“On the rate of convergence of the empirical covariance matrix.”

Based on a joint work with

R. Adamczak, A. Pajor, N. Tomczak-Jaegermann

papers available at my webpage:
http://www.math.ualberta.ca/~alexandr/papers

$\langle \cdot, \cdot \rangle$ denotes the canonical inner product on $\mathbb{R}^n$

$| \cdot |$ denotes the canonical Euclidean norm on $\mathbb{R}^n$ or the cardinality of a set.

A random vector $X \in \mathbb{R}^n$ is called isotropic if

$\mathbb{E}\langle X, y\rangle = 0, \quad \mathbb{E}\,|\langle X, y\rangle|^2 = |y|^2 \quad$ for all $y \in \mathbb{R}^n$.

In other words, if X is centered and its covariance matrix is the identity:

$$\mathbb{E}\, X \otimes X = Id.$$

(recall: $(X \otimes Y)(z) = \langle X, z\rangle Y$).

A measure μ on $\mathbb{R}^n$ is log-concave if for every measurable $A, B \subset \mathbb{R}^n$ and every $\theta \in [0, 1]$,

$$\mu(\theta A + (1 - \theta)B) \geq \mu(A)^{\theta} \mu(B)^{(1-\theta)}$$

Examples.

1. Let $K \subset \mathbb{R}^n$ be a convex body and X be a random vector uniformly distributed in K. The corresponding probability measure is

$$\mu_K(A) = \frac{\text{vol}(K \cap A)}{\text{vol}(K)}$$

is log-concave. Moreover, for every convex body K there exists an affine transform T such that μ_{TK} is isotropic.

2. The Gaussian vector $G = (g_1, ..., g_n)$, where g_i's have $\mathcal{N}(0,1)$ distribution, is isotropic and log-concave.

3. Similarly the vector $X = (\xi_1, ..., \xi_n)$, where ξ_i's have exponential distribution, is isotropic and log-concave.

Let $X \in \mathbb{R}^n$ be a centered random vector with covariance matrix

$$\Sigma = \mathbb{E}\, X \otimes X.$$

Consider N independent random vectors $(X_i)_{i \le N}$ distributed as X. By the law of large numbers, the empirical covariance matrix

$$A := \frac{1}{N} \sum_{i=1}^{N} X_i \otimes X_i \to \Sigma \qquad \text{as} \qquad N \to \infty.$$

Our aim is to give quantitative estimate of the rate of this convergence, that is, to estimate the size $N = N(n, \varepsilon)$ of the sample for which

$$\left\| \frac{1}{N} \sum_{i=1}^{N} X_i \otimes X_i - \Sigma \right\| \le \varepsilon \|\Sigma\|$$

holds with high probability.

Equivalently, one can ask what is $\varepsilon = \varepsilon(n, N)$?

Passing to isotropic distributions we have to estimate relations between N, n, ε in

$$\left\| \frac{1}{N} \sum_{i=1}^{N} X_i \otimes X_i - \mathrm{Id} \right\| \le \varepsilon.$$

Kannan-Lovász-Simonovits question.
Given $\varepsilon > 0$ how many independent vectors X_i are needed for the empirical covariance matrix $\frac{1}{N}\sum_{i=1}^{N} X_i \otimes X_i$ to approximate the identity up to ε with overwhelming probability?

(Initially the question was asked about vectors uniformly distributed in an isotropic convex body. It was motivated by estimating the complexity in computing the volume of the body).

KLS 95/97: $N \sim C(\varepsilon, \delta)n^2$ with $\mathbf{Pr} \geq 1 - \delta$.

Bourgain 96/99: $N \sim C(\varepsilon, \delta)n \ln^3 n$.

Improved to $N \sim C(\varepsilon, \delta)n \ln^2 n$ by Rudelson and to $N \sim C(\varepsilon, \delta)n \ln n$ by Giannopoulos, Hartzoulaki, Tsolomitis and by Paouris.

ALPT: $N \sim C(\varepsilon)n$ with $\delta \sim \exp(-c\sqrt{n})$.

Question: $\varepsilon = \varepsilon(N/n)$ – ?

Bai, Yin: Let A be an $n \times N$ random matrix with i.i.d. entries whose 4-th moments are bounded. Let s_1 and and s_n be the largest and the smallest random values of $A/\sqrt{n}$. Let

$$\beta = \lim_{n\to\infty} \frac{n}{N} < 1.$$

Then

$$1 - \sqrt{\beta} \le \lim_{n\to\infty} s_n \le \lim_{n\to\infty} s_1 \le 1 + \sqrt{\beta}.$$

Remark. It corresponds to $\varepsilon = \sqrt{n/N}$ in

$$\left\| \frac{1}{N} \sum_{i=1}^{N} X_i \otimes X_i - \mathrm{Id} \right\| \le \varepsilon.$$

ALPT: $$\left\| \frac{1}{N} \sum_{i=1}^{N} X_i \otimes X_i - \mathrm{Id} \right\| \le C \sqrt{\frac{n}{N} \ \ln \frac{N}{n}}.$$

Mendelson noticed that (in fact it follows from theorems by **Rudelson** and **Paouris**)

$$\left\| \frac{1}{N} \sum_{i=1}^{N} X_i \otimes X_i - \mathrm{Id} \right\| \le C \sqrt{\frac{n}{N} \ \ln n}$$

(better than ours if N is very big: if $N >> n^{\alpha}$).

Proof.

We have to show that $\forall x \in S^{n-1}$

$$S(x) := \left| \frac{1}{N} \sum_{i=1}^{N} (\langle X_i, x\rangle^2 - \mathbb{E}\langle X_i, x\rangle^2) \right| \leq \varepsilon.$$

Fix a parameter B (we choose later $B \sim \ln^2(N/n)$) and observe that $S(x)$ is bounded by

$$\left| \frac{1}{N} \sum_{i=1}^{N} \Big((|\langle X_i, x\rangle| \wedge B)^2 - \mathbb{E}\,(|\langle X_i, x\rangle| \wedge B)^2 \Big) \right|$$

$$+ \frac{1}{N} \sum_{i=1}^{N} \Big(|\langle X_i, x\rangle|^2 - B^2 \Big) \mathbf{1}_{\{|\langle X_i, x\rangle| \geq B\}}$$

$$+ \frac{1}{N} \mathbb{E} \sum_{i=1}^{N} \Big(|\langle X_i, x\rangle|^2 - B^2 \Big) \mathbf{1}_{\{|\langle X_i, x\rangle| \geq B\}}$$

We denote summands under the supremum by $S_1 = S_1(x)$, $S_2 = S_2(x)$, $S_3 = S_3(x)$ and estimate them separately.

Estimate for S_1:

Given $x \in S^{n-1}$ let

$Z_i = Z_i(x) = (|\langle X_i, x\rangle| \wedge B)^2 - \mathbb{E}\,(|\langle X_i, x\rangle| \wedge B)^2.$

Then $|Z_i| \le B^2$, and

$$\mathrm{Var}(Z_i) \le \mathbb{E}\,(|\langle X_i, x\rangle| \wedge B)^4 \le \mathbb{E}|\langle X_i, x\rangle|^4 \le M_4.$$

Thus, by Bernstein's inequality

$$\mathbf{P}\left(\frac{1}{N}\sum_{i=1}^{N} Z_i \ge \theta\right) \le \exp\left(-\frac{\theta^2 N}{2(M_4 + B^2\theta/3)}\right)$$

Let $\mathcal{N}$ be a θ-net in S^{n-1} of cardinality $(3/\theta)^n$. Then by the union bound if

$$\theta^2 N > 4M_4 n \ln\frac{3}{\theta} \quad \text{and} \quad \theta N > (4/3)B^2 n \ln\frac{3}{\theta}$$

then

$$\mathbf{P}\left(\sup_{x\in\mathcal{N}} S_1(x) \ge \theta\right) \le \exp\left(-\frac{\theta^2 N}{4(M_4 + B^2\theta/3)}\right).$$

Estimates for S_2 and S_3: Denote

$$E_B = E_B(x) = \{i \le N : |\langle X_i, x\rangle| \ge B\},$$

$$m = \sup_{x \in S^{n-1}} |E_B(x)|$$

and

$$A_m = \sup\{|Ax| \ \mid \ x \in S^{N-1}, \ \ |\text{supp}\ x| \le m\},$$

where A is the matrix, whose columns are X_i's.

Then $$B^2|E_B| \le \sum_{i \in E_B} |\langle X_i, x\rangle|^2 \le A_m^2,$$

which yields $$B^2 m \le A_m^2$$

ALPT's Theorem implies that with probability at least $1 - \exp(-c\sqrt{n})$ one has

$$B^2 m \le C\left(n + m \ln^2 \frac{N}{m}\right).$$

Choosing $B = 2\sqrt{2C}\ln(3N/n)$, we observe

$$m \le 2Cn/B^2 \le n/4 \ln^2 \frac{3N}{n}.$$

Using the definition of A_m again we observe

$$\sup_{x\in S^{n-1}} \sum_{i=1}^{n} |\langle X_i, x\rangle|^2 \mathbf{1}_{\{|\langle X_i,x\rangle|\geq B\}}$$

$$= \sup_{x\in S^{n-1}} \sum_{i\in E_B} |\langle X_i, x\rangle|^2$$

$$\leq \sup_{x\in S^{n-1}} \sup_{|E|\leq m} \sum_{i\in E} |\langle X_i, x\rangle|^2 \leq A_m^2$$

and similarly

$$\sup_{x\in S^{n-1}} \mathbb{E} \sum_{i=1}^{n} |\langle X_i, x\rangle|^2 \mathbf{1}_{\{|\langle X_i,x\rangle|\geq B\}} \leq \mathbb{E} A_m^2.$$

ALPT's Theorem implies that with probability at least $1 - \exp(-c\sqrt{n})$

$$\sup_{x\in S^{n-1}} (S_1 + S_2) \leq C \left(\frac{n}{N} + \frac{m}{N} \ln^2 \frac{N}{m}\right) \leq 2C \ \frac{n}{N}.$$

Combining estimates we obtain

$$\sup_{x\in\mathcal{N}} S(x) \le \theta + 2C\frac{n}{N}$$

with probability at least

$$1 - 2\exp(-c\sqrt{n}) - \exp\left(-\frac{\theta^2 N}{4(M_4 + B^2\theta/3)}\right).$$

Choosing

$$\theta = C_2\sqrt{M_4\ \frac{n}{N}\ \ln\frac{3N}{n}},$$

(for big enough C_2) we obtain

$$\sup_{x\in\mathcal{N}} S(x) \le (C_2 + 2C)\sqrt{M_4\ \frac{n}{N}\ \ln\frac{3N}{n}}$$

with probability at least

$$1 - 2\exp(-c\sqrt{n}) - \exp\left(-c_1 n \ln\frac{3N}{n}\right)$$

$$\ge 1 - 3\exp(-c_2\sqrt{n}),$$

where c_1 and c_2 are small constants.

The passing from the θ-net $\mathcal{N}$ to the whole sphere is standard.