Bernoulli **16**(4), 2010, 1385–1414
DOI: 10.3150/10-BEJ254

Concentration of empirical distribution functions with applications to non-i.i.d. models

S.G. BOBKOV[1] and F. GÖTZE[2]

[1]*School of Mathematics, University of Minnesota, 127 Vincent Hall, 206 Church St. S.E., Minneapolis, MN 55455, USA. E-mail: bobkov@math.umn.edu*
[2]*Fakultät für Mathematik, Universität Bielefeld, Postfach 100131, 33501 Bielefeld, Germany. E-mail: goetze@mathematik.uni-bielefeld.de*

The concentration of empirical measures is studied for dependent data, whose joint distribution satisfies Poincaré-type or logarithmic Sobolev inequalities. The general concentration results are then applied to spectral empirical distribution functions associated with high-dimensional random matrices.

Keywords: empirical measures; logarithmic Sobolev inequalities; Poincaré-type inequalities; random matrices; spectral distributions

1. Introduction

Let $(X_1, \dots, X_n)$ be a random vector in $\mathbf{R}^n$ with distribution μ. We study rates of approximation of the average marginal distribution function

$$F(x) = \mathbf{E} F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{P}\{X_i \le x\}$$

by the empirical distribution function

$$F_n(x) = \frac{1}{n} \operatorname{card}\{i \le n : X_i \le x\}, \qquad x \in \mathbf{R}.$$

We shall measure the distance between F and F_n by means of the (uniform) Kolmogorov metric $\|F_n - F\| = \sup_x |F_n(x) - F(x)|$, as well as by means of the L^1-metric

$$W_1(F_n, F) = \int_{-\infty}^{+\infty} |F_n(x) - F(x)| \, \mathrm{d}x.$$

The latter, also called the Kantorovich–Rubinstein distance, may be interpreted as the minimal cost needed to transport the empirical measure F_n to F with cost function $d(x, y) = |x - y|$ (the price paid to transport the point x to the point y).

The classical example is the case where all X_i's are independent and identically distributed (i.i.d.), that is, when μ represents a product measure on $\mathbf{R}^n$ with equal marginals, say, F. If it has

1350-7265

no atoms, the distributions of the random variables $T_n = \sqrt{n}\|F_n - F\|$ are weakly convergent to the Kolmogorov law. Moreover, by the Dvoretzky–Kiefer–Wolfowitz theorem, the r.v.'s T_n are uniformly sub-Gaussian and, in particular, $\mathbf{E}\|F_n - F\| \le \frac{C}{\sqrt{n}}$ up to a universal factor C ([17]; cf. [32] for history and sharp bounds). This result, together with the related invariance principle, has a number of extensions to the case of dependent observations, mainly in terms of mixing conditions imposed on a stationary process; see, for example, [28,38,44].

On the other hand, the observations $X_1, \dots, X_n$ may also be generated by non-trivial functions of independent random variables. Of particular importance are random symmetric matrices $(\frac{1}{\sqrt{n}}\xi_{jk})$, $1 \le j, k \le n$, with i.i.d. entries above and on the diagonal. Arranging their eigenvalues $X_1 \le \cdots \le X_n$ in increasing order, we arrive at the spectral empirical measures F_n. In this case, the mean $F = \mathbf{E}F_n$ also depends on n and converges to the semicircle law under appropriate moment assumptions on ξ_{jk} (cf., for example, [20,36]).

The example of matrices strongly motivates the study of deviations of F_n from the mean F under general analytical hypotheses on the joint distribution of the observations, such as Poincaré or logarithmic Sobolev inequalities. A probability measure μ on $\mathbf{R}^n$ is said to satisfy a *Poincaré-type* or *spectral gap inequality with constant* σ^2 ($\sigma \ge 0$) if, for any bounded smooth function g on $\mathbf{R}^n$ with gradient ∇g,

$$\mathrm{Var}_\mu(g) \le \sigma^2 \int |\nabla g|^2 \,\mathrm{d}\mu. \tag{1.1}$$

In this case, we write $\mathrm{PI}(\sigma^2)$ for short. Similarly, μ satisfies a logarithmic Sobolev inequality with constant σ^2 and we write $\mathrm{LSI}(\sigma^2)$ if, for all bounded smooth g,

$$\mathrm{Ent}_\mu(g^2) \le 2\sigma^2 \int |\nabla g|^2 \,\mathrm{d}\mu. \tag{1.2}$$

Here, as usual, $\mathrm{Var}_\mu(g) = \int g^2 \,\mathrm{d}\mu - (\int g \,\mathrm{d}\mu)^2$ stands for the variance of g and $\mathrm{Ent}_\mu(g) = \int g \log g \,\mathrm{d}\mu - \int g \,\mathrm{d}\mu \log \int g \,\mathrm{d}\mu$ denotes the entropy of $g \ge 0$ under the measure μ. It is well known that $\mathrm{LSI}(\sigma^2)$ implies $\mathrm{PI}(\sigma^2)$.

These hypotheses are crucial in the study of concentration of the spectral empirical distributions, especially of the linear functionals $\int f \,\mathrm{d}F_n$ with individual smooth f on the line; see, for example, the results by Guionnet and Zeitouni [24], Chatterjee and Bose [14], Davidson and Szarek [15] and Ledoux [30]. A remarkable feature of this approach to spectral analysis is that no specific knowledge about the non-explicit mapping from a random matrix to its spectral empirical measure is required. Instead, one may use general Lipschitz properties only, which are satisfied by this mapping. As for the general (not necessarily matrix) scheme, we shall only require the hypotheses (1.1) or (1.2). In particular, we derive the following from (1.1).

Theorem 1.1. *Under* $\mathrm{PI}(\sigma^2)$ *on* $\mathbf{R}^n$ $(n \ge 2)$,

$$\mathbf{E}\int_{-\infty}^{+\infty} |F_n(x) - F(x)| \,\mathrm{d}x \le C\sigma \left(\frac{A + \log n}{n}\right)^{1/3}, \tag{1.3}$$

where $A = \frac{1}{\sigma}\max_{i,j} |\mathbf{E}X_i - \mathbf{E}X_j|$ *and C is an absolute constant.*

Note that the Poincaré-type inequality (1.1) is invariant under shifts of the measure μ, while the left-hand side of (1.3) is not. This is why the bound on the right-hand side of (1.3) should also depend on the means of the observations.

In terms of the ordered statistics $X_1^* \le \cdots \le X_n^*$ of the random vector $(X_1, \dots, X_n)$, there is a general two-sided estimate for the mean of the Kantorovich–Rubinstein distance:

$$\frac{1}{2n} \sum_{i=1} \mathbf{E}|X_i^* - \mathbf{E}X_i^*| \le \mathbf{E}W_1(F_n, F) \le \frac{2}{n} \sum_{i=1} \mathbf{E}|X_i^* - \mathbf{E}X_i^*| \tag{1.4}$$

(see remarks at the end of Section 4). Hence, under the conditions of Theorem 1.1, one may control the local fluctuations of X_i^* (on average), which typically deviate from their mean by not more than $C\sigma(\frac{A+\log n}{n})^{1/3}$.

Under a stronger hypothesis, such as (1.2), one can obtain more information about the fluctuations of $F_n(x) - F(x)$ for individual points x and thus get some control of the Kolmogorov distance. Similarly to the bound (1.3), such fluctuations will, on average, be shown to be at most

$$\beta = \frac{(M\sigma)^{2/3}}{n^{1/3}},$$

in the sense that $\mathbf{E}|F_n(x) - F(x)| \le C\beta$, where M is the Lipschitz seminorm of F (see Proposition 6.3). As for the Kolmogorov distance, we prove the following theorem.

Theorem 1.2. *Assume that F has a density, bounded by a number M. Under* $\mathrm{LSI}(\sigma^2)$, *for any* $r > 0$,

$$\mathbf{P}\{\|F_n - F\| \ge r\} \le \frac{4}{r} \mathrm{e}^{-c(r/\beta)^3}. \tag{1.5}$$

In particular,

$$\mathbf{E}\|F_n - F\| \le C\beta \log^{1/3}\left(1 + \frac{1}{\beta}\right), \tag{1.6}$$

where c and C are positive absolute constants.

In both cases, the stated bounds are of order $n^{-1/3}$ up to a $\log n$ term with respect to the dimension n. Thus, they are not as sharp as in the classical i.i.d. case. Indeed, our assumptions are much weaker and may naturally lead to weaker conclusions. Let us look at two examples illustrating the bounds obtained in the cases that essentially differ from the i.i.d. case.

Example 1. Let X_i be independent and uniformly distributed in the intervals $(i-1, i)$, $i = 1, \dots, n$. Their joint distribution is a product measure, satisfying (1.1) and (1.2) with some absolute constant σ. Clearly, F is the uniform distribution in $(0, n)$ so $M = \frac{1}{n}$ and β is of order $\frac{1}{n}$. As is easy to see, $\mathbf{E}\|F_n - F\|$ is also of order $\frac{1}{n}$, that is, the bound (1.6) is sharp up to a $\log^{1/3} n$ term. Also, since A is of order n, both sides of (1.3) are of order 1. In particular, this shows that the quantity A cannot be removed from (1.3).

Example 2. Let all $X_i = \xi$, where ξ is uniformly distributed in $[-1, 1]$. Note that all random variables are identically distributed with $\mathbf{E}X_i = 0$. The joint distribution μ represents a uniform distribution on the main diagonal of the cube $[-1, 1]^n$, so it satisfies (1.1) and (1.2) with $\sigma = c\sqrt{n}$, where c is absolute. In this case, F is a uniform distribution on $[-1, 1]$, so $M = 1/2$ and β is of order 1. Hence, both sides of (1.6) are of order 1.

Next, we restrict the above statements to the empirical spectral measures F_n of the n eigenvalues $X_1 \leq \cdots \leq X_n$ of a random symmetric matrix $(\frac{1}{\sqrt{n}}\xi_{jk})$, $1 \leq j, k \leq n$, with independent entries above and on the diagonal ($n \geq 2$). Assume that $\mathbf{E}\xi_{jk} = 0$ and $\mathrm{Var}(\xi_{jk}) = 1$ so that the means $F = \mathbf{E}F_n$ converge to the semicircle law G with mean zero and variance one. The boundedness of moments of ξ_{jk} of any order will be guaranteed by (1.1).

Theorem 1.3. *If the distributions of the ξ_{jk}'s satisfy the Poincaré-type inequality* $\mathrm{PI}(\sigma^2)$ *on the real line*, *then*

$$\mathbf{E}\int_{-\infty}^{+\infty} |F_n(x) - F(x)|\,\mathrm{d}x \leq \frac{C\sigma}{n^{2/3}}, \tag{1.7}$$

where C is an absolute constant. *Moreover*, *under* $\mathrm{LSI}(\sigma^2)$,

$$\mathbf{E}\|F_n - G\| \leq C\left(\frac{\sigma}{n}\right)^{2/3} \log^{1/3} n + \|F - G\|. \tag{1.8}$$

By the convexity of the distance, we always have $\mathbf{E}\|F_n - G\| \geq \|F - G\|$. In some random matrix models, the Kolmogorov distance $\|F - G\|$ is known to tend to zero at rate at most $n^{-2/3+\varepsilon}$. For instance, it is true when the distributions of the $\xi_{j,k}$'s have a non-trivial Gaussian component (see [22]). Hence, if, additionally, $\mathrm{LSI}(\sigma^2)$ is satisfied, then we get that for any $\varepsilon > 0$,

$$\mathbf{E}\|F_n - G\| \leq C_{\varepsilon,\sigma} n^{-2/3+\varepsilon}.$$

It is unknown whether this bound is optimal. Note, however, that in the case of Gaussian $\xi_{j,k}$, the distance $\|F - G\|$ is known to be of order $1/n$ [21]. Therefore,

$$\mathbf{E}\|F_n - G\| \leq C\frac{\log^{1/3} n}{n^{2/3}}, \tag{1.9}$$

which is a slight improvement of a bound obtained in [42]. In fact, as was recently shown in [8], we have $\|F - G\| \leq C_\sigma n^{-2/3}$ in the presence of the $\mathrm{PI}(\sigma^2)$-hypothesis. Hence, the bound (1.9) always holds under $\mathrm{LSI}(\sigma^2)$ with constants depending only on σ.

It seems natural to try to relax the $\mathrm{LSI}(\sigma^2)$-hypothesis in (1.8) and (1.9) to $\mathrm{PI}(\sigma^2)$. In this context, let us mention a result of Chatterjee and Bose [14], who used Fourier transforms to derive from $\mathrm{PI}(\sigma^2)$ a similar bound,

$$\mathbf{E}\|F_n - G\| \leq \frac{C\sigma^{1/4}}{n^{1/2}} + 2\|F - G\|.$$

As for (1.7), let us return to the two-sided bound (1.4) which holds with $X_i^* = X_i$ by the convention that the eigenvalues are listed in increasing order. The asymptotic behavior of distributions of X_i with fixed or varying indices has been studied by many authors, especially in the standard Gaussian case. In particular, if i is fixed, while n grows, $n^{2/3}(X_i - \mathbf{E}X_i)$ converges in distribution to (a variant of) the Tracy–Widom law so the $\mathbf{E}|X_i - \mathbf{E}X_i|$ are of order $n^{-2/3}$. This property still holds when ξ_{jk} are symmetric and have sub-Gaussian tails; see [39] and [31] for the history and related results. Although this rate is consistent with the bound (1.7), the main contribution in the normalized sum (1.4) is due to the intermediate terms (in the bulk) and their rate might be different. It was shown by Gustavsson [25] for the GUE model that if $\frac{i}{n} \to t \in (0,1)$, then X_i is asymptotically normal with variance of order $\frac{C(t)\log n}{n^2}$. Hence, it is not surprising that $\mathbf{E}W_1(F_n, F) \leq \frac{C(\log n)^{1/2}}{n}$, see [42].

The paper is organized as follows. In Section 2, we collect a few direct applications of the Poincaré-type inequality to linear functionals of empirical measures. They are used in Section 3 to complete the proof of Theorem 1.1. In the next section, we discuss deviations of $W_1(F_n, F)$ from its mean. In Section 5, we turn to logarithmic Sobolev inequalities. Here, we shall adapt infimum-convolution operators to empirical measures and apply a result of [5] on the relationship between infimum-convolution and log-Sobolev inequalities. In Section 6, we illustrate this approach in the problem of dispersion of the values of the empirical distribution functions at a fixed point. In Section 7, we derive bounds on the uniform distance similar to (1.5) and (1.6) and give a somewhat more general form of Theorem 1.2. In Section 8, we apply the previous results to high-dimensional random matrices to prove Theorem 1.3 and obtain some refinements. Finally, since the hypotheses (1.1) and (1.2) play a crucial role in this investigation, we collect in the Appendix a few results on sufficient conditions for a measure to satisfy PI and LSI.

2. Empirical Poincaré inequalities

We assume that the random variables $X_1, \dots, X_n$ have a joint distribution μ on $\mathbf{R}^n$, satisfying the Poincaré-type inequality (1.1). For a bounded smooth function f on the real line, we apply it to

$$g(x_1, \dots, x_n) = \frac{f(x_1) + \cdots + f(x_n)}{n} = \int f \, \mathrm{d}F_n, \tag{2.1}$$

where F_n is the empirical measure, defined for the 'observations' $X_1 = x_1, \dots, X_n = x_n$. Since

$$|\nabla g(x_1, \dots, x_n)|^2 = \frac{f'(x_1)^2 + \cdots + f'(x_n)^2}{n^2} = \frac{1}{n} \int f'^2 \, \mathrm{d}F_n, \tag{2.2}$$

we obtain an integro-differential inequality, which may viewed as an empirical Poincaré-type inequality for the measure μ.

Proposition 2.1. *Under* $\mathrm{PI}(\sigma^2)$, *for any smooth F-integrable function f on $\mathbf{R}$ such that f' belongs to $L^2(\mathbf{R}, \mathrm{d}F)$, we have*

$$\mathbf{E}\left| \int f \, \mathrm{d}F_n - \int f \, \mathrm{d}F \right|^2 \leq \frac{\sigma^2}{n} \int f'^2 \, \mathrm{d}F. \tag{2.3}$$

Recall that $F = \mathbf{E}F_n$ denotes the mean of the empirical measures. The inequality continues to hold for all locally Lipschitz functions with the modulus of the derivative, understood in the generalized sense, that is, $|f'(x)| = \limsup_{y\to x} \frac{|f(x)-f(y)|}{|x-y|}$. As long as $\int f'^2\,\mathrm{d}F$ is finite, $\int f^2\,\mathrm{d}F$ is also finite and (2.3) holds.

The latter may be extended to all L^p-spaces by applying the following general lemma.

Lemma 2.2. *Under* $\mathrm{PI}(\sigma^2)$, *any Lipschitz function g on $\mathbf{R}^n$ has a finite exponential moment*: *if* $\int g\,\mathrm{d}\mu = 0$ *and* $\|g\|_{\mathrm{Lip}} \le 1$, *then*

$$\int \mathrm{e}^{tg/\sigma}\,\mathrm{d}\mu \le \frac{2+t}{2-t}, \qquad 0 < t < 2. \tag{2.4}$$

Moreover, for any locally Lipschitz g on $\mathbf{R}^n$ with μ-mean zero,

$$\|g\|_p \le \sigma p \|\nabla g\|_p, \qquad p \ge 2. \tag{2.5}$$

More precisely, if $|\nabla g|$ is in $L^p(\mu)$, then so is g and (2.5) holds true with the standard notation $\|g\|_p = (\int |g|^p\,\mathrm{d}\mu)^{1/p}$ and $\|\nabla g\|_p = (\int |\nabla g|^p\,\mathrm{d}\mu)^{1/p}$ for $L^p(\mu)$-norms. The property of being locally Lipschitz means that the function g has a finite Lipschitz seminorm on every compact subset of $\mathbf{R}^n$.

In the concentration context, a variant of the first part of the lemma was first established by Gromov and Milman in [23] and independently in dimension 1 by Borovkov and Utev [12]. Here, we follow [10], Proposition 4.1, to state (2.4). The second inequality, (2.5), may be derived by similar arguments; see also [11], Theorem 4.1, for an extension to the case of Poincaré-type inequalities with weight.

Now, for functions $g = \int f\,\mathrm{d}F_n$ as in (2.1), in view of (2.2), we may write

$$|\nabla g|^p = \frac{1}{n^{p/2}} \left(\int f'^2\,\mathrm{d}F_n\right)^{p/2} \le \frac{1}{n^{p/2}} \int |f'|^p\,\mathrm{d}F_n$$

so that $\mathbf{E}_\mu |\nabla g|^p \le \frac{1}{n^{p/2}} \int |f'|^p\,\mathrm{d}F$. Applying (2.5) and (2.4) with $t = 1$, we obtain the following proposition.

Proposition 2.3. *Under* $\mathrm{PI}(\sigma^2)$, *for any smooth function f on $\mathbf{R}$ such that f' belongs to* $L^p(\mathbf{R}, \mathrm{d}F)$, $p \ge 2$,

$$\mathbf{E}\left|\int f\,\mathrm{d}F_n - \int f\,\mathrm{d}F\right|^p \le \frac{(\sigma p)^p}{n^{p/2}} \int |f'|^p\,\mathrm{d}F.$$

In addition, if $|f'| \le 1$, for all $h > 0$,

$$\mu\left\{\left|\int f\,\mathrm{d}F_n - \int f\,\mathrm{d}F\right| \ge h\right\} \le 6\mathrm{e}^{-nh/\sigma}.$$

The empirical Poincaré-type inequality (2.3) can be rewritten equivalently if we integrate by parts the first integral as $\int f\,\mathrm{d}F_n - \int f\,\mathrm{d}F = -\int f'(x)(F_n(x) - F(x))\,\mathrm{d}x$. At this step, it is safe to assume that f is continuously differentiable and is constant near $-\infty$ and $+\infty$. Replacing f' with f, we arrive at

$$\mathbf{E}\left|\int f(x)\big(F_n(x)-F(x)\big)\,\mathrm{d}x\right|^2 \le \frac{\sigma^2}{n}\int f^2\,\mathrm{d}F \tag{2.6}$$

for any continuous, compactly supported function f on the line. In other words, the integral operator $Kf(x)=\int_{-\infty}^{+\infty}K(x,y)f(y)\,\mathrm{d}y$ with a (positive definite) kernel

$$K(x,y)=\mathbf{E}\big(F_n(x)-F(x)\big)\big(F_n(y)-F(y)\big)=\mathrm{cov}(F_n(x),F_n(y))$$

is continuous and defined on a dense subset of $L^2(\mathbf{R},\mathrm{d}F(x))$, taking values in $L^2(\mathbf{R},\mathrm{d}x)$. It has the operator norm $\|K\|\le\frac{\sigma}{\sqrt{n}}$, so it may be continuously extended to the space $L^2(\mathbf{R},\mathrm{d}F)$ without a change of the norm. In the following, we will use a particular case of (2.6).

Corollary 2.4. *Under* $\mathrm{PI}(\sigma^2)$, *whenever* $a<b$, *we have*

$$\mathbf{E}\left|\int_a^b\big(F_n(x)-F(x)\big)\,\mathrm{d}x\right|\le\frac{\sigma}{\sqrt{n}}\sqrt{F(b)-F(a)}.$$

3. Proof of Theorem 1.1

We shall now study the concentration properties of empirical measures F_n around their mean F based on Poincaré-type inequalities. In particular, we shall prove Theorem 1.1, which provides a bound on the mean of the Kantorovich–Rubinstein distance

$$W_1(F_n,F)=\int_{-\infty}^{+\infty}|F_n(x)-F(x)|\,\mathrm{d}x.$$

Note that it is homogeneous of order 1 with respect to the random vector $(X_1,\dots,X_n)$.

We first need a general observation.

Lemma 3.1. *Given distribution functions F and G, for all real $a<b$ and a natural number N,*

$$\int_a^b|F(x)-G(x)|\,\mathrm{d}x\le\sum_{k=1}^N\left|\int_{a_{k-1}}^{a_k}\big(F(x)-G(x)\big)\,\mathrm{d}x\right|+\frac{2(b-a)}{N},$$

where $a_k=a+(b-a)\frac{k}{N}$.

Proof. Let I denote the collection of those indices k such that in the kth subinterval $\Delta_k=(a_{k-1},a_k)$, the function $\varphi(x)=F(x)-G(x)$ does not change sign. Let J denote the collection of the remaining indices. Then, for $k\in I$,

$$\int_{\Delta_k}|F(x)-G(x)|\,\mathrm{d}x=\left|\int_{\Delta_k}\big(F(x)-G(x)\big)\,\mathrm{d}x\right|.$$

In the other case $k \in J$, since φ changes sign on Δ_k, we may write

$$\sup_{x \in \Delta_k} |\varphi(x)| \le \mathrm{Osc}_{\Delta_k}(\varphi) \equiv \sup_{x,y \in \Delta_k} \big(\varphi(x) - \varphi(y)\big)$$
$$\le \mathrm{Osc}_{\Delta_k}(F) + \mathrm{Osc}_{\Delta_k}(G) = F(\Delta_k) + G(\Delta_k),$$

where, in the last step, F and G are treated as probability measures. Hence, in this case, $\int_{\Delta_k} |F(x) - G(x)|\,\mathrm{d}x \le (F(\Delta_k) + G(\Delta_k))|\Delta_k|$. Combining the two bounds and using $|\Delta_k| = \frac{b-a}{N}$, we get that

$$\begin{aligned}
&\int_a^b |F(x) - G(x)|\,\mathrm{d}x \\
&\quad \le \sum_{k \in I} \left| \int_{\Delta_k} \big(F(x) - G(x)\big)\,\mathrm{d}x \right| + \sum_{k \in J} \big(F(\Delta_k) + G(\Delta_k)\big)|\Delta_k| \\
&\quad \le \sum_{k=1}^N \left| \int_{\Delta_k} \big(F(x) - G(x)\big)\,\mathrm{d}x \right| + \frac{b-a}{N} \sum_{k=1}^N \big(F(\Delta_k) + G(\Delta_k)\big).
\end{aligned}$$

□

Remark. As the proof shows, the lemma may be extended to an arbitrary partition $a = a_0 < a_1 < \cdots < a_N = b$, as follows:

$$\int_a^b |F(x) - G(x)|\,\mathrm{d}x \le \sum_{k=1}^N \left| \int_{a_{k-1}}^{a_k} \big(F(x) - G(x)\big)\,\mathrm{d}x \right| + 2 \max_{1 \le k \le N} (a_k - a_{k-1}).$$

Let us now apply the lemma to the space $(\mathbf{R}^n, \mu)$ satisfying a Poincaré-type inequality. Consider the partition of the interval $[a, b]$ with $\Delta_k = (a_{k-1}, a_k)$, as in Lemma 3.1. By Corollary 2.4,

$$\begin{aligned}
\mathbf{E} \int_a^b |F_n(x) - F(x)|\,\mathrm{d}x &\le \sum_{k=1}^N \mathbf{E} \left| \int_{\Delta_k} \big(F_n(x) - F(x)\big)\,\mathrm{d}x \right| + \frac{2(b-a)}{N} \\
&\le \frac{\sigma}{\sqrt{n}} \sum_{k=1}^N \sqrt{F(\Delta_k)} + \frac{2(b-a)}{N}.
\end{aligned}$$

By Cauchy's inequality, $\sum_{k=1}^N \sqrt{F(\Delta_k)} \le \sqrt{N} (\sum_{k=1}^N F(\Delta_k))^{1/2} \le \sqrt{N}$, hence,

$$\mathbf{E} \int_a^b |F_n(x) - F(x)|\,\mathrm{d}x \le \frac{\sigma\sqrt{N}}{\sqrt{n}} + \frac{2(b-a)}{N}.$$

Now, let us rewrite the right-hand side as $\frac{\sigma}{\sqrt{n}}(\sqrt{N} + \frac{c}{N})$ with parameter $c = \frac{2(b-a)}{\sigma/\sqrt{n}}$ and optimize it over N. On the half-axis $x > 0$, introduce the function $\psi(x) = \sqrt{x} + \frac{c}{x}$ $(c > 0)$. It has derivative $\psi'(x) = \frac{1}{2\sqrt{x}} - \frac{c}{x^2}$, therefore ψ is decreasing on $(0, x_0]$ and is increasing on $[x_0, +\infty)$, where $x_0 = (2c)^{2/3}$. Hence, if $c \le \frac{1}{2}$, we have

$$\inf_N \psi(N) = \psi(1) = 1 + c \le 1 + c^{1/3}.$$

If $c \ge \frac{1}{2}$, then the argmin lies in $[1, +\infty)$. Choose $N = [x_0] + 1 = [(2c)^{2/3}] + 1$ so that $N \ge 2$ and $N - 1 \le x_0 < N \le x_0 + 1$. Hence, we get

$$\psi(N) \le \left(\sqrt{x_0} + 1\right) + \frac{c}{x_0} = 1 + \psi(x_0) = 1 + \frac{3}{2^{2/3}} c^{1/3}.$$

Thus, in both cases, $\inf_N \psi(N) \le 1 + \frac{3}{2^{2/3}} c^{1/3} \le 1 + 3(\frac{b-a}{\sigma/\sqrt{n}})^{1/3}$ and we arrive at the following corollary.

Corollary 3.2. *Under* $\mathrm{PI}(\sigma^2)$, *for all* $a < b$,

$$\mathbf{E}\int_a^b |F_n(x) - F(x)|\,\mathrm{d}x \le \frac{\sigma}{\sqrt{n}}\left[1 + 3\left(\frac{b-a}{\sigma/\sqrt{n}}\right)^{1/3}\right].$$

The next step is to extend the above inequality to the whole real line. Here, we shall use the exponential integrability of the measure F.

Proof of Theorem 1.1. Recall that the measure μ is controlled by using two independent parameters: the constant σ^2 and A, defined by

$$|\mathbf{E}X_i - \mathbf{E}X_j| \le A\sigma, \qquad 1 \le i, j \le n.$$

One may assume, without loss of generality, that $-A\sigma \le \mathbf{E}X_i \le A\sigma$ for all $i \le n$.

Lemma 2.2 with $g(x) = x_i$, $t = 1$ and Chebyshev's inequality give, for all $h > 0$,

$$\mathbf{P}\{X_i - \mathbf{E}X_i \ge h\} \le 3\mathrm{e}^{-h/\sigma}, \qquad \mathbf{P}\{X_i - \mathbf{E}X_i \le -h\} \le 3\mathrm{e}^{-h/\sigma}.$$

Therefore, whenever $h \ge A\sigma$,

$$\mathbf{P}\{X_i \ge h\} \le 3\mathrm{e}^{-(h-A\sigma)/\sigma}, \qquad \mathbf{P}\{X_i \le -h\} \le 3\mathrm{e}^{-(h-A\sigma)/\sigma}.$$

Averaging over all i's, we obtain similar bounds for the measure F, that is, $1 - F(h) \le 3\mathrm{e}^{-(h-A\sigma)/\sigma}$ and $F(-h) \le 3\mathrm{e}^{-(h-A\sigma)/\sigma}$. After integration, we get

$$\int_h^{+\infty} \big(1 - F(x)\big)\,\mathrm{d}x \le 3\sigma\mathrm{e}^{-(h-A\sigma)/\sigma}, \qquad \int_{-\infty}^{-h} F(x)\,\mathrm{d}x \le 3\sigma\mathrm{e}^{-(h-A\sigma)/\sigma}.$$

Using $|F_n(x) - F(x)| \le (1 - F_n(x)) + (1 - F(x))$ so that $\mathbf{E}|F_n(x) - F(x)| \le 2(1 - F(x))$, we get that

$$\mathbf{E}\int_h^{+\infty} |F_n(x) - F(x)|\,\mathrm{d}x \le 6\sigma\mathrm{e}^{-(h-A\sigma)/\sigma}$$

and similarly for the half-axis $(-\infty, -h)$. Combining this bound with Corollary 3.2, with $[a,b] = [-h,h]$, we obtain that, for all $h \geq A\sigma$,

$$\mathbf{E}\int_{-\infty}^{+\infty} |F_n(x) - F(x)|\,\mathrm{d}x \leq \frac{\sigma}{\sqrt{n}}\left[1 + 6\left(\frac{h}{\sigma/\sqrt{n}}\right)^{1/3}\right] + 12\sigma\mathrm{e}^{-(h-A\sigma)/\sigma}.$$

Substituting $h = (A+t)\sigma$ with arbitrary $t \geq 0$, we get that

$$\mathbf{E}\int_{-\infty}^{+\infty} |F_n(x) - F(x)|\,\mathrm{d}x \leq \frac{\sigma}{\sqrt{n}}\left[1 + 6\big((A+t)\sqrt{n}\big)^{1/3} + 12\sqrt{n}\mathrm{e}^{-t}\right].$$

Finally, the choice $t = \log n$ leads to the desired estimate

$$\mathbf{E}\int_{-\infty}^{+\infty} |F_n(x) - F(x)|\,\mathrm{d}x \leq C\sigma\left(\frac{A + \log n}{n}\right)^{1/3}. \qquad \square$$

4. Large deviations above the mean

In addition to the upper bound on the mean of the Kantorovich–Rubinstein distance $W_1(F_n, F)$, one may wonder how to bound large deviations of $W_1(F_n, F)$ above the mean. To this end, the following general observation may be helpful.

Lemma 4.1. *For all points $x = (x_1, \dots, x_n)$, $x' = (x'_1, \dots, x'_n)$ in $\mathbf{R}^n$, we have*

$$W_1(F_n, F'_n) \leq \frac{1}{\sqrt{n}}\|x - x'\|,$$

where $F_n = \frac{\delta_{x_1} + \cdots + \delta_{x_n}}{n}$, $F'_n = \frac{\delta_{x'_1} + \cdots + \delta_{x'_n}}{n}$.

In other words, the canonical map T from $\mathbf{R}^n$ to the space of all probability measures on the line, which assigns to each point an associated empirical measure, has a Lipschitz seminorm $\leq \frac{1}{\sqrt{n}}$ with respect to the Kantorovich–Rubinstein distance. As usual, the Euclidean space $\mathbf{R}^n$ is equipped with the Euclidean metric

$$\|x - x'\| = \sqrt{|x_1 - x'_1|^2 + \cdots + |x_n - x'_n|^2}.$$

Denote by Z_1 the collection of all (Borel) probability measures on the real line with finite first moment. The Kantorovich–Rubinstein distance in Z_1 may equivalently be defined (cf. [16,43]) by

$$W_1(G, G') = \inf_{\pi} \int |u - u'|\,\mathrm{d}\pi(u, u'),$$

where the infimum is taken over all (Borel) probability measures π on $\mathbf{R} \times \mathbf{R}$ with marginal distributions G and G'. In case of empirical measures $G = F_n$, $G' = F'_n$, associated to the points

$x, x' \in \mathbf{R}^n$, let π_0 be the discrete measure on the pairs (x_i, x'_i), $1 \le i \le n$, with point masses $\frac{1}{n}$. Therefore, by Cauchy's inequality,

$$W_1(Tx, Tx') \le \int |u - u'| \, \mathrm{d}\pi_0(u, u') = \frac{1}{n} \sum_{i=1} |x_i - x'_i| \le \frac{1}{\sqrt{n}} \left(\sum_{i=1} |x_i - x'_i|^2 \right)^{1/2}.$$

This proves Lemma 4.1.

Thus, the map $T : \mathbf{R}^n \to Z_1$ has the Lipschitz seminorm $\|T\|_{\mathrm{Lip}} \le \frac{1}{\sqrt{n}}$. As a consequence, given a probability measure μ on $\mathbf{R}^n$, this map transports many potential properties of μ, such as concentration, to the space Z_1, equipped with the Borel probability measure $\Lambda = \mu T^{-1}$. Note that it is supported on the set of all probability measures with at most n atoms. In particular, if μ satisfies a concentration inequality of the form

$$1 - \mu(A^h) \le \alpha(h), \qquad h > 0, \tag{4.1}$$

in the class of all Borel sets A in $\mathbf{R}^n$ with measure $\mu(A) \ge \frac{1}{2}$ (where A^h denotes an open Euclidean h-neighborhood of A), then Λ satisfies a similar (and in fact stronger) property

$$1 - \Lambda(B^h) \le \alpha\big(h\sqrt{n}\big), \qquad h > 0,$$

in the class of all Borel sets B in Z_1 with measure $\Lambda(B) \ge \frac{1}{2}$ (with respect to the Kantorovich–Rubinstein distance). In other words, an optimal so-called concentration function $\alpha = \alpha_\mu$ in (4.1) for the measure μ is related to the concentration function of Λ by

$$\alpha_\Lambda(h) \le \alpha_\mu\big(h\sqrt{n}\big), \qquad h > 0. \tag{4.2}$$

Now, in general, the concentration function has a simple functional description as

$$\alpha_\mu(h) = \sup \mu\{g - m(g) \ge h\},$$

where the sup is taken over all Lipschitz functions g on $\mathbf{R}^n$ with $\|g\|_{\mathrm{Lip}} \le 1$ and where $m(g)$ stands for a median of g under μ. (Actually, this holds for abstract metric spaces.) The concentration function may therefore be controlled by Poincaré-type inequalities in terms of σ^2 (the Gromov–Milman theorem). Indeed, since the quantity $g - m(g)$ is translation invariant, one may assume that g has mean zero. By Lemma 2.2 with $t = 1$, we get $\mu\{g \le -\sigma h\} \le 3\mathrm{e}^{-h} < \frac{1}{2}$, provided that $h > \log 6$, which means that any median of g satisfies $m(g) \ge -\sigma \log 6$. Therefore, again by Lemma 2.2, for any $h > \log 6$,

$$\mu\{g - m(g) \ge \sigma h\} \le \mu\{g \ge \sigma(h - \log 6)\} \le 3 \cdot 6 \cdot \mathrm{e}^{-h}$$

so that

$$\alpha_\mu(\sigma h) \le 18\mathrm{e}^{-h}. \tag{4.3}$$

The latter also automatically holds in the interval $0 \le h \le \log 6$. In fact, by a more careful application of the Poincaré-type inequality, the concentration bound (4.3) may be further improved to $\alpha_\mu(\sigma h) \le C\mathrm{e}^{-2h}$ (see [3]), but this is not crucial for our purposes.

Thus, combining (4.2) with (4.3), we may conclude that under PI(σ^2),

$$\alpha_\Lambda(h) \le 18\mathrm{e}^{-h\sqrt{n}/\sigma}, \qquad h > 0.$$

Now, in the setting of Theorem 1.1, consider on Z_1 the distance function $g(H) = W_1(H, F)$. It is Lipschitz (with Lipschitz seminorm 1) and has the mean $\mathbf{E}_\Lambda g = \mathbf{E}_\mu W_1(F_n, F) \le a$, where $a = C\sigma(\frac{A+\log n}{n})^{1/3}$. Hence, $m(g) \le 2a$ under the measure Λ and for any $h > 0$,

$$\Lambda\{g \ge 2a + h\} \le \Lambda\{g - m(g) \ge h\} \le \alpha_\Lambda(h) \le 18\mathrm{e}^{-h\sqrt{n}/\sigma}.$$

We can summarize as follows.

Proposition 4.2. *If a random vector* $(X_1, \dots, X_n)$ *in* $\mathbf{R}^n$, $n \ge 2$, *has distribution satisfying a Poincaré-type inequality with constant* σ^2, *then, for all* $h > 0$,

$$\mathbf{P}\left\{W_1(F_n, F) \ge C\sigma\left(\frac{A + \log n}{n}\right)^{1/3} + h\right\} \le C\mathrm{e}^{-h\sqrt{n}/\sigma}, \tag{4.4}$$

where $A = \frac{1}{\sigma}\max_{i,j} |\mathbf{E}X_i - \mathbf{E}X_j|$ *and where* C *is an absolute constant.*

Bounds such as (4.4) may be used to prove that the convergence holds almost surely at a certain rate. Here is a simple example, corresponding to non-varying values of the Poincaré constants. (One should properly modify the conclusion when applying this to the matrix scheme; see Section 7.) Let $(X_n)_{n\ge1}$ be a random sequence such that for each n, $(X_1, \dots, X_n)$ has distribution on $\mathbf{R}^n$ satisfying PI(σ^2) with some common σ.

Corollary 4.3. *If* $\max_{i,j\le n} |\mathbf{E}X_i - \mathbf{E}X_j| = \mathrm{O}(\log n)$, *then* $W_1(F_n, F) = \mathrm{O}(\frac{\log n}{n})^{1/3}$ *with probability* 1.

Note, however, that in the scheme of sequences such as in Corollary 4.3, the mean distribution function $F = \mathbf{E}F_n$ might also depend on n.

By a similar contraction argument, the upper bound (4.4) may be sharpened, when the distribution of $(X_1, \dots, X_n)$ satisfies a logarithmic Sobolev inequality. We turn to this type of (stronger) hypothesis in the next section.

Remarks. Let Ω be a metric space and let $d = d(u, u')$ be a non-negative continuous function on the product space $\Omega \times \Omega$. Given Borel probability measures G and G' on Ω, the generalized Kantorovich–Rubinstein or Wasserstein 'distance' with cost function d is defined by

$$W(G, G') = \inf_\pi \int d(u, u')\,\mathrm{d}\pi(u, u'),$$

where the infimum is taken over all probability measures π on $\Omega \times \Omega$ with marginal distributions G and G'. In the case of the real line $\Omega = \mathbf{R}$ with cost function of the form $d(u, u') =$

$\varphi(u-u')$, where φ is convex, this quantity has a simple description,

$$W(G,G') = \int_0^1 \varphi\big(G^{-1}(t) - G'^{-1}(t)\big)\,\mathrm{d}t, \tag{4.5}$$

in terms of the inverse distribution functions $G^{-1}(t) = \min\{x \in \mathbf{R} : G(x) \ge t\}$; see, for example, [13] and [37], Theorem 2.

If $\varphi(u,u') = |u-u'|$, then we also have the L^1-representation for $W_1(G,G')$, which we use from the very beginning as our definition. Moreover, for arbitrary discrete measures $G = F_n = (\delta_{x_1} + \cdots + \delta_{x_n})/n$ and $G' = F'_n = (\delta_{x'_1} + \cdots + \delta_{x'_n})/n$, as in Lemma 4.1, the expression (4.5) is reduced to

$$W_1(F_n, F'_n) = \frac{1}{n}\sum_{i=1}^n |x_i - x'_i|, \tag{4.6}$$

where we assume that $x_1 \le \cdots \le x_n$ and $x'_1 \le \cdots \le x'_n$.

Now, for an arbitrary random vector $X = (X_1,\ldots,X_n)$ in $\mathbf{R}^n$, consider the ordered statistics $X_1^* \le \cdots \le X_n^*$. Equation (4.6) then yields

$$\mathbf{E}W_1(F_n, F'_n) = \frac{1}{n}\sum_{i=1}^n \mathbf{E}|X_i^* - (X'_i)^*|, \tag{4.7}$$

where $(X'_1)^* \le \cdots \le (X'_n)^*$ are ordered statistics generated by an independent copy of X and where F'_n are independent copies of the (random) empirical measures F_n associated with X. By the triangle inequality for the metric W_1, we have

$$\mathbf{E}W_1(F_n, F'_n) \le \mathbf{E}W_1(F_n, F) + \mathbf{E}W_1(F, F'_n) = 2\mathbf{E}W_1(F_n, F).$$

It is applied with the mean distribution function $F = \mathbf{E}F_n$. On the other hand, any function of the form $H \to W_1(G,H)$ is convex on the convex set Z_1, so, by Jensen's inequality, $\mathbf{E}W_1(F_n, F'_n) \ge \mathbf{E}W_1(F_n, \mathbf{E}F'_n) = \mathbf{E}W_1(F_n, F)$. The two bounds give

$$\mathbf{E}W_1(F_n, F) \le \mathbf{E}W_1(F_n, F'_n) \le 2\mathbf{E}W_1(F_n, F). \tag{4.8}$$

By a similar argument,

$$\mathbf{E}|X_i^* - \mathbf{E}X_i^*| \le \mathbf{E}|X_i^* - (X'_i)^*| \le 2\mathbf{E}|X_i^* - \mathbf{E}X_i^*|. \tag{4.9}$$

Combining (4.8) and (4.9) and recalling (4.7), we arrive at the two-sided estimate

$$\frac{1}{2n}\sum_{i=1} \mathbf{E}|X_i^* - \mathbf{E}X_i^*| \le \mathbf{E}W_1(F_n, F) \le \frac{2}{n}\sum_{i=1} \mathbf{E}|X_i^* - \mathbf{E}X_i^*|,$$

which is exactly the inequality (1.4) mentioned in the Introduction. Similar two-sided estimates also hold for other cost functions in the Wasserstein distance.

5. Empirical log-Sobolev inequalities

As before, let $(X_1, \ldots, X_n)$ be a random vector in $\mathbf{R}^n$ with joint distribution μ. Similarly to Proposition 2.1, now using a log-Sobolev inequality for μ, we arrive at the following 'empirical' log-Sobolev inequality.

Proposition 5.1. *Under* $\mathrm{LSI}(\sigma^2)$, *for any bounded, smooth function* f *on* $\mathbf{R}$,

$$\mathrm{Ent}_\mu\bigg[\bigg(\int f\,\mathrm{d}F_n\bigg)^2\bigg] \le \frac{2\sigma^2}{n}\int f'^2\,\mathrm{d}F.$$

In analogy with Poincaré-type inequalities, one may also develop refined applications to the rate of growth of moments and to large deviations of various functionals of empirical measures. In particular, we have the following proposition.

Proposition 5.2. *Under* $\mathrm{LSI}(\sigma^2)$, *for any smooth function* f *on* $\mathbf{R}$ *such that* f' *belongs to* $L^p(\mathbf{R}, \mathrm{d}F)$, $p \ge 2$,

$$\mathbf{E}\bigg|\int f\,\mathrm{d}F_n - \int f\,\mathrm{d}F\bigg|^p \le \frac{(\sigma\sqrt{p})^p}{n^{p/2}}\int |f'|^p\,\mathrm{d}F. \tag{5.1}$$

In addition, if $|f'| \le 1$, *then, for all* $h > 0$,

$$\mu\bigg\{\bigg|\int f\,\mathrm{d}F_n - \int f\,\mathrm{d}F\bigg| \ge h\bigg\} \le 2\mathrm{e}^{-nh^2/2\sigma^2}. \tag{5.2}$$

The proof of the second bound, (5.2), which was already noticed in [24] in the context of random matrices, follows the standard Herbst's argument; see [29] and [6]. The first family of moment inequalities, (5.1), can be sharpened by one inequality on the Laplace transform, such as

$$\mathbf{E}\exp\bigg\{\int f\,\mathrm{d}F_n - \int f\,\mathrm{d}F\bigg\} \le \mathbf{E}\exp\bigg\{\frac{\sigma^2}{n}\int |f'|^2\,\mathrm{d}F_n\bigg\}.$$

The proof is immediate, by [6], Theorem 1.2.

However, a major weak point in both Poincaré and log-Sobolev inequalities, including their direct consequences, as in Proposition 5.2, is that they may not be applied to indicator and other non-smooth functions. In particular, we cannot estimate directly at fixed points the variance $\mathrm{Var}(F_n(x))$ or other similar quantities like the higher moments of $|F_n(x) - F(x)|$. Therefore, we need another family of analytic inequalities. Fortunately, the so-called infimum-convolution operator and associated relations concerning arbitrary measurable functions perfectly fit our purposes. Moreover, some of the important relations hold true and may be controlled in terms of the constant involved in the logarithmic Sobolev inequalities.

Let us now turn to the important concept of infimum- and supremum-convolution inequalities. They were proposed in 1991 by Maurey [33] as a functional approach to some of Talagrand's concentration results concerning product measures. Given a parameter $t > 0$ and a real-valued function g on $\mathbf{R}^n$ (possibly taking the values $\pm\infty$), put

$$Q_t g(x) = \inf_{y \in \mathbf{R}^n} \left[g(y) + \frac{1}{2t} \|x - y\|^2 \right],$$

$$P_t g(x) = \sup_{y \in \mathbf{R}^n} \left[g(y) - \frac{1}{2t} \|x - y\|^2 \right].$$

$Q_t g$ and $P_t g$ then represent, respectively, the infimum- and supremum-convolution of g with cost function being the normalized square of the Euclidean norm in $\mathbf{R}^n$. By definition, one puts $Q_0 g = P_0 g = g$.

For basic definitions and basic properties of the infimum- and supremum-convolution operators, we refer the reader to [19] and [5], mentioning just some of them here. These operators are dually related by the property that for any functions f and g on $\mathbf{R}^n$, $g \geq P_t f \iff f \leq Q_t g$. Clearly, $P_t(-g) = -Q_t g$. Thus, in many statements, it is sufficient to consider only one of these operators. The basic semigroup property of both operators is that for any g on $\mathbf{R}^n$ and $t, s \geq 0$,

$$Q_{t+s} g = Q_t Q_s g, \qquad P_{t+s} g = P_t P_s g.$$

For any function g and $t > 0$, the function $P_t g$ is always lower semicontinuous, while $Q_t g$ is upper semicontinuous. If g is bounded, then $P_t g$ and $Q_t g$ are bounded and have finite Lipschitz seminorms. In particular, both are differentiable almost everywhere.

Given a bounded function g and $t > 0$, for almost all $x \in \mathbf{R}^n$, the functions $t \to P_t g(x)$ and $t \to Q_t g(x)$ are differentiable at t and

$$\frac{\partial P_t g(x)}{\partial t} = \frac{1}{2} \|\nabla P_t g(x)\|^2, \qquad \frac{\partial Q_t g(x)}{\partial t} = -\frac{1}{2} \|\nabla Q_t g(x)\|^2 \qquad \text{a.e.}$$

In other words, the operator $\Gamma g = \frac{1}{2} |\nabla g|^2$ appears as the generator for the semigroup P_t, while $-\Gamma$ appears as the generator for Q_t. As a result, $u(x, t) = Q_t g(x)$ represents the solution to the Hamilton–Jacobi equation $\frac{\partial u}{\partial t} = -\frac{1}{2} \|\nabla u\|^2$ with initial condition $u(x, 0) = g(x)$.

Below, we separately formulate a principal result of [5] which relates logarithmic Sobolev inequalities to supremum- and infimum-convolution operators.

Lemma 5.3. *Let μ be a probability measure on $\mathbf{R}^n$ satisfying* $\mathrm{LSI}(\sigma^2)$. *For any μ-integrable Borel-measurable function g on $\mathbf{R}^n$, we have*

$$\int P_{\sigma^2} g \, \mathrm{d}\mu \geq \log \int \mathrm{e}^g \, \mathrm{d}\mu \tag{5.3}$$

and, equivalently,

$$\int g \, \mathrm{d}\mu \geq \log \int \mathrm{e}^{Q_{\sigma^2} g} \, \mathrm{d}\mu. \tag{5.4}$$

Alternatively, for further applications to empirical measures, one could start from the infimum-convolution inequalities (5.3) and (5.4), taking them as the main hypothesis on the measure μ. They take an intermediate position between Poincaré and logarithmic Sobolev inequalities. However, logarithmic Sobolev inequalities have been much better studied, with a variety of sufficient conditions having been derived.

Now, as in Section 2, we apply the relations (5.3) and (5.4) to functions $g(x_1,\dots,x_n) = \int f\,\mathrm{d}F_n$, where F_n is the empirical measure defined for 'observations' $x_1,\dots,x_n$. By the very definition, for any $t>0$,

$$\begin{aligned} P_t g(x_1,\dots,x_n) &= \sup_{y_1,\dots,y_n\in\mathbf{R}} \left[g(y_1,\dots,y_n) - \frac{1}{2t}\sum_{i=1}^n |x_i-y_i|^2 \right] \\ &= \frac{1}{n}\sup_{y_1,\dots,y_n\in\mathbf{R}} \sum_{i=1}^n \left[f(y_i) - \frac{1}{2t/n}|x_i-y_i|^2 \right] = \int P_{t/n} f\,\mathrm{d}F_n. \end{aligned}$$

Similarly, $Q_t g = \int Q_{t/n} f\,\mathrm{d}F_n$. Therefore, after integration with respect to μ and using the identity $P_a(tf) = tP_{ta}f$, we arrive at corresponding empirical supremum- and infimum-convolution inequalities, as follows.

Proposition 5.4. *Under* LSI(σ^2), *for any F-integrable Borel-measurable function f on* $\mathbf{R}$ *and for any $t>0$,*

$$\log \mathbf{E} e^{t(\int f\,\mathrm{d}F_n - \int f\,\mathrm{d}F)} \le t\int [P_{t\sigma^2/n} f - f]\,\mathrm{d}F, \tag{5.5}$$

$$\log \mathbf{E} e^{t(\int f\,\mathrm{d}F - \int f\,\mathrm{d}F_n)} \le t\int [f - Q_{t\sigma^2/n} f]\,\mathrm{d}F. \tag{5.6}$$

Note that the second inequality may be derived from the first by changing f to $-f$.

6. Local behavior of empirical distributions

In this section, we develop a few direct applications of Proposition 5.4 to the behavior of empirical distribution functions $F_n(x)$ at a fixed point. Such functionals are linear, that is, of the form $\int f\,\mathrm{d}F_n$, corresponding to the indicator function of the half-axis $f = (-\infty, x]$. When f is smooth, Proposition 5.2 tells us that the deviations of $L_n f = \int f\,\mathrm{d}F_n - \int f\,\mathrm{d}F$ are of order $\sigma/\sqrt{n}$. In the general non-smooth case, the infimum- and supremum-convolution operators $P_t f$ and $Q_t f$ behave differently for small values of t and this results in a different rate of fluctuation for $L_n f$.

To see this, let $f = (-\infty, x]$. In this case, the functions $P_t f$ and $Q_t f$ may easily be computed explicitly, but we do not lose much by using the obvious bounds

$$1_{(-\infty, x-\sqrt{2t}]} \le Q_t f \le P_t f \le 1_{(-\infty, x+\sqrt{2t}]}.$$

Therefore, (5.5) and (5.6) yield the following proposition.

Proposition 6.1. *Under* $\mathrm{LSI}(\sigma^2)$, *for any* $x \in \mathbf{R}$ *and* $t > 0$, *with* $h = \sqrt{\frac{2\sigma^2 t}{n}}$,

$$\log \mathbf{E} e^{t(F_n(x)-F(x))} \le t\big(F(x+h) - F(x)\big), \tag{6.1}$$

$$\log \mathbf{E} e^{t(F(x)-F_n(x))} \le t\big(F(x) - F(x-h)\big). \tag{6.2}$$

These estimates may be used to sharpen Corollary 3.2 and therefore to recover Theorem 1.1 (under the stronger hypothesis on the joint distribution μ, however). Indeed, for any $t > 0$,

$$\mathbf{E} e^{t|F_n(x)-F(x)|} \le \mathbf{E} e^{t(F_n(x)-F(x))} + \mathbf{E} e^{-t(F(x)-F_n(x))} \le 2 e^{t(F(x+h)-F(x-h))}.$$

Taking the logarithm and applying Jensen's inequality, we arrive at

$$\mathbf{E}|F_n(x) - F(x)| \le \big(F(x+h) - F(x-h)\big) + \frac{\log 2}{t}.$$

Now, just integrate this inequality over an arbitrary interval (a, b), $a < b$, and use the general relation $\int_{-\infty}^{+\infty} (F(x+h) - F(x-h))\,\mathrm{d}x \le 2h$ to obtain that

$$\mathbf{E} \int_a^b |F_n(x) - F(x)|\,\mathrm{d}x \le 2h + \frac{\log 2}{t}(b-a) = 2\sqrt{\frac{2\sigma^2 t}{n}} + \frac{\log 2}{t}(b-a).$$

Optimization over t leads to an improved version of Corollary 3.2.

Corollary 6.2. *Under* $\mathrm{LSI}(\sigma^2)$, *for all* $a < b$,

$$\mathbf{E} \int_a^b |F_n(x) - F(x)|\,\mathrm{d}x \le 4\left(\frac{\sigma^2(b-a)}{n}\right)^{1/3}.$$

Note that in both cases of Proposition 6.1, for any $t \in \mathbf{R}$,

$$\log \mathbf{E} e^{t(F_n(x)-F(x))} \le |t|\big(F(x+h) - F(x-h)\big), \qquad h = \sqrt{2\sigma^2|t|/n}.$$

Hence, the local behavior of the distribution function F near a given point x turns out to be responsible for the large deviation behavior at this point of the empirical distribution function F_n around its mean.

For a quantitative statement, assume that F has a finite Lipschitz constant $M = \|F\|_{\mathrm{Lip}}$, so it is absolutely continuous with respect to Lebesgue measure on the real line and has a density, bounded by M. It follows from (6.1), with $t = (\alpha n^{1/3})\lambda$ and $\alpha^3 = \frac{2}{9M^2\sigma^2}$, that

$$\mathbf{E} e^{\lambda\xi} \le e^{(2/3)|\lambda|^{3/2}}, \qquad \lambda \in \mathbf{R},$$

where $\xi = \alpha n^{1/3}(F_n(x) - F(x))$. By Chebyshev's inequality, for any $r > 0$,

$$\mu\{\xi \ge r\} \le e^{(2/3)\lambda^{3/2} - \lambda r} = e^{-r^3/3}, \qquad \text{where } \lambda = r^2.$$

Similarly, $\mu\{\xi \le -r\} \le e^{-r^3/3}$. Therefore, $\mu\{\alpha n^{1/3}|F_n(x) - F(x)| \ge r\} \le 2e^{-r^3/3}$. Changing the variable, we are finished.

Recall that we use the quantity

$$\beta = \frac{(M\sigma)^{2/3}}{n^{1/3}}.$$

Proposition 6.3. *Assume that F has a density, bounded by a number M. Under* $\mathrm{LSI}(\sigma^2)$, *for any $x \in \mathbf{R}$ and $r > 0$,*

$$\mathbf{P}\{|F_n(x) - F(x)| \ge \beta r\} \le 2e^{-2r^3/27}. \tag{6.3}$$

In particular, with some absolute constant C, we have

$$\mathbf{E}|F_n(x) - F(x)| \le C\beta. \tag{6.4}$$

Note that (6.4) is consistent with the estimate of Theorem 1.1. To derive similar bounds on the uniform (Kolmogorov) distance $\|F_n - F\| = \sup_x |F_n(x) - F(x)|$ (which we discuss in the next section), it is better to split the bound (6.3) into the two parts,

$$\mathbf{P}\{F_n(x) - F(x) \ge \beta r\} \le e^{-2r^3/27}, \tag{6.5}$$

$$\mathbf{P}\{F(x) - F_n(x) \ge \beta r\} \le e^{-2r^3/27}, \tag{6.6}$$

which were obtained in the last step of the proof of (6.3).

However, since one might not know whether F is Lipschitz or how it behaves locally, and since one might want to approximate this measure itself by some canonical distribution G, it is reasonable to provide a more general statement. By Proposition 6.1, for any $t \ge 0$,

$$\begin{aligned}\log \mathbf{E}e^{t(F_n(x) - G(x))} &\le t\big(F(x+h) - G(x)\big) \\ &\le t\big(G(x+h) - G(x)\big) + t\|F - G\|\end{aligned}$$

and, similarly,

$$\log \mathbf{E}e^{t(G(x) - F_n(x))} \le t\big(G(x) - G(x-h)\big) + t\|F - G\|.$$

Repeating the preceding argument with the random variable

$$\xi = \alpha n^{1/3}\big(F_n(x) - G(x) - \|F - G\|\big)$$

and then interchanging F_n and G, we get a more general version of Proposition 6.3.

Proposition 6.4. *Under* $\mathrm{LSI}(\sigma^2)$, *for any distribution function G with finite Lipschitz seminorm $M = \|G\|_{\mathrm{Lip}}$, for any $x \in \mathbf{R}$ and $r > 0$,*

$$\mathbf{P}\{|F_n(x) - G(x)| \ge \beta r + \|F - G\|\} \le 2e^{-2r^3/27},$$

where $\beta = (M\sigma)^{2/3} n^{-1/3}$. *In particular, up to some absolute constant* C,

$$\mathbf{E}|F_n(x) - G(x)| \le C\beta + \|F - G\|. \tag{6.7}$$

Let us stress that in all of these applications of Proposition 5.4, only the indicator functions $f = 1_{(-\infty,x]}$ were used. One may therefore try to get more information about deviations of the empirical distributions F_n from the mean F by applying the basic bounds (5.5) and (5.6) with different (non-smooth) functions f.

For example, of considerable interest is the so-called local regime, where one tries to estimate the number

$$N_I = \mathrm{card}\{i \le n : X_i \in I\}$$

of observations inside a small interval $I = [x, x + \varepsilon]$ and to take into account the size of the increment $\varepsilon = |I|$. In case of i.i.d. observations, this may done using various tools; already, the formula

$$\mathrm{Var}(F_n(I)) = \frac{1}{n} F(I)\big(1 - F(I)\big) \le \frac{F(x+\varepsilon) - F(x)}{n}$$

suggests that when F is Lipschitz, $F_n(I)$ has small oscillations for small ε (where F_n and F are treated as measures).

However, the infimum- and supremum-convolution operators $P_t f$ and $Q_t f$ do not provide such information. Indeed, for the indicator function $f = 1_I$, by (5.5), we only have, similarly to Proposition 6.1, that

$$\log \mathbf{E} e^{t(F_n(I) - F(I))} \le t\big[\big(F(x+\varepsilon+h) - F(x)\big) - \big(F(x) - F(x-h)\big)\big],$$

where $t > 0$ and $h = \sqrt{\frac{2\sigma^2 t}{n}}$. Here, when h is fixed and $\varepsilon \to 0$, the right-hand side is not vanishing, in contrast with the i.i.d. case. This also shows that standard chaining arguments, such as Dudley's entropy bound or more delicate majorizing measure techniques (described, e.g., in [40]), do not properly work through the infimum-convolution approach.

Nevertheless, the above estimate is still effective for ε of order h, so we can control deviations of $F_n(I) - F(I)$ relative to $|I|$ when the intervals are not too small. This can be done with the arguments used in the proof of Proposition 6.3 or, alternatively (although with worse absolute constants), one can use the inequality (6.3), by applying it to the points x and $x + \varepsilon$. This immediately gives that

$$\mathbf{P}\{|F_n(I) - F(I)| \ge 2\beta r\} \le 4e^{-2r^3/27}.$$

Changing variables, one may rewrite the above in terms of N_I as

$$\mathbf{P}\{|N_I - nF(I)| \ge n\delta|I|\} \le 4\exp\left\{-c\left(\frac{\delta|I|}{\beta}\right)^3\right\}$$

with $c = 1/112$. Note that the right-hand side is small only when $|I| \gg \beta/\delta$, which is of order $n^{-1/3}$ with respect to the number of observations.

This can further be generalized if we apply Proposition 6.4.

Corollary 6.5. *Let G be a distribution function with density $g(x)$ bounded by a number M. Under* $\mathrm{LSI}(\sigma^2)$, *for any $\delta > 0$ and any interval I of length $|I| \geq 4\|F - G\|/\delta$,*

$$\mathbf{P}\left\{\left|N_I - n\int_I g(x)\,\mathrm{d}x\right| \geq n\delta|I|\right\} \leq 4\exp\left\{-c\left(\frac{\delta|I|}{\beta}\right)^3\right\},$$

where $\beta = (M\sigma)^{2/3}n^{-1/3}$ and $c > 0$ is an absolute constant.

Hence, if

$$|I| \geq \frac{C}{\delta}\max\{\beta, \|F - G\|\}$$

and $C > 0$ is large, then, with high probability, we have that $|\frac{N_I}{n} - \int_I g(x)\,\mathrm{d}x| \leq \delta|I|$.

7. Bounds on the Kolmogorov distance. Proof of Theorem 1.2

As before, let F_n denote the empirical measure associated with observations $x_1, \ldots, x_n$ and $F = \mathbf{E}F_n$ their mean with respect to a given probability measure μ on $\mathbf{R}^n$. In this section, we derive uniform bounds on $F_n(x) - F(x)$, based on Proposition 6.3, and thus prove Theorem 1.2. For applications to the matrix scheme, we shall also replace F, which may be difficult to determine, by the well-behaving limit law G (with the argument relying on Proposition 6.4).

Let the random variables $X_1, \ldots, X_n$ have joint distribution μ, satisfying $\mathrm{LSI}(\sigma^2)$, and assume that F has a finite Lipschitz seminorm $M = \|F\|_{\mathrm{Lip}}$. Define

$$\beta = \frac{(M\sigma)^{2/3}}{n^{1/3}}.$$

Proof of Theorem 1.2. We use the inequalities (6.5) and (6.6) to derive an upper bound on $\|F_n - F\| = \sup_x |F_n(x) - F(x)|$. (For the sake of extension of Theorem 1.2 to Theorem 7.1 below, we relax the argument and do not assume that F is continuous.)

So, fix $r > 0$ and an integer $N \geq 2$. One can always pick up points $-\infty = x_0 \leq x_1 \leq \cdots \leq x_{N-1} \leq x_N = +\infty$ with the property that

$$F(x_i-) - F(x_{i-1}) \leq \frac{1}{N}, \qquad i = 1, \ldots, N. \tag{7.1}$$

Note that $F_n(x_0) = F_n(x_0) = 0$ and $F_n(x_N) = F_n(x_N) = 1$. It then follows from (6.5) that

$$\mathbf{P}\left\{\max_{1\leq i\leq N}[F_n(x_i-) - F(x_i-)] \geq \beta r\right\} \leq (N-1)\mathrm{e}^{-2r^3/27}$$

and, similarly, by (6.6),

$$\mathbf{P}\left\{\max_{1\leq i\leq N}[F(x_{i-1}) - F_n(x_{i-1})] \geq \beta r\right\} \leq (N-1)\mathrm{e}^{-2r^3/27}.$$

Hence, for the random variable

$$\xi_N = \max\Big\{\max_{1\le i\le N}[F_n(x_i-) - F(x_i-)],\ \max_{1\le i\le N}[F(x_{i-1}) - F_n(x_{i-1})]\Big\},$$

we have that

$$\mathbf{P}\{\xi_N \ge \beta r\} \le 2(N-1)\mathrm{e}^{-2r^3/27}. \tag{7.2}$$

Now, take any point $x \in \mathbf{R}$ different from all of the x_j's and select i from $1, \dots, n$ such that $x_{i-1} < x < x_i$. Then, by (7.1),

$$\begin{aligned} F_n(x) - F(x) &\le F_n(x_i-) - F(x_{i-1}) \\ &= [F_n(x_i-) - F(x_i-)] + [F(x_i-) - F(x_{i-1})] \le \xi_N + \frac{1}{N}. \end{aligned}$$

Similarly,

$$\begin{aligned} F(x) - F_n(x) &\le F(x_i-) - F_n(x_{i-1}) \\ &= [F(x_i-) - F(x_{i-1})] + [F(x_{i-1}) - F_n(x_{i-1})] \le \xi_N + \frac{1}{N}. \end{aligned}$$

Therefore, $|F_n(x) - F(x)| \le \xi_N + \frac{1}{N}$, which also extends by continuity from the right to all points x_j. Thus, $\|F_n - F\| \le \xi_N + \frac{1}{N}$ and, by (7.2),

$$\mathbf{P}\Big\{\|F_n - F\| > \beta r + \frac{1}{N}\Big\} \le 2(N-1)\mathrm{e}^{-2r^3/27}.$$

Note that this also holds automatically in the case $N = 1$. Choose $N = [\frac{1}{\beta r}] + 1$. We then have $\frac{1}{N} \le \beta r$ and get

$$\mathbf{P}\{\|F_n - F\| > 2\beta r\} \le \frac{2}{\beta r}\mathrm{e}^{-2r^3/27}.$$

Finally, changing $2\beta r$ into r, we arrive at the bound (1.5) of Theorem 1.2,

$$\mathbf{P}\{\|F_n - F\| > r\} \le \frac{4}{r}\exp\Big\{-\frac{2}{27}\Big(\frac{r}{\beta}\Big)^3\Big\}, \tag{7.3}$$

and so the constant $c = 2/27$.

It remains to derive the bound on the mean $\mathbf{E}\|F_n - F\|$. Given $0 \le r_0 \le 1$, we can write, using (7.3),

$$\begin{aligned} \mathbf{E}\|F_n - F\| &= \int_0^1 \mu\{\|F_n - F\| > r\}\,\mathrm{d}r = \int_0^{r_0} + \int_{r_0}^1 \\ &\le r_0 + \frac{4}{r_0}\exp\Big\{-\frac{2}{27}\Big(\frac{r_0}{\beta}\Big)^3\Big\}. \end{aligned} \tag{7.4}$$

This bound also holds for $r_0 > 1$.

First, assume that $0 < \beta \le 1$ and choose $r_0 = 3\beta\log(1+\frac{1}{\beta})$. Then, for the last term in (7.4), we have

$$\begin{aligned}\frac{4}{r_0}\exp\left\{-\frac{2}{27}\left(\frac{r_0}{\beta}\right)^3\right\} &= \frac{4}{3\beta\log(1+1/\beta)}\mathrm{e}^{-2\log(1+1/\beta)}\\ &= \frac{4}{3}\frac{\beta}{(1+\beta)^2\log^{1/3}(1+1/\beta)}\\ &\le B\beta\log^{1/3}\left(1+\frac{1}{\beta}\right)\end{aligned}$$

with some constant B satisfying $(1+\beta)^3\log(1+\frac{1}{\beta}) \ge (\frac{4}{3B})^{3/2}$. For example, we can take $B = 2$ and then, by (7.4), we have

$$\mathbf{E}\|F_n - F\| \le 5\beta\log^{1/3}\left(1+\frac{1}{\beta}\right). \tag{7.5}$$

As for the values $\beta \ge 1$, simple calculations show that the right-hand side of (7.5) is greater than 1, so the inequality (1.6) is fulfilled with $C = 5$.

Theorem 1.2 is therefore proved. □

Remark. If $M\sigma$ in Theorem 1.2 were of order 1, then $\mathbf{E}\|F_n - F\|$ would be of order at most $(\frac{\log n}{n})^{1/3}$. Note, however, that under $\mathrm{PI}(\sigma^2)$, and if all $\mathbf{E}X_i = 0$, the quantity $M\sigma$ is separated from zero and, more precisely, $M\sigma \ge \frac{1}{\sqrt{12}}$.

Indeed, by Hensley's theorem in dimension 1 [1,26], in the class of all probability densities $p(x)$ on the line, the expression $(\int x^2 p(x)\,\mathrm{d}x)^{1/2}\operatorname{ess\,sup}_x p(x)$ is minimized for the uniform distribution on symmetric intervals and is therefore bounded from below by $1/\sqrt{12}$. Since F is Lipschitz, it has a density p with $M = \operatorname{ess\,sup}_x p(x)$. On the other hand, it follows from the Poincaré-type inequality that $\sigma^2 \ge \mathrm{Var}(X_i) = \mathbf{E}X_i^2$. Averaging over all i's, we get $\sigma^2 \ge \int x^2\,\mathrm{d}F(x)$, so $M\sigma \ge (\int x^2 p(x)\,\mathrm{d}x)^{1/2}\operatorname{ess\,sup}_x p(x)$.

With similar arguments based on Proposition 6.4, we also obtain the following generalization of Theorem 1.2.

Theorem 7.1. *Assume that $X_1, \dots, X_n$ have a distribution on $\mathbf{R}^n$ satisfying* $\mathrm{LSI}(\sigma^2)$. *Let G be a distribution function with finite Lipschitz seminorm M. Then, for all $r > 0$,*

$$\mathbf{P}\{\|F_n - G\| \ge r + \|F - G\|\} \le \frac{4}{r}\exp\left\{-\frac{2}{27}\left(\frac{r}{\beta}\right)^3\right\},$$

where $\beta = (M\sigma)^{2/3}n^{-1/3}$. In particular,

$$\mathbf{E}\|F_n - G\| \le 5\beta\log^{1/3}\left(1+\frac{1}{\beta}\right) + \|F - G\|.$$

Remarks. It remains unclear whether or not one can involve the i.i.d. case in the scheme of Poincaré or logarithmic Sobolev inequalities to recover the rate $1/\sqrt{n}$ for $\mathbf{E}\|F_n - F\|$, even if some further natural assumptions are imposed (which are necessary, as we know from Examples 1 and 2). In particular, one may assume that the quantities M and σ are of order 1, and that $\mathbf{E}X_i = 0$, $\mathrm{Var}(X_i) = 1$. The question is the following: under, say, LSI(σ^2), is it true that

$$\mathbf{E}\|F_n - F\| \le \frac{C}{\sqrt{n}}$$

with some absolute C? Or at least $\mathbf{E}|F_n(x) - F(x)| \le \frac{C}{\sqrt{n}}$ for individual points?

8. High-dimensional random matrices

We shall now apply the bounds obtained in Theorems 1.1 and 7.1, to the case of the spectral empirical distributions. Let $\{\xi_{jk}\}_{1\le j\le k\le n}$ be a family of independent random variables on some probability space with mean $\mathbf{E}\xi_{jk} = 0$ and variance $\mathrm{Var}(\xi_{jk}) = 1$. Put $\xi_{jk} = \xi_{kj}$ for $1 \le k < j \le n$ and introduce a symmetric $n \times n$ random matrix,

$$\Xi = \frac{1}{\sqrt{n}} \begin{pmatrix} \xi_{11} & \xi_{12} & \cdots & \xi_{1n} \\ \xi_{21} & \xi_{22} & \cdots & \xi_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \xi_{n1} & \xi_{n2} & \cdots & \xi_{nn} \end{pmatrix}.$$

Arrange its (real random) eigenvalues in increasing order: $X_1 \le \cdots \le X_n$. As before, we associate with particular values $X_1 = x_1, \ldots, X_n = x_n$ an empirical (spectral) measure F_n with mean (expected) measure $F = \mathbf{E}F_n$.

An important point in this scheme is that the joint distribution μ of the spectral values, as a probability measure on $\mathbf{R}^n$, represents the image of the joint distribution of ξ_{jk}'s under a Lipschitz map T with Lipschitz seminorm $\|T\|_{\mathrm{Lip}} = \frac{\sqrt{2}}{\sqrt{n}}$. More precisely, by the Hoffman–Wielandt theorem with respect to the Hilbert–Schmidt norm, we have

$$\sum_{i=1}^n |X_i - X_i'|^2 \le \|\Xi - \Xi'\|_{\mathrm{HS}}^2 = \frac{1}{n} \sum_{j,k=1}^n |\xi_{jk} - \xi_{jk}'|^2 \le \frac{2}{n} \sum_{1\le j\le k\le n} |\xi_{jk} - \xi_{jk}'|^2$$

for any collections $\{\xi_{jk}\}_{j\le k}$ and $\{\xi_{jk}\}_{j\le k}'$ with eigenvalues $(X_1, \ldots, X_n)$, $(X_1', \ldots, X_n')$, respectively. This is a well-known fact ([2], page 165) which may be used in concentration problems; see, for example, [15,30].

In particular (see Proposition A1 in the Appendix), if the distributions of ξ_{jk}'s satisfy a one-dimensional Poincaré-type inequality with common constant σ^2, then μ satisfies a Poincaré-type inequality with an asymptotically much better constant $\sigma_n^2 = \frac{2\sigma^2}{n}$. According to Theorem 1.1,

$$\mathbf{E}\int_{-\infty}^{+\infty} |F_n(x) - F(x)|\,\mathrm{d}x \le C\sigma_n \left(\frac{A_n + \log n}{n}\right)^{1/3},$$

where C is an absolute constant and $A_n = \frac{1}{\sigma_n} \max_{i,j} |\mathbf{E}X_i - \mathbf{E}X_j|$. Since $\max_i |\mathbf{E}X_i|$ is of order at most σ, A_n is at most $\sqrt{n}$ and we arrive at the bound (1.7) in Theorem 1.3:

$$\mathbf{E}\int_{-\infty}^{+\infty} |F_n(x) - F(x)|\,\mathrm{d}x \le \frac{C\sigma}{n^{2/3}}.$$

Now, let us explain the second statement of Theorem 1.3 for the case where the ξ_{jk}'s satisfy a logarithmic Sobolev inequality with a common constant σ^2, in addition to the normalizing conditions $\mathbf{E}\xi_{jk} = 0$, $\mathrm{Var}(\xi_{jk}) = 1$ (which implies that $\sigma \ge 1$). Let G denote the standard semicircle law with variance 1, that is, with density $g(x) = \frac{1}{2\pi}\sqrt{4 - x^2}$, $-2 < x < 2$. In this case, the Lipschitz seminorm is $M = \|G\|_{\mathrm{Lip}} = \frac{1}{\pi}$. Also,

$$\beta_n = \frac{(M\sigma_n)^{2/3}}{n^{1/3}} = C'\left(\frac{\sigma}{n}\right)^{2/3}$$

for some absolute C'. Therefore, applying Theorem 7.1 and using $\sigma \ge 1$, we arrive at the bound (1.8):

$$\mathbf{E}\sup_{x\in\mathbf{R}^n} |F_n(x) - G(x)| \le C\sigma^{2/3}\frac{\log^{1/3} n}{n^{2/3}} + \sup_{x\in\mathbf{R}} |F(x) - G(x)|. \tag{8.1}$$

Thus, Theorem 1.3 is proved. For individual points that are close to the end-points $x = \pm 2$ of the supporting interval of the semicircle law, we may get improved bounds in comparison with (8.1). Namely, by Proposition 6.1 (and repeating the argument from the proof of the inequality of Corollary 6.2), for all $t > 0$,

$$\begin{aligned}
\mathbf{E}\mathrm{e}^{t|F_n(x)-G(x)|} &\le \mathbf{E}\mathrm{e}^{t(F_n(x)-G(x))} + \mathbf{E}\mathrm{e}^{-t(F_n(x)-G(x))} \\
&\le \mathrm{e}^{t(F(x+h)-G(x))} + \mathrm{e}^{t(G(x)-F(x-h))} \\
&\le 2\mathrm{e}^{t(G(x+h)-G(x-h))+t\|F-G\|},
\end{aligned}$$

where $h = \sqrt{\frac{2\sigma_n^2 t}{n}} = \frac{2\sigma\sqrt{t}}{n}$. Taking the logarithm and applying Jensen's inequality, we arrive at

$$\mathbf{E}|F_n(x) - G(x)| \le \|F - G\| + \bigl(G(x+h) - G(x-h)\bigr) + \frac{\log 2}{t}. \tag{8.2}$$

Using the Lipschitz property of G only (that is, $G(x+h) - G(x-h) \le \frac{h}{\pi}$) would yield the previous bound, such as the one in the estimate (6.7) of Proposition 6.4,

$$\mathbf{E}|F_n(x) - G(x)| \le \|F - G\| + C\left(\frac{\sigma}{n}\right)^{2/3}. \tag{8.3}$$

However, the real size of increments $G(x+h) - G(x-h)$ with respect to the parameter h essentially depends on the point x. To be more careful in the analysis of the right-hand side of (8.2), we may use the following elementary calculus bound, whose proof we omit.

Lemma 8.1. *$G(x+h) - G(x-h) \le 2g(x)h + \frac{4}{3\pi}h^{3/2}$ for all $x \in \mathbf{R}$ and $h > 0$.*

Since G is concentrated on the interval $[-2, 2]$, for $|x| \ge 2$, we have a simple bound $G(x+h) - G(x-h) \le \frac{4}{3\pi}h^{3/2}$. As a result, one may derive from (8.2) an improved variant of (8.3). In particular, if $|x| \ge 2$, then

$$\mathbf{E}|F_n(x) - G(x)| \le \|F - G\| + C\left(\frac{\sigma}{n}\right)^{6/7}.$$

The more general statement for all $x \in \mathbf{R}$ is given by the following result.

Theorem 8.2. *Let ξ_{jk} $(1 \le j \le k \le n)$ be independent and satisfy a logarithmic Sobolev inequality with constant σ^2, with $\mathbf{E}\xi_{jk} = 0$ and $\mathrm{Var}(\xi_{jk}) = 1$. For all $x \in \mathbf{R}$,*

$$\mathbf{E}|F_n(x) - G(x)| \le \|F - G\| + C\left[\left(\frac{\sigma}{n}\right)^{6/7} + g(x)^{2/3}\left(\frac{\sigma}{n}\right)^{2/3}\right], \tag{8.4}$$

where C is an absolute constant.

A similar uniform bound may also be shown to hold for $\mathbf{E}\sup_{y \le x}|F_n(y) - G(y)|$ $(x \le 0)$ and $\mathbf{E}\sup_{y \ge x}|F_n(y) - G(y)|$ $(x \ge 0)$. Note that in comparison with (8.3), there is an improvement for the points x at distance not more than $(\frac{\sigma}{n})^{4/7}$ from ± 2.

Proof of Theorem 8.2. According to the bound (8.2) and Lemma 8.1, for any $h > 0$, we may write $\mathbf{E}|F_n(x) - G(x)| \le \|F - G\| + 3\varphi(h)$, where $\varphi(h) = g(x)h + h^{3/2} + \frac{\varepsilon}{h^2}$, $\varepsilon = (\frac{\sigma}{n})^2$.

We shall now estimate the minimum of this function. Write $h = (\frac{\varepsilon}{1+\alpha})^{2/7}$ with parameter $\alpha > 0$ to be specified later on. If $g(x) \le \alpha\sqrt{h}$, then

$$\varphi(h) \le (1+\alpha)h^{3/2} + \frac{\varepsilon}{h^2} = 2(1+\alpha)^{4/7}\varepsilon^{3/7}. \tag{8.5}$$

Note that the requirement on $g(x)$ is equivalent to $\frac{g(x)^7}{\varepsilon} \le \frac{\alpha^7}{1+\alpha}$. Thus, we set $A = \frac{g(x)^7}{\varepsilon}$ and take $\alpha = 1 + 2A^{1/6}$. Since $\alpha \ge 1$, we get $\frac{\alpha^7}{1+\alpha} \ge \frac{\alpha^6}{2} \ge A$. Hence, we may apply (8.5). Using $(1+\alpha)^{4/7} \le (2\alpha)^{4/7}$ and $\alpha^{4/7} \le 1 + ((2A)^{1/6})^{4/7} = 1 + (2A)^{2/21}$, we finally get that

$$\varphi(h) \le 2 \cdot 2^{4/7}\big(1 + (2A)^{2/21}\big)\varepsilon^{3/7} \le 4(\varepsilon^{3/7} + A^{2/21}\varepsilon^{3/7}).$$

This is the desired expression in square brackets in (8.4) and Theorem 8.2 follows. $\square$

Finally, let us comment on the meaning of the general Corollary 6.5 in the matrix model above. To every interval I on the real line, we associate the number

$$N_I = \mathrm{card}\{i \le n : X_i \in I\}$$

of eigenvalues X_i inside it. Again, Corollary 6.5 may be applied to the standard semicircle law G with density $g(x)$, in which case $\beta = C(\frac{\sigma}{n})^{2/3}$. This gives that, under LSI(σ^2), imposed on the entries ξ_{jk}, for any $\delta > 0$ and any interval I of length $|I| \ge 4\|F - G\|/\delta$, we have that

$$\left| \frac{N_I}{n} - \int_I g(x)\,\mathrm{d}x \right| \le \delta |I| \tag{8.6}$$

with probability at least $1 - 4\exp\{-\frac{cn^2\delta^3}{\sigma^2}|I|^3\}$. As we have already mentioned, under PI(σ^2), one can show that $\|F - G\| \le Cn^{-2/3}$ ([8], Theorem 1.1). Therefore, (8.6) holds true with high probability, provided that

$$|I| \ge Cn^{-2/3}/\delta \tag{8.7}$$

with large C (of order, say, $\log^{\varepsilon} n$).

Such properties have been intensively studied in recent years in connection with the universality problem. In particular, it is shown in [18] and [41] that the restriction (8.7) may be weakened to $|I| \ge C_\varepsilon(\log^\alpha n)/n$ under the assumption that the intervals I are contained in $[-2-\varepsilon, 2+\varepsilon]$, $\varepsilon > 0$, that is, ‘in the bulk’.

Appendix

Here we recall some facts about Poincaré-type and log-Sobolev inequalities. While Lemmas 2.2 and 5.3 list some of their consequences, one might wonder which measures actually satisfy these analytic inequalities. Many interesting examples can be constructed with the help of the following elementary proposition.

Proposition A1. *Let $\mu_1, \dots, \mu_N$ be probability measures on* $\mathbf{R}$ *satisfying* PI(σ^2) (*resp.*, LSI(σ^2)). *The image μ of the product measure $\mu_1 \otimes \cdots \otimes \mu_N$ under any map $T : \mathbf{R}^N \to \mathbf{R}^n$ with finite Lipschitz seminorm satisfies* PI($\sigma^2\|T\|_{\mathrm{Lip}}^2$) (*resp.*, LSI($\sigma^2\|T\|_{\mathrm{Lip}}^2$)).

On the real line, disregarding the problem of optimal constants, Poincaré-type inequalities may be reduced to Hardy-type inequalities with weights. Necessary and sufficient conditions for a measure on the positive half-axis to satisfy a Hardy-type inequality with general weights were found in the late 1950s in the work of Kac and Krein [27]. We refer the interested reader to [35] and [34] for a full characterization and an account of the history; here, we just recall the principal result (see also [7]).

Let μ be a probability measure on the line with median m, that is, $\mu(-\infty, m) \le \frac{1}{2}$ and $\mu(m, +\infty) \le \frac{1}{2}$. Define the quantities

$$A_0(\mu) = \sup_{x<m} \left[\mu(-\infty, x) \int_{-\infty}^{x} \frac{\mathrm{d}t}{p_\mu(t)} \right],$$

$$A_1(\mu) = \sup_{x>m} \left[\mu(x, +\infty) \int_{x}^{+\infty} \frac{\mathrm{d}t}{p_\mu(t)} \right],$$

where p_μ denotes the density of the absolutely continuous component of μ (with respect to Lebesgue measure) and where we set $A_0 = 0$ (resp., $A_1 = 0$) if $\mu(-\infty, m) = 0$ (resp., $\mu(m, +\infty) = 0$). We then have the following proposition.

Proposition A2. *The measure μ on $\mathbf{R}$ satisfies* PI(σ^2) *with some finite constant if and only if both $A_0(\mu)$ and $A_1(\mu)$ are finite. Moreover, the optimal value of σ^2 satisfies*

$$c_0\big(A_0(\mu) + A_1(\mu)\big) \le \sigma^2 \le c_1\big(A_0(\mu) + A_1(\mu)\big),$$

where c_0 and c_1 are positive universal constants.

Necessarily, μ must have a non-trivial absolutely continuous part with density which is positive almost everywhere on the supporting interval.

For example, the two-sided exponential measure μ_0, with density $\frac{1}{2}\mathrm{e}^{-|x|}$, satisfies PI($\sigma^2$) with $\sigma^2 = 4$. Therefore, any Lipschitz transform $\mu = \mu_0 T^{-1}$ of μ_0 satisfies PI(σ^2) with $\sigma^2 = 4\|T\|_{\mathrm{Lip}}^2$. The latter property may be expressed analytically in terms of the reciprocal to the so-called isoperimetric constant,

$$H(\mu) = \operatorname*{ess\,inf}_x \frac{p_\mu(x)}{\min\{F_\mu(x), 1 - F_\mu(x)\}},$$

where $F_\mu(x) = \mu(-\infty, x]$ denotes the distribution function of μ and p_μ the density of its absolutely continuous component. Namely, as a variant of the Mazya–Cheeger theorem, we have that PI(σ^2) is valid with $\sigma^2 = 4/H(\mu)^2$; see [9], Theorem 1.3.

To roughly describe the class of measures in the case, where μ is absolutely continuous and has a positive, continuous well-behaving density, one may note that $H(\mu)$ and the Poincaré constant are finite, provided that the measure has a finite exponential moment. In particular, any probability measure with a logarithmically concave density satisfies PI(σ^2) with a finite σ; see [4].

As for logarithmic Sobolev inequalities, we have a similar picture, where the standard Gaussian measure represents a basic example and plays a similar role as the two-sided exponential distribution for Poincaré-type inequalities. A full description on the real line, resembling Proposition A2, was given in [6]. Namely, for one-dimensional probability measure μ, with previous notation, we define the quantities

$$B_0(\mu) = \sup_{x<m}\left[\mu(-\infty, x)\log\frac{1}{\mu(-\infty, x)}\int_{-\infty}^{x}\frac{\mathrm{d}t}{p_\mu(t)}\right],$$

$$B_1(\mu) = \sup_{x>m}\left[\mu(x, +\infty)\log\frac{1}{\mu(x, +\infty)}\int_{x}^{+\infty}\frac{\mathrm{d}t}{p_\mu(t)}\right].$$

We then have the following proposition.

Proposition A3. *The measure μ on $\mathbf{R}$ satisfies* LSI(σ^2) *with some finite constant if and only if $B_0(\mu)$ and $B_1(\mu)$ are finite. Moreover, the optimal value of σ^2 satisfies*

$$c_0\big(B_0(\mu) + B_1(\mu)\big) \le \sigma^2 \le c_1\big(B_0(\mu) + B_1(\mu)\big),$$

where c_0 and c_1 are positive universal constants.

In particular, if μ has a log-concave density, then LSI(σ^2) is satisfied with some finite constant if and only if μ has sub-Gaussian tails.

Acknowledgements

This research was supported by NSF Grants DMS-07-06866 and SFB 701.

References

[1] Ball, K. (1988). Logarithmically concave functions and sections of convex sets in R^n. *Studia Math.* **88** 69–84. MR0932007

[2] Bhatria, R. (1997). *Matrix Analysis. Graduate Texts in Mathematics* **169**. New York: Springer. MR1477662

[3] Bobkov, S.G. (1999). Remarks on the Gromov–Milman inequality. *Vestn. Syktyvkar. Univ. Ser. 1 Mat. Mekh. Inform.* **3** 15–22. MR1716649

[4] Bobkov, S.G. (1999). Isoperimetric and analytic inequalities for log-concave probability measures. *Ann. Probab.* **27** 1903–1921. MR1742893

[5] Bobkov, S.G., Gentil, I. and Ledoux, M. (2001). Hypercontractivity of Hamilton–Jacobi equations. *J. Math. Pures Appl.* **88** 669–696. MR1846020

[6] Bobkov, S.G. and Götze, F. (1999). Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.* **1** 1–28. MR1682772

[7] Bobkov, S.G. and Götze, F. (2008). Hardy-type inequalities via Riccati and Sturm–Liouville equations. In *Sobolev Spaces in Mathematics, I* (V. Maz'ya, ed.). *Intern. Math. Series* **8** 69–86. New York: Springer. MR2508839

[8] Bobkov, S.G., Götze, F. and Tikhomirov, A.N. (2010). On concentration of empirical measures and convergence to the semi-circle law. Bielefeld University. Preprint. *J. Theor. Probab.* To appear.

[9] Bobkov, S.G. and Houdré, C. (1997). Isoperimetric constants for product probability measures. *Ann. Probab.* **25** 184–205. MR1428505

[10] Bobkov, S.G. and Ledoux, M. (1997). Poincare's inequalities and Talagrand's concentration phenomenon for the exponential distribution. *Probab. Theory Related Fields* **107** 383–400. MR1440138

[11] Bobkov, S.G. and Ledoux, M. (2009). Weighted Poincaré-type inequalities for Cauchy and other convex measures. *Ann. Probab.* **37** 403–427. MR2510011

[12] Borovkov, A.A. and Utev, S.A. (1983). On an inequality and a characterization of the normal distribution connected with it. *Probab. Theory Appl.* **28** 209–218. MR0700206

[13] Cambanis, S., Simons, G. and Stout, W. (1976). Inequalities for $Ek(X, Y)$ when the marginals are fixed. *Z. Wahrsch. Verw. Gebiete* **36** 285–294. MR0420778

[14] Chatterjee, S. and Bose, A. (2004). A new method for bounding rates of convergence of empirical spectral distributions. *J. Theoret. Probab.* **17** 1003–1019. MR2105745

[15] Davidson, K.R. and Szarek, S.J. (2001). Local operator theory, random matrices and Banach spaces. In *Handbook of the Geometry of Banach Spaces* **I** 317–366. Amsterdam: North-Holland. MR1863696

[16] Dudley, R.M. (1989). *Real Analysis and Probability*. Pacific Grove, CA: Wadsworth & Brooks/Cole Advanced Books & Software. MR0982264

[17] Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27** 642–669. MR0083864
[18] Erdös, L., Schlein, B. and Yau, H.-T. (2009). Local semicircle law and complete delocalization for Wigner random matrices. *Comm. Math. Phys.* **287** 641–655. MR2481753
[19] Evans, L.C. (1997). *Partial Differential Equations*. *Graduate Studies in Math.* **19**. Providence, RI: Amer. Math. Soc. MR1625845
[20] Götze, F. and Tikhomirov, A.N. (2003). Rate of convergence to the semi-circular law. *Probab. Theory Related Fields* **127** 228–276. MR2013983
[21] Götze, F. and Tikhomirov, A.N. (2005). The rate of convergence for spectra of GUE and LUE matrix ensembles. *Cent. Eur. J. Math.* **3** 666–704 (electronic). MR2171668
[22] Götze, F., Tikhomirov, A.N. and Timushev, D.A. (2007). Rate of convergence to the semi-circle law for the deformed Gaussian unitary ensemble. *Cent. Eur. J. Math.* **5** 305–334 (electronic). MR2300275
[23] Gromov, M. and Milman, V.D. (1983). A topological application of the isoperimetric inequality. *Amer. J. Math.* **105** 843–854. MR0708367
[24] Guionnet, A. and Zeitouni, O. (2000). Concentration of the spectral measure for large matrices. *Electron. Comm. Probab.* **5** 119–136. MR1781846
[25] Gustavsson, J. (2005). Gaussian fluctuations of eigenvalues in the GUE. *Ann. Inst. H. Poincaré Probab. Statist.* **41** 151–178. MR2124079
[26] Hensley, D. (1980). Slicing convex bodies – bounds for slice area in terms of the body's covariance. *Proc. Amer. Math. Soc.* **79** 619–625. MR0572315
[27] Kac, I.S. and Krein, M.G. (1958). Criteria for the discreteness of the spectrum of a singular string. *Izv. Vysš. Učebn. Zaved. Matematika* **2** 136–153. MR0139804
[28] Kim, T.Y. (1999). On tail probabilities of Kolmogorov–Smirnov statistics based on uniform mixing processes. *Statist. Probab. Lett.* **43** 217–223. MR1708089
[29] Ledoux, M. (1999). Concentration of measure and logarithmic Sobolev inequalities. In *Séminaire de Probabilités XXXIII*. *Lecture Notes in Math.* **1709** 120–216. Berlin: Springer. MR1767995
[30] Ledoux, M. (2001). *The Concentration of Measure Phenomenon*. *Math. Surveys and Monographs* **89**. Providence, RI: Amer. Math. Soc. MR1849347
[31] Ledoux, M. (2007). Deviation inequalities on largest eigenvalues. In *Geom. Aspects of Funct. Anal., Israel Seminar 2004–2005*. *Lecture Notes in Math.* **1910** 167–219. Berlin: Springer. MR2349607
[32] Massart, P. (1990). The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. *Ann. Probab.* **18** 1269–1283. MR1062069
[33] Maurey, B. (1991). Some deviation inequalities. *Geom. Funct. Anal.* **1** 188–197. MR1097258
[34] Maz'ya, V.G. (1985). *Sobolev Spaces*. Berlin: Springer. MR0817985
[35] Muckenhoupt, B. (1972). Hardy's inequality with weights. *Studia Math.* **XLIV** 31–38. MR0311856
[36] Pastur, L.A. (1973). Spectra of random selfadjoint operators. *Uspehi Mat. Nauk* **28** 3–64. MR0406251
[37] Ruschendorf, L. (1985). The Wasserstein distance and approximation theorems. *Z. Wahrsch. Verw. Gebiete* **70** 117–129. MR0795791
[38] Sen, P.K. (1974). Weak convergence of multidimensional empirical processes for stationary φ-mixing processes. *Ann. Probab.* **2** 147–154. MR0402845
[39] Soshnikov, A. (1999). Universality at the edge of the spectrum in Wigner random matrices. *Comm. Math. Phys.* **207** 697–733. MR1727234
[40] Talagrand, M. (1996). Majorizing measures: The generic chaining. *Ann. Probab.* **24** 1049–1103. MR1411488
[41] Tao, T. and Vu, V. (2009). Random matrices: Universality of local eigenvalue statistics. Preprint.
[42] Timushev, D.A. (2006). On the rate of convergence in probability of the spectral distribution function of a random matrix. *Teor. Veroyatn. Primen.* **51** 618–622. MR2325551

[43] Vallander, S.S. (1973). Calculations of the Vasserstein distance between probability distributions on the line. *Teor. Verojatn. Primen.* **18** 824–827. MR0328982

[44] Yoshihara, K. (1975/76). Weak convergence of multidimensional empirical processes for strong mixing sequences of stochastic vectors. *Z. Wahrsch. Verw. Gebiete* **33** 133–137. MR0385962

Received July 2008 and revised September 2009