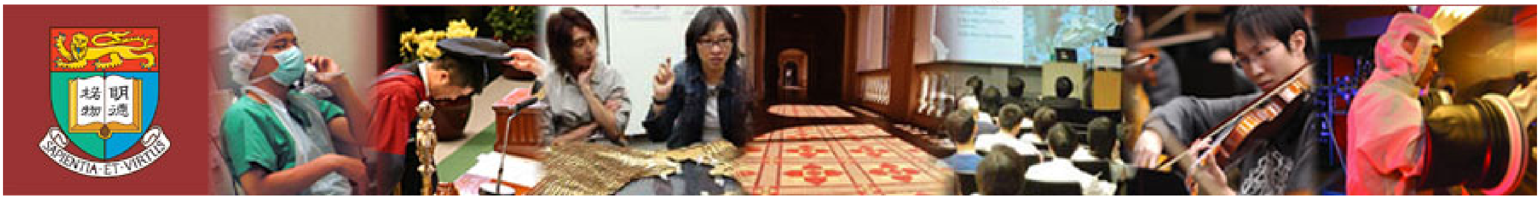

ON ESTIMATION OF THE NOISE VARIANCE IN A HIGH-DIMENSIONAL SIGNAL DETECTION MODEL

Jianfeng Yao and Damien Passemier

Department of Statistics and Actuarial Science - Department of Electronic and Computer Engineering
The University of Hong Kong -Hong Kong University of Science and Technology

ABSTRACT

When the number of receivers p is large compared to the sample size n, it has been widely observed that standard inference solutions are no longer efficient. In this paper, we address such high-dimensional issues related to the estimation of the noise variance. Several authors have reported that the classical maximum likelihood estimator of the noise variance tends to have a downward bias and this bias is increasingly important when p increases. Using recent results of random matrix theory, we are able to identify the bias. Moreover, a bias-corrected estimator is proposed using this knowledge. The asymptotic normality of the estimator in the high-dimensional context is established.

Index Terms— High-dimensional signal detection, noise variance estimator, random matrix theory.

1. INTRODUCTION

Consider the following *signal detection* model. Signals are received using p receivers in order to detect an unknown number of m *source signals*. As a first approximation, the recorded signals can be thought as linear combinations of the *source signals*. If we denote by $\mathbf{x}_t = (x_{t1}, \dots, x_{tp})'$ the p signals recorded at time t, and by $\mathbf{s}_t = (x_{t1}, \dots, x_{tm})'$ the source signals emitted at time t, we have

$$\mathbf{x}_t = \mathbf{A}\mathbf{s}_t + \varepsilon_t \ , \tag{1}$$

where $\mathbf{A}$ is a $p \times m$ *mixing matrix* representing the source-recording mechanism and ε_t a measurement error. It is reasonable to assume that (i) the noise and the source signal are independent; (ii) the noise is centered with a covariance matrix $\text{cov}(\boldsymbol{\varepsilon}_t) = \sigma^2 \mathbf{I}_p$. Then

$$\boldsymbol{\Sigma} = \text{cov}(\mathbf{x}_t) = \mathbf{A}\text{cov}(\mathbf{s}_t)\mathbf{A}' + \sigma^2 \mathbf{I}_p \ .$$

It is clear that the rank of $\mathbf{A}\text{cov}(\mathbf{s}_t)\mathbf{A}'$ does not exceed m and if we denote its eigenvalues by α_j with respective multiplicity numbers m_j ($\sum m_j = m$), then clearly

$$\text{spec}(\boldsymbol{\Sigma}) = (\underbrace{\alpha_1, \dots, \alpha_1}_{m_1}, \dots, \underbrace{\alpha_K, \dots, \alpha_K}_{m_K}, \underbrace{0, \dots, 0}_{p-m}) + \sigma^2(\underbrace{1, \dots, 1}_{p}). \tag{2}$$

Notice that the spectrum can be re-written as

$$\text{spec}(\boldsymbol{\Sigma}) = \sigma^2(\underbrace{\alpha'_1, \dots, \alpha'_1}_{m_1}, \dots, \underbrace{\alpha'_K, \dots, \alpha'_K}_{m_K}, \underbrace{1, \cdots, 1}_{p-m}). \tag{3}$$

This shows that the signal detection model coincides with a *spiked population model* introduced in [1].

Detecting the number of signal m is one of the most important inference problems in the model. This question has been addressed recently by several authors, see e.g. [2] and [3]. Another important issue is to find a good estimate of the variance σ^2.

Let $\bar{\mathbf{x}}$ be the sample mean and define the *sample covariance matrix*

$$\mathbf{S}_n = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'.$$

Let $\lambda_{n,1} \geq \lambda_{n,2} \geq \cdots \geq \lambda_{n,p}$ be the eigenvalues of $\mathbf{S}_n$. Then a commonly used estimator for σ^2 is

$$\widehat{\sigma}^2 = \frac{1}{p-m}\sum_{i=m+1}^{p}\lambda_{n,i}. \tag{4}$$

This is in fact the likelihood estimator of σ^2 if both the signal and the noise are Gaussian [4]. In a low-dimensional setting where p is fixed while $n \to \infty$, it is known that [5]

$$\sqrt{n}(\widehat{\sigma}^2 - \sigma^2) \xrightarrow{\mathcal{D}} N(0, s^2), \quad s^2 = \frac{2\sigma^4}{p-m}. \tag{5}$$

When p is large compared to the sample size n, the above asymptotic normality is no longer a good approximation. Indeed, it has been widely observed in the literature that $\widehat{\sigma}^2$ seriously underestimates the true noise variance σ^2 in such situ-

Funding from GRF Grant HKU 705413p is acknowledged.

978-1-4799-4975-5/14/$31.00 ©2014 IEEE

ation. For the correction of the bias, two approaches are proposed in [2] and [6]. Both estimators are assessed by Monte-Carlo experiments and their theoretic properties (bias, consistency or asymptotic normality) are unknown.

In this paper, we identify completely the aforementioned negative bias of the variance estimator $\hat{\sigma}^2$. Next using this identification, a bias-corrected estimator is found and its asymptotic normality proved under the high-dimensional scenario. Interestingly enough, this new asymptotic limit coincide with the classical low-dimensional limit (5) when the dimension to sample size ratio $c_n := p/n$ shrinks to zero. Therefore, the new asymptotic limit is a natural extension of the classical result to the high-dimensional context.

2. HIGH-DIMENSIONAL CENTRAL LIMIT THEOREM (CLT) FOR THE VARIANCE ESTIMATOR

As explained in Introduction, when the dimension p is large compared to the sample size n, the m.l.e. $\widehat{\sigma}^2$ in (4) has a negative bias. In this section, we identify this bias and establishes its asymptotic normality under the high-dimensional scheme.

Theorem 1. *Assume that (i) both the signal and the noise are Gaussian; (ii) $p \wedge n \to \infty$ and $c_n := p/n \to c > 0$. Then,*

$$\frac{(p-m)}{\sigma^2\sqrt{2c}}(\widehat{\sigma}^2 - \sigma^2) + b(\sigma^2) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1),$$

where $b(\sigma^2) = \sqrt{\frac{c}{2}}\left(m + \sigma^2 \sum_{i=1}^m \frac{1}{\alpha_i}\right)$.

Therefore for high-dimensional data, the m.l.e. $\widehat{\sigma}^2$ has an asymptotic bias $-b(\sigma^2)$ (after normalization). This bias is a complex function of the noise variance and the m non-null eigenvalues of the loading matrix $\mathbf{A}\text{cov}(\mathbf{s}_t)\mathbf{A}'$. It is worth noticing that the above CLT is still valid if $\tilde{c}_n = (p-m)/n$ is substituted for c. Now if we let $p \ll n$ so that $\tilde{c}_n \simeq 0$ and $b(\sigma^2) \simeq 0$, and hence

$$\frac{(p-m)}{\sigma^2\sqrt{2c}}(\widehat{\sigma}^2 - \sigma^2) + b(\sigma^2) \simeq \frac{\sqrt{p-m}}{\sigma^2\sqrt{2}}(\widehat{\sigma}^2 - \sigma^2) \ .$$

This is nothing but the CLT (5) for $\widehat{\sigma}^2$ known under the classical low-dimensional scheme. From this point of view, Theorem 1 constitutes a natural extension of the classical CLT to the high-dimensional context.

Sketched proof of Theorem 1. First we introduce some notations following [7]. Let $F_n = p^{-1}\sum_{i=1}^p \delta_{\lambda_{n,i}}$ be the empirical distribution of the sample eigenvalues $\delta_{\lambda_{n,i}}$ and

$$H_n = \frac{1}{p}\left\{\sum_{i=1}^m \delta_{\alpha_i+\sigma^2} + (p-m)\delta_{\sigma^2}\right\},$$

be the empirical distribution of the (population) eigenvalues of $\mathbf{\Sigma}$. The related Marčenko-Pastur distribution of index (c_n, H_n) is denoted F_{c_n,H_n}.

We have

$$(p-m)\widehat{\sigma}^2 = \sum_{i=1}^p \lambda_{n,i} - \sum_{i=1}^m \lambda_{n,i}.$$

By [8],

$$\sum_{i=1}^m \lambda_{n,i} \longrightarrow \sum_{i=1}^m \left(\alpha_i + \frac{c\sigma^4}{\alpha_i}\right) + \sigma^2 m(1+c) \text{ a.s.} \quad (6)$$

For the first term, we have

$$\begin{aligned}\sum_{i=1}^p \lambda_i &= p\int x dF_n(x) \\ &= p\int x\, \mathrm{d}(F_n - F_{c_n,H_n})(x) + p\int x\, \mathrm{d}F_{c_n,H_n}(x) \\ &= G_n(x) + p\int x\, \mathrm{d}F_{c_n,H_n}(x).\end{aligned}$$

By [7], the first term is asymptotically normal

$$G_n(x) = \sum_{i=1}^p \lambda_{n,i} - p\int x\, \mathrm{d}F_{c_n,H_n}(x) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 2c\sigma^4).$$

Furthermore, by Lemma 1 of [9],

$$\int x\, \mathrm{d}F_{c_n,H_n}(x) = \int t\, \mathrm{d}H_n(t) = \sigma^2 + \frac{1}{p}\sum_{i=1}^m \alpha_i.$$

So we have

$$\sum_{i=1}^p \lambda_{n,i} - p\sigma^2 - \sum_{i=1}^m \alpha_i \xrightarrow{\mathcal{L}} \mathcal{N}(0, 2c\sigma^4). \quad (7)$$

By (6) and (7) and using Slutsky's lemma, we obtain

$$(p-m)(\widehat{\sigma}^2 - \sigma^2) + c\sigma^2\left(m + \sigma^2\sum_{i=1}^m \frac{1}{\alpha_i}\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 2c\sigma^4).$$

□

2.1. Monte-Carlo experiments

We consider an i.i.d. Gaussian sample of size n in three different settings:

- Model 1: $\text{spec}(\mathbf{\Sigma}) = (25, 16, 9, 0, \ldots, 0) + \sigma^2(1, \ldots, 1)$, $\sigma^2 = 4$, $c = 1$;
- Model 2: $\text{spec}(\mathbf{\Sigma}) = (4, 3, 0, \ldots, 0) + \sigma^2(1, \ldots, 1)$, $\sigma^2 = 2$, $c = 0.2$;
- Model 3: $\text{spec}(\mathbf{\Sigma}) = (12, 10, 8, 8, 0, \ldots, 0) + \sigma^2(1, \ldots, 1)$, $\sigma^2 = 3$, $c = 1.5$.

Figure 1 presents the histograms from 1000 replications of

$$\frac{(p-m)}{\sigma^2\sqrt{2c}}(\widehat{\sigma}^2-\sigma^2)+b(\sigma^2)$$

for the three models above, with different sample size n and $p = c \times n$, compared to the density of the standard normal probability law. Even for a moderate sample size like $n = 100$, the distribution is almost normal.

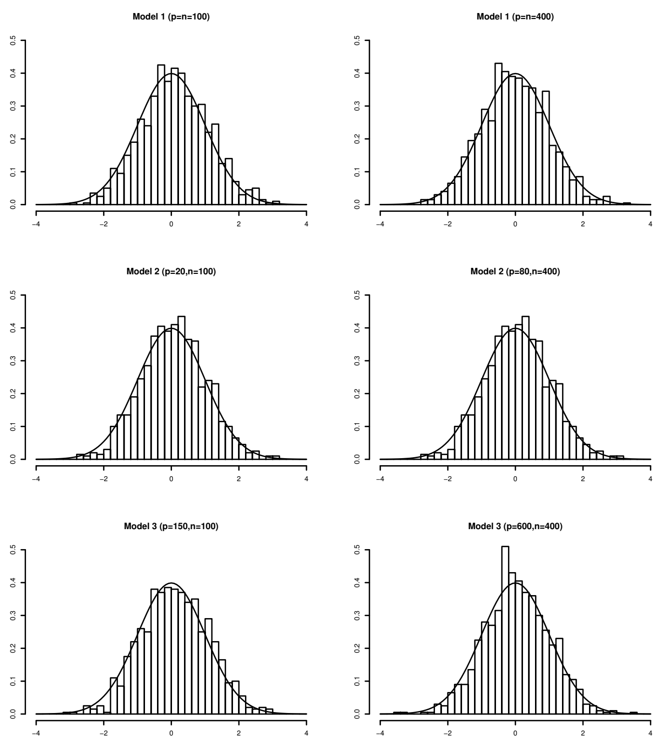

Fig. 1. Histogram of $\frac{(p-m)}{\sigma^2\sqrt{2c}}(\widehat{\sigma}^2-\sigma^2)+b(\sigma^2)$ compared with the density of a standard Gaussian law.

In Table 1, we compare the empirical bias of $\widehat{\sigma}^2$ (i.e. the empirical mean $\hat{b}$ of $\sigma^2 - \widehat{\sigma}^2 = \sigma^2 - \frac{1}{p-m}\sum_{i=m+1}^{p}\lambda_{n,i}$) over 1000 replications with the theoretical one $b = -\sigma^2\sqrt{2c}b(\sigma^2)/(p-m)$ in different settings. In all the three models, the empirical and theoretical bias are close each other. As expected, their difference vanishes when p and n increase.

3. A BIAS-CORRECTED ESTIMATOR

The previous theory recommends to correct the negative bias of $\widehat{\sigma}^2$. However, the bias $b(\sigma^2)$ depends on the number m and the values α_i of the spikes. These parameters could not be known in real-data applications and they need to be first estimated. In the literature, consistent estimators of m have been proposed, e.g. in [2], [10] and [11]. For the values of the spikes α_i, consistent estimators are proposed in [12].

As the bias depends on σ^2 which we want to estimate, a natural correction is to use the plug-in estimator

$$\widehat{\sigma}_*^2 = \widehat{\sigma}^2 + \frac{b(\widehat{\sigma}^2)}{p-m}\widehat{\sigma}^2\sqrt{2c}.$$

Table 1. Comparison between the empirical bias $\hat{b}$ and the theoretical bias b in various settings. Upper-middle-lower block: Model 1-2-3, respectively.

Settings	$\hat{b}$	b	$\lvert\hat{b}-b\rvert$
$(p, n) = (100, 100)$	-0.1556	-0.1589	0.0023
$(p, n) = (400, 400)$	-0.0379	-0.0388	0.0009
$(p, n) = 800, 800)$	-0.0189	-0.0193	0.0004
$(p, n) = 20, 100)$	-0.0654	-0.0704	0.0050
$(p, n) = 80, 400)$	-0.0150	-0.0162	0.0012
$(p, n) = 200, 1000)$	-0.0064	-0.0063	0.0001
$(p, n) = 150, 100)$	-0.0801	-0.0795	0.0006
$(p, n) = 600, 400)$	-0.0400	-0.0397	0.0003
$(p, n) = 1500, 1000)$	-0.0157	-0.0159	0.0002

Notice that in this formula, the number of factors m can be replaces by any consistent estimate as discussed above without affecting its limiting distribution. Using Theorem 1 and the delta-method, we obtain the following CLT

Theorem 2. *Under the same conditions as in Theorem 1, we have*

$$\frac{p-m}{\sigma^2\sqrt{2c}}\left(\widehat{\sigma}_*^2-\sigma^2\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)\ .$$

Compared to the m.l.e. $\widehat{\sigma}^2$ in Theorem 1, the new estimator has no more a bias after normalization by $\frac{p-m}{\sigma^2\sqrt{2c}}$.

To assess the quality of this bias-corrected estimator $\widehat{\sigma}_*^2$, we conduct some simulation experiments using the previous settings: Tables 2 and 3 give the empirical mean of $\widehat{\sigma}_*^2$ over 1000 replications compared with the empirical mean of $\widehat{\sigma}^2$, as well as the mean squared errors and mean absolute deviations. For comparison, the same statistics are also given for the estimator $\widehat{\sigma}_{\text{KN}}^2$ of [2] and the estimator $\widehat{\sigma}_{\text{US}}^2$ of [6]. These two estimators are defined as follow:

- $\widehat{\sigma}_{\text{KN}}^2$ is the solution of the following non-linear system of $m+1$ equations involving the $m+1$ unknowns $\widehat{\rho}_1, \ldots, \widehat{\rho}_m$ and $\widehat{\sigma}_{\text{KN}}^2$

$$\widehat{\sigma}_{\text{KN}}^2 - \frac{1}{p-m}\left[\sum_{j=m+1}^{p}\lambda_{n,j} + \sum_{j=1}^{m}(\lambda_{n,j}-\widehat{\rho}_j)\right] = 0,$$

$$\widehat{\rho}_j^2 - \widehat{\rho}_j\left(\lambda_{n,j} + \widehat{\sigma}_{\text{KN}}^2 - \widehat{\sigma}_{\text{KN}}^2\frac{p-m}{n}\right) + \lambda_{n,j}\widehat{\sigma}_{\text{KN}}^2 = 0.$$

We used the computing code available on the author's web-page to carry out the simulations.

- $\widehat{\sigma}^2_{\text{US}}$ is defined as

$$\widehat{\sigma}^2_{\text{US}} = \frac{\text{median}(\lambda_{n,m+1}, \ldots, \lambda_{n,p})}{p^{-1}_{c,1}(0.5)},$$

where $p^{-1}_{c,1}$ is quantile function of the Marčenko-Pastur distribution $F_{c,1}$.

Table 2. Mean absolute deviation of $\widehat{\sigma}^2$ and $\widehat{\sigma}^2_*$ in various settings.

Mod.	p	n	σ^2	$\lvert\sigma^2 - \widehat{\sigma}^2\rvert$	$\lvert\sigma^2 - \widehat{\sigma}^2_*\rvert$
1	100	100	4	0.1536	0.0021
	400	400		0.0384	$< 10^{-5}$
	800	800		0.0191	0.0002
2	20	100	2	0.0660	0.0012
	80	400		0.0159	0.0001
	200	1000		0.0061	0.0002
3	150	100	3	0.1600	0.0074
	600	400		0.0395	0.0001
	1500	1000		0.0161	0.0002

Table 3. Mean absolute deviation of $\widehat{\sigma}^2_{\text{KN}}$ and $\widehat{\sigma}^2_{\text{US}}$ in various settings.

Mod.	p	n	σ^2	$\widehat{\sigma}^2_{\text{US}}$	$\lvert\sigma^2 - \widehat{\sigma}^2_{\text{US}}\rvert$
1	100	100	4	0.0030	0.1616
	400	400		0.0003	0.0415
	800	800		0.0002	0.0206
2	20	100	2	0.0003	0.0600
	80	400		0.0001	0.0149
	200	1000		0.0002	0.0058
3	150	100	3	0.0065	0.2250
	600	400		0.0006	0.0550
	1500	1000		0.0001	0.0227

In all three models considered, the bias-corrected estimator $\widehat{\sigma}^2_*$ is far much better than the original m.l.e. $\widehat{\sigma}^2$: here mean absolute deviations are reduced by 95% at least. The performances of $\widehat{\sigma}^2_*$ and $\widehat{\sigma}^2_{\text{KN}}$ are similar. The estimator $\widehat{\sigma^2}_{\text{US}}$ shows slightly better performance than the m.l.e. $\widehat{\sigma}^2$, but performs poorly compared to $\widehat{\sigma}^2_*$ and $\widehat{\sigma}^2_{\text{KN}}$. Notice however the theoretic properties of $\widehat{\sigma}^2_{\text{KN}}$ and $\widehat{\sigma^2}_{\text{US}}$ are unknown and so far there have been checked via simulations only.

4. REFERENCES

[1] I. Johnstone, "On the distribution of the largest eigenvalue in principal components analysis," *Ann. Statistics*, vol. 29, no. 2, pp. 295–327, 2001.

[2] S. Kritchman and B. Nadler, "Determining the number of components in a factor model from limited noisy data," *Chem. Int. Lab. Syst.*, vol. 94, 2008.

[3] D. Passemier and J. Yao, "On the detection of the number of spikes, possibly equal, in the high-dimensional case," *J. Multivariate Analysis*, vol. 127, pp. 173–183, 2014.

[4] T. W. Anderson and H. Rubin, "Statistical inference in factor analysis," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. V*, Berkeley and Los Angeles, 1956, pp. 111–150, University of California Press.

[5] T. W. Anderson and Y. Amemiya, "The asymptotic normal distribution of estimators in factor analysis under general conditions," *Ann. Statist.*, vol. 16, no. 2, pp. 759–771, 1988.

[6] M. O. Ulfarsson and V. Solo, "Dimension estimation in noisy PCA with SURE and random matrix theory," *IEEE Trans. Signal Process.*, vol. 56, no. 12, pp. 5804–5816, 2008.

[7] Z. Bai and J. Silverstein, "CLT for linear spectral statistics of large-dimensional sample covariance matrices," *Ann. Probab.*, vol. 32, pp. 553–605, 2004.

[8] J. Baik and J.W. Silverstein, "Eigenvalues of large sample covariance matrices of spiked population models," *J. Multivariate. Anal.*, vol. 97, pp. 1382–1408, 2006.

[9] Z. Bai, J. Chen, and J. Yao, "On estimation of the population spectral distribution from a high-dimensional sample covariance matrix," *Aust. N. Z. J. Stat.*, vol. 52, no. 4, pp. 423–437, 2010.

[10] A. Onatski, "Testing hypotheses about the number of factors in large factor models," *Econometrica*, vol. 77, no. 5, pp. 1447–1479, 2009.

[11] D. Passemier and J. Yao, "On determining the number of spikes in a high-dimensional spiked population model," *Random Matrices: Theory and Applications*, vol. 1, no. 1, pp. 1150002, 2012.

[12] Z. Bai and X. Ding, "Estimation of spiked eigenvalues in spiked models," *Random Matrices Theory Appl.*, vol. 1, no. 2, pp. 1150011, 21, 2012.