ISSN 1064-5624, Doklady Mathematics, 2018, Vol. 98, No. 2, pp. 511–514. © Pleiades Publishing, Ltd., 2018.
Original Russian Text © A.A. Naumov, V.G. Spokoiny, V.V. Ulyanov, 2018, published in Doklady Akademii Nauk, 2018, Vol. 482, No. 6.

MATHEMATICS

Confidence Sets for Spectral Projectors of Covariance Matrices

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

Presented by Academician of the RAS A.N. Shiryaev April 18, 2018

Received May 7, 2018

Abstract—A sample $X_1, \dots, X_n$ consisting of independent identically distributed vectors in $\mathbb{R}^p$ with zero mean and a covariance matrix $\mathbf{\Sigma}$ is considered. The recovery of spectral projectors of high-dimensional covariance matrices from a sample of observations is a key problem in statistics arising in numerous applications. In their 2015 work, V. Koltchinskii and K. Lounici obtained nonasymptotic bounds for the Frobenius norm $\|\mathbf{P}_r - \hat{\mathbf{P}}_r\|_2$ of the distance between sample and true projectors and studied asymptotic behavior for large samples. More specifically, asymptotic confidence sets for the true projector $\mathbf{P}_r$ were constructed assuming that the moment characteristics of the observations are known. This paper describes a bootstrap procedure for constructing confidence sets for the spectral projector $\mathbf{P}_r$ of the covariance matrix $\mathbf{\Sigma}$ from given data. This approach does not use the asymptotical distribution of $\|\mathbf{P}_r - \hat{\mathbf{P}}_r\|_2$ and does not require the computation of its moment characteristics. The performance of the bootstrap approximation procedure is analyzed.

DOI: 10.1134/S1064562418060285

Consider a sequence of independent identically distributed (i.i.d.) random vectors $X, X_1, \dots, X_n$, taking values in $\mathbb{R}^p$. Assume that $\mathbb{E}X = 0$ and $\mathbb{E}\|X\|^2 < \infty$. Let $\Sigma \overset{\text{def}}{=} \mathbb{E}(XX^\top)$ be the covariance matrix of the vector X. Along with the true covariance matrix $\mathbf{\Sigma}$, we consider the sample covariance matrix $\hat{\mathbf{\Sigma}}$ constructed using the observations $X_1, \dots, X_n$:

$$\hat{\mathbf{\Sigma}} \overset{\text{def}}{=} \frac{1}{n}\sum_{j=1}^{n} X_j X_j^\top = \frac{1}{n}\mathbf{X}\mathbf{X}^\top.$$

Here and below, $\mathbf{X} \overset{\text{def}}{=} [X_1, \dots, X_n] \in \mathbb{R}^{p\times n}$.

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

In most statistical applications, the true covariance matrix $\mathbf{\Sigma}$ is typically unknown and is replaced by its sample counterpart $\hat{\mathbf{\Sigma}}$. The accuracy of estimating $\mathbf{\Sigma}$ by $\hat{\mathbf{\Sigma}}$, in particular, when p is much greater than n, has been extensively studied (see [1–3]). A bound in terms of the effective rank $\mathrm{r}(\mathbf{\Sigma}) \overset{\text{def}}{=} \frac{\mathrm{Tr}(\mathbf{\Sigma})}{\|\mathbf{\Sigma}\|}$ was obtained recently in [4]. Such bounds can be used to recover individual eigenvalues of $\mathbf{\Sigma}$ in the case when there are spectral gaps between these eigenvalues.

In this paper, we recover spectral projectors onto the subspace spanned by the eigenvectors corresponding to certain eigenvalues of $\mathbf{\Sigma}$. The recovery of spectral projectors, as well as eigenvectors and eigenspaces of high-dimensional covariance matrices from a sample of observations is a key problem in statistics that is directly related to dimensionality reduction. For example, in principal component analysis, high-dimensional data are projected onto the subspaces spanned by the eigenvectors corresponding to the largest eigenvalues. However, the problem of recovering spectral projectors of high-dimensional covariance matrices has been poorly investigated. In [5] nonasymptotic bounds for the Frobenius norm $\|\mathbf{P}_r - \hat{\mathbf{P}}_r\|_2$ of the distance between sample and true projectors were obtained and asymptotic behavior for large-size samples was studied. According to [5], given the moments of observed random variables, one can construct asymptotic confidence sets for the true projector $\mathbf{P}_r$. However, these moment characteristics are typically unknown. On the other hand, it is well known that such asymptotic results could be applied only to samples of very large sizes. In particular, this is due to the fact that the normalized U-statistics arising in the given problem converge extremely slowly to a limiting law.

Before stating the main result, we introduce the concept of a spectral gap and define the sample counterpart of the spectral projector $\mathbf{P}_r$. Let σ_j, $j = 1, 2, \ldots, p$, denote the eigenvalues of $\mathbf{\Sigma}$ arranged in nonincreasing order and $\mathbf{u}_j$, $j = 1, 2, \ldots, p$, be the corresponding eigenvectors. The matrix $\mathbf{\Sigma}$ can be written in terms of its spectral decomposition, namely, $\mathbf{\Sigma} = \sum_{j=1}^{p} \sigma_j \mathbf{u}_j \mathbf{u}_j^\top$. Now, let μ_j, $j = 1, 2, \ldots, q$, $q \le p$, be different eigenvalues of $\mathbf{\Sigma}$ and $\mathbf{P}_r, r = 1, 2, \ldots, q$, be the corresponding spectral projectors, i.e., $\mathbf{P}_r = \sum_{j:\sigma_j = \mu_r} \mathbf{u}_j \mathbf{u}_j^\top$. Then

$$\mathbf{\Sigma} = \sum_{r=1}^{q} \mu_r \mathbf{P}_r.$$

Let $\Delta_r \overset{\text{def}}{=} \{j: \sigma_j = \mu_r\}$. It is easy to see that $|\Delta_r| = m_r$, where $m_r \overset{\text{def}}{=} \mathrm{Rank}(\mathbf{P}_r)$. For all $r \ge 1$, we define $g_r \overset{\text{def}}{=} \mu_r - \mu_{r+1} > 0$. Let $\overline{g}_r \overset{\text{def}}{=} \min(g_{r-1}, g_r)$ for $r \ge 2$ and $\overline{g}_1 \overset{\text{def}}{=} g_1$. Here, $\overline{g}_r$ is called the rth spectral gap corresponding to the eigenvalue μ_r.

Now we write $\hat{\mathbf{\Sigma}}$ in terms of its spectral decomposition: $\hat{\mathbf{\Sigma}} = \sum_{j=1}^{p} \hat{\sigma}_j \hat{\mathbf{u}}_j \hat{\mathbf{u}}_j^\top$, where $\hat{\sigma}_1 \ge \hat{\sigma}_2 \ge \ldots \ge \hat{\sigma}_p$ and $\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_p$ are the eigenvalues and corresponding eigenvectors of the matrix $\hat{\mathbf{\Sigma}}$. Following [5], we define clusters of eigenvalues $\hat{\sigma}_j$, $j \in \Delta_r$. Let $\widehat{\mathbf{E}} \overset{\text{def}}{=} \hat{\mathbf{\Sigma}} - \mathbf{\Sigma}$. It is easy to show that

$$\inf_{j \notin \Delta_r} |\hat{\sigma}_j - \mu_r| \ge \overline{g}_r - \|\widehat{\mathbf{E}}\|, \quad \sup_{j \in \Delta_r} |\hat{\sigma}_j - \mu_r| \le \|\widehat{\mathbf{E}}\|.$$

Assume that $\|\widehat{\mathbf{E}}\| \le \frac{\overline{g}_r}{2}$. Then all $\hat{\sigma}_j$, $j \in \Delta_r$, belong to the interval $(\mu_r - \|\widehat{\mathbf{E}}\|, \mu_r + \|\widehat{\mathbf{E}}\|) \subset \left(\mu_r - \frac{\overline{g}_r}{2}, \mu_r + \frac{\overline{g}_r}{2}\right)$, while the other eigenvalues of $\hat{\mathbf{\Sigma}}$ lie outside the interval $(\mu_r - (\overline{g}_r - \|\widehat{\mathbf{E}}\|), \mu_r + (\overline{g}_r - \|\widehat{\mathbf{E}}\|)) \supset \left[\mu_r - \frac{\overline{g}_r}{2}, \mu_r + \frac{\overline{g}_r}{2}\right]$. If we additionally require that $\|\widehat{\mathbf{E}}\| < \frac{1}{4} \min_{1 \le s \le r} \overline{g}_s =: \overline{\delta}_r$, then the set $\left\{\hat{\sigma}_j, j \in \bigcup_{s=1}^{r} \Delta_s\right\}$ consists of r clusters. The diameter of each cluster is strictly less than $2\overline{\delta}_r$, and the distance between any two clusters is greater than $2\overline{\delta}_r$. Let $\hat{\mathbf{P}}_r$ denote the projector onto the subspace spanned by $\hat{\mathbf{u}}_j$, $j \in \Delta_r$.

As was noted above, the asymptotic normality of $\|\hat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2$ was proved in [5]. Relying on this result, one can construct asymptotic confidence sets for an unknown projector $\mathbf{P}_r$ of the form

$$\left\{\mathbf{P}_r : \frac{\|\hat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2 - \mathbb{E}\|\hat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2}{\mathrm{Var}^{1/2}(\|\hat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2)} \le z_\alpha\right\},$$

where z_α is the α-quantile of the normal distribution. The basic drawbacks of this approach are the slow convergence rate to the normal law and the fact that a large sample size is required to achieve a reasonable quality of the approximation. Additionally, we need to estimate $\mathbb{E}\|\hat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2$ and $\mathrm{Var}(\|\hat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2)$, which depend on the unknown matrix $\mathbf{\Sigma}$. In [5] a statistical procedure is proposed, according to which the sample is divided into three equal subsamples. The first and second subsamples are used to estimate the expectation and variance of $\|\hat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2$, while the third one is used to construct a confidence set.

In this paper, the quantile

$$\gamma_\alpha \overset{\text{def}}{=} \inf\{\gamma > 0 : \mathbb{P}(n\|\hat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2 > \gamma) \le \alpha\} \qquad (1)$$

is estimated using a bootstrap procedure. This approach

(i) does not rely on the asymptotical distribution of $\|\hat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2$;

(ii) does not require computing the moments of $\|\hat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2$;

(iii) does not require splitting the sample into subsamples; and

(iv) provides an explicit error bound for the bootstrap approximation.

Note that the bootstrap method is one of the most widespread statistical techniques for constructing confidence sets. However, the existing theory proves the possibility of applying this method basically to parametric models. The generalization to the case when the space dimension is much higher than the sample size encounters various difficulties. In this context, we note [6–8]. In this paper, the bootstrap method is extended to the construction of confidence sets for spectral projectors. Additionally, it should be noted that spectral projectors depend nonlinearly on the covariance matrix, which in turn is a quadratic function of the original distribution; this causes additional difficulties.

The weighted (or bootstrap) version of the matrix $\hat{\mathbf{\Sigma}}$ is defined as

$$\mathbf{\Sigma}^{\circ} \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} w_i X_i X_i^{\top},$$

where $w_1, \ldots, w_n$ are i.i.d. eigenvectors independent of $\mathbf{X} = (X_1, \ldots, X_n)$ such that $\mathbb{E}w_1 = 1$ and $\operatorname{Var} w_1 = 1$. As an example, we can consider i.i.d. Gaussian weights $w_i \sim \mathcal{N}(1,1)$. We introduce the conditional probability $\mathbb{P}^{\circ}(\cdot) \stackrel{\text{def}}{=} \mathbb{P}(\cdot|\mathbf{X})$ and denote the corresponding expectation by $\mathbb{E}^{\circ}$. It is easy to see that, if the sample is fixed and the only random variables are the weights w_i, then the expectation $\mathbf{\Sigma}^{\circ}$ is the known matrix $\hat{\mathbf{\Sigma}}$, i.e., $\mathbb{E}^{\circ}\mathbf{\Sigma}^{\circ} = \hat{\mathbf{\Sigma}}$. This situation is opposite to the case $\mathbb{E}\hat{\mathbf{\Sigma}} = \mathbf{\Sigma}$, where the expectation is unknown.

Writing $\mathbf{\Sigma}^{\circ} = \sum_{j=1}^{p} \sigma_j^{\circ}\mathbf{u}_j^{\circ}\mathbf{u}_j^{\circ\top}$, we define the spectral projectors $\mathbf{P}_r^{\circ}$ as the orthogonal projectors onto the subspace spanned by $\mathbf{u}_j^{\circ}, j \in \Delta_r$. For a given α, the quantile γ_{α}° is defined as

$$\gamma_{\alpha}^{\circ} \stackrel{\text{def}}{=} \min\{\gamma > 0\colon \mathbb{P}^{\circ}(n\|\mathbf{P}_r^{\circ} - \hat{\mathbf{P}}_r\|_2^2 > \gamma) \le \alpha\} \tag{2}$$

(cf. (1)). Note that γ_{α}° depends on the sample $\mathbf{X}$. The idea of the method is to use γ_{α}° for constructing the confidence set $\mathscr{E}(\alpha) \stackrel{\text{def}}{=} \{\mathbf{P}\colon n\|\mathbf{P} - \hat{\mathbf{P}}_r\|_2^2 \le \gamma_{\alpha}^{\circ}\}$. Thus, we need to show that

$$\mathbb{P}(\mathbf{P}_r \notin \mathscr{E}(\alpha)) = \mathbb{P}(n\|\mathbf{P}_r - \hat{\mathbf{P}}_r\|_2^2 > \gamma_{\alpha}^{\circ}) \approx \alpha.$$

Define the block matrix

$$\Gamma_r \stackrel{\text{def}}{=} \begin{pmatrix} \Gamma_{r1} & \mathbf{O} & \ldots & \mathbf{O} \\ \mathbf{O} & \Gamma_{r2} & \mathbf{O}\ldots & \mathbf{O} \\ \ldots & & & \\ \mathbf{O} & \ldots & \mathbf{O} & \Gamma_{rq} \end{pmatrix},$$

where $\Gamma_{rs}, s \neq r$, are diagonal $m_r m_s \times m_r m_s$ matrices with values $\frac{2\mu_r\mu_s}{(\mu_r - \mu_s)^2}$ on the main diagonal. Let $\lambda_1(\Gamma_r) \ge \lambda_2(\Gamma_r) \ge \ldots$ denote the eigenvalues of Γ_r. According to available bounds for the distance between $\mathbf{\Sigma}$ and $\hat{\mathbf{\Sigma}}$, the eigenvalues of $\mathbf{\Sigma}$ can be recovered with accuracy $O\left(\frac{1}{\sqrt{n}}\right)$. Thus, the part of the spectrum of $\mathbf{\Sigma}$ below a threshold of order $O\left(\frac{1}{\sqrt{n}}\right)$ cannot be estimated. The same is true of the matrix Γ_r. The value m is determined by

$$\lambda_{\mathrm{m}}(\Gamma_r) \ge \operatorname{Tr}\Gamma_r\left(\sqrt{\frac{\ln n}{n}} + \sqrt{\frac{\ln p}{n}}\right) \ge \lambda_{\mathrm{m}+1}(\Gamma_r). \tag{3}$$

Let Π_{m} denote the projector onto the subspace spanned by the eigenvectors of Γ_r corresponding to its largest m eigenvalues. Below is the main result of this paper.

Theorem 1. *Assume that observations $X, X_1, \ldots, X_n$ are i.i.d. Gaussian random vectors in $\mathbb{R}^p$ with zero mean and covariance matrix $\mathbb{E}XX^{\top} = \mathbf{\Sigma}$. For any $\alpha\colon 0 < \alpha < 1$, the corresponding quantile γ_{α}° is defined by* (2), *where the weights are additionally assumed to satisfy $w_i \sim \mathcal{N}(1,1)$ for all $i = 1, 2, \ldots, n$. Then*

$$|\alpha - \mathbb{P}(n\|\hat{\mathbf{P}}_r - \mathbf{P}_r\|_2^2 > \gamma_{\alpha}^{\circ})| \lesssim \diamond,$$

where

$$\diamond \stackrel{\text{def}}{=} \frac{\mathrm{m}\operatorname{Tr}\Gamma_r}{\sqrt{\lambda_1(\Gamma_r)\lambda_2(\Gamma_r)}}\left(\sqrt{\frac{\ln n}{n}} + \sqrt{\frac{\ln p}{n}}\right) + \frac{\operatorname{Tr}(\mathbf{I} - \Pi_{\mathrm{m}})\Gamma_r}{\sqrt{\lambda_1(\Gamma_r)\lambda_2(\Gamma_r)}} + \frac{m_r \operatorname{Tr}^3\mathbf{\Sigma}}{\bar{g}_r^3\sqrt{\lambda_1(\Gamma_r)\lambda_2(\Gamma_r)}}\left(\sqrt{\frac{\ln^3 n}{n}} + \sqrt{\frac{\ln^3 p}{n}}\right)$$

and m *is determined by* (3).

For the details of the proof of Theorem 1, see [9].

To conclude, we note that the closeness (on balls) of two centered Gaussian measures with different covariance operators has to be estimated in the course of the proof. A similar problem arises in other statistical problems. As a result, a more general problem was solved, namely, upper bounds were obtained for the closeness of two Gaussian measures with different means and covariance operators in the class of balls in a separable Hilbert space (see [10, 11]). The bounds are optimal with respect to the dependence on the spectra of the covariance operators of Gaussian measures. The inequalities cannot be improved in the general case.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Education and Science of the Russian Federation under the Program for Support of Leading Universities of the Russian Federation, project no. 5-100.

REFERENCES

1. J. Tropp, Found. Comput. Math. **12** (4), 389–434 (2012).
2. R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," in *Compressed Sensing: Theory and Applications* (Cambridge Univ. Press, Cambridge, 2012), pp. 210–268.
3. R. van Handel, Trans. Am. Math. Soc. **369** (11), 8161–8178 (2017).

4. V. Koltchinskii and K. Lounici, “Concentration inequalities and moment bounds for sample covariance operators” (2015). ArXiv:1405.2468.

5. V. Koltchinskii and K. Lounici, Ann. Stat. **45** (1), 121–157 (2017).

6. V. Spokoiny, and M. Zhilova, Ann. Stat. **43** (6), 2653–2675 (2015).

7. V. Chernozhukov, D. Chetverikov, and K. Kato, Ann. Stat. **41** (6), 2786–2819 (2013).

8. V. Chernozhukov, D. Chetverikov, and K. Kato, Ann. Probab. **45** (4), 2309–2352 (2017).

9. A. Naumov, V. Spokoiny, and V. Ulyanov, “Bootstrap confidence sets for spectral projectors of sample covariance” (2017). arXiv:1703.00871.

10. F. Götze, A. Naumov, V. Spokoiny, and V. Ulyanov, “Large ball probabilities, Gaussian comparison and anti-concentration,” Bernoulli **25** (2019). arXiv:1708.08663v2.

11. A. Naumov, V. Spokoiny, Yu.Tavyrikov and V. Ulyanov, “Nonasymptotic estimates for the closeness of Gaussian measures on balls”, Dokl. Math. **98** (2018).

Translated by I. Ruzanova