ISSN 1064-5624, Doklady Mathematics, 2018, Vol. 98, No. 2, pp. 490–493. © Pleiades Publishing, Ltd., 2018.
Original Russian Text © A.A. Naumov, V.G. Spokoiny, Yu.E. Tavyrikov, V.V. Ulyanov, 2018, published in Doklady Akademii Nauk, 2018, Vol. 482, No. 5.

MATHEMATICS

# Nonasymptotic Estimates for the Closeness of Gaussian Measures on Balls

**A. A. Naumov[a,b,*], V. G. Spokoiny[a,b,c,d], Yu. E. Tavyrikov[a], and V. V. Ulyanov[a,e]**

Presented by Academician of the RAS I.A. Ibragimov March 28, 2018

Received May 7, 2018

**Abstract**—Upper bounds for the closeness of two centered Gaussian measures in the class of balls in a separable Hilbert space are obtained. The bounds are optimal with respect to the dependence on the spectra of the covariance operators of the Gaussian measures. The inequalities cannot be improved in the general case.

**DOI:** 10.1134/S1064562418060248

Let $\xi$ and $\eta$ be two Gaussian vectors in $\mathbb{R}^p$ with zero mean and covariance matrices $\Sigma_\xi$ and $\Sigma_\eta$, respectively. We want to estimate

$$\delta(\mathscr{C}) := \sup_{C\in\mathscr{C}} |\mathbb{P}(\xi\in C) - \mathbb{P}(\eta\in C)| \tag{1}$$

for some class $\mathscr{C}$ of measurable subsets of $\mathbb{R}^p$. For this purpose, we can use Pinsker's inequality (see, e.g., [1, Section 2.4]), which, in the case of the class of all Borel sets in $\mathbb{R}^p$ taken as $C$, estimates the total variation distance of measures in terms of the Kullback–Leibler divergence. In the Gaussian case, it can be written in explicit form. However, (1) frequently needs to be estimated much more accurately, but for a smaller class $\mathscr{C}$. In [2–4], for example, the class of all rectangles in $\mathbb{R}^p$ was considered and optimal estimates were found. Additionally, in statistical applications (bootstrap method or Bayesian analysis) it is very important (see, e.g., [5–7]) to obtain $p$-independent estimates of (1) for the class of all centered balls $\mathscr{C}_0 := (\{x\in\mathbb{R}^p : \|x\|\le r\}, r\ge 0)$ in $\mathbb{R}^p$ Here and below, $\|x\|$ is the Euclidean vector norm.

The goal of this paper is to estimate $\delta(\mathscr{C}_0)$ in terms of corresponding covariance operators so that the resulting estimates cannot be improved in the general case. We will consider Gaussian random elements in a separable Hilbert space $\mathbb{H}$ with zero means.

For a self-adjoint nonnegative linear operator $\mathbf{A}$ in $\mathbb{H}$ with nonincreasing eigenvalues $\lambda_1(\mathbf{A}) \ge \lambda_2(\mathbf{A}), \ldots$, let $\lambda(\mathbf{A})$ denote the diagonal operator $\operatorname{diag}(\lambda_1(\mathbf{A}), \lambda_2(\mathbf{A}), \ldots)$. For a self-adjoint linear operator $\mathbf{B}$ in $\mathbb{H}$ the trace norm is defined as $\|\mathbf{B}\|_1 := \operatorname{tr}|\mathbf{B}| := \sum_{k=1}^{\infty} |\lambda_k(\mathbf{B})|$, where $\lambda_k(\mathbf{B})$ are the eigenvalues of $\mathbf{B}$. The following theorem is the main result of this paper.

**Theorem 1.** *Let $\xi$ and $\eta$ be Gaussian elements in $\mathbb{H}$ with zero mean and covariance operators $\Sigma_\xi$ and $\Sigma_\eta$, respectively. Let $\lambda_{1\xi} \ge \lambda_{2\xi} \ge \ldots$ and $\lambda_{1\eta} \ge \lambda_{2\eta} \ge \ldots$ be the respective eigenvalues of $\Sigma_\xi$ and $\Sigma_\eta$. Then there exists an absolute constant $C$ such that*

$$\begin{aligned}\delta(\mathscr{C}_0) &= \sup_{x>0} |\mathbb{P}(\|\xi\|\le x) - \mathbb{P}(\|\eta\|\le x)| \\ &\le C((\Lambda_{1\xi}\Lambda_{2\xi})^{-1/2} + (\Lambda_{1\eta}\Lambda_{2\eta})^{-1/2}) \operatorname{tr}(|\lambda(\Sigma_\xi) - \lambda(\Sigma_\eta)|),\end{aligned} \tag{2}$$

*where $\Lambda_{k\xi}^2 := \sum_{j=k}^{\infty} \lambda_{j\xi}^2$ and $\Lambda_{k\eta}^2 := \sum_{j=k}^{\infty} \lambda_{j\eta}^2$ for $k = 1, 2$.*

Note that the expression $\operatorname{tr}(|\lambda(\Sigma_\xi) - \lambda(\Sigma_\eta)|)$ on the right-hand side of (2) does not exceed $\|\Sigma_\xi - \Sigma_\eta\|_1$ (see, e.g., [7]). Additionally, the right-hand side of (2) involves $(\Lambda_{1\xi}\Lambda_{2\xi})^{-1/2}$ and $(\Lambda_{1\eta}\Lambda_{2\eta})^{-1/2}$, which appear in estimating the density functions of $\|\xi\|^2$ and $\|\eta\|^2$, respectively (see, e.g., formula (6)). Under various conditions on the spectrum of the covariance operator, in particular, taking into account the multiplicity

[a] *National Research University Higher School of Economics, Moscow, Russia*
[b] *Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia*
[c] *Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany*
[d] *Humboldt University of Berlin, Berlin, Germany*
[e] *Faculty of Computational Mathematics and Cybernetics, Moscow State University, Moscow, Russia*
**e-mail: anaumov@hse.ru*

of the largest eigenvalue, estimates for the density of the squared norm of a Gaussian vector, were proved, for example, in [8, 9]. In this paper, however, we obtain an inequality with a more accurate dependence on the covariance matrix. It is well known that an arbitrary Gaussian element in $\mathbb{H}$ with a zero mean can be represented in the form

$$\xi \overset{d}{=} \sum_{k=1}^{\infty} \sqrt{\lambda_{k\xi}} Z_k \mathbf{u}_k, \tag{3}$$

where $Z_k, k \geq 1$, are independent standard normal random variables; $\lambda_{k\xi}, k \geq 1$, are the eigenvalues of $\Sigma_\xi$ arranged in nonincreasing order; and $\{\mathbf{u}_k\}_{k\geq 1}$ are the orthonormalized eigenvectors of $\Sigma_\xi$ corresponding to the eigenvalues $\lambda_{k\xi}$. The following lemma provides an upper bound for the maximum of the density $p_\xi(x)$ of the random variable $\|\xi\|^2$ in terms of the eigenvalues of $\Sigma_\xi$.

**Lemma 1.** *Let $\xi$ be a Gaussian element in a separable Hilbert space $\mathbb{H}$ with zero mean and covariance operator $\Sigma_\xi$. Then, for some constant c*,

$$\max_{x\geq 0} p_\xi(x) \leq c(\Lambda_{1\xi}\Lambda_{2\xi})^{-1/2}. \tag{4}$$

*In particular*,

$$\max_{x\geq 0} p_\xi(x) \leq c(\lambda_{1\xi}\lambda_{2\xi})^{-1/2}. \tag{5}$$

Estimate (5) was proved in [9]. However, (4) is much more accurate. The following three typical situations can be distinguished. In the "one-dimensional" case with $\Lambda_{2\xi} \approx 0$, the assertion of the lemma is, in fact, meaningless, i.e., the effective dimension of the problem has to be at least two. In the "two-dimensional" case with $\Lambda_{2\xi} \approx \lambda_{2\xi}$, results (4) and (5) coincide. Finally, in the "multidimensional" case with $\lambda_{1\xi} \ll \Lambda_{1\xi}$, the quantities $\Lambda_{1\xi}, \Lambda_{2\xi}$ are of the same order and the right-hand side of (4) is inversely proportional to the Frobenius norm $\Lambda_{1\xi}$ of the operator $\Sigma_\xi$. In particular, in the finite-dimensional case $\mathbb{H} = \mathbb{R}^p$ with $p \geq 2$, if $\Sigma_\xi$ is close to the identity matrix, then, according to (4), $\max_{x\geq 0} p_\xi(x) \leq cp^{-1/2}$. This estimate agrees with the maximum of the chi-square probability density function with *p* degrees of freedom.

Theorem 1 implies that at least two largest eigenvalues of each of the operators $\Sigma_\xi$ and $\Sigma_\eta$ have to be involved in the estimate for $\delta(\mathscr{C}_0)$. The example given below shows that this dependence and its form in (4) reflect the substance of the matter. Indeed, let $\xi$ and $\eta$ take values in $\mathbb{R}^3$ and their covariance matrices be diagonal, i.e., $\Sigma_\xi = \operatorname{diag}(\lambda_1, \lambda_2, \lambda_3)$ and $\Sigma_\eta = \operatorname{diag}(\mu_1, \mu_2, \mu_3)$, respectively. Then

$$\delta(\mathscr{C}_0) \geq \left|\mathbb{P}(\|\xi\| \leq \sqrt{R}) - \mathbb{P}\|\eta\| \leq \sqrt{R})\right| =: I(R),$$

where $R$ will be specified later. Define $\mathscr{E}_\xi := \left\{(x_1, x_2, x_3) \in \mathbb{R}^3 : \sum_{j=1}^{3} \lambda_j x_j^2 \leq R\right\}$ and set $\mu_1 = \lambda_1$, $\mu_2 = \lambda_2, \mu_3 = \lambda_3(1+\varepsilon)$, where $0 < \varepsilon < 1$. Obviously, $\operatorname{tr}(|\lambda(\Sigma_\xi) - \lambda(\Sigma_\eta)|) = \|\Sigma_\xi - \Sigma_\eta\|_1 = \varepsilon\lambda_3$. It is easy to see that

$$I(R) = \frac{1}{(2\pi)^{3/2}} \int\limits_{\mathscr{E}_\xi \setminus \mathscr{E}_\eta} \exp\left(-\frac{x_1^2 + x_2^2 + x_3^2}{2}\right) dx_1 dx_2 dx_3$$
$$\geq \frac{1}{(2\pi)^{3/2}} (|\mathscr{E}_\xi| - |\mathscr{E}_\eta|) \exp\left[-\frac{R}{2}\left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \frac{1}{\lambda_3}\right)\right],$$

where $|\mathscr{E}_\xi|, |\mathscr{E}_\eta|$ are the volumes of the ellipsoids $\mathscr{E}_\xi, \mathscr{E}_\eta$. Applying the formula for the volume of an ellipsoid, we obtain

$$|\mathscr{E}_\xi| - |\mathscr{E}_\eta| = \frac{4\pi R^{3/2} \|\Sigma_\xi - \Sigma_\eta\|_1}{3\sqrt{\lambda_1\lambda_2}\lambda_3^{3/2}\sqrt{1+\varepsilon}(1+\sqrt{1+\varepsilon})}$$
$$> \frac{\pi \|\Sigma_\xi - \Sigma_\eta\|_1}{\sqrt{\lambda_1\lambda_2}} \left(\frac{R}{2\lambda_3}\right)^{3/2}.$$

Taking $R := 2\lambda_3$ and using the trivial estimate $\left(\frac{R}{2\lambda_3}\right)^{3/2} \exp\left(-\frac{R}{2\lambda_3}\right) \geq \frac{1}{3}$, we get

$$I(R) \geq \frac{\|\Sigma_\xi - \Sigma_\eta\|_1}{16\sqrt{\lambda_1\lambda_2}} \cdot \exp\left(-\frac{\lambda_3}{\lambda_1} - \frac{\lambda_3}{\lambda_2}\right) \geq c \frac{\|\Sigma_\xi - \Sigma_\eta\|_1}{\sqrt{\lambda_1\lambda_2}},$$

where $c = \frac{\exp\{-2\}}{16}$. The last inequality implies

$$\sup_{x>0} \left|\mathbb{P}(\|\xi\| \leq x) - \mathbb{P}(\|\eta\| \leq R)\right| \geq c \frac{\|\Sigma_\xi - \Sigma_\eta\|_1}{\sqrt{\lambda_1\lambda_2}}$$
$$\geq c' \|\Sigma_\xi - \Sigma_\eta\|_1 ((\Lambda_{1\xi}\Lambda_{2\xi})^{-1/2} + (\Lambda_{1\eta}\Lambda_{2\eta})^{-1/2}).$$

Lemma 1 has the following simple corollary, which provides an estimate for the probability of hitting a $\Delta$-strip.

**Corollary 1.** *Let $\xi$ be a Gaussian element in $\mathbb{H}$ having zero mean and covariance operator $\Sigma_\xi$. Let $\Delta > 0$. Then there exists an absolute constant c for which*

$$\mathbb{P}(x < \|\xi\|^2 < x + \Delta) \leq c(\Lambda_{1\xi}\Lambda_{2\xi})^{-1/2}\Delta.$$

**Proof of Theorem 1.** Without loss of generality, we assume that both covariance operators $\Sigma_\xi, \Sigma_\eta$ have a diagonal form with nonincreasing eigenvalues. For every $s: 0 \leq s \leq 1$, we define a Gaussian element $Z(s)$ in $\mathbb{H}$ with zero mean and a diagonal covariance operator $\mathbf{V}(s)$:

$$\mathbf{V}(s) \overset{\text{def}}{=} s\Sigma_\xi + (1-s)\Sigma_\eta.$$

Let $\lambda_1(s) \geq \lambda_2(s) \geq \dots$ denote the eigenvalues of the operator $\mathbf{V}(s)$. Define the resolvent operator $\mathbf{G}(t, s) \stackrel{\text{def}}{=} (\mathbf{I} - 2it\mathbf{V}(s))^{-1}$. Obviously, $\mathbf{G}(t,s)$ is also a diagonal operator and the characteristic function $f(t,s)$ of $\|Z(s)\|^2$ has the form

$$f(t,s) = \mathbb{E}\exp\{it\|Z(s)\|^2\} = \exp\left\{-\frac{1}{2}\operatorname{tr}\log(\mathbf{I} - 2it\mathbf{V}(s))\right\}.$$

It is well known (see, e.g., [11, Section 6.2, p. 168]) that, for a continuous distribution function $F(x)$ and its characteristic function $f(t)$, the inversion formula

$$F(x) = \frac{1}{2} + \frac{i}{2\pi}\lim_{T\to\infty} \text{V.P.}\int_{|t|\leq T} e^{-itx} f(t)\frac{dt}{t}$$

holds. Let $x > 0$ be fixed. Then

$$\mathbb{P}(\|\xi\|^2 < x) - \mathbb{P}(\|\eta\|^2 < x)$$
$$= \frac{i}{2\pi}\lim_{T\to\infty} \text{V.P.}\int_{|t|\leq T} \frac{f(t,1) - f(t,0)}{t} e^{-itx}\,dt.$$

Using the Newton–Leibniz formula, we obtain

$$f(t,1) - f(t,0) = \int_0^1 \frac{\partial f(t,s)}{\partial s} ds$$
$$= \int_0^1 itf(t,s)\operatorname{tr}\{(\Sigma_\xi - \Sigma_\eta)\mathbf{G}(t,s)\}\,ds.$$

Changing the order of integration yields the equality

$$\mathbb{P}(\|\xi\|^2 < x) - \mathbb{P}(\|\eta\|^2 < x)$$
$$= -\frac{1}{2\pi}\int_0^1\int_{-\infty}^{\infty} \operatorname{tr}\{(\Sigma_\xi - \Sigma_\eta)\mathbf{G}(t,s)\} f(t,s)e^{-itx}dtds.$$

For a fixed $s$, consider the expression

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \mu_j(t,s)f(t,s)e^{-itx}dt,$$

where $\mu_j(t,s) := (1 - 2it\lambda_j(s))^{-1}$ are the eigenvalues of $\mathbf{G}(t,s)$. Let $\bar{Z}_j(s)$ be a random variable having an exponential distribution with parameter $\frac{1}{2\lambda_j(s)}$ and independent of $Z(s)$. Then

$$\mathbb{E}e^{it\bar{Z}_j(s)} = \mu_j(t,s).$$

Moreover, $\mu_j(t,s)f(t,s)$ is the characteristic function of the random variable $\bar{Z}_j(s) + \|Z(s)\|^2$. Let $p_j(x,s)$ be the density function corresponding to $\mu_j(t,s)f(t,s)$. Then

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \mu_j(t,s)f(t,s)e^{-itx}dt = p_j(x,s).$$

Let $\mathbf{P}(x,s)$ denote a diagonal operator with values $p_j(x,s)$ on the main diagonal. Then

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \operatorname{tr}\{(\Sigma_\xi - \Sigma_\eta)\mathbf{G}(t,s)\} f(t,s)e^{-itx}dt$$
$$= \operatorname{tr}\{(\Sigma_\xi - \Sigma_\eta)\mathbf{P}(x,s)\}.$$

It is easy to see that the second term does not exceed

$$\|\Sigma_\xi - \Sigma_\eta\|_1 \quad \|\mathbf{P}(x,s)\|, \tag{6}$$

so it remains to estimate $\|\mathbf{P}(x,s)\|$, i.e., each $p_j(x,s)$. However, for any $j$, $\max_x p_j(x,s)$ does not exceed $\max_x p(x,s)$, where $p(x,s)$ is the density function of $\|Z(s)\|^2$. Applying Lemma 1 and integrating the result with respect to $s$, we obtain the assertion of the theorem.

**Proof of Lemma 1.** For simplicity, we consider the case of $\mathbb{H} = \mathbb{R}^{2p}$ and write $\lambda_k$ instead of $\lambda_{k\xi}$. Without loss of generality, it may be assumed that $2\lambda_1^2 \leq \sum_{k=1}^{p}\lambda_{2k-1}^2$ and $2\lambda_2^2 \leq \sum_{k=1}^{p}\lambda_{2k}^2$. Otherwise, we can apply the convolution formula; inequality (5) was proved in [10] exactly by this method and induction arguments. Using representation (3) and denoting by $f(t)$ and $f_k(t), k = 1,2,\dots,2p$, the characteristic functions of $\|\xi\|^2$ and $\lambda_k Z_k^2$, respectively, we obtain

$$p_\xi(x) \leq \frac{1}{2\pi}\int_{-\infty}^{\infty}|f(t)dt| = \frac{1}{2\pi}\int_{-\infty}^{\infty}\prod_{k=1}^{p}|f_{2k-1}(t)|\prod_{k=1}^{p}|f_{2k}(t)|dt$$
$$\leq \left[\frac{1}{2\pi}\int_{-\infty}^{\infty}\prod_{k=1}^{p}|f_{2k-1}(t)|^2dt\right]^{1/2}\left[\frac{1}{2\pi}\int_{-\infty}^{\infty}\prod_{k=1}^{p}|f_{2k}(t)|^2dt\right]^{1/2},$$

where the Cauchy–Schwarz inequality was applied at the last step. The Hölder inequality yields

$$p_\xi^2(x) \leq \frac{1}{2\pi}\prod_{k=1}^{p}\|f_{2k-1}^2\|_{p_{2k-1}}\prod_{k=1}^{p}\|f_{2k}^2\|_{p_{2k}},$$

where $\sum_{k=1}^{p} p_{2k-1}^{-1} = 1$, $\sum_{k=1}^{p} p_{2k}^{-1} = 1$, $p_k \geq 2, k = 1,2,\dots,2p$, and $\|f\|_q^q = \int_{\mathbb{R}}|f(x)|^q dx, q \geq 1$. The exact values of $p_k$ will be specified later. Using the inequality $(1 + x)^a \geq 1 + ax$, where $a \geq 1$ and $x \geq 0$, we get

$$\|f_k^2\|_{p_k}^{p_k} = \int_{-\infty}^{\infty}\frac{1}{(1 + 4\lambda_k^2t^2)^{p_k/2}}dt \leq \int_{-\infty}^{\infty}\frac{dt}{1 + 2p_k\lambda_k^2t^2} \leq \frac{\pi}{\lambda_k\sqrt{2p_k}}.$$

Now let

$$p_{2k-1}^{-1} := \frac{\lambda_{2k-1}^2}{\sum_{k=1}^{p} \lambda_{2k-1}^2}, \quad p_{2k}^{-1} := \frac{\lambda_{2k}^2}{\sum_{k=1}^{p} \lambda_{2k}^2}.$$

Then, for all $k = 1, 2, \ldots, p$,

$$\|f_{2k-1}^2\|_{p_{2k-1}}^{p_{2k-1}} \le \frac{\pi}{\sqrt{2\sum_{k=1}^{p} \lambda_{2k-1}^2}}.$$

Similar inequalities hold for even indices, which leads to the required result.

In this paper we considered Gaussian random elements in a separable Hilbert space $\mathbb{H}$ with zero means. A possible extension of the present results to Gaussian measures with different means and covariance operators can be found, for example, in [12].

## ACKNOWLEDGMENTS

This work was supported by the Russian Science Foundation, grant no. 18-11-00132.

## REFERENCES

1. A. Tsybakov, *Introduction to Nonparametric Estimation* (Springer, New York, 2008).
2. V. Chernozhukov, D. Chetverikov, and K. Kato, Ann. Stat. **41** (6), 2786–2819 (2013).
3. V. Chernozhukov, D. Chetverikov, and K. Kato, Ann. Probab. **45** (4), 2309–2352 (2017).
4. V. Chernozhukov, D. Chetverikov, and K. Kato, Probab. Theory Related Fields **162** (1–2), 47–70 (2015).
5. V. Spokoiny, and M. Zhilova, Ann. Stat. **43** (6), 2653–2675 (2015).
6. A. Naumov, V. Spokoiny, and V. Ulyanov, "Bootstrap confidence sets for spectral projectors of sample covariance" (2017). arXiv:1703.00871.
7. A. Naumov, V. Spokoiny, and V. Ulyanov, "Confidence sets for spectral projectors of covariance matrices," Dokl. Math. **98** (2018).
8. A. Markus, Russ. Math. Surv. **19** (4), 91–120 (1964).
9. G. Christoph, Yu. V. Prokhorov, and V. V. Ulyanov, Theor. Probab. Appl. **40** (2), 250–260 (1996).
10. V. V. Ulyanov, "On Gaussian measure of balls in $H$," in *Frontiers in Pure and Applied Probability, Proceedings of the 4th Russian–Finnish Symposium on Probability Theory and Mathematical Statistics* (TVP Science, Moscow, 1995).
11. K. Chung, *A Course in Probability Theory*, 3rd ed. (Academic, San Diego, CA, 2001).
12. F. Götze, A. Naumov, V. Spokoiny, and V. Ulyanov, "Large ball probabilities, Gaussian comparison and anti-concentration", Bernoulli **25** (2019). arXiv:1708.08663v2.

*Translated by I. Ruzanova*