Concentration Inequalities on the Multislice and for Sampling Without Replacement

Holger Sambale[1] · Arthur Sinulis[1]

Abstract

We present concentration inequalities on the multislice which are based on (modified) log-Sobolev inequalities. This includes bounds for convex functions and multilinear polynomials. As an application, we show concentration results for the triangle count in the $G(n, M)$ Erdős–Rényi model resembling known bounds in the $G(n, p)$ case. Moreover, we give a proof of Talagrand's convex distance inequality for the multislice. Interpreting the multislice in a sampling without replacement context, we furthermore present concentration results for n out of N sampling without replacement. Based on a bounded difference inequality involving the finite-sampling correction factor $1 - (n/N)$, we present an easy proof of Serfling's inequality with a slightly worse factor in the exponent, as well as a sub-Gaussian right tail for the Kolmogorov distance between the empirical measure and the true distribution of the sample.

Keywords Concentration of measure · Convex distance inequality · Erdős–Rényi graphs · Multislice · Sampling without replacement

Mathematics Subject Classification 2020 60E15 · 62D05

✉ Holger Sambale
hsambale@math.uni-bielefeld.de

Arthur Sinulis
asinulis@math.uni-bielefeld.de

1   Fakultät für Mathematik, Universität Bielefeld, Postfach 10 01 31, 33501 Bielefeld, Germany

1 Introduction

In the past few years, in particular in the analysis of Boolean functions, a model which has found emerging interest is the *multislice*. It can be regarded as a natural generalization of several well-known models like slices of the hypercube. In detail, let $L \geq 2$ be a natural number, $\kappa = (\kappa_1, \ldots, \kappa_L) \in \mathbb{N}^L$ (where by convention, $0 \notin \mathbb{N}$), $N := \kappa_1 + \cdots + \kappa_L$, and let $\mathcal{X} = \{x_1, \ldots, x_L\} \subset \mathbb{R}$ be a set of L distinct real numbers. Typically, $\mathcal{X} = \{0, 1, \ldots, L-1\}$ or $\mathcal{X} = \{1, 2, \ldots, L\}$, but we prefer not to specify $\mathcal{X}$ since the most natural choice usually depends on the situation under consideration. The multislice is defined as

$$\Omega_\kappa := \left\{ \omega = (\omega_1, \ldots, \omega_N) \in \mathcal{X}^N : \sum_{i=1}^N \mathbb{1}_{\{\omega_i = x_\ell\}} = \kappa_\ell \text{ for } \ell = 1, \ldots, L \right\}.$$

In other words, any $\omega \in \Omega_\kappa$ is a sequence of elements from $\{x_1, \ldots, x_L\}$ in which each feature x_ℓ appears exactly κ_ℓ times. In the context of sampling without replacement, it describes the procedure of (fully) sampling from a population with a set of characteristics $\{x_1, \ldots, x_L\}$, such that a proportion of κ_ℓ/N of the population has characteristic x_ℓ. We discuss and extend this relation in Sect. 1.2.

To gain an intuition into the multislice, let us consider some special choices of L and κ. For $L = 2$, $\kappa = (k, N-k)$ and $\mathcal{X} = \{0, 1\}$, the multislice reduces to k-slices on the hypercube, while the case of $L = N$, $\kappa = (1, \ldots, 1)$ and $\mathcal{X} = \{1, \ldots, N\}$ can be interpreted as the symmetric group S_N. If $L = 2$, Ω_κ can be interpreted as all possible realizations of an Erdős–Rényi random graph (see Corollary 1 for more details). Moreover, the multislice gives rise to a Markov chain known as the *multi-urn Bernoulli–Laplace diffusion model*, but we will not pursue this aspect. For examples, see [25].

We equip Ω_κ with the uniform distribution which we denote by $\mathbb{P}_\kappa$ or sometimes also simply $\mathbb{P}$. In other words,

$$\mathbb{P}_\kappa(\{\omega\}) = |\Omega_\kappa|^{-1} = \binom{N}{\kappa_1, \ldots, \kappa_L}^{-1} = \frac{\kappa_1! \cdots \kappa_L!}{N!}$$

for any $\omega \in \Omega_\kappa$. If f is any real-valued function on Ω_κ, we write $\mathbb{E}_\kappa f$ or $\mathbb{E} f$ for its expectation with respect to $\mathbb{P}_\kappa$. Moreover, to fix some conventions, we always assume the x_ℓ to be ordered such that $x_1 < x_2 < \ldots < x_L$. In particular, $|\mathcal{X}| := x_L - x_1$ denotes the diameter of $\mathcal{X}$. Furthermore, we shall write $\kappa_{\min} := \min\{\kappa_1, \ldots, \kappa_L\}$. Finally, for any $1 \leq i \neq j \leq N$, let τ_{ij} be the "switch" operator which switches the ith and jth component of the vector ω. In other words, τ_{ij} transforms ω into the vector $\tau_{ij}\omega$ given by

$$\tau_{ij}\omega = \left(\omega_1, \ldots, \omega_{i-1}, \omega_j, \omega_{i+1}, \ldots, \omega_{j-1}, \omega_i, \omega_{j+1}, \ldots, \omega_N \right). \tag{1}$$

Multislices equipped with the uniform measure were also considered in earlier works. Logarithmic Sobolev inequalities were proven in [16,25], while in [15], the

Friedgut–Kalai–Naor (FKN) theorem was extended to the multislice. We shall make use of the functional inequalities proven by Salez [25] to apply the entropy method and prove concentration inequalities in the above-mentioned settings.

1.1 Concentration Inequalities for Various Types of Functionals

In the first section, we present concentration inequalities for some functions on the multislice which are comparable to known concentration results in the independent case. We begin with a number of elementary inequalities.

Proposition 1 *1. Let $f : \Omega_\kappa \to \mathbb{R}$ be a function such that $|f(\omega) - f(\tau_{ij}\omega)| \leq c_{ij}$ for all $\omega \in \Omega_\kappa$, all $1 \leq i < j \leq N$ and suitable constants $c_{ij} \geq 0$. For any $t \geq 0$, we have*

$$\mathbb{P}_\kappa \left(f - \mathbb{E}_\kappa \, f \geq t \right) \leq \exp \left(-\frac{Nt^2}{4 \sum_{1 \leq i < j \leq N} c_{ij}^2} \right). \tag{2}$$

2. Let $f : [x_1, x_L]^N \to \mathbb{R}$ be convex and 1-Lipschitz. Then, for any $t \geq 0$ we have

$$\mathbb{P}_\kappa \left(f - \mathbb{E}_\kappa \, f \geq t \right) \leq \exp \left(-\frac{t^2}{16|\mathcal{X}|^2} \right). \tag{3}$$

Proposition 1 follows by a classic approach of Ledoux [23] (the entropy method), i.e., by exploiting suitable log-Sobolev-type inequalities, some of which might be of independent interest (cf. Propositions 4 and 5). Note that the bounded differences-type inequality (2) is invariant under the change $f \mapsto -f$, so that in particular, this result extends to the concentration inequality

$$\mathbb{P}_\kappa \left(|f - \mathbb{E}_\kappa \, f| \geq t \right) \leq 2 \exp \left(-\frac{Nt^2}{4 \sum_{1 \leq i < j \leq N} c_{ij}^2} \right). \tag{4}$$

By contrast, the assumption of convexity used in (3) is clearly not invariant under the change $f \mapsto -f$, but by different techniques discussed in Sect. 1.3, (3) can be extended to the lower tails as well.

While results for Lipschitz-type functions as in Proposition 1 are fairly standard in concentration of measure theory, in the past decade there has been increasing interest in non-Lipschitz functions. A case in point is so-called *multilinear polynomials*, i.e., polynomials which are affine with respect to every variable. Clearly, any multilinear polynomial $f = f(\omega)$ of degree d may be written as

$$f(\omega) = a_0 + \sum_{i_1=1}^{N} a_{i_1} \omega_{i_1} + \cdots + \sum_{i_1 < \ldots < i_d} a_{i_1 \ldots i_d} \omega_{i_1} \cdots \omega_{i_d}. \tag{5}$$

Typically, multilinear polynomials of degree $d \geq 2$ no longer have sub-Gaussian tails, but the tails show different regimes or levels of decay, corresponding to a larger

family of norms of the tensors of derivatives $\nabla^k f$, $k = 1, \ldots, d$. For t large, terms of the form $\exp(-(t/\beta_d)^{2/d})$ dominate, where β_d depends on the dth order derivatives. Tail inequalities of this type are also called *multilevel tail inequalities*, a term phrased by Adamczak [2,3].

In detail, we need a family of norms $\|\cdot\|_{\mathcal{I}}$ on the space of d-tensors for each partition $\mathcal{I} = \{I_1, \ldots, I_k\} \in P_d$, where P_d denotes the set of all partitions of $\{1, \ldots, d\}$. For any $1 \le i_1, \ldots, i_d \le N$ and any subset $I \subset \{1, \ldots, d\}$, we write $i_I = (i_k)_{k \in I}$, and for each $\ell = 1, \ldots, k$ we denote by $x^{(\ell)}$ a vector in $\mathbb{R}^{N^{I_\ell}}$. Then, for a d-tensor $A = (a_{i_1, \ldots, i_d})$ and a partition $\mathcal{I} \in P_d$, we set

$$\|A\|_{\mathcal{I}} := \sup \Big\{ \sum_{i_1, \ldots, i_d} a_{i_1 \ldots i_d} \prod_{\ell=1}^k x_{i_{I_\ell}}^{(\ell)} : \sum_{i_{I_\ell}} \Big(x_{i_{I_\ell}}^{(\ell)} \Big)^2 \le 1 \text{ for all } \ell = 1, \ldots, k \Big\}.$$

The family $\|\cdot\|_{\mathcal{I}}$ was first introduced in [22], where it was used to prove two-sided estimates for L^p norms of Gaussian chaos, and the definitions given above agree with the ones from [22] as well as [3,5]. We can regard the $\|A\|_{\mathcal{I}}$ as a family of operator-type norms. In particular, it is easy to see that $\|A\|_{\{1,\ldots,d\}} = \|A\|_{\mathrm{HS}} := (\sum_{i_1, \ldots, i_d} a_{i_1 \ldots i_d}^2)^{1/2}$ (Hilbert–Schmidt norm) and $\|A\|_{\{\{1\}, \ldots, \{d\}\}} = \|A\|_{\mathrm{op}} := \sup\{\sum_{i_1, \ldots, i_d} a_{i_1 \ldots i_d} x_{i_1}^{(1)} \cdots x_{i_d}^{(d)} : |x_{i^{(\ell)}}| \le 1 \text{ for all } \ell = 1, \ldots, d\}$ (operator norm).

Theorem 1 *Let $f = f(\omega)$ be a multilinear polynomial* (5) *of degree d. There exists a constant $c = c(d)$ such that*

$$\mathbb{P}_\kappa \left(|f - \mathbb{E}_\kappa f| \ge t \right) \le 2 \exp\left(-c \min_{1 \le k \le d} \min_{\mathcal{I} \in P_k} \left(\frac{t}{|\mathcal{X}|^k \|\mathbb{E}_\kappa \nabla^k f\|_{\mathcal{I}}} \right)^{2/|\mathcal{I}|} \right).$$

Theorem 1 is an analogue of [5, Theorem 1.4] (independent sub-Gaussian random variables), [3, Theorem 2.2] (the Ising model), [18, Theorem 5] (in the presence of certain discrete log-Sobolev inequalities), and [4, Corollary 5.4] (modified log-Sobolev inequalities for Glauber dynamics) for the multislice.

For the sake of illustration, consider the case of $d = 2$ and a quadratic form $f(\omega) = \sum_{i<j} a_{ij} \omega_i \omega_j = \omega^T A \omega / 2$, where A is a symmetric matrix with vanishing diagonal and entries $A_{ij} = a_{ij} = A_{ji}$ for any $i < j$. Let us additionally assume that $\mathbb{E}_\kappa \omega_i = 0$ for any i. In this case, we obviously have $\mathbb{E}_\kappa \nabla f = 0$ and $\mathbb{E}_\kappa \nabla^2 f = A$. Consequently, the conclusion of Theorem 1 reads

$$\mathbb{P}_\kappa \left(|f - \mathbb{E}_\kappa f| \ge t \right) \le 2 \exp\left(-c \min \left(\frac{t^2}{|\mathcal{X}|^4 \|A\|_{\mathrm{HS}}^2}, \frac{t}{|\mathcal{X}|^2 \|A\|_{\mathrm{op}}} \right) \right),$$

showing a version of the famous Hanson–Wright inequality for the multislice (cf. [20]). As an alternate strategy of proving the Hanson–Wright inequality, in Sect. 1.3 we derive Talagrand's convex distance inequality for the multislice, which in particular yields Hanson–Wright inequalities by [1] (where results of this type have already been

established for sampling without replacement along the lines, cf. Remark 2.3 therein). Theorem 1 may be seen as a generalization of these bounds to any order $d \in \mathbb{N}$.

Possible applications of Theorem 1 include the Erdős–Rényi model, which features random graphs with a fixed number of vertices n. There are two variants of the Erdős–Rényi model which are often labeled $G(n, p)$ and $G(n, M)$. In the $G(n, p)$ model, each possible edge between the n vertices is included with probability p independently of the other edges, while in the $G(n, M)$ model, the graph is chosen uniformly at random from the collection of all graphs with n vertices and M edges. In the following, we study $G(n, M)$.

Write $E = \{(i, j) : 1 \leq i < j \leq n\}$ for the set of possible edges, so that $\text{card}(E) = n(n - 1)/2 =: N$. Clearly, any edge $e \in E$ is included with probability $M/N =: p$. However, unlike in the $G(n, p)$ model, the events of the edges being included are not independent. Any configuration ω in $G(n, M)$ can be written as a vector $\omega = (\omega_e)_{e \in E} \in \{0, 1\}^E$ such that $\omega_e = 1$ for exactly M entries. In particular, $G(n, M)$ can be regarded as a multislice with $L = 2, \kappa = (N - M, M)$ and $\mathcal{X} = \{0, 1\}$.

One problem which has attracted considerable attention over the last two decades is the number of copies of certain subgraphs, e.g., triangles, in the Erdős–Rényi model. There is extensive literature on concentration inequalities for the triangle count, such as [12,14,21]. In particular, in [5, Proposition 5.5], bounds for the $G(n, p)$ model are derived using higher-order concentration results for multilinear polynomials in independent random variables. As Theorem 1 provides analogous higher-order concentration results in a dependent situation, we are able to show corresponding bounds for the $G(n, M)$ model by our methods.

Corollary 1 *Consider the $G(n, M)$ Erdős–Rényi model, and consider the number of triangles defined as $f(\omega) := \sum_{i<j<k} \omega_{ij}\omega_{jk}\omega_{ik}$. Then, for any $t \geq 0$,*

$$\mathbb{P}(|f - \mathbb{E}\,f| \geq t) \leq 2 \exp\left(-c \min\left(\frac{t^2}{n^3 + p^2 n^3 + p^4 n^4}, \frac{t}{n^{1/2} + pn}, t^{2/3}\right)\right).$$

Comparing Corollary 1 to [5, Proposition 5.5], we see that we arrive at essentially the same tail bounds despite the dependencies in the $G(n, M)$ model, with the only difference of an additional logarithmic factor $L_p := (\log(2/p))^{-1/2}$ in [5]. This logarithmic factor stems from the use of sub-Gaussian norms for independent Bernoulli random variables (which tend to 0 as $p \to 0$), which is not mirrored in the log-Sobolev tools we use.

Typically, the main interest is to study fluctuations which scale with the expected value of f. In this case, setting $t := \varepsilon\,\mathbb{E}\,f = \varepsilon\binom{n}{3}M(M - 1)(M - 2)/(N(N - 1)(N - 2))$, Corollary 1 reads

$$\mathbb{P}(|f - \mathbb{E}\,f| \geq \varepsilon\,\mathbb{E}\,f) \leq 2 \exp\left(-c \min\left(\varepsilon^2 n^3 p^6, \left(\varepsilon^2 \wedge \varepsilon^{2/3}\right) n^2 p^2\right)\right).$$

In particular, this shows that the optimal exponent $n^2 p^2$ known from the $G(n, p)$ setting also shows up for a suitable range of p, cf. the discussion in [5].

In a similar way, we may also count cycles as in [5, Proposition 5.6], but we do not pursue this in this note.

1.2 Sampling Without Replacement

In this section, we interpret the multislice in the sampling without replacement context, where we sample N times from a population of N individuals $\omega_1, \ldots, \omega_N$, so that the uniform distribution $\mathbb{P}_\kappa$ describes the sampling of all its elements. In applications, one does not sample the entire population, but chooses some sample size $n \leq N$, i.e., for each $\omega \in \Omega_\kappa$, and considers the first n coordinates only. Formally, if pr_n denotes the projection onto the first n coordinates, we may define $\Omega_{\kappa,n} := pr_n(\Omega_\kappa)$. We, again, equip $\Omega_{\kappa,n}$ with the uniform distribution $\mathbb{P}_{\kappa,n}$, which agrees with the push-forward of $\mathbb{P}_\kappa$ under pr_n. As above, we denote the expectation with respect to $\mathbb{P}_{\kappa,n}$ by $\mathbb{E}_{\kappa,n} f$, where f is any real-valued function.

Our first result is a bounded differences inequality for sampling without replacement involving the finite-sampling correction factor $1 - n/N$. In the sequel, $(\omega_{i^c}, \omega_i')$ denotes a vector which agrees with ω in all coordinates but the ith one, while ω_i is replaced by some admissible ω_i' (in the sense that $(\omega_{i^c}, \omega_i') \in \Omega_{\kappa,n}$). Moreover, for any $\sigma \in S_n$ we may define $\sigma\omega \in \Omega_{\kappa,n}$ by noting that σ acts on ω by permuting its indices.

Proposition 2 *Let $f : \Omega_{\kappa,n} \to \mathbb{R}$ be an arbitrary function and $(c_i)_{i=1,\ldots,n}$ such that $|f(\omega) - f(\omega_{i^c}, \omega_i')| \leq c_i$ for all $\omega \in \Omega_{\kappa,n}, \omega_i' \in \mathcal{X}$. For any $t \geq 0$, it holds*

$$\mathbb{P}_{\kappa,n}\left(\frac{1}{n!}\sum_{\sigma \in S_n} f(\sigma\omega) - \mathbb{E}_{\kappa,n} f \geq t\right) \leq \exp\left(-\frac{t^2}{4(1 - \frac{n}{N})\sum_{i=1}^n c_i^2}\right). \quad (6)$$

In particular, if f is symmetric, i.e., $f(\omega) = f(\sigma\omega)$ for any $\sigma \in S_n$ and any $\omega \in \Omega_{\kappa,n}$, and satisfies $|f(\omega) - f(\omega_1', \omega_2, \ldots, \omega_n)| \leq c$ for some $c > 0$, this implies

$$\mathbb{P}_{\kappa,n}\left(f - \mathbb{E}_{\kappa,n} f \geq t\right) \leq \exp\left(-\frac{t^2}{4(1 - \frac{n}{N})c^2 n}\right). \quad (7)$$

Note that equation (6) is invariant under the change $f \mapsto -f$, which yields a two-sided concentration inequality as in (4). To express it in terms of deviation probabilities, for any $\delta \in (0, 1]$ we have with probability at least $1 - \delta$

$$\left|\frac{1}{n!}\sum_{\sigma \in S_n} f(\sigma\omega) - \mathbb{E}_{\kappa,n} f\right| \leq \sqrt{4(1 - n/N)\log\left(\frac{2}{\delta}\right)\sum_{i=1}^n c_i^2}.$$

Concentration inequalities of this type have also been proven in [31, Lemma 2] and [13, Theorem 5] by different methods, and our results agree with these bounds up to constants.

Let us apply Proposition 2 to some known statistics in sampling without replacement. One of the most famous concentration results for sampling without replacement is Serfling's inequality [27], which can be regarded as a strengthening of Hoeffding's inequality for n out of N sampling due to the inclusion of the finite-sampling correction factor $1 - n/N$. For a discussion and some newer results, we refer to [6,19,30].

We can deduce Serfling's inequality with a slightly worse constant from Proposition 2.

Corollary 2 *In the situation above, we have for any $t \geq 0$*

$$\mathbb{P}_{\kappa,n}\left(\frac{1}{n}\sum_{i=1}^{n}\omega_i - \mathbb{E}_{\kappa,n}\,\omega_1 \geq t\right) \leq \exp\left(-\frac{nt^2}{4\left(1 - \frac{n}{N}\right)|\mathcal{X}|^2}\right).$$

The same estimate holds for $\mathbb{P}_{\kappa,n}(\frac{1}{n}\sum_{i=1}^{n}\omega_i - \mathbb{E}_{\kappa,n}\,\omega_1 \leq -t)$.

In the original version of Serfling's inequality, the right-hand side is replaced by $\exp(-2nt^2/((1-(n-1)/N)|\mathcal{X}|^2))$.

As a second example, consider the approximation of the uniform distribution on all the points from which the ω_i are sampled using the empirical measure, measured in terms of the Kolmogorov distance. Formally, we put

$$g_{n,t}(\omega_1, \ldots, \omega_n) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{(-\infty,t]}(\omega_i)$$

and

$$f(\omega) := \sup_{t\in\mathbb{R}}\left(g_{n,t}(\omega) - \mathbb{E}_{\kappa,n}\,g_{n,t}\right).$$

In [19], it was conjectured that $\sqrt{n}\,f$ has sub-Gaussian tails with variance $1 - n/N$. The next result states that after centering around the expectation, this is indeed the case.

Corollary 3 *With the above notation, we have for any $t \geq 0$*

$$\mathbb{P}_{\kappa,n}(\sqrt{n}|f - \mathbb{E}_{\kappa,n}\,f| \geq t) \leq 2\exp\left(-\frac{t^2}{4\left(1 - \frac{n}{N}\right)}\right).$$

1.3 Talagrand's Convex Distance Inequality

Let Ω be any measurable space, $\omega = (\omega_1, \ldots, \omega_N) \in \Omega^N$ and $A \subset \Omega^N$ a measurable set. In his landmark paper [29], Talagrand defined the convex distance between ω and A

$$d_T(\omega, A) := \sup_{\alpha\in\mathbb{R}^N:|\alpha|=1} d_\alpha(\omega, A),$$

where

$$d_\alpha(\omega, A) := \inf_{\omega'\in A} d_\alpha(\omega, \omega') := \inf_{\omega'\in A}\sum_{i=1}^{N}|\alpha_i|\mathbb{1}_{\omega_i\neq\omega_i'}.$$

Talagrand proved concentration inequalities for the convex distance of random permutations and product measures which have attracted continuous interest since then. For product measures, an alternate proof based on the entropy method was given in [10]. In [26], the entropy method was used to reprove the convex distance inequality for random permutations as well, and this proof was extended to slices of the hypercube. In the present article, we further generalize this proof to the multislice, encompassing both situations discussed in [26].

Proposition 3 *For any $A \subseteq \Omega_\kappa$, it holds*

$$\mathbb{P}_\kappa(A) \, \mathbb{E}_\kappa \exp\left(\frac{d_T(\cdot, A)^2}{144}\right) \leq 1.$$

Note that in [24], convex distance inequalities for certain types of dependent random variables are proven. This includes sampling without replacement. In this sense, the result of Proposition 3 is not new, but we present a different strategy of proof solely based on the entropy method.

A famous corollary of Talagrand's convex distance inequality is sub-Gaussian concentration inequalities for convex Lipschitz functions, as first proven in [28]. Thus, Proposition 3 implies the following corollary, which can be regarded as an extension of Proposition 1 to upper *and* lower tails (ignoring the subtle issue of concentration around the mean or the median of a function).

Corollary 4 *Let $f : \mathbb{R}^N \to \mathbb{R}$ be convex and L-Lipschitz. Then, for any $t \geq 0$ it holds*

$$\mathbb{P}_\kappa \left(|f - \mathrm{med}(f)| \geq t\right) \leq 4 \exp\left(-\frac{t^2}{144 L^2 |\mathcal{X}|^2}\right),$$

where $\mathrm{med}(f)$ *is a median for* f.

As a simple application of Corollary 4, we show the following bound on the largest eigenvalue of symmetric matrices whose entries have distribution $\mathbb{P}_\kappa$:

Corollary 5 *Let $X = (X_{ij})_{i,j}$ be a symmetric $n \times n$ random matrix. Let $N := n(n+1)/2$ and assume that the common distribution of the entries $(X_{ij})_{i \leq j}$ on $\mathbb{R}^N$ is given by $\mathbb{P}_\kappa$ for some κ, $L \geq 2$ and $\mathcal{X}$. Let $\lambda_{\max} := \lambda_{\max}(X) := \max\{|\lambda(X)| : \lambda(X) \text{ eigenvalue of } X\}$. We have for any $t \geq 0$*

$$\mathbb{P}\left(|\lambda_{\max}(X) - \mathrm{med}(\lambda_{\max}(X))| \geq t\right) \leq 4 \exp\left(-\frac{t^2}{144 |\mathcal{X}|^2}\right).$$

In particular, this result shows that $\lambda_{\max}$ has sub-Gaussian tails independently of the dimension n. A possible choice of X is the adjacency matrix of a $G(n, M)$ Erdős–Rényi random graph. Corollary 5 is an adaptation of a classical example for independent random variables, see, e.g., [11, Example 6.8].

Furthermore, we are able to prove a somewhat weaker version of the convex distance inequality for n out of N sampling. Here, we consider symmetric sets, i.e., sets $A \subset \Omega_{\kappa,n}$, such that $\omega \in A$ implies $\sigma\omega \in A$ for any permutation $\sigma \in S_n$. Obviously,

assuming A to be symmetric is increasingly restrictive if n tends to N. This is mirrored in the additional finite-sampling correction factor $1 - n/N$ in the following theorem (which sharpens the convex distance inequality in [24]).

Theorem 2 *For any symmetric set $A \in \Omega_{\kappa,n}$ with $\mathbb{P}_{\kappa,n}(A) \geq \frac{1}{2}$ and any $t \geq 0$, we have*

$$\mathbb{P}_{\kappa,n}\left(d_T(\cdot, A) \geq t\right) \leq e \exp\left(-\frac{t^2}{16(1 - \frac{n}{N})}\right).$$

As above, Theorem 2 implies the following result.

Corollary 6 *Let f be a convex and symmetric L-Lipschitz function. Then, for any $t \geq 0$ we have*

$$\mathbb{P}_{\kappa,n}\left(|f - \mathrm{med}(f)| \geq t\right) \leq 2e \exp\left(-\frac{t^2}{16(1 - n/N)L^2|\mathcal{X}|^2}\right).$$

Examples of functions to which Corollary 6 may be applied are the estimators for the mean and the standard deviation given by $f(\omega) = \bar{\omega} = n^{-1}\sum_{i=1}^{n}\omega_i$ (sample mean) and $f(\omega) = s(\omega) = (\frac{1}{n-1}\sum_{i=1}^{n}(\omega_i - \bar{\omega})^2)^{1/2} = (\frac{1}{n(n-1)}\sum_{i<j}(\omega_i - \omega_j)^2)^{1/2}$ (sample standard deviation), having Lipschitz constants $L = n^{-1/2}$ and $L = (2n)^{-1/2}$, respectively. In particular, for any $\delta \in (0, 1]$ we have with probability at least $1 - \delta$ for any of the two estimators

$$|f - \mathrm{med}(f)| \leq \sqrt{16(1 - n/N)L^2|\mathcal{X}|^2 \log(2e/\delta)}.$$

It is well known that concentration results centered around the expectation and the median differ only by a constant. Indeed, in our case, for any convex, symmetric L-Lipschitz function

$$
\begin{aligned}
|\mathbb{E}_{\kappa,n} f - \mathrm{med}(f)| &\leq \mathbb{E}_{\kappa,n}|f - \mathrm{med}(f)| = \int_0^\infty \mathbb{P}_{\kappa,n}\left(|f - \mathrm{med}(f)| \geq t\right) dt \\
&\leq 2e \int_0^\infty \exp\left(-\frac{t^2}{16(1 - n/N)\mathcal{X}^2 L^2}\right) dt \\
&= 2e\sqrt{4\pi(1 - n/N)|\mathcal{X}|^2 L^2} \\
&\approx 19.27|\mathcal{X}|L\sqrt{1 - n/N}.
\end{aligned}
$$

2 Logarithmic Sobolev Inequalities for the Multislice

The main tool for establishing concentration inequalities in this note is the entropy method, which is based on the use of logarithmic Sobolev-type inequalities. Let us recall some basic facts and definitions especially adapted to discrete spaces. A key object is a suitable difference operator, i.e., a kind of "discrete derivative." Given a

probability space $(\mathcal{Y}, \mathcal{F}, \mu)$, we call any operator $\Gamma : L^\infty(\mu) \to L^\infty(\mu)$ satisfying $|\Gamma(af + b)| = a |\Gamma f|$ for all $a > 0, b \in \mathbb{R}$ a difference operator. Moreover, by $\mathbb{E}_\mu$ we denote integration with respect to μ.

Definition 1 1. We say that μ satisfies a logarithmic Sobolev inequality $\Gamma-\mathrm{LSI}(\sigma^2)$ if for all bounded measurable functions f, we have

$$\mathrm{Ent}_\mu(f^2) \leq 2\sigma^2 \, \mathbb{E}_\mu \, \Gamma(f)^2,$$

where $\mathrm{Ent}_\mu(f) := \mathbb{E}_\mu \, f \log(f) - \mathbb{E}_\mu \, f \log \mathbb{E}_\mu \, f$ (for any positive function f) denotes the entropy functional.

2. We say that μ satisfies a modified logarithmic Sobolev inequality $\Gamma-\mathrm{mLSI}(\sigma^2)$ if for all bounded measurable functions f, we have

$$\mathrm{Ent}_\mu(e^f) \leq \frac{\sigma^2}{2} \, \mathbb{E}_\mu \, \Gamma(f)^2 e^f.$$

3. We say that μ satisfies a Poincaré inequality $\Gamma-\mathrm{PI}(\sigma^2)$ if for all bounded measurable functions f, we have

$$\mathrm{Var}_\mu(f) \leq \sigma^2 \, \mathbb{E}_\mu \, \Gamma(f)^2,$$

where $\mathrm{Var}(f) := \mathbb{E}_\mu \, f^2 - (\mathbb{E}_\mu \, f)^2$ is the variance.

4. If any of these functional inequalities does not hold for *all* bounded measurable functions but for some subclass $\mathcal{A} \subset L^\infty(\mu)$, we say that μ satisfies a $\Gamma-\mathrm{LSI}(\sigma^2)$ (PI, mLSI) *on* $\mathcal{A}$.

If Γ satisfies the chain rule (as the ordinary gradient ∇ does), $\Gamma-\mathrm{LSIs}$ and $\Gamma-\mathrm{mLSIs}$ are equivalent concepts, but in the examples we consider in this note, this is usually not true. Moreover, it is well known that a $\Gamma-\mathrm{LSI}(\sigma^2)$ and a $\Gamma-\mathrm{mLSI}(\sigma^2)$ both imply a $\Gamma-\mathrm{PI}(\sigma^2)$, cf. e.g., [8, Proposition 3.6].

For the multislice, we mostly consider the following canonical difference operator. Recalling the "switch" operator from (1), for any function $f : \Omega_\kappa \to \mathbb{R}$ we set

$$\Gamma_{ij}(f)(\omega) := \Gamma_{ij} f(\omega) := f(\omega) - f(\tau_{ij}\omega) =: f(\omega) - \tau_{ij} f(\omega)$$

and define the difference operator Γ by

$$\Gamma(f) := \left(\frac{1}{2N} \sum_{1 \leq i < j \leq N} \Gamma_{ij}(f)^2 \right)^{1/2}.$$

Note that $\Gamma_{ij}(f)^2$ might be interpreted as a sort of "local variance." Indeed, it is easy to verify that

$$\Gamma_{ij}(f)^2(\omega) = 2 \int (f(\omega) - f(\omega_{\{i,j\}^c}, \eta_{ij}))^2 d \, \mathbb{P}_\kappa(\eta_{ij} \mid \omega_{\{i,j\}^c}), \tag{8}$$

where $\omega_{\{i,j\}^c} = (\omega_k)_{k \notin \{i,j\}}$ and $\eta_{ij} = (\eta_i, \eta_j)$. Therefore, we have $\Gamma(f)^2 = 2N^{-1}|\partial f|^2$ for the difference operator $|\partial f|$ introduced in [17].

Sometimes (and typically for auxiliary purposes), we shall also need a second, closely related difference operator which we denote by Γ^+. Here, we simply set

$$\Gamma_{ij}^+(f)(\omega) := (f(\omega) - f(\tau_{ij}\omega))_+,$$

where $x_+ := \max(x, 0)$ denotes the positive part of a real number, and define Γ^+ accordingly.

Recently, in [25] sharp (modified) logarithmic Sobolev inequalities for the multi-slice were established. Rewriting these results in accordance with our notation and slightly extending them immediately leads to the following proposition, serving as the basis for our arguments:

Proposition 4 *With the above definitions of Γ and Γ^+, $\mathbb{P}_\kappa$ satisfies the following functional inequalities:*

- *Γ−LSI($2 \log(N/\kappa_{\min})/\log(2)$),*
- *Γ−mLSI(4),*
- *Γ^+−mLSI(8).*

Proof of Proposition 4 The Γ−LSI directly follows from [25, Theorem 5]. Moreover, by [25, Lemma 1] (substituting $f \geq 0$ by e^f), we have

$$\mathrm{Ent}_{\mathbb{P}_\kappa}(e^f) \leq \frac{1}{N} \sum_{i<j} \mathbb{E}_\kappa (e^{f(\tau_{ij}\omega)} - e^{f(\omega)})(f(\tau_{ij}\omega) - f(\omega)) \tag{9}$$

for any $f: \Omega_\kappa \to \mathbb{R}$. Using the fact that $\omega \mapsto \tau_{ij}\omega$ is an automorphism of Ω_κ and applying the inequality $(a-b)(e^a - e^b) \leq \frac{1}{2}(e^a + e^b)(a-b)^2$ leads to the Γ−mLSI(4). By similar arguments, we may also deduce the Γ^+−mLSI(8). In particular, we note that the expected values on the right-hand side of (9) are symmetric in ω and $\tau_{ij}\omega$ and use the inequality $(a-b)_+(e^a - e^b) \leq (a-b)_+^2 e^a$. $\qquad\square$

From Proposition 4, we may derive a convex ∇ − (m)LSI on the multislice, where ∇ denotes the usual Euclidean gradient.

Proposition 5 *For any $f \in \mathcal{A}_c := \{f: [x_1, x_L]^N \to \mathbb{R} \mid f \text{ convex}\}$, we have*

$$\mathrm{Ent}_{\mathbb{P}_\kappa}(e^f) \leq 4|\mathcal{X}|^2 \, \mathbb{E}_\kappa \, e^f |\nabla f|^2.$$

In other words, $\mathbb{P}_\kappa$ satisfies a ∇ − mLSI($8|\mathcal{X}|^2$) on $\mathcal{A}_c$.

Proof Using convexity in the first step and the inequality $(a-b)^2 \leq 2a^2 + 2b^2$ yields

$$\Gamma^+(f)^2(\omega) = \frac{1}{4N} \sum_{i \neq j} (f(\omega) - f(\tau_{ij}\omega))_+^2 \leq \frac{1}{4N} \sum_{i \neq j} \langle \nabla f(\omega), \omega - \tau_{ij}\omega \rangle^2$$

$$
= \frac{1}{4N} \sum_{i \neq j} (\omega_i - \omega_j)^2 \big(\partial_i f(\omega) - \partial_j f(\omega) \big)^2
$$

$$
\leq \frac{|\mathcal{X}|^2}{2N} \sum_{i \neq j} \big(\partial_i f(\omega)^2 + \partial_j f(\omega)^2 \big)
$$

$$
= \frac{|\mathcal{X}|^2}{N} (N-1) \sum_{i=1}^{N} \partial_i f(\omega)^2 \leq |\mathcal{X}|^2 |\nabla f|^2(\omega).
$$

As $\mathbb{P}_\kappa$ satisfies a $\Gamma^+ - \mathrm{mLSI}(8)$ by Proposition 4, the claim follows. □

Another class of functional inequalities we address in this note is Beckner inequalities. Restricting ourselves to the multislice (rather than providing a general definition), $\mathbb{P}_\kappa$ satisfies a Beckner inequality with parameter $p \in (1, 2]$ (Bec-p) if there exists some constant $\beta_p > 0$ such that

$$
\mathbb{E}_\kappa f^p - (\mathbb{E}_\kappa f)^p \leq \frac{\beta_p p}{2} \mathcal{E}_\kappa(f, f^{p-1}) \tag{10}
$$

for any non-negative function f. Here,

$$
\mathcal{E}_\kappa(f, g) := \frac{1}{2N} \sum_{1 \leq i < j \leq N} \mathbb{E}_\kappa (\Gamma_{ij} f)(\Gamma_{ij} g)
$$

for any functions f, g on Ω_κ (which is the Dirichlet form of the underlying Markov chain).

Recently, in [4] it was shown that in the context of general Markov semigroups, Beckner inequalities with constants bounded away from zero as $p \downarrow 1$ and modified log-Sobolev inequalities are equivalent. In their article, the authors provide numerous examples and applications, also briefly discussing the multislice. Since we need results of this type for our purposes, we include a somewhat more detailed discussion in the present note.

Proposition 6 *For any $p \in (1, 2]$, $\mathbb{P}_\kappa$ satisfies a Beckner inequality Bec-p with constant $\beta_p = \frac{4N}{p(N+2)}$.*

Proof First note that the result holds true for $\kappa = (1, \ldots, 1)$ and $L = N$ as proven in [8, Proposition 4.8], with the difference in the constant being due to different normalizations. To extend this result to general κ, we apply a "projection" or "coarsening" argument, cf. [25, Section 3.4]. Indeed, consider the map $\Psi : \{1, \ldots, N\} \to \{1, \ldots, L\}$ given by $\Psi(i) = \ell$ iff $i \in \{\kappa_1 + \cdots + \kappa_{\ell-1} + 1, \ldots, \kappa_1 + \cdots + \kappa_\ell\}$ and extend it to the multislice (with $\mathcal{X} = \{1, \ldots L\}$) by coordinate-wise application, i.e., $\Psi(\omega_1, \ldots, \omega_N) := (\Psi(\omega_1), \ldots, \Psi(\omega_N))$. Moreover, to address a general choice of $\mathcal{X}$, let $\Phi : \{1, \ldots, L\} \to \mathcal{X}$ be the "canonical identification" $\Phi(i) := x_i$. Then, by [25, Lemma 4] applied to $\tilde{f} := f \circ \Phi$ and $\tilde{g} := g \circ \Phi$,

$$
\mathbb{E}_\kappa f = \mathbb{E}_{(1,\ldots,1)}(\tilde{f} \circ \Psi), \qquad \mathcal{E}_\kappa(f, g) = \mathcal{E}_{(1,\ldots,1)}(\tilde{f} \circ \Psi, \tilde{g} \circ \Psi)
$$

for any functions f, g. From these identities, we immediately obtain the result. □

Finally, we may also derive logarithmic Sobolev inequalities for symmetric functions of sampling without replacement. Here, we use other types of difference operators. Let $f: \Omega_{\kappa,n} \to \mathbb{R}$ be any (not necessarily symmetric) function. Then, writing $(\omega_1, \ldots, \omega_i', \ldots, \omega_n) = (\omega_1, \ldots, \omega_{i-1}, \omega_i', \omega_{i+1}, \ldots, \omega_n)$, we set

$$\mathfrak{h}(f)^2(\omega_1, \ldots, \omega_n) = \frac{1}{2} \sum_{i=1}^{n} (\sup_{\omega_i} f(\omega_1, \ldots, \omega_n) - \inf_{\omega_i'} f(\omega_1, \ldots, \omega_i', \ldots, \omega_n))^2,$$

$$\mathfrak{h}^+(f)^2(\omega_1, \ldots, \omega_n) = \frac{1}{2} \sum_{i=1}^{n} (f(\omega_1, \ldots, \omega_n) - \inf_{\omega_i'} f(\omega_1, \ldots, \omega_i', \ldots, \omega_n))^2.$$

Here, the supremum and the infimum have to be interpreted as extending over all admissible configurations, i.e., such that $(\omega_{i^c}, \omega_i), (\omega_{i^c}, \omega_i') \in \Omega_{\kappa,n}$.

Proposition 7 *Let $\mathcal{A}_{n,s} := \{f: \Omega_{\kappa,n} \to \mathbb{R} \mid f \text{ symmetric}\}$. With the above definitions of $\mathfrak{h}$ and $\mathfrak{h}^+$, $\mathbb{P}_{\kappa,n}$ satisfies the following functional inequalities on $\mathcal{A}_{n,s}$:*

- $\mathfrak{h}-\text{LSI}(2\log(N/\kappa_{\min})(1 - \frac{n}{N})/\log(2))$,
- $\mathfrak{h}-\text{mLSI}(4(1 - \frac{n}{N}))$,
- $\mathfrak{h}^+-\text{mLSI}(8(1 - \frac{n}{N}))$.

Proof We only prove the $\mathfrak{h}^+-\text{mLSI}$. The proofs of the other two inequalities follow by a modification of the arguments below.

First note that any function f on $\Omega_{\kappa,n}$ can be extended to a function F on Ω_κ which only depends on the first n coordinates by setting $F(\omega_1, \ldots, \omega_N) := f(\omega_1, \ldots, \omega_n)$, which may be rewritten as $F = f \circ pr_n$. We now apply Proposition 4 to F. Obviously, $\text{Ent}_{\mathbb{P}_\kappa}(e^F) = \text{Ent}_{\mathbb{P}_{\kappa,n}}(e^f)$. It therefore remains to consider the right-hand side of the mLSI. Here, we obtain

$$\frac{1}{2N} \sum_{i<j} \mathbb{E}_\kappa (F(\omega) - F(\tau_{ij}\omega))_+^2 e^{F(\omega)}$$

$$= \frac{1}{2N} \sum_{i=1}^{n} \sum_{j=n+1}^{N} \mathbb{E}_\kappa (F(\omega) - F(\tau_{ij}\omega))_+^2 e^{F(\omega)}$$

$$\leq \frac{1}{2N} \sum_{i=1}^{n} \sum_{j=n+1}^{N} \mathbb{E}_\kappa (F(\omega) - \inf_{\omega_i'} F(\omega_{i^c}, \omega_i'))_+^2 e^{F(\omega)}$$

$$= \frac{N-n}{2N} \sum_{i=1}^{n} \mathbb{E}_{\kappa,n}(f(\omega_1, \ldots, \omega_n) - \inf_{x_i'} f(\omega_1, \ldots, x_i', \ldots, \omega_n))_+^2 e^{f(\omega_1, \ldots, \omega_n)}.$$

Here, the first equality follows by symmetry of f with respect to the symmetric group S_n, and the fact that f does not depend on $(\omega_{n+1}, \ldots, \omega_N)$. The first inequality is due to the monotonicity of $x \mapsto x_+$, and the last equality follows as $\mathbb{P}_{\kappa,n}$ is the push-forward

of $\mathbb{P}_\kappa$ under pr_n. Thus, for any $f \in \mathcal{A}_{n,s}$ it holds

$$\text{Ent}_{\mathbb{P}_{\kappa,n}}(e^f) \leq \frac{4}{2N} \sum_{i<j} \mathbb{E}_\kappa (F(\omega) - F(\tau_{ij}\omega))_+^2 e^{F(\omega)}$$

$$\leq 4\frac{N-n}{N} \mathbb{E}_{\kappa,n} \mathfrak{h}^+(f)^2(\omega_1, \ldots, \omega_n) e^{f(\omega_1, \ldots, \omega_n)},$$

which finishes the proof.　　□

3 Proofs of the Concentration Inequalities

3.1 Proofs of Section 1.1

Proof of Proposition 1 Recall that if a probability measure μ satisfies a $\Gamma - \text{mLSI}(\sigma^2)$ on $\mathcal{A}$ (where Γ denotes some difference operator), we have for any $f \in \mathcal{A}$ such that $\Gamma(f) \leq L$,

$$\mu(f - \mathbb{E} f \geq t) \leq \exp\left(-\frac{t^2}{2\sigma^2 L^2}\right) \tag{11}$$

for any $t \geq 0$. For a reference, see, e.g., [7] or [26, (1.2)]. Combining this fact with Proposition 4 and noting that by definition,

$$\Gamma(f) \leq \left(\frac{1}{2N} \sum_{1 \leq i < j \leq N} c_{ij}^2\right)^{1/2},$$

we arrive at (2). In the same way, we may derive (3) using Proposition 5.　　□

The proof of Theorem 1 is more advanced. The basic idea is to follow the steps of the proof of [3, Theorem 2.2] and its refinements as presented in [4, Section 5.3]. First, we derive moment estimates for functions on the multislice.

Lemma 1 *For any $f : \Omega_\kappa \to \mathbb{R}$ and any $p \geq 2$,*

$$\|f - \mathbb{E}_\kappa f\|_{L^p(\mathbb{P}_\kappa)} \leq \sqrt{4\theta p} \|\Gamma(f)\|_{L^p(\mathbb{P}_\kappa)},$$

where $\theta := \sqrt{e}/(\sqrt{e} - 1) < 2.5415.$

Proof This follows immediately from Proposition 6 and [4, Proposition 3.3]. (Note that the notation used therein differs from ours; in particular, no square root is taken in the definition of Γ.) To apply the latter result, we have to check that the constants of the Beckner inequalities Bec-p satisfy

$$\beta_p^{-1} = \frac{p(N+2)}{4N} \geq a(p-1)^s$$

for some $a > 0$, $s \geq 0$ and any $p \in (1, 2]$. Clearly, we may take $a = 1/4$ and $s = 0$, which finishes the proof. □

Note that alternatively, we could apply [17, Proposition 2.4], using (8) and Proposition 4, which yields

$$\|f - \mathbb{E}_\kappa f\|_{L^p(\mathbb{P}_\kappa)} \leq \sqrt{\frac{8\log(N/\kappa_{\min})}{\log(2)}} \sqrt{p-1} \|\Gamma(f)\|_{L^p(\mathbb{P}_\kappa)}.$$

As a result of using the $\Gamma - \mathrm{LSI}$, we arrive at a substantially weaker constant, however.

Next, we have to relate differences of multilinear polynomials to (formal) derivatives, which is typically achieved by an inequality of the form $\Gamma(f) \leq c|\nabla f|$ for some absolute constant $c > 0$. However, it comes out that such an inequality cannot be true in our setting. For instance, taking $N = 3$, $\mathcal{X} = \{0, 1\}$ and $f(\omega) = \omega_1\omega_2 - \omega_1\omega_3$, it is easy to check that for $\omega = (0, 1, 1)$, we have $0 = |\nabla f(\omega)| < \Gamma(f)(\omega)$. The same problem arises if we take Γ^+ instead of Γ. It is possible to prove an inequality of this type with $c := |\mathcal{X}|$ for multilinear polynomials with non-negative coefficients and $\mathcal{X} \subset [0, \infty)$. (This can be seen by slightly modifying the proof of Proposition 8.) However, the proof of Theorem 1 also includes an iteration and linearization procedure, and if we only allow for non-negative coefficients, we get stuck at $d = 2$.

The following proposition provides us with the estimate we need to get the recursion going, at the cost of also involving second-order derivatives.

Proposition 8 *Let $f = f(\omega)$ be a multilinear polynomial as in Theorem 1. Then, we have*

$$\Gamma(f)^2 \leq \frac{3|\mathcal{X}|^2}{2}|\nabla f|^2 + \frac{3|\mathcal{X}|^4}{4N}\|\nabla^2 f\|_{\mathrm{HS}}^2. \tag{12}$$

In particular, for any $p \geq 2$ we have

$$\|f - \mathbb{E}_\kappa f\|_{L^p(\mathbb{P}_\kappa)} \leq \sqrt{6\theta|\mathcal{X}|^2 p}\||\nabla f|\|_{L^p(\mathbb{P}_\kappa)} + \sqrt{3\theta|\mathcal{X}|^4 p/N}\|\|\nabla^2 f\|_{\mathrm{HS}}\|_{L^p(\mathbb{P}_\kappa)} \tag{13}$$

with θ as in Lemma 1.

Proof In the proof, we additionally assume f to be d-homogeneous, i.e.,

$$f(\omega) = \sum_{i_1 < \ldots < i_d} a_{i_1 \ldots i_d} \omega_{i_1} \cdots \omega_{i_d}.$$

This is done in order to ease notation, and it is no problem to extend our proof to the non-homogeneous case. For notational convenience, for any $i_1 < \ldots < i_d$ and any permutation $\sigma \in S_d$, we define $a_{i_{\sigma(1)} \ldots i_{\sigma(d)}} := a_{i_1 \ldots i_d}$, and we set $a_{i_1 \ldots i_d} = 0$ if $i_j = i_k$ for some $j \neq k$. Finally, note that some of the notation below has to be interpreted accordingly for small values of d, e.g., summation over $i_1 < \ldots < i_{d-1}$ reduces to

summation over i_1 for $d = 2$. Observe that for any $k, \ell \in \{1, \dots, N\}, k \neq \ell$, we have

$$
\Gamma_{k\ell}(f)(\omega)^2 = \left(\sum_{\substack{i_1 < \dots < i_{d-1} \\ \ell \notin \{i_1, \dots, i_{d-1}\}}} a_{i_1 \dots i_{d-1}k} \omega_{i_1} \cdots \omega_{i_{d-1}} (\omega_k - \omega_\ell) \right.
$$

$$
\left. + \sum_{\substack{i_1 < \dots < i_{d-1} \\ k \notin \{i_1, \dots, i_{d-1}\}}} a_{i_1 \dots i_{d-1}\ell} \omega_{i_1} \cdots \omega_{i_{d-1}} (\omega_\ell - \omega_k) \right)^2
$$

$$
= \left(\sum_{i_1 < \dots < i_{d-1}} a_{i_1 \dots i_{d-1}k} \omega_{i_1} \cdots \omega_{i_{d-1}} (\omega_k - \omega_\ell) \right.
$$

$$
+ \sum_{i_1 < \dots < i_{d-1}} a_{i_1 \dots i_{d-1}\ell} \omega_{i_1} \cdots \omega_{i_{d-1}} (\omega_\ell - \omega_k)
$$

$$
\left. + \sum_{i_1 < \dots < i_{d-2}} a_{i_1 \dots i_{d-2}k\ell} \omega_{i_1} \cdots \omega_{i_{d-2}} (\omega_k - \omega_\ell)^2 \right)^2
$$

$$
\leq 3|\mathcal{X}|^2 \left(\left(\sum_{i_1 < \dots < i_{d-1}} a_{i_1 \dots i_{d-1}k} \omega_{i_1} \cdots \omega_{i_{d-1}} \right)^2 \right.
$$

$$
\left. + \left(\sum_{i_1 < \dots < i_{d-1}} a_{i_1 \dots i_{d-1}\ell} \omega_{i_1} \cdots \omega_{i_{d-1}} \right)^2 \right)
$$

$$
+ 3|\mathcal{X}|^4 \left(\sum_{i_1 < \dots < i_{d-2}} a_{i_1 \dots i_{d-2}k\ell} \omega_{i_1} \cdots \omega_{i_{d-2}} \right)^2
$$

$$
= 3|\mathcal{X}|^2 (\partial_k f(\omega)^2 + \partial_\ell f(\omega)^2) + 3|\mathcal{X}|^4 \partial_{k\ell} f(\omega)^2.
$$

Consequently, it holds

$$
\Gamma(f)^2 = \frac{1}{4N} \sum_{k \neq \ell} \Gamma_{k\ell}(f)^2 \leq \frac{3|\mathcal{X}|^2}{4N} \sum_{k \neq \ell} ((\partial_k f)^2 + (\partial_\ell f)^2) + \frac{3|\mathcal{X}|^4}{4N} \sum_{k \neq \ell} (\partial_{kl} f)^2
$$

$$
\leq \frac{3|\mathcal{X}|^2}{2} |\nabla f|^2 + \frac{3|\mathcal{X}|^4}{4N} \|\nabla^2 f\|_{\mathrm{HS}}^2,
$$

proving equation (12). Finally, combining (12) with Lemma 1, we immediately arrive at (13). □

With the help of Proposition 8, we may now prove Theorem 1. To this end, let us introduce some additional notation. If $A = (a_{i_1 \dots i_k})_{i_1, \dots, i_k \leq N}$, $B = (b_{i_1 \dots i_k})_{i_1, \dots, i_k \leq N}$

are two k-tensors, we define an inner product $\langle \cdot, \cdot \rangle$ by

$$\langle A, B \rangle := \sum_{i_1,\ldots,i_k \leq N} a_{i_1\ldots i_k} b_{i_1\ldots i_k}.$$

Moreover, if $x^j = (x_1^j, \ldots, x_N^j)$, $j = 1, \ldots, k$, are any vectors, we set $x^1 \otimes \cdots \otimes x^k := (x_{i_1}^1 \cdots x_{i_k}^k)_{i_1,\ldots,i_k \leq N}$. We also extend this notation to the situation in which some of these vectors may be N^2-dimensional. Indeed, let $x^1, \ldots, x^k$ be N-dimensional vectors as above, and let $y^1, \ldots, y^\ell$ be N^2-dimensional, $y^j = (y_{v_1,v_2}^j)_{v_1,v_2 \leq N}$. In this case, we set

$$x^1 \otimes \cdots \otimes x^k \otimes y^1 \otimes \cdots \otimes y^\ell := (x_{i_1}^1 \cdots x_{i_k}^k y_{i_{k+1},i_{k+2}}^1 \cdots y_{i_{k+2\ell-1},i_{k+2\ell}}^\ell)_{i_1,\ldots,i_{k+2\ell} \leq N},$$

which we regard as a rectangular $(k + \ell)$-tensor whose first k components are N-dimensional and whose last ℓ components are N^2-dimensional.

Proof of Theorem 1 To ease notation, we assume $|\mathcal{X}| = 1$ in the sequel. The general case follows in the same way with only minor changes. Recall the fact that for a standard Gaussian g in $\mathbb{R}^k$ for some $k \in \mathbb{N}$ and $x \in \mathbb{R}^k$ we have $\sqrt{p}M^{-1}|x| \leq \|\langle x, g \rangle\|_{L^p} \leq M\sqrt{p}|x|$ for all $p \geq 1$ and some universal constant $M > 1$. Combining this and equation (13), we arrive at

$$\|f - \mathbb{E}_\kappa f\|_{L^p(\mathbb{P}_\kappa)} \leq K \left(\|\langle \nabla f, G \rangle\|_{L^p} + (2N)^{-1/2}\|\langle \nabla^2 f, H \rangle\|_{L^p} \right), \quad (14)$$

for $K := \sqrt{6\theta}M$. Here, G is an N-dimensional standard Gaussian and H is an N^2-dimensional standard Gaussian such that G and H are independent of each other and of the ω_i, and the L^p norms on the right-hand side are taken with respect to the product measure of $\mathbb{P}_\kappa$ and the Gaussians.

Note that $\langle \nabla f, G \rangle$ and $\langle \nabla^2 f, H \rangle$ are again multilinear polynomials in the ω_i. Moreover, $\langle \nabla \langle \nabla f, G_1 \rangle, G_2 \rangle = \langle \nabla^2 f, G_1 \otimes G_2 \rangle$ and $\langle \nabla^2 \langle \nabla f, G \rangle, H \rangle = \langle \nabla^3 f, G \otimes H \rangle$. In the last expression, we regard $\nabla^3 f$ as a 2-tensor whose second component is N^2-dimensional. Similar relations also hold for the other terms in (14).

The proof now follows by iterating (14). For simplicity of presentation, let us consider the case of $d = 2$ first. Here, we apply the triangle inequality (in the form $\|\langle \nabla f, G \rangle\|_{L^p} \leq \|\langle \mathbb{E}_\kappa \nabla f, G \rangle\|_{L^p} + \|\langle \nabla f - \mathbb{E}_\kappa \nabla f, G \rangle\|_{L^p}$ and similarly for $\langle \nabla^2 f, H \rangle$) to (14). We may then apply (14) to $\langle \nabla f - \mathbb{E}_\kappa \nabla f, G \rangle$ and $\langle \nabla^2 f - \mathbb{E}_\kappa \nabla^2 f, H \rangle$ again. This leads to

$$\begin{aligned}
\|f &- \mathbb{E}_\kappa f\|_{L^p(\mathbb{P}_\kappa)} \\
&\leq K\|\langle \mathbb{E}_\kappa \nabla f, G \rangle\|_{L^p} + K(2N)^{-1/2}\|\langle \mathbb{E}_\kappa \nabla^2 f, H \rangle\|_{L^p} \\
&\quad + K^2\|\langle \nabla^2 f, G_1 \otimes G_2 \rangle\|_{L^p} + 2K^2(2N)^{-1/2}\|\langle \mathbb{E}_\kappa \nabla^3 f, G \otimes H \rangle\|_{L^p} \\
&\quad + K^2(2N)^{-1}\|\langle \mathbb{E}_\kappa \nabla^4 f, H_1 \otimes H_2 \rangle\|_{L^p} \\
&= K\|\langle \mathbb{E}_\kappa \nabla f, G \rangle\|_{L^p} + K(2N)^{-1/2}\|\langle \mathbb{E}_\kappa \nabla^2 f, H \rangle\|_{L^p} \\
&\quad + K^2\|\langle \mathbb{E}_\kappa \nabla^2 f, G_1 \otimes G_2 \rangle\|_{L^p}.
\end{aligned} \quad (15)$$

In the last step, we have used that since f is a multilinear polynomial of degree 2, its second-order derivatives are constant and all derivatives of order larger than 2 vanish.

Next, we use that by [22], there are constants C_k depending on k only such that for any (possibly rectangular) k-tensor A and any $p \geq 2$,

$$\|\langle A, g_1 \otimes \cdots \otimes g_k \rangle\|_{L^p} \leq C_k \sum_{\mathcal{I} \in P_k} p^{|\mathcal{I}|/2} \|A\|_{\mathcal{I}}, \tag{16}$$

where $g_1, \ldots, g_k$ are standard Gaussians. Applying (16) to (15), we obtain for some absolute constant C

$$
\begin{aligned}
\|f &- \mathbb{E}_\kappa\, f\|_{L^p(\mathbb{P}_\kappa)} \\
&\leq\; K \|\langle \mathbb{E}_\kappa \nabla f, G \rangle\|_{L^p} + K(2N)^{-1/2} \|\langle \mathbb{E}_\kappa \nabla^2 f, H \rangle\|_{L^p} + K^2 \|\langle \mathbb{E}_\kappa \nabla^2 f, G_1 \otimes G_2 \rangle\|_{L^p} \\
&\leq\; C_1 K p^{1/2} |\mathbb{E}_\kappa \nabla f| + C_1 K(2N)^{-1/2} p^{1/2} \|\mathbb{E}_\kappa \nabla^2 f\|_{\mathrm{HS}} + C_2 K^2 p^{1/2} \|\mathbb{E}_\kappa \nabla^2 f\|_{\mathrm{HS}} \\
&\quad + C_2 K^2 p \|\mathbb{E}_\kappa \nabla^2 f\|_{\mathrm{op}} \\
&\leq\; C\big(p^{1/2} |\mathbb{E}_\kappa \nabla f| + p^{1/2} \|\mathbb{E}_\kappa \nabla^2 f\|_{\mathrm{HS}} + p \|\mathbb{E}_\kappa \nabla^2 f\|_{\mathrm{op}} \big).
\end{aligned}
$$

From here, the assertion follows by standard arguments, cf., e.g., [18, Proposition 4].

Finally, we consider an arbitrary $d \geq 2$ and explain how the proof given above generalizes. First, we apply the triangle inequality to (14) and iterate $d - 1$ times. This yields

$$\|f - \mathbb{E}_\kappa\, f\|_{L^p(\mathbb{P}_\kappa)} \leq \psi_d + \sum_{i=1}^{d-1} \psi_i, \tag{17}$$

where we have

$$\psi_d := \sum_{\ell=0}^{d} \binom{d}{\ell} K^d (2N)^{-\ell/2} \|\langle \nabla^{d+\ell} f, G_1 \otimes \cdots \otimes G_{d-\ell} \otimes H_1 \otimes \cdots \otimes H_\ell \rangle\|_{L^p},$$

$$\psi_i := \sum_{\ell=0}^{i} \binom{i}{\ell} K^i (2N)^{-\ell/2} \|\langle \mathbb{E}_\kappa \nabla^{i+\ell} f, G_1 \otimes \cdots \otimes G_{i-\ell} \otimes H_1 \otimes \cdots \otimes H_\ell \rangle\|_{L^p} \tag{18}$$

for any $i = 1, \ldots, d - 1$. As f is a multilinear polynomial of degree d, these expressions simplify since the derivatives of order d are constant and all derivatives of higher order vanish. In particular,

$$\psi_d = K^d \|\langle \mathbb{E}_\kappa \nabla^d f, G_1 \otimes \cdots \otimes G_d \rangle\|_{L^p}.$$

Now, as above we apply (16) to (17) (or rather the L^p norms appearing in (18)) to arrive at

$$\|f - \mathbb{E}_\kappa\, f\|_{L^p(\mathbb{P}_\kappa)} \leq C \sum_{k=1}^d \sum_{\mathcal{I}\in P_k} p^{|\mathcal{I}|/2} \|\mathbb{E}_\kappa\, \nabla^k f\|_{\mathcal{I}}$$

for some absolute constant $C > 0$ depending on d only. In particular, we use that if we apply (16) to some $\ell \geq 1$ term in ψ_i in (18), the norms which arise reappear in the norms corresponding to $\ell = 0$ in the $\psi_{i+\ell}$ terms. The proof is concluded by recalling [18, Proposition 4] again.　□

Proof of Corollary 1 The proof works by calculating $\|\mathbb{E}_\kappa\, \nabla^k f\|_{\mathcal{I}}$ for $k = 1, 2, 3$ and applying Theorem 1. In the sequel, we use the convention $\omega_{ji} := \omega_{ij}$ whenever $j > i$. It is easy to see that for any edge $e = \{i, j\}$, we have

$$\frac{\partial}{\partial \omega_e} f(\omega) = \sum_{k\in\{1,\dots,n\}\setminus\{i,j\}} \omega_{ik}\omega_{jk}.$$

Moreover, the second-order derivatives $\partial^2 f/(\partial\omega_{e_1}\partial\omega_{e_2})$ are zero unless e_1 and e_2 share exactly one vertex, in which case it is ω_{ij} if i and j are the two vertices distinct from the common one. Finally, the third-order derivatives $\partial^3 f/(\partial\omega_{e_1}\partial\omega_{e_2}\partial\omega_{e_3})$ are 1 if e_1, e_2, e_3 form a triangle and zero if not.

Using that

$$\mathbb{E}\,\omega_{e_1}\cdots\omega_{e_k} = \frac{M(M-1)\cdots(M-k+1)}{N(N-1)\cdots(N-k+1)},$$

for any $k = 1, \dots, N$ and any pairwise distinct set of edges $e_1, \dots, e_k$, we therefore obtain

$$\|\mathbb{E}\,\nabla f\|_{\{1\}} = \sqrt{N}(n-2)\frac{M(M-1)}{N(N-1)} \leq n^2 p^2.$$

Moreover, we have $\mathbb{E}\,\nabla^2 f = p(\mathbb{1}_{|e_1\cap e_2|=1})_{e_1,e_2}$, where $|e_1 \cap e_2|$ denotes the number of common vertices of e_1 and e_2. Therefore, we may use the calculations from the proof of [5, Proposition 5.5], which yield

$$\|\mathbb{E}\,\nabla^2 f\|_{\{1,2\}} \leq pn^{3/2}, \qquad \|\mathbb{E}\,\nabla^2 f\|_{\{1\},\{2\}} \leq 2pn,$$
$$\|\mathbb{E}\,\nabla^3 f\|_{\{1,2,3\}} \leq n^{3/2}, \qquad \|\mathbb{E}\,\nabla^3 f\|_{\{1\},\{2\},\{3\}} \leq 2^{3/2},$$
$$\|\mathbb{E}\,\nabla^3 f\|_{\{1,2\},\{3\}} = \|\mathbb{E}\,\nabla^3 f\|_{\{1,3\},\{2\}} = \|\mathbb{E}\,\nabla^3 f\|_{\{2,3\},\{1\}} \leq \sqrt{2n}.$$

The proof now follows by plugging in.　□

3.2 Proofs of Section 1.2

The results of Sect. 1.2 follow from the logarithmic Sobolev inequalities established in Sect. 2 by standard means.

Proof of Proposition 2 Noting that

$$
\mathfrak{h}(f) = \left(\frac{1}{2} \sum_{i=1}^{n} \left(\sup_{\omega_1} f(\omega) - \inf_{\omega_1'} f(\omega_1', \omega_2, \ldots, \omega_n) \right)^2 \right)^{1/2} \le \left(\frac{c^2 n}{2} \right)^{1/2},
$$

(7) follows from Proposition 7 using the arguments from the proof of Proposition 1. To prove (6) define the symmetric function $g(\omega) := \frac{1}{n!} \sum_{\sigma \in S_n} f(\sigma \omega)$, and observe that by exchangeability of the ω_i we have $\mathbb{E}_{\kappa,n} f = \mathbb{E}_{\kappa,n} g$. Moreover,

$$
|g(\omega) - g(\omega_1', \omega_2, \ldots, \omega_n)| \le \frac{1}{n!} \sum_{\sigma \in S_n} |f(\sigma \omega) - f(\sigma(\omega_1', \omega_2, \ldots, \omega_n))|
$$

$$
\le \frac{1}{n!} \sum_{\sigma \in S_n} \sum_{i=1}^{n} \mathbb{1}_{\sigma(1)=i} c_i \le \frac{1}{n} \sum_{i=1}^{n} c_i.
$$

Applying equation (7) to g and using Jensen's inequality yield

$$
\mathbb{P}_{\kappa,n}(g - \mathbb{E}_{\kappa,n} g \ge t) \le \exp\left(-\frac{t^2}{4(1 - n/N)n(n^{-1}\sum_i c_i)^2} \right)
$$

$$
\le \exp\left(-\frac{t^2}{4(1 - n/N)\sum_i c_i^2} \right)
$$

as claimed. □

Proof of Corollary 2 This follows immediately from Proposition 2, as $f(\omega) = \frac{1}{n}\sum_{i=1}^{n} \omega_i$ is a symmetric function satisfying $|f(\omega) - f(\omega_1', \omega_2, \ldots, \omega_n)| \le |\mathcal{X}|/n$. □

Proof of Corollary 3 This is a consequence of Proposition 2, as for any $\omega \in \Omega_{\kappa,n}$ and ω_1' we have by the reverse triangle inequality

$$
|f(\omega) - f(\omega_1', \omega_2, \ldots, \omega_n)| \le n^{-1} \sup_{t \in \mathbb{R}} |\mathbb{1}_{(-\infty,t]}(\omega_1) - \mathbb{1}_{(-\infty,t]}(\omega_1')| \le n^{-1}.
$$

□

3.3 Proofs of Section 1.3

To prove Talagrand's convex distance inequality on the multislice, we follow the approach by Boucheron, Lugosi and Massart [9], see also [26, Proposition 1.9]. A key step in the proof is the following lemma.

Lemma 2 *Let $f : \Omega_\kappa \to \mathbb{R}$ be a non-negative function such that*

1. $\Gamma^+(f)^2 \le f$,
2. $|f(\omega) - f(\tau_{ij}\omega)| \le 1$ for all ω, i, j.

Then, for all $t \in [0, \mathbb{E}_\kappa f]$ we have

$$\mathbb{P}_\kappa \left(\mathbb{E}_\kappa f - f \ge t\right) \le \exp\left(-\frac{t^2}{32 \, \mathbb{E}_\kappa f}\right).$$

Particularly, we have

$$\mathbb{P}_\kappa(f = 0) \exp\left(\frac{\mathbb{E}_\kappa f}{32}\right) \le 1.$$

In particular, this holds for $f(\omega) = \frac{1}{4} d_T(\omega, A)^2$, where $A \subset S_n$ is any set.

We defer the proof of Lemma 2 until the end of the section and first show how to apply it to prove Talagrand's convex distance inequality.

Proof of Proposition 3 The difference operator Γ^+ clearly satisfies $\Gamma^+(g^2) \le 2g\Gamma^+(g)$ for all positive functions g, as well as a $\Gamma^+ - \text{mLSI}(8)$. Moreover, as we will see in the proof of Lemma 2, we have $\Gamma^+(d_T(\cdot, A)) \le 1$. Thus, by [26, (3.6)] it holds for $\lambda \in [0, 1/16)$

$$\mathbb{P}_\kappa(A) \, \mathbb{E}_\kappa \exp\left(\lambda d_T(\cdot, A)^2\right) \le \mathbb{P}_\kappa(A) \exp\left(\frac{\lambda}{1 - 16\lambda} \, \mathbb{E}_\kappa \, d_T(\cdot, A)^2\right).$$

Furthermore, Lemma 2 shows that

$$\mathbb{P}_\kappa(A) \exp\left(\frac{\mathbb{E}_\kappa \, d_T(\cdot, A)^2}{128}\right) \le 1.$$

So, for $\lambda = 1/144$ we have

$$\mathbb{P}_\kappa(A) \, \mathbb{E}_\kappa \exp\left(\frac{d_T(\cdot, A)^2}{144}\right) \le \mathbb{P}_\kappa(A) \exp\left(\frac{1}{128} \, \mathbb{E}_\kappa \, d_T(\cdot, A)^2\right) \le 1.$$

$\square$

Proofs of Corollaries 4 and 6 These corollaries follow in exactly the same way as the proof of [28, Theorem 3]. The only difference is to note that for any $x, y \in \{f \le \text{med}(f)\}$ such that $f(x) \ge \text{med}(f) + t$ we have

$$t \le \text{med}(f) + t - f(y) \le f(x) - f(y) \le L|x - y| \le L|\mathcal{X}| \sup_{\alpha \in \mathbb{R}^n : |\alpha| = 1} \sum_{i=1}^n \alpha_i \mathbb{1}_{x_i \ne y_i},$$

so that

$$f(x) \geq \mathrm{med}(f) + t \Rightarrow d_T(x, A) \geq t/(|\mathcal{X}|L).$$

$\square$

Proof of Corollary 5 Since $\lambda_{\max} = \|X\|_{\mathrm{op}}$, it is clear by triangular inequality that $\lambda_{\max}$ is a convex function of the X_{ij}, $i \leq j$. Moreover, due to Lidskii's inequality (cf. [11, Exercise 3.16]), $\lambda_{\max}$ is 1-Lipschitz. It therefore remains to apply Corollary 4. $\square$

Proof of Lemma 2 Rewriting [25, Lemma 1], for any positive function g it holds

$$\mathrm{Ent}_\kappa(g) \leq \frac{1}{2N} \sum_{i,j} \mathbb{E}_\kappa(g(\tau_{ij}\omega) - g(\omega))(\log g(\tau_{ij}\omega) - \log g(\omega))$$

$$= \frac{1}{N} \sum_{i,j} \mathbb{E}_\kappa(g(\tau_{ij}\omega) - g(\omega))(\log g(\tau_{ij}\omega) - \log g(\omega))_+.$$

Using this, we obtain for any $\lambda \in [0, 1]$

$$\mathrm{Ent}_\kappa(e^{-\lambda f}) \leq \frac{\lambda}{N} \mathbb{E}_\kappa \sum_{i,j} (f(\omega) - f(\tau_{ij}\omega))_+ \left(\exp(-\lambda f(\tau_{ij}\omega)) - \exp(-\lambda f(\omega))\right)$$

$$= \frac{\lambda}{N} \mathbb{E}_\kappa \sum_{i,j} (f(\omega) - f(\tau_{ij}\omega))_+ (\exp(\lambda(f(\omega) - f(\tau_{ij}\omega))) - 1)e^{-\lambda f(\omega)}$$

$$= \frac{\lambda}{N} \mathbb{E}_\kappa \sum_{i,j} (f(\omega) - f(\tau_{ij}\omega))_+ \Psi(\lambda(f(\omega) - f(\tau_{ij}\omega)))e^{-\lambda f(\omega)},$$

where $\Psi(x) := e^x - 1$. By a Taylor expansion, it can easily be seen that $\Psi(x) \leq 2x$ for all $x \in [0, 1]$. Therefore, recalling that by assumption 2 we have $f(\omega) - f(\tau_{ij}\omega) \leq 1$, and $f(\omega) - f(\tau_{ij}\omega) \geq 0$ due to the positive part, and using assumption 1 in the last step, we obtain

$$\mathrm{Ent}_\kappa\left(e^{-\lambda f}\right) \leq \frac{2\lambda^2}{N} \mathbb{E}_\kappa \sum_{i,j} \left(f(\omega) - f(\tau_{ij}\omega)\right)_+^2 e^{-\lambda f(\omega)}$$

$$= 8\lambda^2 \mathbb{E}_\kappa \Gamma^+(f)^2 e^{-\lambda f} \leq 8\lambda^2 \mathbb{E}_\kappa f e^{-\lambda f}.$$

The covariance of $fe^{-\lambda f}$ is non-positive (i.e., $\mathbb{E} f e^{-\lambda f} \leq \mathbb{E} f \mathbb{E} e^{-\lambda f}$), which yields

$$\mathrm{Ent}_\kappa(e^{-\lambda f}) \leq 8\lambda^2 \mathbb{E}_\kappa f \mathbb{E}_\kappa e^{-\lambda f}.$$

In other terms, if we set $h(\lambda) := \mathbb{E}_\kappa e^{-\lambda f}$, we have

$$\left(\frac{\log h(\lambda)}{\lambda}\right)' \leq 8 \mathbb{E}_\kappa f,$$

which by the fundamental theorem of calculus implies for all $\lambda \in [0, 1]$

$$\mathbb{E}_\kappa \exp\left(\lambda(\mathbb{E}_\kappa\, f - f)\right) \leq \exp\left(8\lambda^2\, \mathbb{E}_\kappa\, f\right).$$

So, for any $t \in [0, \mathbb{E}_\kappa\, f]$, by Markov's inequality and setting $\lambda = \frac{t}{16\,\mathbb{E}_\kappa\, f}$

$$\mathbb{P}_\kappa(\mathbb{E}_\kappa\, f - f \geq t) \leq \exp\left(-\lambda t + 8\lambda^2\, \mathbb{E}_\kappa\, f\right) = \exp\left(-\frac{t^2}{32\,\mathbb{E}_\kappa\, f}\right).$$

The second part follows by non-negativity and $t = \mathbb{E}_\kappa\, f$.

It remains to check that $f(\omega) = \frac{1}{4}d_T(\omega, A)^2$ satisfies the two conditions of this lemma. To this end, we first show that $\Gamma^+(d_T(\cdot, A))^2 \leq 1$. Writing $g(\omega) := d_T(\omega, A)$, it is well known (see [9]) that by Sion's minimax theorem, we have

$$g(\omega) = \inf_{\nu \in \mathcal{M}(A)} \sup_{\alpha \in \mathbb{R}^N : |\alpha| = 1} \sum_{k=1}^N \alpha_k \nu\left(\omega' : \omega'_k \neq \omega_k\right), \tag{19}$$

where $\mathcal{M}(A)$ is the set of all probability measures on A. To estimate $\Gamma^+(g)^2(\omega)$, one has to compare $g(\omega)$ and $g(\tau_{ij}\omega)$. To this end, for any $\omega \in \Omega_\kappa$ fixed, let $\tilde{\alpha}$, $\tilde{\nu}$ be parameters for which the value $g(\omega)$ is attained, and let $\hat{\nu} = \hat{\nu}_{ij}$ be a minimizer of $\inf_{\nu \in \mathcal{M}(A)} \sum_{k=1}^N \tilde{\alpha}_k \nu(\omega' : \omega'_k \neq (\tau_{ij}\omega)_k)$. This leads to

$$\Gamma^+(g)(\omega)^2 \leq \frac{1}{4N} \sum_{i,j=1}^N \left(\sum_{k=1}^N \tilde{\alpha}_k \left(\hat{\nu}\left(\omega'_k \neq \omega_k\right) - \hat{\nu}\left(\omega'_k \neq \left(\tau_{ij}\omega\right)_k\right)\right)\right)^2_+$$

$$\leq \frac{1}{2N} \sum_{i,j=1}^N \left(\tilde{\alpha}_i^2 + \tilde{\alpha}_j^2\right) \leq 1.$$

Using this as well as $\Gamma^+(g^2) \leq 2g\Gamma^+(g)$ for all positive functions g, we have

$$\Gamma^+(f)^2 = \frac{1}{16}\Gamma^+\left(d_T(\cdot, A)^2\right)^2 \leq \frac{1}{4}d_T(\cdot, A)^2\Gamma^+(d_T(\cdot, A))^2 \leq f.$$

To show the second property, we proceed similarly to [10, Proof of Lemma 1]. By (19) and the Cauchy–Schwarz inequality, we have

$$f(\omega) = \frac{1}{4} \inf_{\nu \in \mathcal{M}(A)} \sum_{k=1}^N \nu\left(\omega' : \omega'_k \neq \omega_k\right)^2.$$

Assuming without loss of generality that $f(\omega) \geq f(\tau_{ij}\omega)$, choose $\hat{v} = \hat{v}_{ij} \in \mathcal{M}(A)$ such that the value of $f(\tau_{ij}\omega)$ is attained. It follows that

$$f(\omega) - f\left(\tau_{ij}\omega\right) \leq \frac{1}{4}\sum_{k=1}^{N} \hat{v}\left(\omega_k' \neq \omega_k\right)^2 - \hat{v}\left(\omega_k' \neq \left(\tau_{ij}\omega\right)_k\right)^2 \leq \frac{2}{4},$$

which finishes the proof. □

Proof of Theorem 2 Since A is a symmetric set, $\omega \mapsto d_T(\omega, A)$ is a symmetric function, which follows by the definition

$$d_T(\omega, A) = \sup_{\alpha \in \mathbb{R}^n : |\alpha|=1} \inf_{\omega' \in A} \sum_{i=1}^{n} |\alpha_i| \mathbb{1}_{\omega_i \neq \omega_i'}.$$

As in (19), we may use Sion's minimax theorem to rewrite d_T as

$$d_T(\omega, A) = \inf_{v \in \mathcal{M}(A)} \sup_{\alpha \in \mathbb{R}^n : |\alpha|=1} \sum_{k=1}^{n} \alpha_k v(\omega' : \omega_k' \neq \omega_k).$$

As in the proof of Proposition 3, let $\tilde{v}, \tilde{\alpha}$ be the parameters for which the value $d_T(\omega, A)$ is attained, and let $\hat{v}, \hat{\omega}_i'$ be minimizers of $\inf_{\omega_i'} \inf_{v \in \mathcal{M}(A)} \sum_{j=1}^{n} \tilde{\alpha}_k v(\eta : \eta_k \neq (\omega_{i^c}, \omega_i')_k)$. We then have

$$\begin{aligned}
\mathfrak{h}^+(d_T(\omega, A))^2 &= \frac{1}{2}\sum_{i=1}^{n}\left(d_T(\omega, A) - \inf_{\omega_i'} d_T\left((\omega_{i^c}, \omega_i'), A\right)^2_+\right. \\
&\leq \frac{1}{2}\sum_{i=1}^{n}\left(\sum_{k=1}^{n}\tilde{\alpha}_k \hat{v}\left(\eta : \eta_k \neq \omega_k\right) - \sum_{k=1}^{n}\tilde{\alpha}_k \hat{v}\left(\eta : \eta_k \neq \left(\omega_{i^c}, \hat{\omega}_i'\right)_k\right)\right)^2_+ \\
&\leq \frac{1}{2}\sum_{i=1}^{n}\tilde{\alpha}_i^2 = \frac{1}{2}.
\end{aligned}$$

Recall that by Proposition 7, $\mathbb{P}_{\kappa,n}$ satisfies an $\mathfrak{h}^+ - \mathrm{LSI}(8(1-\frac{n}{N}))$ on the set of all symmetric functions. As a consequence, using (11) again, we obtain the sub-Gaussian estimate

$$\mathbb{P}_{\kappa,n}\left(d_T(\cdot, A) - \mathbb{E}_{\kappa,n} d_T(\cdot, A) \geq t\right) \leq \exp\left(-\frac{t^2}{8(1-n/N)}\right).$$

In the next step, we observe that by the Poincaré inequality we have

$$\mathrm{Var}(d_T(\cdot, A)) \leq 8(1-n/N)\,\mathbb{E}_{\kappa,n}\,\mathfrak{h}^+(d_T(\cdot, A))^2 \leq 4(1-n/N).$$

Hence, Chebyshev's inequality leads to

$$\left(\mathbb{E}_{\kappa,n}\, d_T(\cdot, A)\right)^2 \mathbb{P}_{\kappa,n}\left(d_T(\cdot, A) - \mathbb{E}_{\kappa,n}\, d_T(\cdot, A) \leq -\mathbb{E}_{\kappa,n}\, d_T(\cdot, A)\right) \leq 4(1 - n/N).$$

Using that $\mathbb{P}_{\kappa,n}(A) \geq 1/2$, we therefore have $\mathbb{E}_{\kappa,n}\, d_T(\cdot, A) \leq \sqrt{8(1 - n/N)}$. Finally, since $(t - a)^2 \geq t^2/2 - a^2$ for any $a \in \mathbb{R}$ we obtain for $t \geq \sqrt{8(1 - n/N)}$

$$\mathbb{P}_{\kappa,n}(d_T(\cdot, A) \geq t) \leq \exp\left(-\frac{(t - \sqrt{8(1 - n/N)})^2}{8(1 - n/N)}\right) \leq \exp\left(-\frac{t^2}{16(1 - n/N)} + 1\right).$$

For $t \leq \sqrt{8(1 - n/N)}$, the inequality holds trivially, which finishes the proof. $\qquad\square$

Data availability. Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

References

1. Adamczak, R.: A note on the Hanson-Wright inequality for random vectors with dependencies. Electron. Commun. Probab. **20**(72), 13 (2015)
2. Adamczak, R., Bednorz, W., Wolff, P.: Moment estimates implied by modified log-Sobolev inequalities. ESAIM Probab. Stat. **21**, 467–494 (2017)
3. Adamczak, R., Kotowski, M., Polaczyk, B., Strzelecki, M.: A note on concentration for polynomials in the Ising model. Electron. J. Probab. **24**(42), 22 (2019)
4. Adamczak, R., Polaczyk, B., Strzelecki, M.: Modified log-Sobolev inequalities, Beckner inequalities and moment estimates. arXiv preprint (2020), arXiv:2007.10209
5. Adamczak, R., Wolff, P.: Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order. Probab. Theory Related Fields **162**(3–4), 531–586 (2015)
6. Bardenet, R., Maillard, O.-A.: Concentration inequalities for sampling without replacement. Bernoulli **21**(3), 1361–1385 (2015)
7. Bobkov, S.G., Götze, F.: Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. J. Funct. Anal. **163**(1), 1–28 (1999)
8. Bobkov, S.G., Tetali, P.: Modified logarithmic Sobolev inequalities in discrete settings. J. Theoret. Probab. **19**(2), 289–336 (2006)
9. Boucheron, S., Lugosi, G., Massart, P.: Concentration inequalities using the entropy method. Ann. Probab. **31**(3), 1583–1614 (2003)
10. Boucheron, S., Lugosi, G., Massart, P.: On concentration of self-bounding functions. Electron. J. Probab. **14**(64), 1884–1899 (2009)
11. Boucheron, S., Lugosi, G., Massart, P.: Concentration inequalities. Oxford University Press, Oxford (2013)
12. Chatterjee, S.: The missing log in large deviations for triangle counts. Random Structures Algorithms **40**(4), 437–451 (2012)

13. Cortes, C., Mohri, M., Pechyony, D., Rastogi, A.: Stability analysis and learning bounds for transductive regression algorithms. arXiv preprint (2009), arXiv:0904.0814
14. DeMarco, B., Kahn, J.: Upper tails for triangles. Random Structures Algorithms **40**(4), 452–459 (2012)
15. Filmus, Y.: FKN theorem for the multislice, with applications. Combin. Probab. Comput. **29**(2), 200–212 (2020)
16. Filmus, Y., O'Donnell, R., Wu, X.: A Log-Sobolev Inequality for the Multislice, with Applications. LIPIcs **124**, 34:1–34:12, Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2019)
17. Götze, F., Sambale, H., Sinulis, A.: Higher order concentration for functions of weakly dependent random variables. Electron. J. Probab. **24**, no. 85, 19 (2019)
18. Götze, F., Sambale, H., Sinulis, A.: Concentration inequalities for bounded functionals via generalized log-Sobolev inequalities. J. Theoret. Probab. **34**(3), 1623–1652 (2021)
19. Greene, E., Wellner, J.A.: Exponential bounds for the hypergeometric distribution. Bernoulli **23**(3), 1911–1950 (2017)
20. Hanson, D.L., Wright, F.T.: A bound on tail probabilities for quadratic forms in independent random variables. Ann. Math. Statist. **42**, 1079–1083 (1971)
21. Janson, S., Ruciński, A.: The infamous upper tail. Random Structures Algorithms **20**(3), 317–342 (2002)
22. Latała, R.: Estimates of moments and tails of Gaussian chaoses. Ann. Probab. **34**(6), 2315–2331 (2006)
23. Ledoux, M.: On Talagrand's deviation inequalities for product measures. ESAIM Probab. Statist. **1**, 63–87 (1997)
24. Paulin, D.: The convex distance inequality for dependent random variables, with applications to the stochastic travelling salesman and other problems. Electron. J. Probab. **19**(68), 34 (2014)
25. Salez, J.: A sharp log-Sobolev inequality for the multislice. arXiv preprint (2020), arXiv:2004.05833
26. Sambale, H., Sinulis, A.: Modified log-Sobolev inequalities and two-level concentration. ALEA Lat. Am. J. Probab. Math. Stat. **18**, 855–885 (2021)
27. Serfling, R.J.: Probability inequalities for the sum in sampling without replacement. Ann. Statist. **2**, 39–48 (1974)
28. Talagrand, M.: An isoperimetric theorem on the cube and the Kintchine-Kahane inequalities. Proc. Am. Math. Soc. **104**(3), 905–909 (1988)
29. Talagrand, M.: Concentration of measure and isoperimetric inequalities in product spaces. Inst. Hautes Études Sci. Publ. Math. **81**, 73–205 (1995)
30. Tolstikhin, I.O.: Concentration inequalities for samples without replacement. Theory Probab. Appl. **61**(3), 462–481 (2017)
31. El-Yaniv, R., Pechyony, D.: Transductive Rademacher complexity and its applications. J. Artificial Intelligence Res. **35**, 193–234 (2009)