i^{(l)} - m_0^{(l)}])^2 (\sum_{i=1}^n [\tilde{\xi}_i^{(h)} - m_0^{(h)}])^2}{n^3 \text{Log}_2(n)} \right]. \end{aligned} \tag{4.16}$$

As to the first summand on the RHS of (4.15), it is enough to notice that  $\{(V_0)_{l,l}^{-1}(\tilde{\xi}_i^{(l)} - m_0^{(l)})^2\}_{i \geq 1}$  is a sequence of i.i.d.,  $\chi^2(1)$ -distributed, real random variables, so that an application of Proposition 2.1 with  $r = 4$ , in combination with the Lyapunov inequality, yields

$$\mu_{\theta_0}^\infty \left[ \left( \sup_{n \geq d+1} \frac{|\sum_{i=1}^n [(\tilde{\xi}_i^{(l)} - m_0^{(l)})^2 - (V_0)_{l,l}]|}{\sqrt{n \text{Log}_2(n)}} \right)^2 \right] \leq 2(V_0)_{l,l}^2 [\alpha_0(4) + 15^4 \alpha_1(4)]^{1/2}.$$

For the second summand on the RHS of (4.15), observe that

$$\frac{(\sum_{i=1}^n [\tilde{\xi}_i^{(l)} - m_0^{(l)}])^4}{n^3 \text{Log}_2(n)} = \left( \frac{|\sum_{i=1}^n [\tilde{\xi}_i^{(l)} - m_0^{(l)}]|}{\sqrt{n \text{Log}_2(n)}} \right)^4 \frac{[n \text{Log}_2(n)]^2}{n^3 \text{Log}_2(n)} \leq \left( \frac{|\sum_{i=1}^n [\tilde{\xi}_i^{(l)} - m_0^{(l)}]|}{\sqrt{n \text{Log}_2(n)}} \right)^4$$

yielding, by virtue of Proposition 2.1,

$$\mu_{\theta_0}^\infty \left[ \sup_{n \geq d+1} \frac{(\sum_{i=1}^n [\tilde{\xi}_i^{(l)} - m_0^{(l)}])^4}{n^3 \text{Log}_2(n)} \right] \leq (V_0)_{l,l}^2 [\alpha_0(4) + 3\alpha_1(4)].$$

To bound the first summand on the RHS of (4.16), introduce, as a preliminary step, the auxiliary function  $\omega_k(\rho) := \mathbb{E}[(XY - \rho)^k]$ , where  $k \in \mathbb{N}$  and  $(X, Y) \in \mathbb{R}^2$  is a Gaussian random vector with

zero means, unit variances and correlation coefficient  $\rho$ . Then, apply again Proposition 2.1 with  $r = 4$ , in combination with the Lyapunov inequality, and observe that  $1 \leq \omega_2(\rho) = 1 + \rho^2 \leq 2$   $\omega_4(\rho) = 9 - 33\rho^2 + 84\rho^4 \leq 60$  lead to

$$\begin{aligned} & \mu_{\theta_0}^\infty \left[ \left( \sup_{n \geq d+1} \frac{|\sum_{i=1}^n [\tilde{\xi}_i^{(l)} - m_0^{(l)}](\tilde{\xi}_i^{(h)} - m_0^{(h)}) - (V_0)_{l,h}|}{\sqrt{n \text{Log}_2(n)}} \right)^2 \right] \\ & \leq (V_0)_{l,l}(V_0)_{h,h}\omega_2(\rho_{l,h}) \left[ \alpha_0(4) + \alpha_1(4) \left( \frac{\omega_4(\rho_{l,h})}{\omega_2(\rho_{l,h})} \right)^4 \right]^{1/2} \\ & \leq \sqrt{\alpha_0(4) + 60^4 \alpha_1(4)} [(V_0)_{l,l}^2 + (V_0)_{h,h}^2], \end{aligned}$$

where  $\rho_{l,h} := (V_0)_{l,h} / \sqrt{(V_0)_{l,l}(V_0)_{h,h}}$ . Finally, for the second summand on the RHS of (4.16), it is enough to notice that

$$\begin{aligned} & \left( \sum_{i=1}^n [\tilde{\xi}_i^{(l)} - m_0^{(l)}] \right)^2 \cdot \left( \sum_{i=1}^n [\tilde{\xi}_i^{(h)} - m_0^{(h)}] \right)^2 \\ & \leq \frac{1}{2} \left( \sum_{i=1}^n [\tilde{\xi}_i^{(l)} - m_0^{(l)}] \right)^4 + \frac{1}{2} \left( \sum_{i=1}^n [\tilde{\xi}_i^{(h)} - m_0^{(h)}] \right)^4, \end{aligned}$$

so that, by resorting to Proposition 2.1 with  $r = 4$ , one gets

$$\begin{aligned} & \mu_{\theta_0}^\infty \left[ \sup_{n \geq 1} \frac{(\sum_{i=1}^n [\tilde{\xi}_i^{(l)} - m_0^{(l)}])^2 (\sum_{i=1}^n [\tilde{\xi}_i^{(h)} - m_0^{(h)}])^2}{n^3 \text{Log}_2(n)} \right] \\ & \leq \frac{1}{2} [(V_0)_{l,l}^2 + (V_0)_{h,h}^2] \cdot [\alpha_0(4) + 3\alpha_1(4)]. \end{aligned}$$

At this stage, there are all the elements to deduce that

$$\mu_{\theta_0}^\infty \left[ \sup_{n \geq d+1} b_n^2 \text{tr}[V_0 + \hat{V}_n - 2(V_0 \hat{V}_n)^{1/2}] \mathbb{1}\{\det(V_0^{-1} \hat{V}_n) \geq \varepsilon\} \right] \leq c_* K(\varepsilon, d) \|V_0\|_F \quad (4.17)$$

for a suitable numerical constant  $c_*$  independent of  $V_0$  and even of the dimension  $d$ .

It remains to analyze  $\mu_{\theta_0}^\infty[\sup_{n \geq d+1} b_n^2 \text{tr}[V_0 + \hat{V}_n - 2(V_0 \hat{V}_n)^{1/2}] \mathbb{1}\{\det(V_0^{-1} \hat{V}_n) < \varepsilon\}]$  which, in view of the Boole inequality, can be bounded from above by

$$\sum_{n \geq d+1} b_n^2 \int_0^{+\infty} \mu_{\theta_0}^\infty(\{\text{tr}[V_0 + \hat{V}_n - 2(V_0 \hat{V}_n)^{1/2}] \mathbb{1}\{\det(V_0^{-1} \hat{V}_n) < \varepsilon\} > z\}) dz.$$

Preliminarily, given any  $A > 0$ , one gets

$$\int_0^A \mu_{\theta_0}^\infty(\{\text{tr}[V_0 + \hat{V}_n - 2(V_0 \hat{V}_n)^{1/2}] \mathbb{1}\{\det(V_0^{-1} \hat{V}_n) < \varepsilon\} > z\}) dz \leq A \mu_{\theta_0}^\infty(\{\det(V_0^{-1} \hat{V}_n) < \varepsilon\})$$

for every  $n \in \mathbb{N}$ . Now, the series  $\sum_{n \geq d+1} b_n^2 \mu_{\theta_0}^\infty(\{\det(V_0^{-1} \hat{V}_n) < \varepsilon\})$  turns out to be convergent for any choice of  $\varepsilon \in (0, 1)$  since, in this case, one can prove that  $\mu_{\theta_0}^\infty(\{\det(V_0^{-1} \hat{V}_n) < \varepsilon\})$  goes to zero exponentially fast, with respect to  $n$ . Indeed, a well-known result concerning the Wishart distribution (see, e.g., Theorem 7.5.3 in [5]) states that, for  $n > d$ ,  $\det(V_0^{-1} \hat{V}_n)$  has the same distribution (under  $\mu_{\theta_0}^\infty$ ) as the product of  $d$  independent real random variables, say  $Y_1, \dots, Y_d$ , where  $Y_i$  has a Gamma distribution with scale parameter  $\frac{n}{2}$  and shape parameter  $\frac{n-i}{2}$ . Since  $\mu_{\theta_0}^\infty(\{\det(V_0^{-1} \hat{V}_n) < \varepsilon\}) \leq \sum_{i=1}^d \mu_{\theta_0}^\infty(\{Y_i < \varepsilon^{1/d}\})$  holds in view of the Boole inequality, then one can resort to the well-known Chernoff bounds to obtain

$$\mu_{\theta_0}^\infty(\{Y_i < \varepsilon^{1/d}\}) \leq \exp\left\{\frac{n-i}{2} \left[1 - \frac{n\varepsilon^{1/d}}{n-i} + \log\left(\frac{n\varepsilon^{1/d}}{n-i}\right)\right]\right\}$$

provided that  $\varepsilon^{1/d} < \frac{n-i}{n}$  for any  $i \in \{1, \dots, d\}$  and  $n \geq d+1$ . Since  $\sup_{i \leq d, n \geq d+1} (\frac{n}{n-i}) \leq d+1$ , choosing  $\varepsilon \leq [2(d+1)]^{-d}$  yields  $1 - \frac{n\varepsilon^{1/d}}{n-i} + \log(\frac{n\varepsilon^{1/d}}{n-i}) \leq \frac{1}{2} - \log 2 < 0$  and, hence,

$$\mu_{\theta_0}^\infty(\{\det(V_0^{-1} \hat{V}_n) < \varepsilon\}) \leq d \exp\left\{\frac{n-d}{2} (1/2 - \log 2)\right\} \tag{4.18}$$

for  $n \geq d+1$ . Now, one studies  $\int_A^{+\infty} \mu_{\theta_0}^\infty(\{\mathbb{1}\{\det(V_0^{-1} \hat{V}_n) < \varepsilon\} \sum_{j=1}^d [\lambda_j^{(n)} - \log \lambda_j^{(n)} - 1] > z\}) dz$  through a splitting of the above integrand into the sum of the following two terms:

$$\mu_{\theta_0}^\infty\left(\left\{\mathbb{1}\{\det(V_0^{-1} \hat{V}_n) < \varepsilon, \text{tr}(V_0^{-1} \hat{V}_n) > -\eta \log(\det(V_0^{-1} \hat{V}_n))\} \times \sum_{j=1}^d [\lambda_j^{(n)} - \log \lambda_j^{(n)} - 1] > z/2\right\}\right)$$

and

$$\mu_{\theta_0}^\infty\left(\left\{\mathbb{1}\{\det(V_0^{-1} \hat{V}_n) < \varepsilon, \text{tr}(V_0^{-1} \hat{V}_n) \leq -\eta \log(\det(V_0^{-1} \hat{V}_n))\} \times \sum_{j=1}^d [\lambda_j^{(n)} - \log \lambda_j^{(n)} - 1] > z/2\right\}\right)$$

with  $\eta > 0$ . The former can be bounded by  $\mu_{\theta_0}^\infty(\{(1+1/\eta) \text{tr}(V_0^{-1} \hat{V}_n) > z/2\})$  while the latter can be bounded by  $\mu_{\theta_0}^\infty(\{\det(V_0^{-1} \hat{V}_n) < \varepsilon, \det(V_0^{-1} \hat{V}_n) < \exp\{-\frac{z}{2(1+\eta)}\}\})$ . At this stage, recall Theorem 7.3.5 in [5], which states that  $\text{tr}(V_0^{-1} \hat{V}_n)$  has a Gamma distribution with scale parameter

$\frac{n}{2}$  and shape parameter  $\frac{(n-1)d}{2}$ , and apply again the Chernoff bounds to get

$$\begin{aligned} &\mu_{\theta_0}^\infty \left( \left\{ \text{tr}(V_0^{-1} \hat{V}_n) > \frac{z\eta}{2(1+\eta)} \right\} \right) \\ &\leq \exp \left\{ \frac{(n-1)d}{2} \left[ 1 - \frac{zn\eta}{2(1+\eta)(n-1)d} + \log \left( \frac{zn\eta}{2(1+\eta)(n-1)d} \right) \right] \right\} \end{aligned}$$

provided that  $\frac{zn\eta}{2(1+\eta)} > \frac{(n-1)d}{n}$  holds for any  $z \geq A$  and  $n \geq d + 1$ . In particular, this condition is in force if  $A \geq 4d(1 + 1/\eta)$ . In such a case, one has

$$\left[ 1 - \frac{zn\eta}{2(1+\eta)(n-1)d} + \log \left( \frac{zn\eta}{2(1+\eta)(n-1)d} \right) \right] \leq -\frac{(e-2)\eta}{4de(1+\eta)}z$$

for any  $z \geq A$  and  $n \geq d + 1$ . Whence,

$$\begin{aligned} &\int_A^{+\infty} \mu_{\theta_0}^\infty \left( \left\{ \text{tr}(V_0^{-1} \hat{V}_n) > \frac{z\eta}{2(1+\eta)} \right\} \right) dz \\ &\leq \frac{8e(1+\eta)}{\eta(e-2)(n-1)} \exp \left\{ -\frac{\eta(e-2)A}{8e(1+\eta)}(n-1) \right\}. \end{aligned} \tag{4.19}$$

Apropos of  $\mu_{\theta_0}^\infty(\{\det(V_0^{-1} \hat{V}_n) < \varepsilon, \det(V_0^{-1} \hat{V}_n) < \exp\{-\frac{z}{2(1+\eta)}\}\})$ , taking  $A \geq -2(1+\eta) \log \varepsilon$  implies that the probability at issue coincides with  $\mu_{\theta_0}^\infty(\{\det(V_0^{-1} \hat{V}_n) < \exp\{-\frac{z}{2(1+\eta)}\}\})$ , for all  $z \geq A$ . Therefore, the same reasoning that led to (4.18) shows that

$$\begin{aligned} &\mu_{\theta_0}^\infty \left( \left\{ \det(V_0^{-1} \hat{V}_n) < \exp \left\{ -\frac{z}{2(1+\eta)} \right\} \right\} \right) \\ &\leq d \exp \left\{ \frac{n-d}{2} \left[ 1 + \log(d+1) - (d+1) \exp \left\{ -\frac{z}{2d(1+\eta)} \right\} - \frac{z}{2d(1+\eta)} \right] \right\}. \end{aligned}$$

Now, recalling that  $A$  has been chosen in such a way that,  $A \geq 4d(1 + 1/\eta)$ , one gets

$$\begin{aligned} &1 + \log(d+1) - (d+1) \exp \left\{ -\frac{z}{2d(1+\eta)} \right\} - \frac{z}{2d(1+\eta)} \\ &\leq \left[ \frac{(1 + \log(d+1))\eta}{4d(1+\eta)} - \frac{1}{2d(1+\eta)} \right] z \end{aligned}$$

which, upon choosing  $\eta \leq 1/(1 + \log(d + 1))$ , leads to

$$\begin{aligned} &\int_A^{+\infty} \mu_{\theta_0}^\infty \left( \left\{ \det(V_0^{-1} \hat{V}_n) < \exp \left\{ -\frac{z}{2(1+\eta)} \right\} \right\} \right) dz \\ &\leq \frac{8d^2(1+\eta)}{(n-d)} \exp \left\{ -\frac{A}{8d(1+\eta)}(n-d) \right\}. \end{aligned} \tag{4.20}$$

In conclusion, for any  $\varepsilon \leq [2(d + 1)]^{-d}$ , it is possible to combine (4.18)–(4.20) and optimize, with respect to the choices of  $A$  and  $\eta$ , to deduce the existence of a positive constant  $H(\varepsilon, d)$  such that

$$\sum_{n \geq d+1} b_n^2 \int_0^{+\infty} \mu_{\theta_0}^\infty(\{\text{tr}[V_0 + \hat{V}_n - 2(V_0 \hat{V}_n)^{1/2}] \mathbb{1}\{\det(V_0^{-1} \hat{V}_n) < \varepsilon\} > z\}) dz \leq H(\varepsilon, d) \tag{4.21}$$

concluding the proof of (1.2), along with the determination of  $C_2(\mathbf{p})$  in (2.6).

As for the validity of (1.3), one considers again (4.11) and starts by applying the  $d$ -dimensional version of the Hartman–Wintner LIL (see, e.g., in Theorem 3.1 of [36]) to write

$$p_0^\infty \left( \left\{ \limsup_{n \rightarrow +\infty} b_n |\mathbf{m}_0 - \hat{\mathbf{m}}_n| = \sqrt{2} \sigma_{\max}(V_0) \right\} \right) = 1,$$

where  $\sigma_{\max}^2(V_0)$  stands for the largest eigenvalue of  $V_0$ . To deal with the latter summand on the right of (4.11), one resorts again to the combination of (4.13) with (4.10) to obtain

$$\begin{aligned} \sqrt{\text{tr}[V_0 + \hat{V}_n - 2(V_0 \hat{V}_n)^{1/2}] \mathbb{1}\{\det(V_0^{-1} \hat{V}_n) \geq \varepsilon\}} &\leq \sqrt{2K(\varepsilon, d) \|V_0\|_F} \cdot \|V_0^{-1} \hat{V}_n - \text{Id}_d\|_F \\ &\leq \sqrt{2K(\varepsilon, d) \|V_0^{-1}\|_F} \cdot \|\hat{V}_n - V_0\|_F \end{aligned}$$

for any  $\varepsilon \in (0, 1)$ . Then, observe that  $b_n \|\hat{V}_n - V_0\|_F \leq \sum_{l,h=1}^d \frac{|\sum_{i=1}^n [\tilde{\eta}_i^{(l,h)} - (V_0)_{l,h}]|}{\sqrt{n \text{Log}_2(n)}}$ , by elementary algebra. At this stage, after recalling that

$$\begin{aligned} \sum_{i=1}^n [\tilde{\eta}_i^{(l,h)} - (V_0)_{l,h}] &= \sum_{i=1}^n [(\tilde{\xi}_i^{(l)} - m_0^{(l)})(\tilde{\xi}_i^{(h)} - m_0^{(h)}) - (V_0)_{l,h}] \\ &\quad - \frac{1}{n} \left( \sum_{i=1}^n [\tilde{\xi}_i^{(l)} - m_0^{(l)}] \right) \left( \sum_{i=1}^n [\tilde{\xi}_i^{(h)} - m_0^{(h)}] \right), \end{aligned}$$

apply the classical Hartman–Wintner LIL to conclude that

$$p_0^\infty \left( \left\{ \limsup_{n \rightarrow +\infty} \frac{|\sum_{i=1}^n [\tilde{\eta}_i^{(l,h)} - (V_0)_{l,h}]|}{\sqrt{n \text{Log}_2(n)}} = \sigma_l \sigma_h \sqrt{2\omega_2(\rho_{l,h})} \right\} \right) = 1$$

since

$$p_0^\infty \left( \left\{ \limsup_{n \rightarrow +\infty} \frac{|(\sum_{i=1}^n [\tilde{\xi}_i^{(l)} - m_0^{(l)}])(\sum_{i=1}^n [\tilde{\xi}_i^{(h)} - m_0^{(h)}])|}{n \sqrt{n \text{Log}_2(n)}} = 0 \right\} \right) = 1.$$

Noticing that  $\sum_{l,h=1}^d \sigma_l \sigma_h \sqrt{2\omega_2(\rho_{l,h})} \leq d \text{tr}[V_0]$  and that  $\sqrt{\|V_0^{-1}\|_F} \cdot \text{tr}[V_0] \leq \sqrt{\text{tr}[V_0]}$ , one can conclude the proof after taking into account that  $p_0^\infty(\{\limsup_{n \rightarrow +\infty} \mathbb{1}\{\det(V_0^{-1} \hat{V}_n) \geq \varepsilon\} = 1\}) = 1$ , by strong consistency of  $\hat{V}_n$ .

### 4.4. Proof of Theorem 2.3

From the definition of the Kullback–Leibler information, one gets

$$\begin{aligned} \mathbb{K}(\mu_{\hat{\theta}_n} \mid \mu_{\theta_0}) &= \hat{\theta}_n \cdot [V^{-1}(\hat{\theta}_n) - V^{-1}(\theta_0)] - [M(V^{-1}(\hat{\theta}_n)) - M(V^{-1}(\theta_0))] \\ &= \int_0^1 (1-s) {}^t(\hat{\theta}_n - \theta_0) \text{Hess}[I_\mu](\theta_0 + s(\hat{\theta}_n - \theta_0))(\hat{\theta}_n - \theta_0) ds, \end{aligned}$$

where the second identity follows from the combination of (2.10) with Bernstein’s representation of the remainder term in Taylor formula. Then, from (2.14),

$$\begin{aligned} d_{[\mathbb{S}]}^{(2)}(\mu_{\theta_0}, \mu_{\hat{\theta}_n}) &\leq \sqrt{C_T(\theta_0) \int_0^1 (1-s) {}^t(\hat{\theta}_n - \theta_0) \text{Hess}[I_\mu](\theta_0 + s(\hat{\theta}_n - \theta_0))(\hat{\theta}_n - \theta_0) ds} \\ &\leq \sqrt{C_T(\theta_0)} |\hat{\theta}_n - \theta_0| \Phi(\hat{\theta}_n) \end{aligned}$$

and hence, in view of the  $k$ -dimensional version of the Hartman–Wintner LIL,

$$\limsup_{n \rightarrow +\infty} \sqrt{\frac{n}{\text{Log}_2(n)}} d_{[\mathbb{S}]}^{(2)}(\mu_{\theta_0}, \mu_{\hat{\theta}_n}) \leq \sqrt{2C_T(\theta_0)} \Phi(\theta_0) \sigma_{\max}(\theta_0), \tag{4.22}$$

where  $\sigma_{\max}^2(\theta_0)$  is the largest eigenvalue of  $\text{Hess}[M](V^{-1}(\theta_0))$ . See, e.g., Theorem 3.1 in [36]. This proves (1.3) with  $b_n = (n/\text{Log}_2(n))^{1/2}$  and  $Y(\mathfrak{p}_0) = \sqrt{2C_T(\theta_0)} \Phi(\theta_0) \sigma_{\max}(\theta_0)$ .

To prove (1.2), define  $\Theta_i := \{\theta \in \mathbb{R}^k \mid |\theta - \theta_0| \leq \delta(\theta_0)\}$  and  $\Theta_e := \Theta \setminus \Theta_i$ , where  $\delta(\theta_0)$  is chosen so that  $\Theta_i$  is a proper subset of  $\Theta$ , and notice that

$$\begin{aligned} \sup_{n \geq 1} b_n |\hat{\theta}_n - \theta_0| \Phi(\hat{\theta}_n) &\leq \sup_{n \geq 1} b_n |\hat{\theta}_n - \theta_0| \Phi(\hat{\theta}_n) \mathbb{1}\{\hat{\theta}_n \in \Theta_i\} \\ &\quad + \sup_{n \geq 1} b_n |\hat{\theta}_n - \theta_0| \Phi(\hat{\theta}_n) \mathbb{1}\{\hat{\theta}_n \in \Theta_e\}. \end{aligned} \tag{4.23}$$

As to the first summand on the right-hand side, for any  $r > 2$ , one can write

$$\mu_{\theta_0}^\infty \left[ \sup_{n \geq 1} b_n |\hat{\theta}_n - \theta_0| \Phi(\hat{\theta}_n) \mathbb{1}\{\hat{\theta}_n \in \Theta_i\} \right] \leq \max_{\eta \in \Theta_i} \Phi(\eta) \sum_{j=1}^k \left( \mu_{\theta_0}^\infty \left[ \sup_{n \geq 1} b_n |\hat{\theta}_n^{(j)} - \theta_0^{(j)}|^r \right] \right)^{1/r},$$

where  $\hat{\theta}_n^{(j)}$  and  $\theta_0^{(j)}$  stand for the  $j$ th component of  $\hat{\theta}_n$  and  $\theta_0$ , respectively. At this stage, invoke Proposition 2.1 to conclude that

$$\mu_{\theta_0}^\infty \left[ \sup_{n \geq 1} b_n |\hat{\theta}_n - \theta_0| \Phi(\hat{\theta}_n) \mathbb{1}\{\hat{\theta}_n \in \Theta_i\} \right] \tag{4.24}$$

$$\leq \max_{\eta \in \Theta_i} \Phi(\eta) \sum_{j=1}^k \sigma_j(\theta_0) \left[ \alpha_0(r) + \alpha_1(r) \left( \frac{\mu_{\theta_0}^\infty [ |t^{(j)}(\tilde{\xi}_1) - \theta_0^{(j)}|^r ]}{\sigma_j(\theta_0)^r} \right)^{r} \right]^{1/r}, \tag{4.25}$$

where  $\sigma_j(\boldsymbol{\theta}_0) := \sqrt{\partial_{j,j}^2 M(V^{-1}(\boldsymbol{\theta}_0))}$  and  $t^{(j)}(\tilde{\xi}_1)$  stands for the  $j$ th component of  $\mathbf{t}(\tilde{\xi}_1)$ . This upper bound represents a first contribution to the determination of  $C(\mathfrak{p}_0)$ , to be now completed by bounding the second summand on the right-hand side of (4.23). Apropos of this, one can start with the following general considerations, valid for any  $n_0 \in \mathbb{N}$ :

$$\begin{aligned} & \mu_{\boldsymbol{\theta}_0}^\infty \left[ \sup_{n \geq n_0} b_n |\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0| \Phi(\hat{\boldsymbol{\theta}}_n) \mathbb{1}\{\hat{\boldsymbol{\theta}}_n \in \Theta_e\} \right] \\ &= \int_0^{+\infty} \mu_{\boldsymbol{\theta}_0}^\infty \left[ \bigcup_{n \geq n_0} \{b_n |\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0| \Phi(\hat{\boldsymbol{\theta}}_n) \mathbb{1}\{\hat{\boldsymbol{\theta}}_n \in \Theta_e\} > t\} \right] dt \\ &\leq \sum_{n \geq n_0} b_n \int_0^{\tau(\boldsymbol{\theta}_0)} \mu_{\boldsymbol{\theta}_0}^\infty [|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0| \Phi(\hat{\boldsymbol{\theta}}_n) \mathbb{1}\{\hat{\boldsymbol{\theta}}_n \in \Theta_e\} > t] dt \end{aligned} \tag{4.26}$$

$$+ \sum_{n \geq n_0} b_n \int_{\tau(\boldsymbol{\theta}_0)}^{+\infty} \mu_{\boldsymbol{\theta}_0}^\infty [|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0| \Phi(\hat{\boldsymbol{\theta}}_n) \mathbb{1}\{\hat{\boldsymbol{\theta}}_n \in \Theta_e\} > t] dt. \tag{4.27}$$

The term in (4.26) can be bounded thanks to the fact that  $\{|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0| \Phi(\hat{\boldsymbol{\theta}}_n) \mathbb{1}\{\hat{\boldsymbol{\theta}}_n \in \Theta_e\} > t\} \subseteq \{\hat{\boldsymbol{\theta}}_n \in \Theta_e\}$ , yielding

$$\begin{aligned} & \sum_{n \geq n_0} b_n \int_0^{\tau(\boldsymbol{\theta}_0)} \mu_{\boldsymbol{\theta}_0}^\infty [|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0| \Phi(\hat{\boldsymbol{\theta}}_n) \mathbb{1}\{\hat{\boldsymbol{\theta}}_n \in \Theta_e\} > t] dt \\ & \leq \tau(\boldsymbol{\theta}_0) \sum_{n \geq n_0} b_n \mu_{\boldsymbol{\theta}_0}^\infty [|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0| > \delta(\boldsymbol{\theta}_0)]. \end{aligned}$$

Here and in other points of the present proof, some results in [68] are invoked to estimate tail probabilities of the type of  $\mu_{\boldsymbol{\theta}_0}^\infty [|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0| > a]$ . More precisely, the analysis of (4.26) is based on the corollary on page 491 of that paper, whose applicability to the present context follows after checking (2.1) therein. Since vectors  $\mathbf{t}(\tilde{\xi}_j) - \boldsymbol{\theta}_0$  correspond to vectors  $\boldsymbol{\xi}_j$  in [68], in the place of  $E[|\boldsymbol{\xi}_j|^m]$  therein one considers

$$\mu_{\boldsymbol{\theta}_0}^\infty [|\mathbf{t}(\tilde{\xi}_j) - \boldsymbol{\theta}_0|^m] \leq k^{m/2-1} \sum_{i=1}^k \mu_{\boldsymbol{\theta}_0}^\infty [ |t^{(i)}(\tilde{\xi}_j) - \theta_0^{(i)}|^m ].$$

Thanks to the existence of the moment generating function of  $\mathbf{t}(\tilde{\xi}_j) - \boldsymbol{\theta}_0$ , the classical Cauchy estimate shows that

$$\mu_{\boldsymbol{\theta}_0}^\infty [ |t^{(i)}(\tilde{\xi}_j) - \theta_0^{(i)}|^m ] \leq m! \frac{C(r(\boldsymbol{\theta}_0))}{r(\boldsymbol{\theta}_0)^m}$$

is valid for all  $m \in \mathbb{N}$  with a suitable chosen  $r(\boldsymbol{\theta}_0) > 0$ , provided that  $C(r(\boldsymbol{\theta}_0))$  is the maximum (as  $i = 1, \dots, k$ ) of the maximum modulus of the moment generating function of  $|t^{(i)}(\tilde{\xi}_j) - \theta_0^{(i)}|$  on  $\{z \in \mathbb{C} \mid |z| = r(\boldsymbol{\theta}_0)\}$ . Plainly, the last inequality entails (2.1) in [68] with specific  $H = H(\boldsymbol{\theta}_0)$

and  $b_j^2 = \mu_{\theta_0}^\infty[|\mathbf{t}(\tilde{\xi}_j) - \theta_0|^2] =: B(\theta_0)^2$ . Putting  $B_n^2 := nB^2(\theta_0)$  and  $x := \sqrt{n}\delta(\theta_0)/B(\theta_0)$ , the thesis of the aforesaid corollary yields

$$\begin{aligned} & \sum_{n \geq n_0} b_n \mu_{\theta_0}^\infty[|\hat{\theta}_n - \theta_0| > \delta(\theta_0)] \\ & \leq \sum_{n \geq n_0} b_n \exp\left\{ \frac{-n\delta(\theta_0)^2}{B(\theta_0)^2 + 1, 62\delta(\theta_0)H(\theta_0)} \right\} \\ & \leq c \left( \frac{\delta(\theta_0)^2}{B(\theta_0)^2 + 1, 62\delta(\theta_0)H(\theta_0)} \right)^{-2} \exp\left\{ \frac{-n_0\delta(\theta_0)^2}{B(\theta_0)^2 + 1, 62\delta(\theta_0)H(\theta_0)} \right\} \end{aligned} \quad (4.28)$$

as a consequence of the elementary inequality

$$\sum_{n \geq n_0} b_n e^{-nz} \leq \frac{c}{z^2} e^{-n_0 z} \quad (4.29)$$

valid for  $z > 0$  with some constant  $c > 0$ . Estimate (4.28) provides the required upper bound for (4.26). As for (4.27), start by writing

$$\begin{aligned} & \mu_{\theta_0}^\infty[|\hat{\theta}_n - \theta_0| \Phi(\hat{\theta}_n) > t] \\ & \leq \mu_{\theta_0}^\infty[|\hat{\theta}_n - \theta_0| \Phi(\hat{\theta}_n) > t, \Phi(\hat{\theta}_n) \leq \sigma(t)] \\ & \quad + \mu_{\theta_0}^\infty[|\hat{\theta}_n - \theta_0| \Phi(\hat{\theta}_n) > t, \Phi(\hat{\theta}_n) > \sigma(t)] \\ & \leq \mu_{\theta_0}^\infty[|\hat{\theta}_n - \theta_0| > t/\sigma(t)] + \mu_{\theta_0}^\infty[\Phi(\hat{\theta}_n) > \sigma(t)] \\ & \leq \mu_{\theta_0}^\infty[|\hat{\theta}_n - \theta_0| > t/\sigma(t)] + \mu_{\theta_0}^\infty[\Phi(\hat{\theta}_n) > \sigma(t), |\hat{\theta}_n - \theta_0| \leq \rho(t)] \\ & \quad + \mu_{\theta_0}^\infty[|\hat{\theta}_n - \theta_0| > \rho(t)] \\ & \leq \mu_{\theta_0}^\infty[|\hat{\theta}_n - \theta_0| > t/\sigma(t)] + \mu_{\theta_0}^\infty\left[\Phi_1(|\hat{\theta}_n - \theta_0|) > \frac{\sigma(t)}{2}\right] \\ & \quad + \mu_{\theta_0}^\infty\left[\Phi_2(\hat{\theta}_n) > \frac{\sigma(t)}{2}, |\hat{\theta}_n - \theta_0| \leq \rho(t)\right] + \mu_{\theta_0}^\infty[|\hat{\theta}_n - \theta_0| > \rho(t)] \end{aligned}$$

and, after introducing the function  $m(t) := \min\{\rho(t), \frac{t}{\sigma(t)}, \Phi_1^{-1}(\frac{\sigma(t)}{2})\}$ , the sum of the first, second and fourth term can be bounded by

$$3 \exp\left\{ \frac{-nm(t)^2}{B(\theta_0)^2 + 1, 62m(t)H(\theta_0)} \right\} \leq 3 \exp\left\{ \frac{-nm(t)}{2H(\theta_0)} \right\} \quad (t > \tau(\theta_0))$$

invoking once again the Corollary in [68] with  $x := \sqrt{nm(t)}/B(\theta_0)$ . The validity of the above inequality rests on the fact that, by properly choosing  $\rho$  and  $\sigma$ ,  $m$  turns out to be monotone and

diverging at infinity. Now, from (4.29),

$$\sum_{n \geq n_0} 3b_n \exp\left\{\frac{-nm(t)}{2H(\theta_0)}\right\} \leq 3c \left(\frac{m(t)}{2H(\theta_0)}\right)^{-2} \exp\left\{\frac{-n_0m(t)}{2H(\theta_0)}\right\} \quad (t > \tau(\theta_0))$$

and one finds that the right-hand side is integrable on  $[\tau(\theta_0), +\infty)$ , thanks to (2.16), after a proper choice of  $n_0$ . At this stage, it remains only to bound  $\mu_{\theta_0}^\infty[\Phi_2(\hat{\theta}_n) > \frac{\sigma(t)}{2}, |\hat{\theta}_n - \theta_0| \leq \rho(t)] = \mu_{\theta_0}^\infty[\hat{\theta}_n \in R_1(t)]$ , a task which will be carried out by an application of some large deviation estimate. In fact, recall that

$$\mu_{\theta_0}^\infty[\hat{\theta}_n \in R] \leq e^{-n \inf_{\theta \in R} I_{\theta_0}(\theta)}$$

holds true for every closed convex subset  $R$  of  $\Theta$ . See, for example, Section 2 of [15]. Therefore, since  $R_1(t)$  is compact, consider a covering  $\{V_i(t)\}_{i=1, \dots, N(t)}$  of  $R_1(t)$ , made by closed and convex subsets of  $\Theta$ , which is entirely contained in  $R_2$ , to obtain

$$\mu_{\theta_0}^\infty[\hat{\theta}_n \in R_1(t)] \leq \mu_{\theta_0}^\infty\left[\hat{\theta}_n \in \bigcup_{i=1}^{N(t)} V_i(t)\right] \leq \sum_{i=1}^{N(t)} \mu_{\theta_0}^\infty[\hat{\theta}_n \in V_i(t)] \leq N(t)e^{-nh(t)},$$

where  $h(t) := \inf\{I_{\theta_0}(\theta) \mid \Phi_2(\theta) \geq \frac{\sigma(t)}{4}, |\theta - \theta_0| \leq \rho(t)\}$  and  $N(t)$  is an abbreviation for  $N_{\Phi_2}(\rho(t), \sigma(t))$ . The conclusion is reached by first applying (4.29), giving

$$\sum_{n \geq n_0} b_n e^{-nh(t)} \leq c \left(\frac{1}{h(t)}\right)^2 e^{-n_0h(t)}$$

and then invoking (2.16) to prove the integrability of  $N(t)\left(\frac{1}{h(t)}\right)^2 e^{-n_0h(t)}$  on  $[\tau(\theta_0), +\infty)$ . After choosing  $n_0$  properly, sums of the type  $\sum_{n=1}^{n_0-1}$  admits trivial bounds which, combined with the rest of the proof, lead to the desired result.

### 4.5. Proof of Theorem 3.1

First, observe that the following two identities

$$[d_{[\mathbb{S}]_\rho}^{(p)}(\pi(\tilde{\xi}^{(n)}), \delta_{\hat{p}_n})]^p = \int_{[\mathbb{S}]} (d_{[\mathbb{S}]}^{(p)}(p, \hat{p}_n))^p \pi(\tilde{\xi}^{(n)}, dp) = \rho[(d_{[\mathbb{S}]}^{(p)}(\tilde{p}, \hat{p}_n))^p \mid \tilde{\xi}^{(n)}]$$

hold with  $\rho$ -probability 1, the former being a direct consequence of the definition (1.1). To derive the latter, suffice it to apply the disintegration theorem (see, e.g., Theorem 6.4 in [44]) after recalling that  $\hat{p}_n : (\Omega, \mathcal{F}_n) \rightarrow (\mathbb{S}, \mathcal{B}([\mathbb{S}]_\rho))$  is measurable. In the next step, paralleling the reasoning

in [10] to prove Theorem 2 therein, one gets

$$\begin{aligned} \limsup_{n \rightarrow \infty} \rho \left[ (b_n d_{[\mathbb{S}]}^{(p)}(\tilde{\mathfrak{p}}, \hat{\mathfrak{p}}_n))^p \mid \tilde{\xi}^{(n)} \right] &\leq \lim_{k \rightarrow +\infty} \sup_{n \geq k, h \geq k} \rho \left[ (b_h d_{[\mathbb{S}]}^{(p)}(\tilde{\mathfrak{p}}, \hat{\mathfrak{p}}_h))^p \mid \tilde{\xi}^{(n)} \right] \\ &\leq \lim_{k \rightarrow +\infty} \sup_{n \geq k} \rho \left[ \sup_{h \geq m} (b_h d_{[\mathbb{S}]}^{(p)}(\tilde{\mathfrak{p}}, \hat{\mathfrak{p}}_h))^p \mid \tilde{\xi}^{(n)} \right] \end{aligned}$$

for every  $m \in \mathbb{N}$ . Then, one has that  $X_m := \sup_{h \geq m} (b_h d_{[\mathbb{S}]}^{(p)}(\tilde{\mathfrak{p}}, \hat{\mathfrak{p}}_h))^p$  belongs to  $L^1(\Omega, \mathcal{F}, \rho)$ , since

$$\rho[X_m] \leq \rho[X_1] = \rho[\rho[X_1 \mid \tilde{\mathfrak{p}}]] \leq \rho[C_p(\tilde{\mathfrak{p}})] < +\infty.$$

Hence, the martingale convergence theorem gives

$$\lim_{k \rightarrow +\infty} \sup_{n \geq k} \rho[X_m \mid \tilde{\xi}^{(n)}] = \rho \left[ X_m \mid \sigma \left( \bigcup_{n \geq 1} \sigma(\tilde{\xi}^{(n)}) \right) \right]$$

with  $\rho$ -probability 1, and now it is enough to observe that  $X_m$  is a  $\sigma(\bigcup_{n \geq 1} \mathcal{F}_n)$ -measurable real random variable, for any  $m \in \mathbb{N}$ , since  $\rho[\tilde{\mathfrak{e}}_n \Rightarrow \tilde{\mathfrak{p}}, \text{asn} \rightarrow +\infty] = 1$ . Whence,

$$\rho \left[ \left\{ \lim_{k \rightarrow +\infty} \sup_{n \geq k} \rho[X_m \mid \tilde{\xi}^{(n)}] = X_m \right\} \right] = 1$$

for every  $m \in \mathbb{N}$ , and  $\limsup_{n \rightarrow \infty} [d_{[[\mathbb{S}]_p]}^{(p)}(\pi(\tilde{\xi}^{(n)}), \delta_{\hat{\mathfrak{p}}_n})]^p \leq \limsup_{m \rightarrow +\infty} X_m \leq [Y_p(\tilde{\mathfrak{p}})]^p$  hold with  $\rho$ -probability 1, yielding (3.1).

### 4.6. Proof of Theorem 3.2

First, observe that (3.3) follows immediately from the combination of (3.2) with (3.1). Thus, given  $p \geq 1$ , one proves (3.2) as a consequence of the following more general statement: for  $\mathfrak{p}_0 \in [\mathbb{S}]_p$  and  $\zeta \in [[\mathbb{S}]_p]$  there holds

$$d_{[[\mathbb{S}]_p]}^{(p)}(\rho[\zeta] \circ \tilde{\mathfrak{e}}_{n,m}^{-1}, \mathfrak{p}_0^\infty \circ \tilde{\mathfrak{e}}_{n,m}^{-1}) \leq d_{[[\mathbb{S}]_p]}^{(p)}(\zeta, \delta_{\mathfrak{p}_0}) \tag{4.30}$$

for every  $m \in \mathbb{N}$ ,  $\rho[\zeta]$  being the p.m. on  $(\Omega, \mathcal{F}) = (\mathbb{S}^\infty, \mathcal{B}(\mathbb{S}^\infty))$  given by  $\rho[\zeta](\cdot) := \int_{[\mathbb{S}]_p} \mathfrak{p}^\infty(\cdot) \zeta(d\mathfrak{p})$ . Now, for simplicity, assume that  $\zeta = \sum_{j=1}^N \alpha_j \delta_{\mathfrak{p}_j}$ , where  $\mathfrak{p}_1, \dots, \mathfrak{p}_N \in [\mathbb{S}]_p$  and  $\alpha_1, \dots, \alpha_N \in [0, 1]$  with  $\sum_{j=1}^N \alpha_j = 1$ , and define  $\bar{\eta}[\mathfrak{p}_0, \mathfrak{p}_j] \in \mathcal{F}(\mathfrak{p}_0, \mathfrak{p}_j)$  to be an optimal coupling for  $d_{[\mathbb{S}]}^{(p)}$ , that is,  $d_{[\mathbb{S}]}^{(p)}(\mathfrak{p}_0, \mathfrak{p}_j) = (\int_{\mathbb{S}^2} d_{\mathbb{S}}(x, y)^p \bar{\eta}[\mathfrak{p}_0, \mathfrak{p}_j](dx dy))^{1/p}$ , for  $j = 1, \dots, N$ . Then, put  $\bar{\gamma}_m(A_1 \times \dots \times A_m \times B_1 \times \dots \times B_m) := \sum_{j=1}^N \alpha_j \{ \prod_{i=1}^m \bar{\eta}[\mathfrak{p}_0, \mathfrak{p}_j](A_i \times B_i) \}$  for  $A_1, \dots, A_m, B_1, \dots, B_m \in \mathcal{B}(\mathbb{S})$ , noticing that  $\bar{\gamma}_m \in \mathcal{F}(\mathfrak{p}_0^m, \rho_m[\zeta])$ , where  $\rho_m[\zeta](\cdot) :=$

$\int_{[[\mathbb{S}]_p]} \mathfrak{p}^m(\cdot)\zeta(d\mathfrak{p})$ . Whence, from the definition of  $d_{[[\mathbb{S}]_p]}^{(p)}$ ,

$$\begin{aligned} & d_{[[\mathbb{S}]_p]}^{(p)}(\rho[\zeta] \circ \tilde{\epsilon}_{n,m}^{-1}, \mathfrak{p}_0^\infty \circ \tilde{\epsilon}_{n,m}^{-1}) \\ & \leq \left[ \int_{\mathbb{S}^{2m}} d_{[[\mathbb{S}]}^{(p)}\left(\frac{1}{m} \sum_{i=1}^m \delta_{x_i}, \frac{1}{m} \sum_{i=1}^m \delta_{y_i}\right)^p \bar{\gamma}_m(dx_1 \cdots dx_m dy_1 \cdots dy_m) \right]^{1/p}. \end{aligned}$$

Since a combination of the convexity of  $d_{[[\mathbb{S}]}^{(p)}$  with the Lyapunov inequality for moments yields  $d_{[[\mathbb{S}]}^{(p)}(\frac{1}{m} \sum_{i=1}^m \delta_{x_i}, \frac{1}{m} \sum_{i=1}^m \delta_{y_i})^p \leq \frac{1}{m} \sum_{i=1}^m d_{[[\mathbb{S}]}^{(p)}(\delta_{x_i}, \delta_{y_i})^p = \frac{1}{m} \sum_{i=1}^m d_{\mathbb{S}}(x_i, y_i)^p$ , one gets

$$\begin{aligned} & d_{[[\mathbb{S}]_p]}^{(p)}(\rho[\zeta] \circ \tilde{\epsilon}_{n,m}^{-1}, \mathfrak{p}_0^\infty \circ \tilde{\epsilon}_{n,m}^{-1}) \\ & \leq \left[ \int_{\mathbb{S}^{2m}} \left\{ \frac{1}{m} \sum_{i=1}^m d_{\mathbb{S}}(x_i, y_i)^p \right\} \bar{\gamma}_m(dx_1 \cdots dx_m dy_1 \cdots dy_m) \right]^{1/p} \\ & = \left[ \sum_{j=1}^N \alpha_j \int_{\mathbb{S}^2} d_{\mathbb{S}}(x, y)^p \bar{\eta}[\mathfrak{p}_0, \mathfrak{p}_j](dx dy) \right]^{1/p} = \left[ \sum_{j=1}^N \alpha_j d_{[[\mathbb{S}]}^{(p)}(\mathfrak{p}_0, \mathfrak{p}_j)^p \right]^{1/p}. \end{aligned}$$

Hence, the identities  $\sum_{j=1}^N \alpha_j d_{[[\mathbb{S}]}^{(p)}(\mathfrak{p}_0, \mathfrak{p}_j)^p = \int_{[[\mathbb{S}]_p]} d_{[[\mathbb{S}]}^{(p)}(\mathfrak{p}_0, \mathfrak{p})^p \zeta(d\mathfrak{p}) = d_{[[\mathbb{S}]_p]}^{(p)}(\zeta, \delta_{\mathfrak{p}_0})^p$  lead to the proof of (4.30), under the assumption that  $\zeta = \sum_{j=1}^N \alpha_j \delta_{\mathfrak{p}_j}$ .

To complete the argument, invoke (the proof of) Lemma 11.8.4 in [33] to show that any p.m.  $\zeta \in [[\mathbb{S}]_p]$  admits an approximating sequence  $\{\zeta^{(N)}\}_{N \geq 1}$  with  $\zeta^{(N)} = \sum_{j=1}^N \alpha_j^{(N)} \delta_{\mathfrak{p}_j^{(N)}}$ , such that  $d_{[[\mathbb{S}]_p]}^{(p)}(\zeta, \zeta^{(N)}) \rightarrow 0$  as  $N \rightarrow \infty$ . Now,  $\rho[\zeta^{(N)}] \circ \tilde{\epsilon}_{n,m}^{-1} \Rightarrow \rho[\zeta] \circ \tilde{\epsilon}_{n,m}^{-1}$  holds for every  $m \in \mathbb{N}$  by virtue of the continuous mapping theorem (cf., e.g., Theorems 4.27 and 4.29 in [44]). Hence, after noticing that  $\{\int_{[[\mathbb{S}]_p]} \int_{\mathbb{S}} d_{\mathbb{S}}(x_0, x)^p \mathfrak{p}(dx) \zeta^{(N)}(d\mathfrak{p})\}_{N \geq 1}$  is a bounded sequence, in view of Lemma 11.8.4 in [33], exploit the lower semicontinuity of  $d_{[[\mathbb{S}]_p]}^{(p)}$  (see Proposition 7.1.3 in [3]), that is,

$$d_{[[\mathbb{S}]_p]}^{(p)}(\rho[\zeta] \circ \tilde{\epsilon}_{n,m}^{-1}, \mathfrak{p}_0^\infty \circ \tilde{\epsilon}_{n,m}^{-1}) \leq \liminf_{N \rightarrow \infty} d_{[[\mathbb{S}]_p]}^{(p)}(\rho[\zeta^{(N)}] \circ \tilde{\epsilon}_{n,m}^{-1}, \mathfrak{p}_0^\infty \circ \tilde{\epsilon}_{n,m}^{-1})$$

to deduce the validity of (4.30) for any  $\zeta \in [[\mathbb{S}]_p]$ .

### Acknowledgements

The authors are obliged to the anonymous referees for helpful comments and suggestions. They are also grateful to Giuseppe Savaré, P. Diaconis and E.A. Carlen for useful conversations. This research was supported by the Italian Ministry of Education, University and Research (MIUR): Dipartimenti di Eccellenza Program (2018–2022) – Dept. of Mathematics “F. Casorati”, University of Pavia.

## References

- [1] Ajtai, M., Komlós, J. and Tusnády, G. (1984). On optimal matchings. *Combinatorica* **4** 259–264. [MR0779885](#)
- [2] Aldous, D.J. (1985). Exchangeability and related topics. In *École D’été de Probabilités de Saint-Flour, XIII – 1983. Lecture Notes in Math.* **1117** 1–198. Berlin: Springer. [MR0883646](#)
- [3] Ambrosio, L., Gigli, N. and Savaré, G. (2008). *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed. *Lectures in Mathematics ETH Zürich*. Basel: Birkhäuser. [MR2401600](#)
- [4] Ambrosio, L., Stra, F. and Trevisan, D. (2018). A PDE approach to a 2-dimensional matching problem. To appear on *Probab. Theory Related Fields*. DOI:[10.1007/s00440-018-0837-x](#).
- [5] Anderson, T.W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. New York: Wiley. [MR0771294](#)
- [6] Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory*. *Wiley Series in Probability and Mathematical Statistics*. Chichester: Wiley. [MR0489333](#)
- [7] Berti, P., Crimaldi, I., Pratelli, L. and Rigo, P. (2009). Rate of convergence of predictive distributions for dependent data. *Bernoulli* **15** 1351–1367. [MR2597596](#)
- [8] Berti, P., Pratelli, L. and Rigo, P. (2012). Limit theorems for empirical processes based on dependent data. *Electron. J. Probab.* **17** no. 9, 18. [MR2878788](#)
- [9] Berti, P., Pratelli, L. and Rigo, P. (2018). Asymptotic predictive inference with exchangeable data. *Braz. J. Probab. Stat.* **32** 815–833. [MR3845031](#)
- [10] Blackwell, D. and Dubins, L. (1962). Merging of opinions with increasing information. *Ann. Math. Stat.* **33** 882–886. [MR0149577](#)
- [11] Bobkov, S. and Ledoux, M. (2016). One-dimensional empirical measures, order statistics, and Kantorovich transport distances. To appear on *Mem. Amer. Math. Soc.*
- [12] Boissard, E. (2011). Simple bounds for convergence of empirical and occupation measures in 1-Wasserstein distance. *Electron. J. Probab.* **16** 2296–2333. [MR2861675](#)
- [13] Boissard, E. and Le Gouic, T. (2014). On the mean speed of convergence of empirical and occupation measures in Wasserstein distance. *Ann. Inst. Henri Poincaré Probab. Stat.* **50** 539–563. [MR3189084](#)
- [14] Bolley, F., Guillin, A. and Villani, C. (2007). Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probab. Theory Related Fields* **137** 541–593. [MR2280433](#)
- [15] Borovkov, A.A. and Mogul’skiĭ, A.A. (2011). Chebyshev-type exponential inequalities for sums of random vectors and for trajectories of random walks. *Teor. Veroyatn. Primen.* **56** 3–29. [MR2848414](#)
- [16] Brown, L.D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory. Institute of Mathematical Statistics Lecture Notes – Monograph Series* **9**. Hayward, CA: IMS. [MR0882001](#)
- [17] Cantelli, F.P. (1933). Sulla determinazione empirica delle leggi di probabilità. *G. Ist. Ital. Attuari* **4** 421–424.
- [18] Caracciolo, S., Lucibello, G. and Parisi, G. (2014). Scaling hypothesis for the Euclidean bipartite matching problem. *Phys. Rev. E* **90** 012118.
- [19] Chow, Y.S. and Teicher, H. (1997). *Probability Theory: Independence, Interchangeability, Martingales*, 3rd ed. *Springer Texts in Statistics*. New York: Springer. [MR1476912](#)
- [20] Cifarelli, D.M., Dolera, E. and Regazzini, E. (2016). Frequentistic approximations to Bayesian prevision of exchangeable random elements. *Internat. J. Approx. Reason.* **78** 138–152. [MR3543878](#)
- [21] Cifarelli, D.M., Dolera, E. and Regazzini, E. (2017). Note on “Frequentistic approximations to Bayesian prevision of exchangeable random elements” [Int. J. Approx. Reason. 78 (2016) 138–152] [[MR3543878](#)]. *Internat. J. Approx. Reason.* **86** 26–27. [MR3653971](#)
- [22] Cox, D.R. (1975). Partial likelihood. *Biometrika* **62** 269–276. [MR0400509](#)

- [23] de Finetti, B. (1930). Funzione caratteristica di un fenomeno aleatorio. *Atti R. Accad. Naz. Lincei, Mem.* **4** 86–133.
- [24] de Finetti, B. (1933). Sull'approssimazione empirica di una legge di probabilità. *G. Ist. Ital. Attuari* **4** 415–420.
- [25] de Finetti, B. (1933). La legge dei grandi numeri nel caso dei numeri aleatori equivalenti. *Atti R. Accad. Naz. Lincei, Rend.* **18** 203–207.
- [26] de Finetti, B. (1937). La prévision: Ses lois logiques, ses sources subjectives. *Ann. Inst. Henri Poincaré* **7** 1–68. [MR1508036](#)
- [27] Dereich, S., Scheutzwow, M. and Schottstedt, R. (2013). Constructive quantization: Approximation by empirical measures. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 1183–1203. [MR3127919](#)
- [28] Devroye, L., Györfi, L. and Lugosi, G. (1996). *A Probabilistic Theory of Pattern Recognition. Applications of Mathematics (New York)* **31**. New York: Springer. [MR1383093](#)
- [29] Diaconis, P. and Freedman, D. (1986). On the consistency of Bayes estimates. *Ann. Statist.* **14** 1–67. [MR0829555](#)
- [30] Dowson, D.C. and Landau, B.V. (1982). The Fréchet distance between multivariate normal distributions. *J. Multivariate Anal.* **12** 450–455. [MR0666017](#)
- [31] Dudley, R.M. (1968). The speed of mean Glivenko–Cantelli convergence. *Ann. Math. Stat.* **40** 40–50. [MR0236977](#)
- [32] Dudley, R.M. (1999). *Uniform Central Limit Theorems. Cambridge Studies in Advanced Mathematics* **63**. Cambridge: Cambridge Univ. Press. [MR1720712](#)
- [33] Dudley, R.M. (2002). *Real Analysis and Probability. Cambridge Studies in Advanced Mathematics* **74**. Cambridge: Cambridge Univ. Press. [MR1932358](#)
- [34] Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26. [MR0515681](#)
- [35] Efron, B. (2010). *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction. Institute of Mathematical Statistics (IMS) Monographs* **1**. Cambridge: Cambridge Univ. Press. [MR2724758](#)
- [36] Einmahl, U. (2016). Law of the iterated logarithm type results for random vectors with infinite second moments. *Math. Appl. (Warsaw)* **44** 167–181. [MR3557096](#)
- [37] Fournier, N. and Guillin, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields* **162** 707–738. [MR3383341](#)
- [38] Glivenko, V.I. (1933). Sulla determinazione empirica delle leggi di probabilità. *G. Ist. Ital. Attuari* **4** 92–99.
- [39] Gozlan, N. (2009). A characterization of dimension free concentration in terms of transportation inequalities. *Ann. Probab.* **37** 2480–2498. [MR2573565](#)
- [40] Gozlan, N. and Léonard, C. (2007). A large deviation approach to some transportation cost inequalities. *Probab. Theory Related Fields* **139** 235–283. [MR2322697](#)
- [41] Gozlan, N. and Léonard, C. (2010). Transport inequalities. A survey. *Markov Process. Related Fields* **16** 635–736. [MR2895086](#)
- [42] Horn, R.A. and Johnson, C.R. (2013). *Matrix Analysis*, 2nd ed. Cambridge: Cambridge Univ. Press. [MR2978290](#)
- [43] Horowitz, J. and Karandikar, R.L. (1994). Mean rates of convergence of empirical measures in the Wasserstein metric. *J. Comput. Appl. Math.* **55** 261–273. [MR1329874](#)
- [44] Kallenberg, O. (2002). *Foundations of Modern Probability*, 2nd ed. *Probability and Its Applications (New York)*. New York: Springer. [MR1876169](#)
- [45] Kolmogorov, A.N. (1933). Sulla determinazione empirica di una legge di distribuzione. *G. Ist. Ital. Attuari* **4** 83–91.
- [46] Maritz, J.S. and Lwin, T. (1989). *Empirical Bayes Methods*, 2nd ed. *Monographs on Statistics and Applied Probability* **35**. London: CRC Press. [MR1019835](#)

- [47] Massart, P. (1988). About the Prohorov [Prokhorov] distance between the uniform distribution over the unit cube in  $\mathbf{R}^d$  and its empirical measure. *Probab. Theory Related Fields* **79** 431–450. [MR0959517](#)
- [48] Morris, C.N. (1983). Parametric empirical Bayes inference: Theory and applications. *J. Amer. Statist. Assoc.* **78** 47–65. [MR0696849](#)
- [49] Olkin, I. and Pukelsheim, F. (1982). The distance between two random vectors with given dispersion matrices. *Linear Algebra Appl.* **48** 257–263. [MR0683223](#)
- [50] Petrone, S., Rizzelli, S., Rousseau, J. and Scricciolo, C. (2014). Empirical Bayes methods in classical and Bayesian inference. *Metron* **72** 201–215. [MR3233149](#)
- [51] Petrone, S., Rousseau, J. and Scricciolo, C. (2014). Bayes and empirical Bayes: Do they merge? *Biometrika* **101** 285–302. [MR3215348](#)
- [52] Robbins, H. (1956). An empirical Bayes approach to statistics. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, Vol. I* 157–163. Berkeley and Los Angeles: Univ. California Press. [MR0084919](#)
- [53] Rockafellar, R.T. (1970). *Convex Analysis. Princeton Mathematical Series* **28**. Princeton, NJ: Princeton Univ. Press. [MR0274683](#)
- [54] Rousseau, J. (2016). On the frequentist properties of Bayesian nonparametric methods. *Annu. Rev. Sci. Stat.* **3** 211–231.
- [55] Severini, T.A. (2000). *Likelihood Methods in Statistics. Oxford Statistical Science Series* **22**. Oxford: Oxford Univ. Press. [MR1854870](#)
- [56] Shor, P.W. and Yukich, J.E. (1991). Minimax grid matching and empirical measures. *Ann. Probab.* **19** 1338–1348. [MR1112419](#)
- [57] Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. New York: Wiley. [MR0838963](#)
- [58] Siegmund, D. (1969). On moments of the maximum of normed partial sums. *Ann. Math. Stat.* **40** 527–531. [MR0239695](#)
- [59] Talagrand, M. (1994). The transportation cost from the uniform measure to the empirical measure in dimension  $\geq 3$ . *Ann. Probab.* **22** 919–959. [MR1288137](#)
- [60] Talagrand, M. (1996). Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.* **6** 587–600. [MR1392331](#)
- [61] Teicher, H. (1971). Completion of a dominated ergodic theorem. *Ann. Math. Stat.* **42** 2156–2158. [MR0300327](#)
- [62] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics. Springer Series in Statistics*. New York: Springer. [MR1385671](#)
- [63] Vapnik, V.N. and Červonenkis, A.J. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Appl.* **16** 264–280.
- [64] Villani, C. (2009). *Optimal Transport: Old and New. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **338**. Berlin: Springer. [MR2459454](#)
- [65] Weed, J. and Bach, F. (2017). Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. Preprint. Available at [arXiv:1707.00087](#).
- [66] Yosida, K. and Kakutani, S. (1939). Birkhoff’s ergodic theorem and the maximal ergodic theorem. *Proc. Imp. Acad. (Tokyo)* **15** 165–168. [MR0000355](#)
- [67] Yukich, J.E. (1989). Optimal matching and empirical measures. *Proc. Amer. Math. Soc.* **107** 1051–1059. [MR1000171](#)
- [68] Yurinskii, V.V. (1976). Exponential inequalities for sums of random vectors. *J. Multivariate Anal.* **6** 473–499. [MR0428401](#)

Received December 2017 and revised July 2018