Math-Net.Ru

Общероссийский математический портал

D. V. Belomestny, V. G. Spokoiny, Concentration inequalities for smooth random fields, *Теория вероятн. и ее примен.*, 2013, том 58, выпуск 2, 401–410

DOI: <https://doi.org/10.4213/tvp4515>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением
<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 54.205.225.156

26 апреля 2023 г., 10:12:26

© 2013 г.

BELOMESTNY D.*, SPOKOINY V.**

CONCENTRATION INEQUALITIES FOR
SMOOTH RANDOM FIELDS¹⁾

В этой заметке выведено точное концентрационное неравенство для максимума гладкого случайного поля по конечномерному множеству параметров. Показано, что этот максимум может быть с большой вероятностью ограничен суммой значения поля в детерминированной точке и внутренней размерности оптимизационной задачи. В качестве приложения мы доказываем экспоненциальное неравенство для функций максимального собственного значения случайной матрицы.

Ключевые слова и фразы: гладкие случайные поля, концентрационное неравенство, максимальное собственное значение случайной матрицы.

1. Introduction. Concentration inequalities is a quite active field of research, which is driven by numerous applications, see [2] and [3] for an overview. Concentration inequalities have been used in many fields of both pure and applied mathematics, including stochastic optimization, random matrix theory, geometric functional analysis, randomized algorithms, statistics, machine learning and compressed sensing. A typical situation, where the concentration inequalities are useful, is the case, where one is interested in probabilistic bounds for a random variable which is the solution of a (stochastic) optimization problem. This type of problems appear in statistics and stochastic optimisation. Many statistical estimators (e.g., the maximum-likelihood estimator) are solutions to random optimization problems. There is a substantial statistical literature dealing with concentration in statistics, see [5] for an overview. On the stochastic optimization side let us mention the bin packing problem and the travelling salesman problem, where the concentration approach leads to rather sharp probabilistic bounds for the quantities of interest. For example, in the bin packing problem we are given n items of sizes in the interval $[0, 1]$ and are required to pack them into the fewest number of unit-capacity bins as possible. In the stochastic version, the item sizes are independent random variables in the interval $[0, 1]$.

In this note we prove rather general and sharp concentration inequality for smooth random fields. As a simple corollary of the main result we get a sharp exponential inequality for a convex function of the maximal eigenvalue of a random matrix.

2. Main setup. Let $G(x, \theta)$, $\theta \in \Theta \subseteq \mathbf{R}^p$, be a family of real valued functions on $\mathbf{R}^n$ and let X be a random vector in $\mathbf{R}^n$. The purpose of this paper is to derive exponential probability bounds for the random variable

$$\sup_{\theta \in \Theta} G(X, \theta).$$

First we make the following assumptions.

* Duisburg-Essen University, Germany; Московский физико-технический институт, Москва, Россия; e-mail: denis.belomestny@uni-due.de

** Weierstrass Institute for Applied analysis and Stochastics, Berlin, Germany; Московский физико-технический институт, Москва, Россия; e-mail: spokoiny@wias-berlin.de

¹⁾ This research was partially supported by the Deutsche Forschungsgemeinschaft through the CRC 823 «Statistics of nonlinear dynamical processes», the CRC 649 «Economic Risk» and by Laboratory for Structural Methods of Data Analysis in Predictive Modeling, MIPT, RF government grant, ag. 11.G34.31.0073.

(GC) The function $G(x, \theta)$ is smooth in θ for any $x \in \mathbf{R}^n$ and the mean function $M(\theta) \stackrel{\text{def}}{=} \mathbf{E} G(X, \theta)$ is three times continuously differentiable in θ . Denote

$$\theta^* \stackrel{\text{def}}{=} \arg \max_{\theta \in \Theta} M(\theta).$$

There are a positive definite symmetric matrix D^* and a positive number $r_0 > 0$ such that

$$\nabla^2 M(\theta) \preceq -D^*, \quad |\theta - \theta^*| = r, \quad r > r_0. \quad (2.1)$$

(VI) There is a symmetric positive definite matrix V_0 such that

$$\mathbf{D} \left\{ \nabla_{\theta} G(X, \theta^*) \right\} \preceq V^2 \quad \text{and} \quad V_0^2 \succeq \varepsilon^{-2} I, \quad \mathbf{a}^2 D_0^2 \succeq V^2$$

with $\nabla^2 M(\theta^*) = -D_0^2$, a small parameter $\varepsilon \in (0, 1/2)$ and $\mathbf{a} \in \mathbf{R}$.

Introduce a centered random field:

$$\zeta(\theta) \stackrel{\text{def}}{=} G(X, \theta) - \mathbf{E} G(X, \theta)$$

and a local elliptic neighborhood of θ^* via

$$\Theta_0(r) \stackrel{\text{def}}{=} \{\theta \in \Theta: \|V_0(\theta - \theta^*)\| \leq r\}.$$

Finally we make two integrability assumptions.

(ED) There exists a constant ω_0 such that it holds for all $\theta \in \Omega_0(r)$ and all $r \leq r_0$,

$$\sup_{\gamma \in \mathbf{R}^p} \ln \mathbf{E} \exp \left\{ \lambda \frac{\gamma^\top \{\nabla \zeta(\theta) - \nabla \zeta(\theta^*)\}}{\omega_0 \varepsilon r \|V_0 \gamma\|} \right\} \leq \frac{1}{2} \nu_0^2 \lambda^2, \quad |\lambda| \leq g.$$

(Er) It holds for any $\lambda > 0$,

$$\sup_{\gamma \in \mathbf{R}^p} \ln \mathbf{E} \exp \left\{ \lambda \frac{\gamma^\top \nabla \zeta(\theta)}{\|V_0 \gamma\|} \right\} \leq \frac{1}{2} \nu_0^2 \lambda^2.$$

Discussion. Under (GC) the second order Taylor expansion of the function $M(\theta)$ at θ^* gives

$$M(\theta) = M(\theta^*) + \frac{1}{2} (\theta - \theta^*)^\top \nabla^2 M(\theta^*) (\theta - \theta^*) + R(\theta)$$

with

$$\frac{R(\theta)}{\|\theta - \theta^*\|^2} = O(\|\theta - \theta^*\|), \quad \|\theta - \theta^*\| \rightarrow 0.$$

Then under (VI)

$$\left| \frac{2(M(\theta) - M(\theta^*))}{\|D_0(\theta - \theta^*)\|^2} + 1 \right| \leq \delta_0 \varepsilon r, \quad \theta \in \Theta_0(r), \quad (2.2)$$

for some $\delta_0 > 0$. Condition (2.1) basically means that M is globally concave and together with the Taylor expansion

$$M(\theta) = M(\theta^*) + (\theta - \theta^*)^\top \nabla^2 M(\theta^* + \alpha(\theta - \theta^*)) (\theta - \theta^*), \quad \alpha \in (0, 1),$$

gives

$$M(\theta) - M(\theta^*) \leq -\frac{\lambda_{\min}(D^*)}{\lambda_{\max}(V_0^2)} r^2 \stackrel{\text{def}}{=} -b^* r^2 \quad (2.3)$$

if $\|\theta - \theta^*\| = r$.

3. Main result. Define for $\mathcal{B} \stackrel{\text{def}}{=} D_0^{-1} V_0^2 D_0^{-1}$

$$p \stackrel{\text{def}}{=} \text{tr}(\mathcal{B}), \quad v^2 \stackrel{\text{def}}{=} 2 \text{tr}(\mathcal{B}^2), \quad \lambda_0 \stackrel{\text{def}}{=} \|\mathcal{B}\|_\infty \stackrel{\text{def}}{=} \lambda_{\max}(\mathcal{B}).$$

Theorem 3.1. Under assumptions (GC), (VI), (ED), and (Er)

$$\mathbf{P} \left(\sup_{\boldsymbol{\theta} \in \Theta} G(X, \boldsymbol{\theta}) > G(X, \boldsymbol{\theta}^*) + \frac{1}{2} \lambda_0 p + c \lambda_0 (v \sqrt{x} + x) \right) \leq e^{-x},$$

for any $x > 0$ satisfying $\varepsilon \sqrt{(x+3p)} < 1$ and some constant c depending on ν_0 , b^* , δ_0 , and ω_0 only.

Applications (maximal eigenvalue). Let $A = (a_{ij})_{i,j=1}^p$ be a Hermitian random matrix with a positive definite symmetric mean matrix $\mathbf{E} A$ and let

$$G(A, \boldsymbol{\theta}) \stackrel{\text{def}}{=} \boldsymbol{\theta}^\top A \boldsymbol{\theta} - f(\|\boldsymbol{\theta}\|^2), \quad \boldsymbol{\theta} \in \Theta,$$

with $\Theta = \{\boldsymbol{\theta} \in \mathbf{R}^p : \|\boldsymbol{\theta}\| < R\}$ for some large enough $R > 0$ and a nonnegative monotone increasing smooth function f . Let f^* be the Legendre transform of f , then

$$\sup_{\boldsymbol{\theta} \in \Theta} G(A, \boldsymbol{\theta}) = f^*(\lambda_{\max}(A)).$$

Since $M(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \mathbf{E} G(A, \boldsymbol{\theta}) = \boldsymbol{\theta}^\top \mathbf{E} A \boldsymbol{\theta} - f(\|\boldsymbol{\theta}\|^2)$, we have

$$\sup_{\boldsymbol{\theta} \in \Theta} M(\boldsymbol{\theta}) = f^*(\lambda_{\max}(\mathbf{E} A))$$

and the maximum is attained at the point $\boldsymbol{\theta}^* = \sqrt{r^*} \mathbf{e}_p$, where $\mathbf{e}_p$ is the eigenvector of the matrix $\mathbf{E} A$ corresponding to its largest eigenvalue and $r^* > 0$ solves the equation $f'(r^*) = \lambda_{\max}(\mathbf{E} A)$. Moreover,

$$\nabla^2 M(\boldsymbol{\theta}) = \mathbf{E} A - f'(\|\boldsymbol{\theta}\|^2) I_p - f''(\|\boldsymbol{\theta}\|^2) \boldsymbol{\theta} \boldsymbol{\theta}^\top$$

and as a result

$$\nabla^2 M(\boldsymbol{\theta}^*) = \mathbf{E} A - \lambda_{\max}(\mathbf{E} A) I_p - f''(r^*) r^* \mathbf{e}_p \mathbf{e}_p^\top \stackrel{\text{def}}{=} -D_0^2$$

for some positive definite matrix D_0 , provided $f''(r^*) > 0$. Hence the assumption (GC) is fulfilled if f is globally convex. Assume

$$\sup_{\|\boldsymbol{\theta}\|=r^*} \sup_{\boldsymbol{\gamma} \in \mathbf{R}^p} \ln \mathbf{E} \exp \left\{ \lambda \frac{\boldsymbol{\gamma}^\top (A - \mathbf{E} A) \boldsymbol{\theta}}{\|V_0 \boldsymbol{\gamma}\|} \right\} \leq \frac{1}{2} \nu_0^2 \lambda^2, \quad (3.1)$$

where $V_0 = \mathbf{D}(A \boldsymbol{\theta}^*)$. Our main result implies

$$\begin{aligned} & \mathbf{P} \left(f^*(\lambda_{\max}(A)) - f^*(\lambda_{\max}(\mathbf{E} A)) \right. \\ & \quad \left. \geq \frac{1}{2} \lambda_0 p + (\boldsymbol{\theta}^*)^\top (A - \mathbf{E} A) \boldsymbol{\theta}^* + c \lambda_0 (v \sqrt{x} + x) \right) \leq e^{-x} \end{aligned} \quad (3.2)$$

with $p = \text{tr}(D_0^{-2} V_0^2)$ and $v^2 = \text{tr}(D_0^{-4} V_0^4)$. Furthermore it follows from (3.1)

$$\mathbf{P} \left((\boldsymbol{\theta}^*)^\top (A - \mathbf{E} A) \boldsymbol{\theta}^* > \sqrt{x} \|V_0 \boldsymbol{\theta}^*\| \right) \leq e^{\nu_0^2/2} e^{-x}. \quad (3.3)$$

Combining (3.2) with (3.3), we get

$$\begin{aligned} & \mathbf{P} \left(f^*(\lambda_{\max}(A)) - f^*(\lambda_{\max}(\mathbf{E} A)) \right. \\ & \quad \left. \geq \frac{1}{2} \lambda_0 p + \sqrt{x} \|V_0 \boldsymbol{\theta}^*\| + c \lambda_0 (v \sqrt{x} + x) \right) \leq (1 + e^{\nu_0^2/2}) e^{-x}. \end{aligned} \quad (3.4)$$

Let us compare the above inequality with the known results on the maximal eigenvalue of a random Hermitian matrix. For example, in [4] an exponential inequality for the spectral norm of a bounded Hermitian random matrix A is derived via the method of exchangeable pairs. In particular, it is shown that if $A = X_1 + \dots + X_n$, where $X_1, \dots, X_n$ are independent and identically distributed Hermitian $p \times p$ matrices satisfying

$$X_k^2 \preceq B^2, \quad k = 1, \dots, n, \quad (3.5)$$

then

$$\mathbf{P}(\lambda_{\max}(A - \mathbf{E}A) > t) \leq p \exp\left(-\frac{t^2}{2\sigma^2}\right) \quad (3.6)$$

with $\sigma^2 = n\|B^2 + \mathbf{D}(X_1)\|/2$. Inequality (3.6) is in fact equivalent to the following one:

$$\mathbf{P}\left(\lambda_{\max}(A - \mathbf{E}A) > \sqrt{2(x + \ln p)}\sigma\right) \leq e^{-x}. \quad (3.7)$$

In our setting with $f(x) = nx^2$ we get $r^* = \lambda_{\max}(\mathbf{E}A)/(2n) = \lambda_{\max}(\mathbf{E}X_1)/2$,

$$\begin{aligned} V_0^2 &= \frac{1}{2n} \lambda_{\max}(\mathbf{E}A) \mathbf{D}(A\mathbf{e}_p) = \frac{n}{2} \lambda_{\max}(\mathbf{E}X_1) \mathbf{D}(X_1\mathbf{e}_p), \\ D_0^2 &= n \cdot \lambda_{\max}(\mathbf{E}A) \left(I_p + 2\mathbf{e}_p\mathbf{e}_p^\top - \frac{\mathbf{E}X_1}{\lambda_{\max}(\mathbf{E}X_1)} \right), \end{aligned}$$

and

$$D_0^{-2}V_0^2 = \left(I_p + 2\mathbf{e}_p\mathbf{e}_p^\top - \frac{\mathbf{E}X_1}{\lambda_{\max}(\mathbf{E}X_1)} \right)^{-1} \frac{\mathbf{D}(X_1\mathbf{e}_p)}{2}.$$

Hence $p = c_1 \cdot p$ and $v = c_2 \cdot p$ for some constants c_1 and c_2 not depending on n and p . Furthermore,

$$\|V_0\boldsymbol{\theta}^*\| = \sqrt{n} \cdot \lambda_{\max}(\mathbf{E}X_1) \left\| \mathbf{e}_p^\top \mathbf{D}^{1/2}(X_1\mathbf{e}_p) \right\|.$$

and inequality (3.4) transforms to

$$\mathbf{P}\left(\lambda_{\max}^2(A) - \lambda_{\max}^2(\mathbf{E}A) \geq c(\sqrt{nx} + x + p)\right) \leq (1 + e^{\nu_0^2/2}) e^{-x} \quad (3.8)$$

with some constant $c > 0$ not depending on p and n .

Note that in the domain $\lambda_{\max}(A + \mathbf{E}A) > 1$, $p/n < 1$ inequality (3.8) is more accurate than (3.7). Moreover, while condition (3.5) basically means that A is bounded with probability 1, our assumption (3.1) only requires a sub-Gaussian behavior of $A - \mathbf{E}A$.

4. Proof of the main result. Denote $Z(\boldsymbol{\theta}, \boldsymbol{\theta}^*) \stackrel{\text{def}}{=} G(X, \boldsymbol{\theta}) - G(X, \boldsymbol{\theta}^*)$. From Proposition 4.1, Lemma 4.2, and Lemma 4.3 stated in what follows we get

$$\begin{aligned} \sup_{\boldsymbol{\theta} \in \Omega_0(r)} Z(\boldsymbol{\theta}, \boldsymbol{\theta}^*) &\leq \sup_{\boldsymbol{\theta} \in \Theta} \mathbf{Z}_\epsilon(\boldsymbol{\theta}, \boldsymbol{\theta}^*) + \diamond_\epsilon(r) \leq \frac{1}{2} \|\boldsymbol{\xi}_\epsilon\|^2 + \diamond_\epsilon(r) \\ &= \frac{1}{2} \|\boldsymbol{\xi}\|^2 + \frac{1}{2} \left\{ \|\boldsymbol{\xi}_\epsilon\|^2 - \|\boldsymbol{\xi}\|^2 \right\} + \diamond_\epsilon(r) \\ &\leq \frac{1}{2} \|\boldsymbol{\xi}\|^2 + \frac{\tau_\epsilon}{2(1 - \tau_\epsilon)} \|\boldsymbol{\xi}\|^2 + \diamond_\epsilon(r) \\ &= \frac{\|\boldsymbol{\xi}\|^2}{2(1 - \tau_\epsilon)} + \diamond_\epsilon(r). \end{aligned}$$

Now Proposition 4.2 implies

$$\mathbf{P}\left(\sup_{\boldsymbol{\theta} \in \Omega_0(r_0)} Z(\boldsymbol{\theta}, \boldsymbol{\theta}^*) > \frac{\lambda_0 \cdot \mathfrak{J}(x, \mathcal{B})}{2(1 - \tau_\epsilon)} + 6\nu_0\omega_0\epsilon r_0(1 + \sqrt{x + 3p})^2\right) \leq 4e^{-x},$$

where $\mathfrak{z}(x, \mathcal{B})$ is given by (4.5). Further, we shall prove that there are $r_0 > 0$ and a deterministic upper function $u(\boldsymbol{\theta}) \geq 0$ such that

$$\mathbf{P}\left(\sup_{\boldsymbol{\theta} \in \Theta \setminus \Omega_0(r)} \{Z(\boldsymbol{\theta}, \boldsymbol{\theta}^*) + u(\boldsymbol{\theta})\} \geq 0\right) \leq e^{-x} \quad (4.1)$$

for $r > r_0$ and $x > 0$. Inequality (4.1) then implies

$$\mathbf{P}\left(\sup_{\boldsymbol{\theta} \notin \Omega_0(r_0)} Z(\boldsymbol{\theta}, \boldsymbol{\theta}^*) \geq 0\right) \leq e^{-x}.$$

A possible way of checking condition (4.1) is based on a lower quadratic bound for the negative expectation $M(\boldsymbol{\theta})$ in the sense of condition (2.3).

Lemma 4.1. *Suppose (GC) and (Er). Let, for $r \geq r_0$,*

$$6\nu_0 \sqrt{x + 3p} \leq rb^*$$

with $x + 3p \geq 2.5$. Then

$$\mathbf{P}\left(\sup_{\boldsymbol{\theta} \notin \Omega_0(r)} Z(\boldsymbol{\theta}, \boldsymbol{\theta}^*) \geq 0\right) \leq e^{-x}.$$

P r o o f. The result follows from Theorem 5.6 with $\mu = b^*/(3\nu_0)$, $t(\mu) \equiv 0$, $\mathcal{U}(\boldsymbol{\theta}) = Z(\boldsymbol{\theta}, \boldsymbol{\theta}^*) - \mathbf{E} Z(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$, and $M(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}^*) \geq b^* \|V_0(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\|^2/2$. Lemma 4.1 is proved.

It follows now from Lemma 4.1 that the inequality

$$\sup_{\boldsymbol{\theta} \in \Theta} Z(\boldsymbol{\theta}, \boldsymbol{\theta}^*) \leq \sup_{\boldsymbol{\theta} \in \Omega_0(r_0)} Z(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$$

holds with probability at least $1 - e^{-x}$. As a result we get the desired inequality.

Auxiliary results Let δ, ϱ be nonnegative constants. Introduce for a vector $\epsilon = (\delta, \varrho)$ the following notation:

$$\begin{aligned} \mathbf{Z}_\epsilon(\boldsymbol{\theta}, \boldsymbol{\theta}^*) &\stackrel{\text{def}}{=} (\boldsymbol{\theta} - \boldsymbol{\theta}^*)^\top \nabla \zeta(\boldsymbol{\theta}^*) - \frac{1}{2} \|D_\epsilon(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\|^2 \\ &= \boldsymbol{\xi}_\epsilon^\top D_\epsilon(\boldsymbol{\theta} - \boldsymbol{\theta}^*) - \frac{1}{2} \|D_\epsilon(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\|^2, \end{aligned} \quad (4.2)$$

where $\nabla \zeta(\boldsymbol{\theta}^*) = \nabla_{\boldsymbol{\theta}} G(X, \boldsymbol{\theta}^*)$ by $\nabla M(\boldsymbol{\theta}^*) = 0$ and

$$D_\epsilon^2 = D_0^2(1 - \delta) - \varrho V_0^2, \quad \boldsymbol{\xi}_\epsilon \stackrel{\text{def}}{=} D_\epsilon^{-1} \nabla G(X, \boldsymbol{\theta}^*).$$

Here we implicitly assume that with the proposed choice of the constants δ and ϱ , the matrix D_ϵ^2 is nonnegative: $D_\epsilon^2 \geq 0$. Representation (4.2) indicates that the process $\mathbf{Z}_\epsilon(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$ has the quadratic local structure. Now, given r , fix some $\delta \geq \delta_0 \varepsilon r$ and $\varrho \geq 3\nu_0 \omega_0 \varepsilon r$ with the value δ_0 from (2.2) and ω_0 from condition (ED). Finally set $\underline{\epsilon} = -\epsilon$, so that $D_{\underline{\epsilon}}^2 = D_0^2(1 + \delta) + \varrho V_0^2$.

Proposition 4.1. *Assume (ED) and (VI). Let for some r , the values $\varrho \geq 3\nu_0 \omega_0 \varepsilon r$ and $\delta \geq \delta_0 \varepsilon r$ be such that $D_0^2(1 - \delta) - \varrho V_0^2 \geq 0$. Then*

$$\mathbf{Z}_{\underline{\epsilon}}(\boldsymbol{\theta}, \boldsymbol{\theta}^*) - \diamond_{\underline{\epsilon}}(r) \leq Z(\boldsymbol{\theta}, \boldsymbol{\theta}^*) \leq \mathbf{Z}_\epsilon(\boldsymbol{\theta}, \boldsymbol{\theta}^*) + \diamond_\epsilon(r), \quad \boldsymbol{\theta} \in \Omega_0(r),$$

with $\mathbf{Z}_\epsilon(\boldsymbol{\theta}, \boldsymbol{\theta}^)$, $\mathbf{Z}_{\underline{\epsilon}}(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$ defined by (4.2). The error terms $\diamond_\epsilon(r)$ and $\diamond_{\underline{\epsilon}}(r)$ satisfy*

$$\mathbf{P}\left(\varrho^{-1} \max\{\diamond_\epsilon(r), \diamond_{\underline{\epsilon}}(r)\} \geq \left(1 + \sqrt{x + 3p}\right)^2\right) \leq e^{-x}. \quad (4.3)$$

P r o o f. Consider for fixed r and $\epsilon = (\delta, \varrho)$ the quantity

$$\diamond_{\epsilon}(r) \stackrel{\text{def}}{=} \sup_{\boldsymbol{\theta} \in \Omega_0(r)} \left\{ Z(\boldsymbol{\theta}, \boldsymbol{\theta}^*) - \mathbf{E} Z(\boldsymbol{\theta}, \boldsymbol{\theta}^*) - (\boldsymbol{\theta} - \boldsymbol{\theta}^*)^\top \nabla Z(\boldsymbol{\theta}, \boldsymbol{\theta}^*) - \frac{\varrho}{2} \|V_0(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\|^2 \right\}.$$

Since $\delta \geq \delta_0 \epsilon r$, it holds $-M(\boldsymbol{\theta}) \geq (1 - \delta)D_0^2$ and $Z(\boldsymbol{\theta}, \boldsymbol{\theta}^*) - \mathbf{Z}_{\epsilon}(\boldsymbol{\theta}, \boldsymbol{\theta}^*) \leq \diamond_{\epsilon}(r)$. Moreover, in view of $\nabla M(\boldsymbol{\theta}^*) = 0$, the definition of $\diamond_{\epsilon}(r)$ can be rewritten as

$$\diamond_{\epsilon}(r) = \sup_{\boldsymbol{\theta} \in \Omega_0(r)} \left\{ \zeta(\boldsymbol{\theta}) - \zeta(\boldsymbol{\theta}^*) - (\boldsymbol{\theta} - \boldsymbol{\theta}^*)^\top \nabla \zeta(\boldsymbol{\theta}^*) - \frac{\varrho}{2} \|V_0(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\|^2 \right\}.$$

Now the claim of the theorem can easily be reduced to an exponential bound for the quantity $\diamond_{\epsilon}(r)$. We apply Theorem 5.4 to the process

$$\mathcal{U}(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = \frac{1}{\omega_0 \epsilon r} \left\{ \zeta(\boldsymbol{\theta}) - \zeta(\boldsymbol{\theta}^*) - (\boldsymbol{\theta} - \boldsymbol{\theta}^*)^\top \nabla \zeta(\boldsymbol{\theta}^*) \right\}, \quad \boldsymbol{\theta} \in \Omega_0(r),$$

and $H_0 = V_0$. Condition $(\mathcal{E}D)$ follows from (ED) with the same ν_0 and g in view of $\nabla \mathcal{U}(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = \{\nabla \zeta(\boldsymbol{\theta}) - \nabla \zeta(\boldsymbol{\theta}^*)\}/(\omega \epsilon r)$. So, the conditions of Theorem 5.4 are fulfilled yielding (4.3) in view of $\varrho \geq 3\nu_0 \omega_0 \epsilon r$. Proposition 4.1 is proved.

Lemma 4.2. *It holds*

$$\sup_{\boldsymbol{\theta} \in \Theta} \mathbf{Z}_{\epsilon}(\boldsymbol{\theta}, \boldsymbol{\theta}^*) \leq \sup_{\boldsymbol{\theta} \in \mathbf{R}^p} \mathbf{Z}_{\epsilon}(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = \frac{1}{2} \|\boldsymbol{\xi}_{\epsilon}\|^2.$$

Lemma 4.3. *Define $\boldsymbol{\xi} \stackrel{\text{def}}{=} D_0^{-1} \nabla \zeta(\boldsymbol{\theta}^*)$. Suppose (VI) and let $\tau_{\epsilon} \stackrel{\text{def}}{=} \epsilon r_0 (\delta_0 + 3\nu_0 \omega_0 a^2) < 1$. Then*

$$\|\boldsymbol{\xi}_{\epsilon}\|^2 - \|\boldsymbol{\xi}\|^2 \leq \frac{\tau_{\epsilon}}{1 - \tau_{\epsilon}} \|\boldsymbol{\xi}\|^2, \quad \|\boldsymbol{\xi}\|^2 - \|\boldsymbol{\xi}_{\epsilon}\|^2 \leq \frac{\tau_{\epsilon}}{1 + \tau_{\epsilon}} \|\boldsymbol{\xi}\|^2.$$

Proposition 4.2. *Let (ED) hold with $\nu_0 = 1$. Then $\mathbf{E} \|\boldsymbol{\xi}\|^2 \leq p$, and for each $x > 0$*

$$\mathbf{P}(\|\boldsymbol{\xi}\|^2 \geq \lambda_0 \cdot \mathfrak{z}(x, \mathcal{B})) \leq 2e^{-x}, \quad (4.4)$$

where $\mathfrak{z}(x, \mathcal{B})$ is defined by

$$\mathfrak{z}(x, \mathcal{B}) \stackrel{\text{def}}{=} \begin{cases} p + 2\sqrt{x}^{1/2}, & x \leq \frac{v}{18}, \\ p + 6x, & \frac{v}{18} < x. \end{cases} \quad (4.5)$$

P r o o f. It holds

$$\begin{aligned} \mathbf{E} \|\boldsymbol{\xi}\|^2 &= \mathbf{E} \operatorname{tr}(\boldsymbol{\xi} \boldsymbol{\xi}^\top) = \operatorname{tr} \left(D_0^{-1} \left[\mathbf{E} \nabla \zeta(\boldsymbol{\theta}^*) \{ \nabla \zeta(\boldsymbol{\theta}^*) \}^\top \right] D_0^{-1} \right) \\ &= \operatorname{tr} \left[D_0^{-2} \mathbf{D} \{ \nabla \zeta(\boldsymbol{\theta}^*) \} \right] \end{aligned}$$

and $\boldsymbol{\gamma}^\top \mathbf{D} \{ \nabla \zeta(\boldsymbol{\theta}^*) \} \boldsymbol{\gamma} \leq \boldsymbol{\gamma}^\top V_0^2 \boldsymbol{\gamma}$ and thus, $\mathbf{E} \|\boldsymbol{\xi}\|^2 \leq p$. The deviation bound (4.4) is proved in Corollary 5.1. Proposition 4.2 is proved.

5. Appendix. The proofs of the results below can be found in [6, Appendix A and Appendix B].

5.1. Deviation probability for quadratic forms. Assume that

$$\ln \mathbf{E} \exp \left\{ \gamma^\top \boldsymbol{\xi} \right\} \leq \frac{1}{2} \|\gamma\|^2, \quad \gamma \in \mathbf{R}^p, \quad \|\gamma\| \leq g. \quad (5.1)$$

This section presents a general exponential bound for the probability $\mathbf{P}(\|\mathcal{B}\boldsymbol{\xi}\| > y)$ with a given matrix $\mathcal{B}$ and a vector $\boldsymbol{\xi}$ obeying condition (5.1). We assume that $\mathcal{B}$ is symmetric. Define important characteristics of $\mathcal{B}$

$$p \stackrel{\text{def}}{=} \text{tr}(\mathcal{B}^2), \quad v^2 \stackrel{\text{def}}{=} 2 \text{tr}(\mathcal{B}^4), \quad \lambda^* \stackrel{\text{def}}{=} \|\mathcal{B}^2\|_\infty \stackrel{\text{def}}{=} \lambda_{\max}(\mathcal{B}^2).$$

For simplicity of formulation we suppose that $\lambda^* = 1$, otherwise one has to replace p and v^2 with p/λ^* and v^2/λ^* . Let g be given in (5.1). Define w_c by the equation

$$\frac{w_c(1 + w_c)}{(1 + w_c^2)^{1/2}} = gp^{-1/2}.$$

Define also $\mu_c = w_c^2/(1 + w_c^2) \wedge 2/3$. Note that $w_c^2 \geq 2$ implies $\mu_c = 2/3$. Further define

$$y_c^2 = (1 + w_c^2)p, \quad 2x_c = \mu_c y_c^2 + \ln \det\{I_p - \mu_c \mathcal{B}^2\}. \quad (5.2)$$

Theorem 5.1. *Let a random vector $\boldsymbol{\xi}$ in $\mathbf{R}^p$ fulfill (5.1). Then for each $x < x_c$*

$$\mathbf{P}\left(\|\mathcal{B}\boldsymbol{\xi}\|^2 - \mathbf{E}\|\mathcal{B}\boldsymbol{\xi}\|^2 > (2vx^{1/2}) \vee (6x), \|\mathcal{B}\boldsymbol{\xi}\| \leq y_c\right) \leq 2e^{-x}.$$

Moreover, for $y \geq y_c$, with $g_c = g - \sqrt{\mu_c p} = gw_c/(1 + w_c)$, it holds

$$\mathbf{P}\left(\|\mathcal{B}\boldsymbol{\xi}\| > y\right) \leq 8.4 \exp\left(-x_c - \frac{1}{2} g_c(y - y_c)\right).$$

Let us now describe the value $\mathfrak{z}(x, \mathcal{B})$ ensuring a small value for the large deviation probability $\mathbf{P}(\|\mathcal{B}\boldsymbol{\xi}\|^2 > \mathfrak{z}(x, \mathcal{B}))$. For ease of formulation, we suppose that $g^2 \geq 2p$ yielding $\mu_c^{-1} \leq 3/2$. The other case can easily be adjusted.

Corollary 5.1. *Let $\boldsymbol{\xi}$ fulfill (5.1) with $g^2 \geq 2p$. Then it holds for $x \leq x_c$ with x_c from (5.2):*

$$\begin{aligned} \mathbf{P}\left(\|\mathcal{B}\boldsymbol{\xi}\|^2 - \mathbf{E}\|\mathcal{B}\boldsymbol{\xi}\|^2 \geq \mathfrak{z}(x, \mathcal{B})\right) &\leq 2e^{-x} + 8.4e^{-x_c}, \\ \mathfrak{z}(x, \mathcal{B}) &\stackrel{\text{def}}{=} \begin{cases} 2vx^{1/2}, & x \leq \frac{v}{18}, \\ 6x, & \frac{v}{18} < x \leq x_c. \end{cases} \end{aligned}$$

For $x > x_c$

$$\mathbf{P}\left(\|\mathcal{B}\boldsymbol{\xi}\|^2 \geq \mathfrak{z}_c(x, \mathcal{B})\right) \leq 8.4e^{-x}, \quad \mathfrak{z}_c(x, \mathcal{B}) \stackrel{\text{def}}{=} \left| y_c + \frac{2(x - x_c)}{g_c} \right|^2.$$

5.2. Some results for empirical processes. This section presents some general results of the theory of empirical processes. We assume some exponential moment conditions on the increments of the process which allow us to apply the well-developed chaining arguments in Orlicz spaces; We, however, follow the more recent approach inspired by the notions of generic chaining and majorizing measures due to M. Talagrand; see, e.g., [7]. The results are close to that of [1]. We state the results in a slightly different form and present an independent and self-contained proof. The first result states a bound for local fluctuations of the process $\mathcal{U}(v)$ given on a metric space Υ . Then this result will be used for bounding the maximum of the negatively drifted process $\mathcal{U}(v) - \mathcal{U}(v_0) - \rho d^2(v, v_0)$ over a vicinity $\Upsilon_0(r)$ of the central point v_0 . The behavior of $\mathcal{U}(v)$ outside of the local central set $\Upsilon_0(r)$ is described using the *upper function* method. Namely, we construct a

multiscale deterministic function $u(\mu, v)$ ensuring that with probability at least $1 - e^{-x}$ it holds $\mu \mathcal{U}(v) + u(\mu, v) \leq \mathfrak{z}(x)$ for all $v \notin \Upsilon_0(r)$ and $\mu \in \mathbf{M}$, where $\mathfrak{z}(x)$ grows linearly in x .

Let $d(v, v_0)$ be a semi-distance on Υ . We suppose the following condition to hold.

($\mathcal{E}d$) *There exist $g > 0$, $r_0 > 0$, $\nu_0 \geq 1$ such that for any $\lambda \leq g$ and $v, v_0 \in \Upsilon$ with $d(v, v_0) \leq r_0$*

$$\ln \mathbf{E} \exp \left\{ \lambda \frac{\mathcal{U}(v) - \mathcal{U}(v_0)}{d(v, v_0)} \right\} \leq \frac{1}{2} \nu_0^2 \lambda^2. \quad (5.3)$$

Formulation of the result involves a sigma-finite measure π on the space Υ which is often called the *majorizing measure* and used in the *generic chaining* device; a typical example of choosing π is the Lebesgue measure on $\mathbf{R}^p$. Let Υ^0 be a subset of Υ , a sequence r_k be fixed with $r_0 = \text{diam}(\Upsilon^0)$ and $r_k = r_0 \cdot 2^{-k}$. Let also $\mathcal{B}_k(v) \stackrel{\text{def}}{=} \{v_0 \in \Upsilon^0 : d(v, v_0) \leq r_k\}$ be the d -ball centered at v of radius r_k and let $\pi_k(v)$ denote its π -measure:

$$\pi_k(v) \stackrel{\text{def}}{=} \int_{\mathcal{B}_k(v)} \pi(dv_0) = \int_{\Upsilon^0} \mathbf{1}(d(v, v_0) \leq r_k) \pi(dv_0).$$

Denote also

$$M_k \stackrel{\text{def}}{=} \max_{v \in \Upsilon^0} \frac{\pi(\Upsilon^0)}{\pi_k(v)}, \quad k \geq 1.$$

Finally set $c_1 = 1/3$, $c_k = 2^{-k+2}/3$ for $k \geq 2$, and define the value $\mathbf{Q}(\Upsilon^0)$ by

$$\mathbf{Q}(\Upsilon^0) \stackrel{\text{def}}{=} \sum_{k=1}^{\infty} c_k \ln(2M_k) = \frac{1}{3} \ln(2M_1) + \frac{4}{3} \sum_{k=2}^{\infty} 2^{-k} \ln(2M_k).$$

Theorem 5.2. *Let $\mathcal{U}$ be a separable process following ($\mathcal{E}d$). If Υ^0 is a d -ball in Υ with the center v^0 and the radius r_0 , i.e., $d(v, v^0) \leq r_0$ for all $v \in \Upsilon^0$, then for $\lambda \leq g_0 \stackrel{\text{def}}{=} \nu_0 g$*

$$\ln \mathbf{E} \exp \left\{ \frac{\lambda}{3\nu_0 r_0} \sup_{v \in \Upsilon^0} |\mathcal{U}(v) - \mathcal{U}(v^0)| \right\} \leq \frac{1}{2} \lambda^2 + \mathbf{Q}(\Upsilon^0).$$

Due to the result of Theorem 5.2, the bound for the maximum of $\mathcal{U}(v, v_0)$ over $v \in \mathcal{B}_r(v_0)$ grows quadratically in r . So, its applications to situations with $r^2 \gg \mathbf{Q}(\Upsilon^0)$ are limited. The following result shows that introducing a negative quadratic drift helps to state a uniform in r local probability bound. Namely, the bound for the process $\mathcal{U}(v, v_0) - \rho d^2(v, v_0)/2$ with some positive ρ over a ball $\mathcal{B}_r(v_0)$ around the point v_0 only depends on the drift coefficient ρ but not on r .

Theorem 5.3. *Let rb be such that ($\mathcal{E}d$) holds on $\mathcal{B}_{rb}(v_0)$. Let also $\mathbf{Q}(\Upsilon^0) \leq \mathbf{Q}$ for $\Upsilon^0 = \mathcal{B}_r(v_0)$ with $r \leq rb$. If $\rho > 0$ and $\mathfrak{z}$ are fixed to ensure $\sqrt{2\rho\mathfrak{z}} \leq g_0 = \nu_0 g$ and $\rho(\mathfrak{z} - 1) \geq 2$, then*

$$\begin{aligned} \ln \mathbf{P} \left(\sup_{v \in \mathcal{B}_{rb}(v_0)} \left\{ \frac{1}{3\nu_0} \mathcal{U}(v, v_0) - \frac{\rho}{2} d^2(v, v_0) \right\} > \mathfrak{z} \right) \\ \leq -\rho(\mathfrak{z} - 1) + \ln(4\mathfrak{z}) + \mathbf{Q}. \end{aligned}$$

Moreover, if $\sqrt{2\rho\mathfrak{z}} > g_0$, then

$$\begin{aligned} \ln \mathbf{P} \left(\sup_{v \in \mathcal{B}_{rb}(v_0)} \left\{ \frac{1}{3\nu_0} \mathcal{U}(v, v_0) - \frac{\rho}{2} d^2(v, v_0) \right\} > \mathfrak{z} \right) \\ \leq -g_0 \sqrt{\rho(\mathfrak{z} - 1)} + \frac{g_0^2}{2} + \ln(4\mathfrak{z}) + \mathbf{Q}. \end{aligned}$$

This result can be used for describing the concentration bound for the maximum of $(3\nu_0)^{-1}\mathcal{U}(v, v_0) - \rho d^2(v, v_0)/2$. Namely, it suffices to find $\mathfrak{z}$ ensuring the prescribed deviation probability. We state the result for a special case with $\rho = 1$ and $g_0 \geq 3$ which simplifies the notation.

Corollary 5.2. *Under the conditions of Theorem 5.3, for any $x \geq 0$ with $x + \mathbf{Q} \geq 4$*

$$\mathbf{P} \left(\sup_{v \in \mathcal{B}_{rb}(v_0)} \left\{ \frac{1}{3\nu_0} \mathcal{U}(v, v_0) - \frac{1}{2} d^2(v, v_0) \right\} > \mathfrak{z}_0(x, \mathbf{Q}) \right) \leq e^{-x},$$

where with $g_0 = \nu_0 g \geq 2$

$$\mathfrak{z}_0(x, \mathbf{Q}) \stackrel{\text{def}}{=} \begin{cases} (1 + \sqrt{x + \mathbf{Q}})^2 & \text{if } 1 + \sqrt{x + \mathbf{Q}} \leq g_0, \\ 1 + \left\{ 2g_0^{-1}(x + \mathbf{Q}) + g_0 \right\}^2 & \text{otherwise.} \end{cases} \quad (5.4)$$

Let us now discuss the special case when Υ is an open subset in $\mathbf{R}^p$, the stochastic process $\mathcal{U}(v)$ is absolutely continuous and its gradient $\nabla \mathcal{U}(v) \stackrel{\text{def}}{=} d\mathcal{U}(v)/dv$ has bounded exponential moments.

($\mathcal{E}D$) *There exist $g > 0$, $\nu_0 \geq 1$, and for each $v \in \Upsilon$, a symmetric nonnegative matrix $H(v)$ such that for any $\lambda \leq g$ and any unit vector $\gamma \in \mathbf{R}^p$,*

$$\ln \mathbf{E} \exp \left\{ \lambda \frac{\gamma^\top \nabla \mathcal{U}(v)}{\|H(v)\gamma\|} \right\} \leq \frac{1}{2} \nu_0^2 \lambda^2.$$

Consider the local sets of the elliptic form $\Upsilon_0(r) \stackrel{\text{def}}{=} \{v: \|H_0(v - v_0)\| \leq r\}$, where H_0 dominates $H(v)$ on this set: $H(v) \preceq H_0$.

Theorem 5.4. *Let ($\mathcal{E}D$) hold with some g and a matrix $H(v)$. Suppose that $H(v) \preceq H_0$ for all $v \in \Upsilon_0(r)$. Then*

$$\mathbf{P} \left(\sup_{v \in \Upsilon_0(r)} \left\{ \frac{1}{3\nu_0} \mathcal{U}(v, v_0) - \frac{1}{2} \|H_0(v - v_0)\|^2 \right\} \geq \mathfrak{z}_0(x, p) \right) \leq e^{-x},$$

where $\mathfrak{z}_0(x, p)$ coincides with $\mathfrak{z}_0(x, \mathbf{Q})$ from (5.4) for $\mathbf{Q} = c_1 p$.

The previous result can be explained as a local upper function for the process $\mathcal{U}(\cdot)$. Indeed, in a vicinity $\mathcal{B}_{r^*}(v_0)$ of the central point v_0 , it holds $(3\nu_0)^{-1}\mathcal{U}(v, v_0) \leq d^2(v, v_0)/2 + \mathfrak{z}$ with a probability exponentially small in $\mathfrak{z}$. Now we extend this local result to the whole set Υ using multiscaling arguments. To simplify the notation assume that $\mathcal{U}(v_0) \equiv 0$. Then $\mathcal{U}(v, v_0) = \mathcal{U}(v)$. We say that $u(\mu, v)$ is a *multiscale upper function* for $\mu\mathcal{U}(\cdot)$ on a subset Υ^0 of Υ if

$$\mathbf{P} \left(\sup_{\mu \in \mathbf{M}} \sup_{v \in \Upsilon^0} \left\{ \mu \mathcal{U}(v) - u(\mu, v) \right\} \geq \mathfrak{z}(x) \right) \leq e^{-x},$$

for some fixed function $\mathfrak{z}(x)$. An upper function can be used for describing the concentration sets of the point of maximum $\tilde{v} = \arg \max_{v \in \Upsilon^0} \mathcal{U}(v)$; see Theorem 5.6 in what follows.

The desired global bound requires an extension of the local exponential moment condition ($\mathcal{E}d$). Below we suppose that the pseudo-metric $d(v, v_0)$ is given on the whole set Υ . For each r this metric defines the ball $\Upsilon_0(r)$ by the constraint $d(v, v_0) \leq r$. Below the condition ($\mathcal{E}d$) is assumed to be fulfilled for any r , however the constant g may be dependent of the radius r .

($\mathcal{E}r$) *For any r , there exists $g(r) > 0$ such that (5.3) holds for all $v, v_0 \in \Upsilon_0(r)$ and all $\lambda \leq g(r)$.*

Condition $(\mathcal{E}r)$ implies a similar condition for the scaled process $\mu\mathcal{U}(v)$ with $g = \mu^{-1}g(r)$ and $d(v, v_0)$ replaced by $\mu d(v, v_0)$. Corollary 5.2 implies for any x with $1 + \sqrt{x + \mathbf{Q}} \leq g_0(r) \stackrel{\text{def}}{=} \nu_0 g(r)/\mu$

$$\mathbf{P}\left(\sup_{v \in \mathcal{B}_r(v_0)} \left\{ \frac{\mu}{3\nu_0} \mathcal{U}(v) - \frac{1}{2} \mu^2 r^2 \right\} > \mathfrak{z}_0(x, \mathbf{Q})\right) \leq e^{-x}. \quad (5.5)$$

Let now a finite or separable set $\mathbf{M}$ and a function $t(\mu) \geq 1$ be fixed such that

$$\sum_{\mu \in \mathbf{M}} e^{-t(\mu)} \leq 2. \quad (5.6)$$

One possible choice of the set $\mathbf{M}$ and the function $t(\mu)$ is to take a geometric sequence $\mu_k = \mu_0 \cdot 2^{-k}$ with any fixed μ_0 and define $t(\mu_k) = k = -\ln_2(\mu_k/\mu_0)$ for $k \geq 0$.

Putting together the bounds (5.5) for different $\mu \in \mathbf{M}$ yields the following result.

Theorem 5.5. *Suppose $(\mathcal{E}r)$ and (5.6). Then for any $x \geq 2$, there exists a random set $A(x)$ of a total probability at least $1 - 2e^{-x}$, such that it holds on $A(x)$ for any r*

$$\sup_{v \in \mathcal{B}_r(v_0)} \sup_{\mu \in \mathbf{M}(r, x)} \left[\frac{\mu}{3\nu_0} \mathcal{U}(v) - \frac{1}{2} \mu^2 r^2 - \left\{ 1 + \sqrt{x + \mathbf{Q} + t(\mu)} \right\}^2 \right] < 0,$$

where

$$\mathbf{M}(r, x) \stackrel{\text{def}}{=} \left\{ \mu \in \mathbf{M}: 1 + \sqrt{x + \mathbf{Q} + t(\mu)} \leq \frac{\nu_0 g(r)}{\mu} \right\}.$$

Let $M(v)$ be a deterministic *boundary* function. We aim at bounding the probability that a process $\mathcal{U}(v)$ hits this boundary on the set Υ . This precisely means the probability that $\sup_{v \in \Upsilon} \{ \mathcal{U}(v) - M(v) \} \geq 0$. An important observation here is that multiplication by any positive factor μ does not change the relation. This allows one to apply the multiscale result from Theorem 5.5. For any fixed x and any $v \in \mathcal{B}_r(v_0)$, define

$$\mathfrak{M}(v) \stackrel{\text{def}}{=} \sup_{\mu \in \mathbf{M}(r, x)} \left\{ \frac{1}{3\nu_0} \mu M(v) - \frac{1}{2} \mu^2 r^2 - 2t(\mu) \right\}.$$

Theorem 5.6. *Suppose $(\mathcal{E}r)$, (5.6), and $x + \mathbf{Q} \geq 2.5$. Let, given x ,*

$$\mathfrak{M}(v) \geq 2(x + \mathbf{Q}), \quad v \in \Upsilon.$$

Then

$$\mathbf{P}\left(\sup_{v \in \Upsilon} \{ \mathcal{U}(v) - M(v) \} \geq 0\right) \leq 2e^{-x}.$$

REFERENCES

1. *Bednorz W.* A theorem on majorizing measures. — Ann. Probab., 2006, v. 34, № 5, p. 1771–1781.
2. *Ledoux M.* The Concentration of Measure Phenomenon. Providence: Amer. Math. Soc., RI, 2001, 181 p.
3. *Lugosi G.* Concentration-of-Measure Inequalities. Lecture Notes, 2005; <http://www.econ.upf.edu/~lugosi/anu.pdf>
4. *Mackey L., Jordan M. I., Chen R. Y., Farrell B., Tropp J. A.* Matrix concentration inequalities via the method of exchangeable pairs. — arXiv:1201.6002.
5. *Massart P.* Some applications of concentration inequalities to statistics. — Ann. Fac. Sci. Toulouse, 2000, v. 9, № 2, p. 245–303.
6. *Spokoiny V.* Parametric estimation. Finite sample theory. — arXiv:1111.3029v4.
7. *Talagrand M.* The Generic Chaining. Berlin: Springer-Verlag, 2005, 222 p.

Поступила в редакцию
19.III.2013