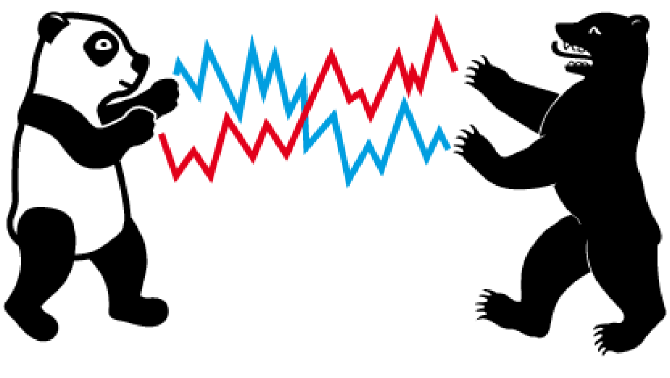
IRTG 1792 Discussion Paper 2018-028

Toolbox: Gaussian comparison on Eucledian balls

Andzhey Koziuk *
Vladimir Spokoiny *



* WIAS, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the International Research Training Group 1792 "High Dimensional Nonstationary Time Series".

http://irtg1792.hu-berlin.de
ISSN 2568-5619

International Research Training Group 1792

Toolbox: Gaussian comparison on Eucledian balls

Andzhey Koziuk

Weierstrass Institute,
Institute for Information Transmission Problems of RAS
Mohrenstrasse 39
10117 Berlin, Germany
andzhey.koziuk@wias-berlin.de

Vladimir Spokoiny

Weierstrass Institute and
Humboldt University Berlin,
Moscow Institute of Physics and Technology
Mohrenstr. 39,
10117 Berlin, Germany
spokoiny@wias-berlin.de

Abstract

In the work a characterization of difference of multivariate Gaussian measures is found on the family of centered Eucledian balls. In particular, it helps to derive

$$\sup_t \left| I\!P\left(\|\boldsymbol{x}_1\|^2 < t\right) - I\!P\left(\|\boldsymbol{x}_0\|^2 < t\right) \right| \leq \sqrt{\left| Tr\left(I - \left(\Sigma_0 \Sigma_1^{-1} + \Sigma_0^{-1}\Sigma_1\right)/2\right)\right|}$$

with vectors $\boldsymbol{x}_1 \sim \mathcal{N}(0, \Sigma_1)$ and $\boldsymbol{x}_0 \sim \mathcal{N}(0, \Sigma_0)$.

Keywords: multivariate Gaussian measure, Kolmogorov distance, Gaussian comparison

1 Introduction

The work is organized as a report and exposition is kept concise - no proof is deferred, introduction reduced to minimum and conclusion contains problems to address in view of the development.

The work centers on comparison of the measures. The prerequisite for a study was an idea that difference of multivariate probabilities being more regular than a probability enjoys an independent structure. In essence, regularity gained through subtraction was channeled into a probability to illuminate possibly existing construct. For an explorer it is inevitable to assume a guideline before the factual verification. Luckily enough such a structure was found to exist on a family of multivariate Gaussian measures on centered Eucledian balls.

2 Associated development

The most tightly related investigation and motivational examples can be found in Götze, F. and Naumov, A. and Spokoiny, V. and Ulyanov, V. [1]. Compared to the work the current development is not based on the pdf estimation and derives an explicit characterization for the difference of multivariate Gaussian measures (see corollary [3.4]) .

3 Gaussian comparison on Euclidean balls

Kolmogorov distance on a class of centered Euclidean balls

$$\mathcal{B}_t \stackrel{\text{def}}{=} \{\boldsymbol{x} \in I\!\!R^p \ : \ \|\boldsymbol{x}\| < \sqrt{t}\}$$

is studied

$$\sup_t |I\!\!P(\boldsymbol{x}_1 \in \mathcal{B}_t) - I\!\!P(\boldsymbol{x}_0 \in \mathcal{B}_t)|$$

For independent $\boldsymbol{x}_0 \sim \mathcal{N}(0, \Sigma_0)$ and $\boldsymbol{x}_1 \sim \mathcal{N}(0, \Sigma_1)$ define a composite vector

$$\boldsymbol{x}_s = \sqrt{1-s}\boldsymbol{x}_0 + \sqrt{s}\boldsymbol{x}_1$$

with $s = [0, 1]$ and following

$$\boldsymbol{x}_s \sim \mathcal{N}(0, \Sigma_s)$$

where

$$\Sigma_s = (1-s)\Sigma_0 + s\Sigma_1.$$

The vector enables construction of a bridge over the regular difference

$$g_\alpha(t) \stackrel{\text{def}}{=} I\!\!E f\left(\alpha\|\boldsymbol{x}_1\|^2 - \alpha t\right) - I\!\!E f\left(\alpha\|\boldsymbol{x}_0\|^2 - \alpha t\right) \tag{3.1}$$

$$\stackrel{\text{def}}{=} \int_0^1 I\!\!E \frac{\partial f_\alpha(\|\boldsymbol{x}_s\|^2 - t)}{\partial s} ds = \alpha \int_0^1 I\!\!E f'_\alpha\left(\|\boldsymbol{x}_s\|^2 - t\right) \boldsymbol{x}_s^T \left(\boldsymbol{x}_1 \frac{1}{\sqrt{s}} - \boldsymbol{x}_0 \frac{1}{\sqrt{1-s}}\right) ds$$

using fundamental theorem of calculus and assuming that f is sufficiently smooth.

The most direct tool to work with Gaussian measures is Stein's identity. In the current work one uses an adaptation of the fact.

Lemma 3.1 (Stein's lemma). *If $\boldsymbol{x} \in I\!\!R^p$ is a centered Gaussian with a covariance $\Sigma \succ 0$ and a vector function $h(\boldsymbol{x}) : I\!\!R^p \to I\!\!R^p$ is an almost differentiable with $I\!\!E\|\frac{\partial h_j(\boldsymbol{x})}{\partial \boldsymbol{x}}\| \leq \infty$ then*

$$I\!\!E \boldsymbol{x} h^T(\boldsymbol{x}) = \Sigma I\!\!E \frac{\partial h^T(\boldsymbol{x})}{\partial \boldsymbol{x}}. \tag{3.2}$$

Proof. In his work Charles M. Stein 1981 [2] (Lemma 2) derived for standard normal $\boldsymbol{y} \sim \mathcal{N}(0, I)$ and a function $h_j(\Sigma^{1/2}\boldsymbol{y}) : \mathbb{R}^p \to \mathbb{R}$

$$\mathbb{E}\boldsymbol{y}h_j(\Sigma^{1/2}\boldsymbol{y}) = \mathbb{E}\frac{\partial h_j(\Sigma^{1/2}\boldsymbol{y})}{\partial \boldsymbol{y}} = \Sigma^{1/2}\mathbb{E}\frac{\partial h_j(\Sigma^{1/2}\boldsymbol{y})}{\partial \Sigma^{1/2}\boldsymbol{y}}.$$

A change of variables $\boldsymbol{x} = \Sigma^{1/2}\boldsymbol{y}$ yields $\mathbb{E}\boldsymbol{x}h_j(\boldsymbol{x}) = \Sigma\mathbb{E}\frac{\partial h_j(\boldsymbol{x})}{\partial \boldsymbol{x}}$. Therefore, the rules of a matrix multiplication dictate

$$\mathbb{E}\boldsymbol{x}h^T(\boldsymbol{x}) = (\mathbb{E}\boldsymbol{x}h_1(\boldsymbol{x}), \mathbb{E}\boldsymbol{x}h_2(\boldsymbol{x}), ..., \mathbb{E}\boldsymbol{x}h_p(\boldsymbol{x}))$$
$$= \left(\Sigma\mathbb{E}\frac{\partial h_1(\boldsymbol{x})}{\partial \boldsymbol{x}}, \Sigma\mathbb{E}\frac{\partial h_2(\boldsymbol{x})}{\partial \boldsymbol{x}}, ..., \Sigma\mathbb{E}\frac{\partial h_p(\boldsymbol{x})}{\partial \boldsymbol{x}}\right) = \Sigma\mathbb{E}\frac{\partial h^T(\boldsymbol{x})}{\partial \boldsymbol{x}}.$$

□

For an expository reasons only let us also include in the note another - less straightforward, however applicable in a most general case - relaxation method of a probability. Introduce a smooth indicator function

$$f_\alpha(x) \stackrel{\text{def}}{=} \mathbb{1}(x > 0) - \frac{1}{2}\mathbf{sign}(x)e^{-\alpha|x|} \tag{3.3}$$

and a regular and limiting object based on an integral operator $\mathbb{L}_{\boldsymbol{x}}(\cdot)$ (e.g. expectation)

$$\phi_\alpha(t) \stackrel{\text{def}}{=} \mathbb{L}_{\boldsymbol{x}}\left(f_\alpha(\|\boldsymbol{x}\|^2 - t)\right) \text{ and } \phi_\infty(t) \stackrel{\text{def}}{=} \mathbb{L}_{\boldsymbol{x}}\left(\mathbb{1}\left(\|\boldsymbol{x}\|^2 > t\right)\right).$$

The relaxation of the indicator - the kernel (3.3) - is characterized structurally in the next lemma.

Lemma 3.2. *Assume an integral operator $\mathbb{L}_{\boldsymbol{x}}(\cdot)$ s.t. $\phi_\alpha(t)$ is smooth bounded and with bounded second derivative. Then $\phi_\alpha(t)$ satisfies an ODE*

$$\phi_\alpha(t) = \phi_\infty(t) + \phi''_\alpha(t)/\alpha^2.$$

Moreover, an ordering holds

$$\forall\alpha > 0 \quad \sup_t |\phi_\alpha(t)| \le \sup_t |\phi_\infty(t)|.$$

Proof. Notice that

$$\mathbb{L}_{\boldsymbol{x}}\left(f_\alpha(\|\boldsymbol{x}\|^2 - t) - \mathbb{1}\left(\|\boldsymbol{x}\|^2 > t\right)\right) = \frac{\left(\mathbb{L}_{\boldsymbol{x}}\left(f_\alpha(\|\boldsymbol{x}\|^2 - t)\right)\right)''_t}{\alpha}.$$

Thus the kernel (3.3) admits an ODE representation

$$\mathbb{L}_{\boldsymbol{x}}\left(f_\alpha(\|\boldsymbol{x}\|^2 - t)\right) = \mathbb{L}_{\boldsymbol{x}}\left(\mathbb{1}\left(\|\boldsymbol{x}\|^2 > t\right)\right) + \frac{1}{\alpha^2}\left(\mathbb{L}_{\boldsymbol{x}}\left(f_\alpha(\|\boldsymbol{x}\|^2 - t)\right)\right)''_t$$

with an inequality

$$\sup_t \left|\mathbb{L}_{\boldsymbol{x}}\left(f_\alpha(\|\boldsymbol{x}\|^2 - t)\right)\right| \le \sup_t \left|\mathbb{L}_{\boldsymbol{x}}\left(\mathbb{1}\left(\|\boldsymbol{x}\|^2 > t\right)\right)\right|$$

following from the characterization of extreme points - second derivative in maximum is negative and positive in minimum - and boundedness of the second derivative. □

The regular difference $g_\alpha(t)$ (3.1) with the kernel (3.3) can be equivalently rewritten by the means of Stein's lemma [3.1].

Lemma 3.3. *Denote $\boldsymbol{y} \sim \mathcal{N}(0, I)$ and uniform $s \sim \mathcal{U}(0,1)$ and define a linear operator*

$$\mathbb{L}_{\boldsymbol{y},s} \stackrel{\text{def}}{=} \frac{1}{2} I\!\!E \left(\boldsymbol{y}^T \log' \Sigma_s \boldsymbol{y} - \log' \det \Sigma_s \right) (\cdot)$$

then the function (3.1) follows

$$g_\alpha(t) = \mathbb{L}_{\boldsymbol{y},s} \left(f_\alpha \left(\boldsymbol{y}^T \Sigma_s \boldsymbol{y} - t \right) \right).$$

Proof. Independence of $\boldsymbol{x}_0$ and $\boldsymbol{x}_1$ and the chain rule for differentiation demonstrate

$$\begin{aligned}
\frac{1}{\sqrt{1-s}} Tr I\!\!E_0 \boldsymbol{x}_0 I\!\!E_1 \boldsymbol{x}_s^T f'_\alpha \left(\|\boldsymbol{x}_s\|^2 - t \right) &\stackrel{\textbf{Stein } \boldsymbol{x}_0}{=} Tr \Sigma_0 I\!\!E_0 \frac{1}{\sqrt{1-s}} \frac{\partial}{\partial \boldsymbol{x}_0} I\!\!E_1 \boldsymbol{x}_s^T f'_\alpha \left(\|\boldsymbol{x}_s\|^2 - t \right) \\
= Tr \Sigma_0 I\!\!E \frac{\partial}{\partial \boldsymbol{x}_0} \frac{\partial \boldsymbol{x}_0}{\partial \boldsymbol{x}_s} \boldsymbol{x}_s^T f'_\alpha \left(\|\boldsymbol{x}_s\|^2 - t \right) &\stackrel{\textbf{chain}}{=} Tr \Sigma_0 I\!\!E \frac{\partial}{\partial \boldsymbol{x}_s} \boldsymbol{x}_s^T f'_\alpha \left(\|\boldsymbol{x}_s\|^2 - t \right) \\
&\stackrel{\textbf{Stein } \boldsymbol{x}_s}{=} Tr \left(\Sigma_0 \Sigma_s^{-1} I\!\!E \boldsymbol{x}_s \boldsymbol{x}_s^T f'_\alpha \left(\|\boldsymbol{x}_s\|^2 - t \right) \right).
\end{aligned}$$

The difference can be smoothed further applying again Stein's identity and differentiation by parts

$$\begin{aligned}
& Tr \Sigma_0 \Sigma_s^{-1} I\!\!E \left(\boldsymbol{x}_s \boldsymbol{x}_s^T f'_\alpha \left(\|\boldsymbol{x}_s\|^2 - t \right) \right) \\
&= \frac{1}{2\alpha} Tr \Sigma_0 \Sigma_s^{-1} I\!\!E \left(\frac{\partial}{\partial \boldsymbol{x}_s} \boldsymbol{x}_s^T f_\alpha \left(\|\boldsymbol{x}_s\|^2 - t \right) - f_\alpha \left(\|\boldsymbol{x}_s\|^2 - t \right) \right) \\
&\stackrel{\textbf{Stein } \boldsymbol{x}_s}{=} \frac{1}{2\alpha} Tr \Sigma_0 \Sigma_s^{-1} I\!\!E \left(\Sigma_s^{-1} \boldsymbol{x}_s \boldsymbol{x}_s^T f_\alpha \left(\|\boldsymbol{x}_s\|^2 - t \right) - f_\alpha \left(\|\boldsymbol{x}_s\|^2 - t \right) \right) \\
&= \frac{1}{2\alpha} Tr \Sigma_0 \Sigma_s^{-1} I\!\!E \left(\Sigma_s^{-1} \boldsymbol{x}_s \boldsymbol{x}_s^T - I \right) f_\alpha \left(\|\boldsymbol{x}_s\|^2 - t \right).
\end{aligned}$$

The same applied to $\boldsymbol{x}_1$ yields analogously

$$\frac{1}{\sqrt{s}} Tr I\!\!E \boldsymbol{x}_1 \boldsymbol{x}_s^T f'_\alpha \left(\|\boldsymbol{x}_s\|^2 - t \right) = \frac{1}{2\alpha} Tr \Sigma_1 \Sigma_s^{-1} I\!\!E \left(\Sigma_s^{-1} \boldsymbol{x}_s \boldsymbol{x}_s^T - I \right) f_\alpha \left(\|\boldsymbol{x}_s\|^2 - t \right).$$

Define $\log' \Sigma_s \stackrel{\text{def}}{=} (\Sigma_1 - \Sigma_0) \Sigma_s^{-1}$ and note that the distribution of a random vector $\boldsymbol{y} \stackrel{\text{def}}{=} \Sigma_s^{-1/2} \boldsymbol{x}_s$ is standard normal then the change of variables concludes the alternative representation for $g_\alpha(t)$ (3.1). □

Moreover, expanding on the lemma it is possible to conclude the characterizing for the difference of multivariate Gaussian measures corollary.

Corollary 3.4. *Assume independent and centered Gaussian vectors $\boldsymbol{x}_0, \boldsymbol{x}_1 \in I\!\!R^p$. Denote $\boldsymbol{y} \sim \mathcal{N}(0, I)$ and uniform $s \sim \mathcal{U}(0,1)$ and define a linear operator*

$$\mathbb{L}_{\boldsymbol{y},s} \stackrel{\text{def}}{=} \frac{1}{2} I\!\!E \left(\boldsymbol{y}^T \log' \Sigma_s \boldsymbol{y} - \log' \det \Sigma_s \right) (\cdot),$$

then

$$I\!\!P \left(\boldsymbol{x}_1 \in \mathcal{B}_t \right) - I\!\!P \left(\boldsymbol{x}_0 \in \mathcal{B}_t \right) = \mathbb{L}_{\boldsymbol{y},s} \left(\mathbb{I} \left(\boldsymbol{y}^T \Sigma_s \boldsymbol{y} < t \right) \right).$$

Proof. Notice that

$$\mathbb{L}_{\boldsymbol{y},s}\left(\mathbb{1}\left(\boldsymbol{y}^T\Sigma_s\boldsymbol{y} > t\right)\right) = \mathbb{L}_{\boldsymbol{y},s}\left(\mathbb{1}\left(\boldsymbol{y}^T\Sigma_s\boldsymbol{y} < t\right)\right),$$

then Lebesgue dominated convergence theorem and the lemma [3.3] enable the statement in the limit $\alpha \to \infty$. □

Informally, one can think of differentiation and integration by parts (Stein's identity) applied to non-smooth functions - indicator and Dirac delta function - which gives a hope to built tight or even optimal bound for Gaussian comparison (see theorem [3.5]).

The corollary helps to shift analysis from hardly accessible difference to a more constructive object.

Theorem 3.5 (Gaussian comparison)**.** *Assume independent and centered Gaussian vectors* $\boldsymbol{x}_0, \boldsymbol{x}_1 \in \mathbb{R}^p$ *with covariance operators* $\Sigma_0, \Sigma_1 \succ 0$ *respectively. Define uniformly upper bounded constant*

$$C_p = \sqrt{\mathbb{P}\left(\boldsymbol{y}^T A_s \boldsymbol{y} < p\right)} < 1$$

for $\boldsymbol{y} \sim \mathcal{N}(0, I)$*, uniformly distributed* $s \sim \mathcal{U}(0,1)$ *and a matrix* $A_s = \frac{p \log' \Sigma_s}{\log' \det \Sigma_s}$*. Then it holds*

$$\sup_t \left|\mathbb{P}\left(\boldsymbol{x}_1 \in \mathcal{B}_t\right) - \mathbb{P}\left(\boldsymbol{x}_0 \in \mathcal{B}_t\right)\right| \le C_p \sqrt{\left|Tr\left(I - \left(\Sigma_0\Sigma_1^{-1} + \Sigma_0^{-1}\Sigma_1\right)/2\right)\right|}.$$

Proof. The corollary [3.4] implies

$$\sup_t \left|\mathbb{P}\left(\boldsymbol{x}_1 \in \mathcal{B}_t\right) - \mathbb{P}\left(\boldsymbol{x}_0 \in \mathcal{B}_t\right)\right| \le \left|\frac{1}{2}\int_0^1 \mathbb{E}\left(\boldsymbol{y}^T \log' \Sigma_s \boldsymbol{y} - \log' \det \Sigma_s\right) \mathbb{1}\left(\boldsymbol{y}^T \log' \Sigma_s \boldsymbol{y} < \log' \det \Sigma_s\right) ds\right|$$

then Cauchy-Schwartz inequality gives

$$\sup_t \left|\mathbb{P}\left(\boldsymbol{x}_1 \in \mathcal{B}_t\right) - \mathbb{P}\left(\boldsymbol{x}_0 \in \mathcal{B}_t\right)\right|$$

$$\le \frac{1}{2}\sqrt{\mathbb{P}\left(\boldsymbol{y}^T \log' \Sigma_s \boldsymbol{y} / \log' \det \Sigma_s < 1\right)} \sqrt{\int_0^1 \mathbb{E}\left(\boldsymbol{y}^T \log' \Sigma_s \boldsymbol{y} - \log' \det \Sigma_s\right)^2 ds}.$$

Observe additionally that $\forall s \in (0,1)$ - $Tr\left(\log' \Sigma_s\right)^2 = -Tr\left(\log'' \Sigma_s\right)$, then cross terms cancel out and one attains

$$\sqrt{2\int_0^1 \mathbb{E}\left(\boldsymbol{y}^T \log' \Sigma_s \boldsymbol{y} - \log' \det \Sigma_s\right)^2 ds} = \sqrt{2\int_0^1 Tr\left(\log' \Sigma_s\right)^2 ds}$$

$$= \sqrt{2\left|Tr \int_0^1 \left(\log \Sigma_s\right)'' ds\right|} = \sqrt{2\left|Tr\left(\log' \Sigma_s\Big|_{s=1} - \log' \Sigma_s\Big|_{s=0}\right)\right|}$$

$$= 2\sqrt{\left|Tr\left(I - \left(\Sigma_0\Sigma_1^{-1} + \Sigma_0^{-1}\Sigma_1\right)/2\right)\right|}.$$

The statement readily follows uniformly in dimension. □

Remark 3.1. *Notice that the theorem has no condition $p > 2$ typical for a density based approximation of Kolmogorov distance.*

Remark 3.2. *Also observe that $Tr A_s = p$ and for a fixed s if there exists $\delta \neq 0$ such that $|A_s - I| \succ \delta I$ then the limit $\lim_{p\to\infty} p^k I\!P\left(\boldsymbol{y}^T A_s \boldsymbol{y} < p\right) = 0$ tends to zero as there is always a gap $\delta_s > 0$ s.t. $I\!P\left(\boldsymbol{y}^T A_s \boldsymbol{y} < p\right) = I\!P\left(\|\boldsymbol{y}\|^2 < p - \delta_s\right)$.*

A useful implication of the result [3.4] is the anti-concentration bound. With an a-priori chosen shift $\Delta > 0$ it is equivalent to bounding Kolmogorov distance with scaled covariances

$$\begin{aligned} & I\!P\left(\boldsymbol{x} \in \mathcal{B}_{t+\Delta}\right) - I\!P\left(\boldsymbol{x} \in \mathcal{B}_{t}\right) \qquad (3.4) \\ \leq \sup_{t'} & \left| I\!P\left(\sqrt{(1+\Delta/t)}\boldsymbol{x} \in \mathcal{B}_{t'}\right) - I\!P\left(\boldsymbol{x} \in \mathcal{B}_{t'}\right)\right| = \sup_{t} \left| I\!P\left(\boldsymbol{x}_0 \in \mathcal{B}_{t}\right) - I\!P\left(\boldsymbol{x}_1 \in \mathcal{B}_{t}\right)\right| \end{aligned}$$

with $\Sigma_0 = \Sigma$ and$\Sigma_1 = (1+\Delta/t)\,\Sigma$. Even though the theorem [3.5] straightforwardly applies let us justify the anti-concentration independently to cross-validate the comparison result.

Theorem 3.6 (Anti-concentration)**.** *For a centered Gaussian vector $\boldsymbol{x} \sim \mathcal{N}(0,\Sigma)$ in $I\!R^p$*

$$I\!P\left(\boldsymbol{x} \in \mathcal{B}_{t+\Delta}\right) - I\!P\left(\boldsymbol{x} \in \mathcal{B}_{t}\right) \leq C p^{1/2}\left(\log\left(1+\Delta/t\right) \wedge \sqrt{\log\left(1+\Delta/t\right)}\right)$$

with a universal constant $C < \infty$.

Proof. Choose $\beta > 1$, a covariance $\Sigma \succ 0$ and define

$$\boldsymbol{x}_0 = \boldsymbol{x},\;\; \boldsymbol{x}_1 = \beta^{1/2}\boldsymbol{x},\;\; \beta_s = (1-s) + s\beta \;\text{ and }\; \Sigma_s = (1-s)\Sigma + s\beta\Sigma.$$

The corollary [3.4] claims that

$$\sup_t \left| I\!P\left(\boldsymbol{x}_1 \in \mathcal{B}_t\right) - I\!P\left(\boldsymbol{x}_0 \in \mathcal{B}_t\right)\right| = \sup_t \left| \frac{1}{2} Tr \int_0^1 I\!E\left(\boldsymbol{y}\boldsymbol{y}^T - I\right)\left(\log \Sigma_s\right)' 1\!\!1\left(\boldsymbol{y}^T\Sigma_s\boldsymbol{y} < t\right) ds \right|,$$

which in the particular case is written as

$$\sup_t \left| I\!P\left(\beta^{1/2}\boldsymbol{x} \in \mathcal{B}_t\right) - I\!P\left(\boldsymbol{x} \in \mathcal{B}_t\right)\right| = \sup_t \left| \frac{1}{2}\int_0^1 \left(\log\beta_s\right)' I\!E\left(\|\boldsymbol{y}\|^2 - p\right) 1\!\!1\left(\boldsymbol{y}^T\Sigma\boldsymbol{y} < t/\beta_s\right) ds \right|.$$

One had to exploit the logarithmic structure $(\log \Sigma_s)' = (\log \beta_s)'$ to attain the expression. Therefore, it holds

$$\begin{aligned} \sup_t \left| I\!P\left(\beta^{1/2}\boldsymbol{x} \in \mathcal{B}_t\right) - I\!P\left(\boldsymbol{x} \in \mathcal{B}_t\right)\right| &\leq \frac{1}{2} I\!E\left|\|\boldsymbol{y}\|^2 - p\right| \left|\int_0^1 (\log\beta_s)'\,ds\right| \\ &\leq \frac{1}{2} I\!E\left|\|\boldsymbol{y}\|^2 - p\right| \log\beta \leq C p^{1/2}\left(\log\beta \wedge \sqrt{\log\beta}\right) \end{aligned}$$

with a universal constant

$$\forall p \;:\; C > I\!E\left|\frac{\|\boldsymbol{y}\|^2 - p}{2\sqrt{p}}\right| \vee I\!E\left|\frac{\|\boldsymbol{y}\|^2\log\beta - p\log\beta}{2\sqrt{p\log\beta}}\right|.$$

To complete the derivation it suffice to insert $\beta \stackrel{\text{def}}{=} 1 + \Delta/t$ and use the observation (3.4) in front of the corollary. $\square$

There exists an alternative technique reproducing derived anti-concentration. It provides a convenient instrument to validate the characterization [3.4].

Theorem 3.7. *For a Gaussian vector $\boldsymbol{x} \in I\!R^p$ with a covariance $\Sigma \succ 0$ it holds*

$$I\!P\left(\boldsymbol{x} \in \mathcal{B}_{t+\Delta}\right) - I\!P\left(\boldsymbol{x} \in \mathcal{B}_t\right) \leq \sqrt{p}\Delta/t.$$

Proof. Pinsker's inequality states

$$I\!P\left(\boldsymbol{x} \in \mathcal{B}_{t+\Delta}\right) - I\!P\left(\boldsymbol{x} \in \mathcal{B}_t\right) \leq \sqrt{\mathcal{K}(I\!P_1, I\!P_2)/2}.$$

Explicitly, the Kullback-Leibler divergence between $I\!P_1 = \mathcal{N}\left(0, \Sigma\right)$ and $I\!P_2 = \mathcal{N}\left(0, \frac{t^2}{(t+\Delta)^2}\Sigma\right)$ is written

$$\mathcal{K}(I\!P_1, I\!P_2) = p/2((\Delta/t)^2 + 2(\Delta/t) - 2\log(1 - \Delta/t)) \leq p(\Delta/t)^2.$$

The combination of the two concludes $I\!P\left(\boldsymbol{x} \in \mathcal{B}_{t+\Delta}\right) - I\!P\left(\boldsymbol{x} \in \mathcal{B}_t\right) \leq \sqrt{p}\Delta/t$. □

The result approximately matches the corollary [3.6] and an exact relation is explained via an inequality

$$I\!P\left(\boldsymbol{x} \in \mathcal{B}_{t+\Delta}\right) - I\!P\left(\boldsymbol{x} \in \mathcal{B}_t\right) \overset{corollary 3.6}{<} C\sqrt{p}\log\left(1 + \Delta/t\right) < C\sqrt{p}\Delta/t.$$

Up to a constant factor the corollary [3.6] is sharper than the theorem [3.7]. However, they match in order and no significant improvement is gained compared to Pinsker's inequality.

Interestingly the theorem [3.6] can be used to bound a pdf $\rho_{\|\boldsymbol{x}\|^2}(t) \overset{\text{def}}{=} I\!E\delta\left(\|\boldsymbol{x}\|^2 - t\right)$ of squared Eucledian norm of a Gaussian vector

$$\rho_{\|\boldsymbol{x}\|^2}(t) = \lim_{\Delta \to 0^+} \frac{I\!P\left(\boldsymbol{x} \in \mathcal{B}_{t+\Delta}\right) - I\!P\left(\boldsymbol{x} \in \mathcal{B}_t\right)}{\Delta} \leq \lim_{\Delta \to 0^+} \frac{Cp^{1/2}\log\left(1 + \Delta/t\right)}{\Delta} = C\frac{\sqrt{p}}{t}.$$

Corollary 3.8. *A Gaussian vector $\boldsymbol{x} \sim \mathcal{N}\left(0, \Sigma\right)$ follows*

$$\rho_{\|\boldsymbol{x}\|^2}(t) \leq C\sqrt{p}/t$$

with a universal constant $C < \infty$.

Moreover, the corollary [3.4] additionally enables a structural characterization of the density.

Theorem 3.9. *A Gaussian vector $\boldsymbol{x} \sim \mathcal{N}\left(0, \Sigma\right)$ follows*

$$\rho_{\|\boldsymbol{x}\|^2}(t) = \frac{1}{2t} I\!E_{\boldsymbol{y}^T\Sigma\boldsymbol{y} < t}\left(p - \|\boldsymbol{y}\|^2\right)$$

with the standard normal $\boldsymbol{y} \sim \mathcal{N}\left(0, I\right)$.

Proof. Once again fix $\beta > 1$, a covariance $\Sigma \succ 0$ and define

$$\boldsymbol{x}_0 = \boldsymbol{x}, \;\; \boldsymbol{x}_1 = \beta^{1/2}\boldsymbol{x}, \;\; \beta_s = (1-s) + s\beta \;\text{ and }\; \Sigma_s = (1-s)\Sigma + s\beta\Sigma.$$

The corollary [3.4] claims that

$$I\!\!P\left(\boldsymbol{x}_1 \in \mathcal{B}_t\right) - I\!\!P\left(\boldsymbol{x}_0 \in \mathcal{B}_t\right) = \frac{1}{2} Tr \int_0^1 I\!\!E\left(\boldsymbol{y}\boldsymbol{y}^T - I\right)\left(\log \Sigma_s\right)' \, 1\!\!I\left(\boldsymbol{y}^T \Sigma_s \boldsymbol{y} < t\right) ds,$$

which is written here as

$$I\!\!P\left(\boldsymbol{x} \in \mathcal{B}_{t+\Delta}\right) - I\!\!P\left(\boldsymbol{x} \in \mathcal{B}_t\right) = I\!\!P\left(\left(1 + \Delta/t\right)^{-1/2} \boldsymbol{x} \in \mathcal{B}_t\right) - I\!\!P\left(\boldsymbol{x} \in \mathcal{B}_t\right)$$

$$= \frac{1}{2}\int_0^1 \left(\log \beta_s\right)' I\!\!E\left(\|\boldsymbol{y}\|^2 - p\right) 1\!\!I\left(\beta_s \boldsymbol{y}^T \Sigma \boldsymbol{y} < t\right) ds.$$

choosing $\beta \overset{\text{def}}{=} (1 + \Delta/t)^{-1}$ and using the logarithmic structure $(\log \Sigma_s)' = (\log \beta_s)'$.

Therefore, taking the limit $\Delta \to 0$ one finds a description of the probability density function

$$\rho_{\|\boldsymbol{x}\|^2}(t) = \lim_{\Delta \to 0^+} \frac{I\!\!P\left(\boldsymbol{x} \in \mathcal{B}_{t+\Delta}\right) - I\!\!P\left(\boldsymbol{x} \in \mathcal{B}_t\right)}{\Delta}$$

$$= \lim_{\beta \to 1^-} \frac{\beta - 1}{2t\left(1/\beta - 1\right)} \int_0^1 \frac{1}{\beta_s} I\!\!E\left(\|\boldsymbol{y}\|^2 - p\right) 1\!\!I\left(\beta_s \boldsymbol{y}^T \Sigma \boldsymbol{y} < t\right) ds = \frac{1}{2t} I\!\!E\left(p - \|\boldsymbol{y}\|^2\right) 1\!\!I\left(\boldsymbol{y}^T \Sigma \boldsymbol{y} < t\right).$$

□

However, the state of the art bound on the density can be found in Götze, F. and Naumov, A. and Spokoiny, V. and Ulyanov, V. [1].

4 Conclusion

The work allows to hope that an advancement is also possible for the density estimation. One can derive a refined version of the corollary [3.8].

Another clear application objective is a bootstrap. It corresponds to a covariance operator empirically estimating the other one. The procedure is generally helpful for a practitioner constructing confidence sets.

Hierarchically even deeper challenge to answer is whether there exist and what are the other structurally stable probabilistic phenomena in multivariate or Hilbert spaces. Particularly, aside from Stein's lemma substituted on Stein-Chen's lemma the core of the analysis remains indicating a possible gain in the direction.

References

[1] F. Götze, A. Naumov, V. Spokoiny, and V. Ulyanov. Large ball probability, Gaussian comparison and anti-concentration. *ArXiv e-prints*, August 2017.

[2] Charles M Stein. Estimation of the mean of a multivariate normal distribution. *The annals of Statistics*, pages 1135–1151, 1981.

IRTG 1792 Discussion Paper Series 2018

For a complete list of Discussion Papers published, please visit irtg1792.hu-berlin.de.

001 "Data Driven Value-at-Risk Forecasting using a SVR-GARCH-KDE Hybrid" by Marius Lux, Wolfgang Karl Härdle and Stefan Lessmann, January 2018.
002 "Nonparametric Variable Selection and Its Application to Additive Models" by Zheng-Hui Feng, Lu Lin, Ruo-Qing Zhu asnd Li-Xing Zhu, January 2018.
003 "Systemic Risk in Global Volatility Spillover Networks: Evidence from Option-implied Volatility Indices " by Zihui Yang and Yinggang Zhou, January 2018.
004 "Pricing Cryptocurrency options: the case of CRIX and Bitcoin" by Cathy YH Chen, Wolfgang Karl Härdle, Ai Jun Hou and Weining Wang, January 2018.
005 "Testing for bubbles in cryptocurrencies with time-varying volatility" by Christian M. Hafner, January 2018.
006 "A Note on Cryptocurrencies and Currency Competition" by Anna Almosova, January 2018.
007 "Knowing me, knowing you: inventor mobility and the formation of technology-oriented alliances" by Stefan Wagner and Martin C. Goossen, February 2018.
008 "A Monetary Model of Blockchain" by Anna Almosova, February 2018.
009 "Deregulated day-ahead electricity markets in Southeast Europe: Price forecasting and comparative structural analysis" by Antanina Hryshchuk, Stefan Lessmann, February 2018.
010 "How Sensitive are Tail-related Risk Measures in a Contamination Neighbourhood?" by Wolfgang Karl Härdle, Chengxiu Ling, February 2018.
011 "How to Measure a Performance of a Collaborative Research Centre" by Alona Zharova, Janine Tellinger-Rice, Wolfgang Karl Härdle, February 2018.
012 "Targeting customers for profit: An ensemble learning framework to support marketing decision making" by Stefan Lessmann, Kristof Coussement, Koen W. De Bock, Johannes Haupt, February 2018.
013 "Improving Crime Count Forecasts Using Twitter and Taxi Data" by Lara Vomfell, Wolfgang Karl Härdle, Stefan Lessmann, February 2018.
014 "Price Discovery on Bitcoin Markets" by Paolo Pagnottoni, Dirk G. Baur, Thomas Dimpfl, March 2018.
015 "Bitcoin is not the New Gold - A Comparison of Volatility, Correlation, and Portfolio Performance" by Tony Klein, Hien Pham Thu, Thomas Walther, March 2018.
016 "Time-varying Limit Order Book Networks" by Wolfgang Karl Härdle, Shi Chen, Chong Liang, Melanie Schienle, April 2018.
017 "Regularization Approach for NetworkModeling of German EnergyMarket" by Shi Chen, Wolfgang Karl Härdle, Brenda López Cabrera, May 2018.
018 "Adaptive Nonparametric Clustering" by Kirill Efimov, Larisa Adamyan, Vladimir Spokoiny, May 2018.
019 "Lasso, knockoff and Gaussian covariates: a comparison" by Laurie Davies, May 2018.

IRTG 1792, Spandauer Straße 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.

IRTG 1792 Discussion Paper Series 2018

For a complete list of Discussion Papers published, please visit irtg1792.hu-berlin.de.

020 "A Regime Shift Model with Nonparametric Switching Mechanism" by Haiqiang Chen, Yingxing Li, Ming Lin and Yanli Zhu, May 2018.

021 "LASSO-Driven Inference in Time and Space" by Victor Chernozhukov, Wolfgang K. Härdle, Chen Huang, Weining Wang, June 2018.

022 " Learning from Errors: The case of monetary and fiscal policy regimes" by Andreas Tryphonides, June 2018.

023 "Textual Sentiment, Option Characteristics, and Stock Return Predictability" by Cathy Yi-Hsuan Chen, Matthias R. Fengler, Wolfgang Karl Härdle, Yanchu Liu, June 2018.

024 "Bootstrap Confidence Sets For Spectral Projectors Of Sample Covariance" by A. Naumov, V. Spokoiny, V. Ulyanov, June 2018.

025 "Construction of Non-asymptotic Confidence Sets in 2 -Wasserstein Space" by Johannes Ebert, Vladimir Spokoiny, Alexandra Suvorikova, June 2018.

026 "Large ball probabilities, Gaussian comparison and anti-concentration" by Friedrich Götze, Alexey Naumov, Vladimir Spokoiny, Vladimir Ulyanov, June 2018.

027 "Bayesian inference for spectral projectors of covariance matrix" by Igor Silin, Vladimir Spokoiny, June 2018.

028 "Toolbox: Gaussian comparison on Eucledian balls" by Andzhey Koziuk, Vladimir Spokoiny, June 2018.

IRTG 1792, Spandauer Straße 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.