Statistica Sinica Preprint No: SS-2017-0296.R1

Title	Edgeworth correction for the largest eigenvalue in a spiked PCA model
Manuscript ID	SS-2017-0296.R1
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202017.0296
Complete List of Authors	Jeha Yang and Iain M. Johnstone
Corresponding Author	Iain M. Johnstone
E-mail	imj@stanford.edu
Notice: Accepted version subject to English editing.	

Statistica Sinica

EDGEWORTH CORRECTION FOR THE LARGEST EIGENVALUE IN A SPIKED PCA MODEL*

Jeha Yang and Iain M. Johnstone

Stanford University

Abstract: We study improved approximations to the distribution of the largest eigenvalue $\hat{\ell}$ of the sample covariance matrix of n zero-mean Gaussian observations in dimension $p+1$. We assume that one population principal component has variance $\ell > 1$ and the remaining 'noise' components have common variance 1. In the high dimensional limit $p/n \to \gamma > 0$, we begin study of Edgeworth corrections to the limiting Gaussian distribution of $\hat{\ell}$ in the supercritical case $\ell > 1 + \sqrt{\gamma}$. The skewness correction involves a quadratic polynomial as in classical settings, but the coefficients reflect the high dimensional structure. The methods involve Edgeworth expansions for sums of independent non-identically distributed variates obtained by conditioning on the sample noise eigenvalues, and limiting bulk properties *and* fluctuations of these noise eigenvalues.

Key words and phrases: Spiked PCA model, Roy's statistic, Edgeworth expansion

1. Introduction

Models for high dimensional data with low dimensional structure are the focus of much current research. This paper considers one of the simplest such settings, the rank one

*Peter Hall visited Stanford many times, including a month long visit with Jeannie in 1988 [ck year]. The second author (IMJ) was generously hosted by Peter even more often both at ANU and Melbourne. Stimulating and enjoyable as those visits predictably were, we never discussed Edgeworth expansions. Fortunately, the clarity of Peter's exposition in his Bootstrap and Edgeworth book and his well-known fondness for the monograph of Petrov (1975) provided exactly what we needed for this project, begun after his most untimely passing.

"spiked model" with Gaussian data, in order to begin the study of Edgeworth expansion approximations for high dimensional data. Specifically, we work with the following simple model.

Model (M). Suppose that we observe $X = [x_1, \cdots, x_n]'$ where $x_1, \ldots, x_n$ are i.i.d from $N_{p+1}(0, \Sigma)$, and the population covariance matrix $\Sigma = I + (\ell - 1)vv'$ for some unit vector v. Suppose also that p increases with n so that $\gamma_n = p/n \to \gamma \in (0, \infty)$ and that $\ell > 1 + \sqrt{\gamma}$.

Thus, one population principal component has variance $\ell > 1$ and the remaining p have common variance 1.

The Baik, Ben Arous and Péché (2005) *phase transition* is an important phenomenon that appears in this high dimensional asymptotic regime. It concerns the largest eigenvalues in spiked models, which are of primary interest in principal components analysis. In the rank one special case, let $\hat{\ell}$ be the largest eigenvalue of the sample covariance matrix $S = n^{-1}X'X$. Below the phase transition, $\ell < 1+\sqrt{\gamma}$, and after a centering and scaling that does not depend on ℓ, asymptotically $n^{2/3}\hat{\ell}$ has a Tracy-Widom distribution. Above the phase transition, the 'super-critical regime', the convergence rate is $n^{1/2}$ and the limit Gaussian:

$$n^{1/2}[\hat{\ell} - \rho(\ell, \gamma_n)]/\sigma(\ell, \gamma_n) \xrightarrow{\mathcal{D}} N(0, 1). \tag{1.1}$$

The centering and scaling functions now depend on ℓ:

$$\rho(\ell, \gamma) = \ell + \gamma\ell/(\ell - 1), \qquad \sigma^2(\ell, \gamma) = 2\ell^2[1 - \gamma/(\ell - 1)^2]. \tag{1.2}$$

Baik, Ben Arous and Péché (2005) proved (1.1) for complex valued data using structure specific to the complex case. The real case was established using different methods by Paul (2007), under the additional assumption $\gamma_n - \gamma = o(n^{-1/2})$ and with γ_n in (1.1) replaced by γ. We will see below that (1.1) holds as stated without this assumption. Consequently, we adopt the abbreviations

$$\rho_n = \rho(\ell, \gamma_n), \qquad \sigma_n = \sigma(\ell, \gamma_n). \tag{1.3}$$

The quality of approximation in asymptotic normality results such as (1.1) is often studied using Edgeworth expansions, e.g. Hall (1992). However, our high dimensional setting appears to lie beyond the standard frameworks for Edgeworth expansions, such as for example the use of smooth functions of a *fixed* dimensional vector of means of independent random variables, as in Hall (1992, Sec. 2.4).

2. Main Result

Our main result is a skewness correction for the normal approximation (1.1) to the largest eigenvalue statistic. The simplest version of the result may be stated as follows. As usual Φ and ϕ denote the standard Gaussian cumulative and density respectively.

Theorem 1. *Adopt* ***Model (M)****, and let $\hat{\ell}$ be the largest eigenvalue of $S = n^{-1}\sum_{i=1}^n x_i x_i'$, and let $R_n = n^{1/2}(\hat{\ell} - \rho_n)/\sigma_n$, where the centering and scaling are defined in* (1.2) *and* (1.3). *Then we have a first order Edgeworth expansion*

$$\mathbb{P}(R_n \le x) = \Phi(x) + n^{-1/2} p_1(x)\phi(x) + o(n^{-1/2}), \tag{2.1}$$

valid uniformly in x, and with

$$p_1(x) = \sqrt{2}\left(\tfrac{1}{3}[(\ell-1)^3 + \gamma](1-x^2) - \tfrac{1}{2}\gamma\ell\right)\left((\ell-1)^2 - \gamma\right)^{-3/2}. \tag{2.2}$$

We compare (2.1) with the previously known expression for dimension p fixed in the next section. The effects of high dimensionality are seen both in the coefficient of the "usual" polynomial $1 - x^2$ as well as in the additional constant term proportional to $\gamma\ell$.

We turn to formulating the version of Theorem 1 that we actually prove, and in the process sketch some elements of our approach in order to give a first indication of the role of high dimensionality in the Edgeworth correction. Building on the approach of Paul (2007), the $n \times (p+1)$ data matrix may be partitioned as $X = [\sqrt{\ell} Z_1, Z_2]$, with the 'signal' in the first column and the remaining p columns containing pure noise: i.i.d. standard normal variates. Now consider the eigen decomposition $n^{-1} Z_2 Z_2' = U \Lambda U'$ in which U is $n \times n$ orthogonal and the diagonal matrix Λ contains the ordered nonzero eigenvalues $\lambda_1 \geq \cdots \geq \lambda_{n \wedge p}$ of $n^{-1} Z_2 Z_2'$, supplemented by zeros in the case $n > p$. It is a special feature of white Gaussian noise that (U, Λ) are mutually independent, with U being uniformly (i.e. Haar) distributed on its respective space. In view of this, if we set $z = U' Z_1$, it follows that the eigenvalues of S depend only on z and Λ, and that

$$z = U' Z_1 \sim N(0, I_n), \quad z \perp \Lambda. \tag{2.3}$$

The vector z provides enough independent randomness for Gaussian limit behavior of $\hat{\ell}$, conditional on Λ. In particular, for a function f on $[0, \infty)$, we define

$$S_n(f) = n^{-1/2} \sum_{i=1}^{n} f(\lambda_i)(z_i^2 - 1). \tag{2.4}$$

As n grows, we may also use the bulk regularity properties of Λ. Thus the empirical distribution F_n of the p sample eigenvalues of $n^{-1} Z_2' Z_2$ converges to the Marchenko-Pastur distribution F_γ supported on $[a(\gamma), b(\gamma)]$ if $\gamma \leq 1$ and with an atom $(1 - \gamma^{-1})$ at 0 if $\gamma > 1$, where

$$a(\gamma) = (1 - \sqrt{\gamma})^2, \quad b(\gamma) = (1 + \sqrt{\gamma})^2.$$

The 'companion' empirical distribution F_n of the n eigenvalues $(\lambda_1, \ldots, \lambda_n)$ of $n^{-1} Z_2 Z_2'$ converges to the companion MP law $\mathsf{F}_\gamma = (1 - \gamma) I_{[0,\infty)} + \gamma F_\gamma$. Integrals against F indicating one of these types of distributions will be written in the form

$$F(f) = \int f(\lambda) F(d\lambda).$$

Paul's Schur complement argument, reviewed in the proof section below, leads to an equation for the fluctuation of $\hat{\ell}$ about its centering ρ_n:

$$n^{1/2}(\hat{\ell} - \rho_n) = \frac{S_n(g_n)}{\mathsf{F}_{\gamma_n}(g_n^2)} + O_p(n^{-1/2}), \tag{2.5}$$

where $g_n(\lambda) = (\rho_n - \lambda)^{-1}$. From (S1.3), $\mathsf{F}_{\gamma_n}(g_n^2) = 2\sigma_n^{-2}$. The sum $S_n(g_n)$ is asymptotically normal given Λ, with asymptotic variance $\mathsf{F}_\gamma(g^2)$, for example via the Lyapounov CLT, and completing this argument yields the asymptotic normality result (1.1).

A more accurate version of (2.5) is needed for a first Edgeworth approximation. Indeed we later show that

$$n^{1/2}(\hat{\ell} - \rho_n) = \frac{S_n(g_n) + n^{-1/2}G_n(g_n)}{\mathsf{F}_{\gamma_n}(g_n^2) + n^{-1/2}G_n(\tilde{g}_n) + O_p(n^{-1})},$$

where $\tilde{g}_n$ is defined later. This expression involves the discrepancy between a trace and its centering:

$$G_n(f) = \sum_{i=1}^n f(\lambda_i) - n\int f(\lambda)\mathsf{F}_{\gamma_n}(d\lambda) = n(\mathsf{F}_n(f) - \mathsf{F}_{\gamma_n}(f)) = p(F_n(f) - F_{\gamma_n}(f)).$$

This centered linear statistic, though unnormalized, is $O_p(1)$, and indeed, according to the CLT of Bai and Silverstein (2004), for suitable f is asymptotically normal:

$$G_n(f) \xrightarrow{\mathcal{D}} N(\mu(f), \sigma^2(f)). \tag{2.6}$$

We use a first term Edgeworth approximation to the distribution of $S_n(g_n)$ conditional on Λ, using results for sums of independent non-identically distributed variables described in Petrov (1975, Ch VI.). This uses the conditional cumulants of S_n for $j = 2, 3$, given by

$$\frac{d^j}{dt^j}\log \mathbb{E}[e^{itS_n}|\Lambda]|_{t=0} = \kappa_j n^{-1}\sum_{i=1}^n g_n^j(\lambda_i),$$

where, in turn, $\kappa_j = 2^{j-1}(j-1)!$ are the cumulants of $z^2 - 1 \sim \chi^2_{(1)} - 1$. A deterministic asymptotic approximation to these conditional cumulants is then given by

$$\kappa_{2,n} = 2\mathsf{F}_{\gamma_n}(g_n^2), \qquad \kappa_{3,n} = 8\mathsf{F}_{\gamma_n}(g_n^3). \tag{2.7}$$

With these preparations we are ready for the main theorem.

Theorem 2. *With the assumptions of Theorem 1, we have the Edgeworth expansion*

$$\mathbb{P}(R_n \le x) = \Phi(x) + n^{-1/2} p_{1,n}(x)\phi(x) + o(n^{-1/2}), \tag{2.8}$$

valid uniformly in x, and with

$$p_{1,n}(x) = \tfrac{1}{6}\kappa_{2,n}^{-3/2}\kappa_{3,n}(1 - x^2) - \kappa_{2,n}^{-1/2}\mu(g_n),$$

for $g_n(\lambda) = (\rho_n - \lambda)^{-1}$ and $\kappa_{j,n}$ defined by (2.7)*, and $\mu(\cdot)$ the asymptotic mean in the Bai-Silverstein limit* (2.6)*.*

The structure of $p_{1,n}(x)$ as an even quadratic polynomial is the same as in the smooth function of means model (Hall, 1992, Theorem 2.2). In our high dimensional setting, the first term in $p_{1,n}(x)$ reflects the Edgeworth approximation to $S_n(g_n)$ conditional on Λ, while the second shows the effects of fluctations of Λ. From (S1.3), (S1.4) and (S1.5), we have more explicit evaluations

$$\begin{aligned}
\kappa_{2,n} &= 2(1-\ell^{-1})^2((\ell-1)^2 - \gamma_n)^{-1} = 4\sigma_n^{-2}, \\
\kappa_{3,n} &= 8(1-\ell^{-1})^3((\ell-1)^3 + \gamma_n)((\ell-1)^2 - \gamma_n)^{-3}, \\
\mu(g_n) &= \gamma_n(\ell-1)((\ell-1)^2 - \gamma_n)^{-2},
\end{aligned}$$

which lead to an explicit form of the first order correction term

$$p_{1,n}(x) = \sqrt{2}\left(\tfrac{1}{3}[(\ell-1)^3 + \gamma_n](1-x^2) - \tfrac{1}{2}\gamma_n\ell\right)((\ell-1)^2 - \gamma_n)^{-3/2}.$$

Since the error term is $o(n^{-1/2})$ and $\gamma_n = \gamma + o(1)$, we may replace γ_n by γ in the previous display and recover Theorem 1.

Remark. To emphasize the advantage of using $\gamma_n = p/n$ rather than γ in the centering and scaling formulas, note that if $\gamma_n = \gamma + an^{-1/2}$, then the limiting distribution of

$$\check{R}_n = n^{1/2}[\hat{\ell} - \rho(\ell, \gamma)]/\sigma(\ell, \gamma)$$

has a non-zero mean $\alpha = \alpha(a, \ell, \gamma)$. The situation is yet more delicate for the skewness correction: if $\gamma_n = \gamma + bn^{-1}$, then

$$\mathbb{P}(\check{R}_n \leq x) - \mathbb{P}(R_n \leq x) = n^{-1/2}(\beta_0 + \beta_1 x)\phi(x) + o(n^{-1/2})$$

for constants β_1, β_0 depending on b, ℓ, γ.

Remark. A parallel result for rank one perturbations of the Gaussian Orthogonal Ensemble is available. Consider a data matrix $X = \theta e_1 e_1^T + Z$ where $\theta > 1$ and Z is $p \times p$ symmetric with $Z_{ii} \sim N(0, 2/p)$ and $Z_{ij} \sim N(0, 1/p)$ for $i > j$, and $p \to \infty$. The largest eigenvalue of X, denoted $\hat{\theta}$, converges a.s. to $\rho = \theta + \theta^{-1}$, and with $\sigma = \sqrt{2(1 - \theta^{-2})}$, the quantity $R_p = \sqrt{p}(\hat{\theta} - \rho)/\sigma$ is asymptotically standard Gaussian (Benaych-Georges, Guionnet, and Maida, 2011, Theorem 5.1). As is well known, the empirical spectral distribution of $Z^{[2:p,2:p]}$ converges weakly to the *semicircle law* F_{sc} with density $\frac{1}{2\pi}\sqrt{4 - x^2}$ on the interval $[-2, 2]$). Our method, along with CLT for linear spectral statistics $F_{sc}(f)$ of Bai and Yao (2005) leads to a first order Edgeworth correction for R_p:

$$p_1(x) = \frac{\sqrt{2}}{(\theta^2 - 1)^{3/2}} \left(\frac{1 - x^2}{3} - \frac{1}{2} \right),$$

which has a structure analogous to that of our main result.

Comparison with fixed p. In classical asymptotic theory, when $n \to \infty$ with p fixed, asymptotically $\hat{\ell} \sim N(\ell, 2\ell^2)$. Introduce therefore $\check{R}_n = \sqrt{n}(\hat{\ell} - \ell)/(\sqrt{2}\ell)$. When specialized to the skewness correction term, Theorem 2.1 of Muirhead and Chikuse (1975) reads

$$\mathbb{P}(\check{R}_n \le x) = \Phi(x) + n^{-1/2}\left(\frac{\sqrt{2}}{3}(1 - x^2) - \frac{p}{\sqrt{2}(\ell - 1)}\right)\phi(x) + O(n^{-1}). \tag{2.9}$$

Formally setting $\gamma = 0$ in (2.2) of Theorem 1, we get only the term $p_1(x) = (\sqrt{2}/3)(1 - x^2)$. To see that the two results are nevertheless consistent, write $\rho_n = \ell(1 + b_n)$ and $\sigma_n = \sqrt{2}\ell c_n$ where $b_n = \gamma_n/(\ell - 1)$ and $c_n = [1 - \gamma_n/(\ell - 1)^2]^{1/2}$, so that

$$R_n = \sqrt{n}\frac{\hat{\ell} - \ell - b_n\ell}{\sqrt{2}\ell c_n} = c_n^{-1}(\check{R}_n - d_n),$$

where $d_n = \sqrt{n/2}b_n = \sqrt{n/2}\gamma_n/(\ell - 1) = (2n)^{-1/2}p/(\ell - 1)$ is the second term in (2.9). Applying (2.9) at $\check{x}_n = c_n x + d_n$, we find

$$\mathbb{P}(R_n \le x) = \mathbb{P}(\check{R}_n \le \check{x}_n) = \Phi(\check{x}_n) + [n^{-1/2}\tfrac{\sqrt{2}}{3}(1 - \check{x}_n^2) - d_n]\phi(\check{x}_n) + O(n^{-1}).$$

Observe that $\Phi(\check{x}_n) - d_n\phi(\check{x}_n) = \Phi(c_n x) + O(d_n^2)$ with $d_n = O(n^{-1/2})$, and $c_n = [1 - \gamma_n/(\ell - 1)^2]^{1/2} = 1 + O(n^{-1})$. Therefore, $\check{x}_n = x + O(n^{-1/2})$ and $c_n x = x + O(n^{-1})$, yielding

$$\mathbb{P}(R_n \le x) = \Phi(x) + n^{-1/2}\tfrac{\sqrt{2}}{3}(1 - x^2)\phi(x) + O(n^{-1}),$$

and so we do recover agreement with $\gamma = 0$ in (2.2).

Hermite polynomials and numerical comparisons. It is helpful to view Edgeworth expansions in terms of Hermite polynomials $H_n(x)$, defined by $H_n(x)\phi(x) = (-d/dx)^n\phi(x)$. In particular, $H_n(x) = 1, x, x^2 - 1$ and $x^3 - 3x$ for $n = 0, 1, 2$ and 3. The Edgeworth approximation of Theorem 2 then becomes

$$F_E = \Phi - n^{-1/2}(\alpha_2 H_2 + \alpha_0)\phi$$

with $h = \ell - 1$ and

$$\alpha_2 = \frac{\sqrt{2}}{3}\frac{h^3 + \gamma_n}{(h^2 - \gamma_n)^{3/2}}, \qquad \alpha_0 = \frac{1}{\sqrt{2}}\frac{\gamma_n l}{(h^2 - \gamma_n)^{3/2}}.$$

Since $(d/dx)H_n(x) = -H_{n+1}(x)$, the Edgeworth corrected density is given by

$$f_E = \phi + n^{-1/2}(\alpha_2 H_3 + \alpha_0 H_1)\phi.$$

The relative error

$$\frac{f_E - \phi}{\phi} = n^{-1/2}q, \qquad q = \alpha_2 H_3 + \alpha_0 H_1,$$

is a cubic polynomial with positive leading coefficient. It is easy to verify that the three roots, namely $0, \pm(3 - \alpha_0/\alpha_2)^{1/2}$ are real when $\ell > 1 + \sqrt{\gamma_n}$. Hence the Edgeworth density approximation is necessarily negative for $\hat{\ell}$ sufficiently small, and intersects the normal density three times.

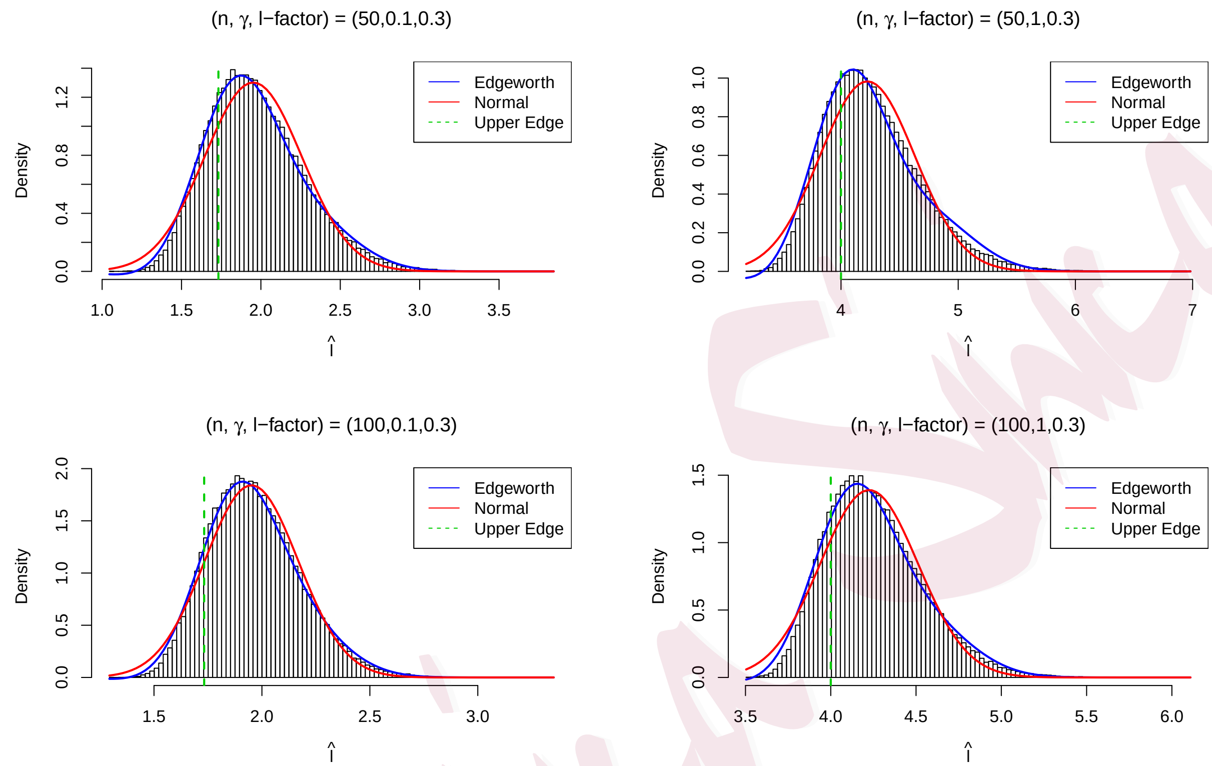
We now show numerical examples in which the Edgeworth corrected 'density' provides a better approximation to the distribution of R_n than does the standard normal. The parameters

$$n \in \{50, 100\}; \qquad \gamma_n \in \{0.1, 1\}; \qquad \ell\text{-factor} := \ell/(1 + \sqrt{\gamma_n}) - 1 \in \{0.3, 0.5\},$$

are chosen so that n is neither too small for asymptotics to be meaningful nor too large to distinguish $f_E(x)$ and $\phi(x)$, γ_n is close to either 0 or 1, and ℓ is moderately separated from the (finite version) critical point $1 + \sqrt{\gamma_n}$.

Figures 1 and 2 in fact show the densities $y \to \sqrt{n/\sigma_n} f_E(\sqrt{n/\sigma_n}(y - \rho_n))$ after shifting and scaling to correspond to $\hat{\ell}$. Superimposed are the corresponding rescaled normal density as well as histograms of $100,000$ simulated replicates of $\hat{\ell}$. The green dashed lines show the upper bulk edge $(1 + \sqrt{\gamma_n})^2$ to emphasize that these settings for $\hat{\ell}$ are not too far above the

Figure 1: Plots for l-factor $= 0.3$



bulk. In the cases shown, the Edgeworth correction provides a (right) skewness correction that matches the simulated histograms reasonably well, though unsurprisingly the small $n = 50$ and large $\gamma = 1$ case, has the least good match.

When ℓ is closer to the phase transition, so that the ℓ-factor is smaller, the skewness correction becomes unsatisfactory due to the singularity in the denominator of α_2 and α_0 as h approaches $\sqrt{\gamma}$. Empirically, we have found that the skewness correction may be reasonable, with a single inflection point visible above the mode, when

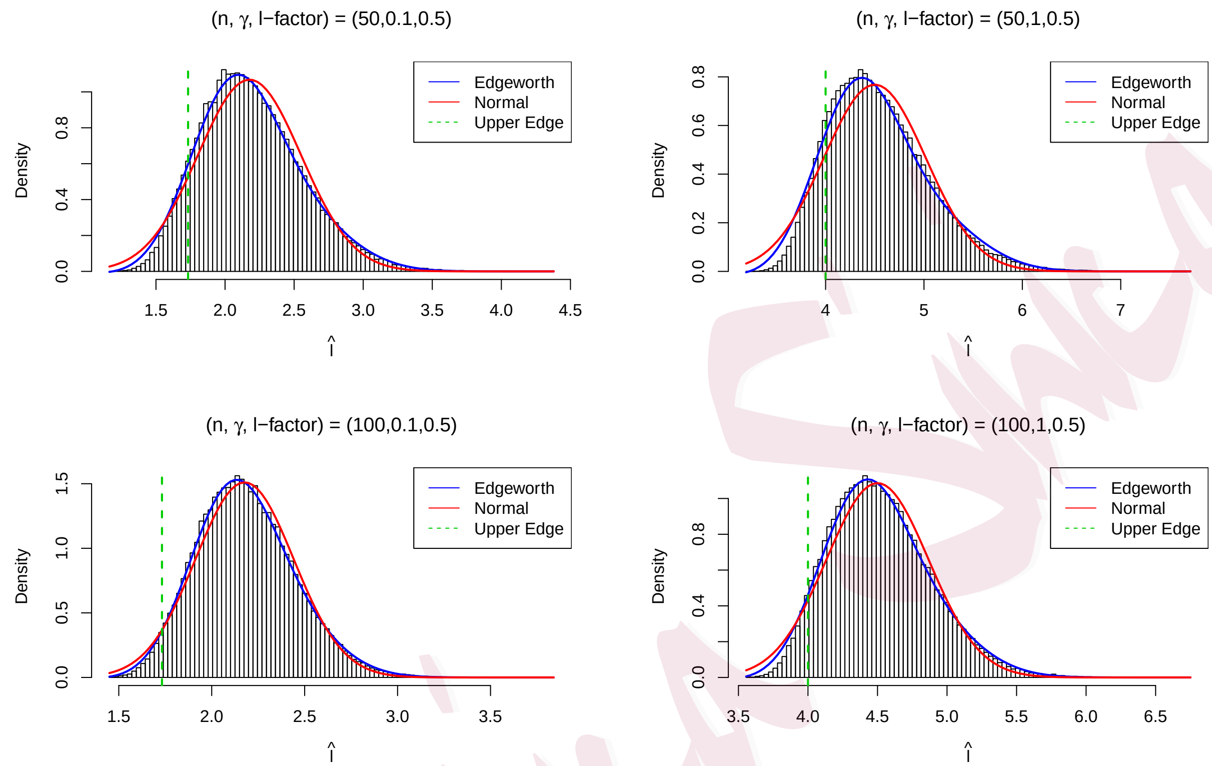
$$\frac{1}{n}(9/2)\alpha_2^2 = \frac{1}{n}\frac{(h^3+\gamma)^2}{(h^2-\gamma)^3} \leq 0.2.$$

3. Proof

3.1. Outline

We start with deriving the useful expression of R_n as introduced in the first section with

Figure 2: Plots for l-factor $= 0.5$



more details. Without loss of generality, we may assume that the population covariance matrix of the distribution of $x_1, \cdots, x_n$ is $\text{diag}(\ell, 1, \cdots, 1)$(by an appropriate rotation, not changing S). Then, we write $X = [\sqrt{\ell} Z_1 \quad Z_2]$ where Z_1, Z_2 are $n \times 1$, $n \times p$ with i.i.d. standard normal elements, respectively. The eigenvalue equation $S\hat{v} = \hat{\ell}\hat{v}$ becomes

$$\begin{pmatrix} \ell Z_1' Z_1 & \sqrt{\ell} Z_1' Z_2 \\ \sqrt{\ell} Z_2' Z_1 & Z_2' Z_2 \end{pmatrix} \begin{pmatrix} \hat{v}_1 \\ \hat{v}_2 \end{pmatrix} = n\hat{\ell} \begin{pmatrix} \hat{v}_1 \\ \hat{v}_2 \end{pmatrix},$$

where $\hat{v}_1, \hat{v}_2$ are the first coordinate and the rest of $\hat{v}$, respectively. As usual, we substitute the second equation into the first, then cancel $\hat{v}_1$ to obtain

$$n\hat{\ell} = \ell Z_1'[I_n + Z_2(n\hat{\ell} I_p - Z_2' Z_2)^{-1} Z_2'] Z_1 = \ell Z_1'[\hat{\ell}(\hat{\ell} I_n - n^{-1} Z_2 Z_2')^{-1}] Z_1 = \ell z'[-\hat{\ell} R(\hat{\ell})] z,$$

whenever $\det(n\hat{\ell} I_p - Z_2' Z_2) \neq 0$, i.e. almost surely. Note that the second equation is a particular case of the Woodbury formula, $z = U' Z_1$ where U is from the eigendecomposition

$n^{-1}Z_2Z_2' = U\Lambda U'$ as introduced before, and the resolvent $R(x) = (\Lambda - xI_n)^{-1}$ is defined for $x \notin \{\lambda_1, \ldots, \lambda_n\}$. Now using the resolvent identity $R(x) = R(y) + (x-y)R(x)R(y)$ for $x, y \notin \{\lambda_1, \cdots, \lambda_n\}$, we obtain

$$n\hat{\ell} = \ell z'[-\rho_n R(\rho_n) - (\hat{\ell} - \rho_n)\Lambda R(\hat{\ell})R(\rho_n)]z,$$

which can be rearranged into a key equation

$$(\hat{\ell} - \rho_n)(1 + \ell n^{-1} z'\Lambda R(\hat{\ell})R(\rho_n)z) = \ell\rho_n(-n^{-1}z'R(\rho_n)z - \ell^{-1}) \tag{3.1}$$

whenever $\hat{\ell}, \rho_n \notin \{\lambda_1, \cdots, \lambda_n\}$ i.e. almost surely ; we assume this from now on. To investigate (3.1) further, we will make frequent use of the *stochastic decomposition*

$$n^{-1}\sum_{i=1}^{n} f(\lambda_i)z_i^2 = \mathsf{F}_{\gamma_n}(f) + n^{-1/2}S_n(f) + n^{-1}G_n(f). \tag{3.2}$$

where $\mathsf{F}_n(\cdot), S_n(\cdot)$ and $G_n(\cdot)$ are defined as above, which are of order $O_p(1)$ as we will see in the proof section. Noting that $-R(\rho_n) = \mathrm{diag}(g_n(\lambda_1), \cdots, g_n(\lambda_n))$ and $\mathsf{F}_{\gamma_n}(g_n) = \ell^{-1}$ (S1.2), we have $-n^{-1}z'R(\rho_n)z = \ell^{-1} + n^{-1/2}S_n(g_n) + n^{-1}G_n(g_n)$ from (3.2). Hence we can rewrite (3.1) as

$$(\hat{\ell} - \rho_n)(1 + \ell n^{-1}z'\Lambda R(\hat{\ell})R(\rho_n)z) = n^{-1/2}\ell\rho_n(S_n(g_n) + n^{-1/2}G_n(g_n)). \tag{3.3}$$

Also, use the resolvent identity to write

$$1 + \ell n^{-1}z'\Lambda R(\hat{\ell})R(\rho_n)z = 1 + \ell n^{-1}z'\Lambda R^2(\rho_n)z - \ell\nu_n, \tag{3.4}$$

where

$$\nu_n = -(\hat{\ell} - \rho_n)n^{-1}z'\Lambda R(\hat{\ell})R^2(\rho_n)z \tag{3.5}$$

will be $O_p(n^{-1/2})$ by (3.3) and tail bounds. One can use (3.2) to write the leading term as

$$1 + \ell n^{-1}z'\Lambda R^2(\rho_n)z = \ell\rho_n\mathsf{F}_{\gamma_n}(g_n^2) + n^{-1/2}\ell S_n(m_1g_n^2) + n^{-1}\ell G_n(m_1g_n^2) \tag{3.6}$$

where $m_k(\lambda) := \lambda^k, k \in \mathbb{N}$ are monomials, since $1+\ell\mathsf{F}_{\gamma_n}(m_1 g_n^2)-\ell\rho_n\mathsf{F}_{\gamma_n}(g_n^2) = 1-\ell\mathsf{F}_{\gamma_n}(g_n) = 0$ again by (S1.2). This allows us to rewrite (3.3) as

$$n^{1/2}(\hat{\ell} - \rho_n) = \frac{S_n(g_n) + O_p(n^{-1/2})}{\mathsf{F}_{\gamma_n}(g_n^2) + O_p(n^{-1/2})}$$

which establishes (2.5). To expand ν_n further, we insert (2.5) into (3.5), yielding

$$\begin{aligned}\nu_n &= n^{-1/2}(S_n(g_n)/\mathsf{F}_{\gamma_n}(g_n^2) + O_p(n^{-1/2}))(\mathsf{F}_{\gamma_n}(m_1 g_n^3) + O_p(n^{-1/2})) \\ &= n^{-1/2} r_n S_n(g_n) + O_p(n^{-1}),\end{aligned} \tag{3.7}$$

where

$$r_n = \ell\rho_n\mathsf{F}_{\gamma_n}(m_1 g_n^3)/(1 + \ell\mathsf{F}_{\gamma_n}(m_1 g_n^2)) = \mathsf{F}_{\gamma_n}(m_1 g_n^3)/\mathsf{F}_{\gamma_n}(g_n^2). \tag{3.8}$$

Putting (3.6), (3.7) and $\mathsf{F}_{\gamma_n}(g_n^2) = 2\sigma_n^{-2}$ (S1.3) into (3.4) gives

$$1 + \ell n^{-1} z'\Lambda R(\hat{\ell})R(\rho_n)z = \ell(2\rho_n\sigma_n^{-2} + n^{-1/2}S_n(m_1 g_n^2 - r_n g_n) + \delta_n) \tag{3.9}$$

where

$$\delta_n = n^{-1}G_n(m_1 g_n^2) - (\nu_n - n^{-1/2} r_n S_n(g_n)) \tag{3.10}$$

is $O_p(n^{-1})$ ignorable ; a rigorous proof of this fact is postponed to the delta method section.

All in all, combining (3.3) and (3.9), we obtain the master equation

$$n^{1/2}(\hat{\ell} - \rho_n) = \frac{\rho_n(S_n(g_n) + n^{-1/2}G_n(g_n))}{2\rho_n\sigma_n^{-2} - n^{-1/2}S_n(g_n h_n) + \delta_n}, \quad \text{with} \quad h_n = r_n - m_1 g_n. \tag{3.11}$$

Now we are ready to see the outline of the main proof. For notational convenience, let $\eta(\ell,\gamma) := \rho(\ell,\gamma) - b(\gamma) = (\ell-1)^{-1}(\ell - 1 - \sqrt{\gamma})^2 > 0$.

Step 1 From *tail bounds*, show that for any fixed $\delta \in (0, \min(1, \eta(\ell,\gamma)/4, \gamma/2))$, the event

$$E_{0,n} = \{\lambda_1 + \delta < \min\{\rho(\ell,\gamma), \rho_n, \hat{\ell}\}, \mathsf{F}_n(m_2) - \mathsf{F}_n(m_1)^2 > \gamma^2/8\} \tag{3.12}$$

is of probability $1-O(\exp(-cn^{1/2}))$ for a positive c depending only on γ, ℓ, δ. Therefore, $\mathbb{P}\left(R_n \leq x\right)-\mathbb{P}\left(E_{0,n} \cap \{R_n \leq x\}\right) = O(\exp(-cn^{1/2}))$ uniformly in $x \in \mathbb{R}$, i.e. it suffices to do the analysis on $E_{0,n}$. Then, for notational convenience, let $\mathbb{E}_n\left[X\right] := \mathbb{E}\left[I(E_{0,n})X\right]$ and $\mathbb{P}_n\left(E\right) := \mathbb{P}\left(E_{0,n} \cap E\right)$ for any random variable X and event E.

Step 2 Using (3.11), *linearize* the event $\{R_n \leq x\}$ as

$$\begin{aligned}\{R_n \leq x\} &= \{\rho_n(S_n(g_n) + n^{-1/2}G_n(g_n)) \leq (2\rho_n\sigma_n^{-2} - n^{-1/2}S_n(g_nh_n) + \delta_n)\sigma_n x\} \\ &= \{M_n - \delta_n x_n \leq 2\sigma_n^{-1}x\} \qquad (3.13)\end{aligned}$$

where $x_n = \rho_n^{-1}\sigma_n x$ and M_n, the main linearized statistic, is defined as

$$M_n := S_n((1 + n^{-1/2}x_n h_n)g_n) + n^{-1/2}G_n(g_n). \qquad (3.14)$$

Step 3 Use the *Edgeworth expansion for sums of independent random variables* to expand $\mathbb{P}\left(M_n \leq 2\sigma_n^{-1}x \mid \Lambda\right)$ on $E_{0,n}$ up to the accuracy of $o(n^{-1/2})$ uniformly in $x \in \mathbb{R}$. Then take its expectation over Λ to obtain the corresponding expansion of $\mathbb{P}_n\left(M_n \leq 2\sigma_n^{-1}x\right)$.

Step 4 Apply the *delta method for Edgeworth expansion* to obtain

$$\mathbb{P}_n\left(R_n \leq x\right) = \mathbb{P}_n\left(M_n \leq 2\sigma_n^{-1}x\right) + o(n^{-1/2}) \qquad (3.15)$$

uniformly on $x \in \mathbb{R}$.

3.2. Bai-Silverstein CLT

As a core component of our analysis, a particular case of the CLT for linear spectral statistics from Bai and Silverstein (2004) is introduced.

Theorem 3. *Suppose that* $Z_n := [z_1 \cdots z_n]$ *with* $z_1, \cdots, z_n \overset{i.i.d.}{\sim} N(0, I_p)$ *and* $\gamma_n := p/n \to \gamma \in \mathbb{R}^+$ *as* $n \to \infty$. *As defined above, let* $F_n(x)$ *and* $F_{\gamma_n}(x)$ *be the empirical spectral distribution* *of* $Z_nZ_n^t/p$ *and the Marchenko-Pastur distribution with the parameter* γ_n *respectively, and* $G_n(x) := p(F_n(x) - F_{\gamma_n}(x))$. *Then, for any real function* f *analytic on an open interval containing* $I(\gamma) := [I(\gamma \in (0,1))a(\gamma), b(\gamma)]$,

$$G_n(f) \xrightarrow{d} N(\mu(f), \sigma^2(f)),$$

where $\mu(f)$ *and* $\sigma^2(f)$ *are finite values determined by* $\{f(x) \mid x \in I(\gamma)\}$. *In particular,* $\mu(f)$ *is given by ((5.13) of Bai and Silverstein (2004))*

$$\mu(f) = \frac{f(a(\gamma)) + f(b(\gamma))}{4} - \frac{1}{2\pi}\int_{a(\gamma)}^{b(\gamma)} \frac{f(x)}{\sqrt{4\gamma - (x-1-\gamma)^2}}dx.$$

It is clear that Bai-Silverstein CLT is applicable for $g(\lambda) := (\rho(\ell,\gamma) - \lambda)^{-1}$, because $\rho(\ell,\gamma) - b(\gamma) = \eta(\ell,\gamma) > 0$.

3.3. Tail bounds

We introduce tail bounds in this section in order to establish **Step 1**, i.e. to separate λ_1 from $\min\{\rho(\ell,\gamma), \rho_n, \hat{\ell}\}$, and $\mathsf{F}_n(m_2)$ from $\mathsf{F}_n(m_1)^2$, with overwhelming probability. All proofs are postponed to the section **S2**.

We start with λ_1 and $\min\{\rho(\ell,\gamma), \rho_n\}$. Note that $\min\{\rho(\ell,\gamma), \rho_n\} - b(\gamma) > \delta$ for some positive δ and all large enough n, so the following proposition is sufficient.

Proposition 4 (Proposition 1 of Paul (2007))**.** *For each* $\delta \in (0, b(\gamma)/2)$, *the event* $E_{1,n} := \{\lambda_1 > b(\gamma) + \delta\}$ *satisfies*

$$\mathbb{P}(E_{1,n}) \le \exp(-3n\delta^2/(64b(\gamma)))$$

for all $n > n_\delta$, *where* $n_\delta \in \mathbb{N}$ *is determined by* δ *and* $\{\gamma_n\}_{n\in\mathbb{N}}$.

Now assume $\delta \in (0, \min(\eta(\ell,\gamma)/3, b(\gamma)/2))$ and choose $n_0(\delta) \in \mathbb{N}$ such that $|\rho_n - \rho(\ell,\gamma)| < \delta$ for all $n > n_0(\delta)$. Then, on $E_{1,n}^c$

$$\lambda_1 + \delta \le b(\gamma) + 2\delta < \rho(\ell,\gamma) - \delta < \min\{\rho(\ell,\gamma), \rho_n\}$$

for all $n > n_0(\delta)$, as desired.

The next 2 propositions are to restrict $|\hat{\ell} - \rho_n|$ on $E_{1,n}^c$, resulting in separation between λ_1 and $\min\{\rho(\ell,\gamma), \rho_n, \hat{\ell}\}$. Observe that

$$\hat{\ell} = \sup_{v \in \mathbb{S}^{p-1}} \|Sv\|_2 > \sup_{w \in \mathbb{S}^{p-2}} \|S^{[2:(p+1),2:(p+1)]}w\|_2 = \lambda_1$$

whenever $\hat{v}_1 \neq 0$, hence $z'\Lambda R(\hat{\ell})R(\rho_n)z \geq 0$ almost surely on $E_{1,n}^c$. This leads to

$$|l\rho_n(S_n(g_n) + n^{-1/2}G_n(g_n))| = (1 + ln^{-1}z'\Lambda R(\hat{\ell})R(\rho_n)z)|n^{1/2}(\hat{\ell} - \rho_n)| \geq |n^{1/2}(\hat{\ell} - \rho_n)| \quad (3.16)$$

almost surely on $E_{1,n}^c$, from (3.3). Therefore, it suffices to find tail bounds for $S_n(g_n)$ and $G_n(g_n)$ on $E_{1,n}^c$. We introduce propositions for more general settings, which will be necessary in the delta method for Edgeworth expansion section.

Proposition 5. *For $M > 0$ and a function f absolutely bounded by U_f on $[0, b(\gamma) + \delta]$, $E_{2,n}(f, M) := \{|S_n(f)| > M\}$ satisfies*

$$\mathbb{P}(E_{1,n}^c \cap E_{2,n}(f, M)) \leq 15 \exp(-M/U_f).$$

Proposition 6. *For functions $\{f_n\}_{n\in\mathbb{N}}$ such that (i) $f_n(x^2), n \in \mathbb{N}$ share a Lipschitz constant L on $[0, (b(\gamma) + \delta)^{1/2}]$ (as functions of x) and (ii) $\{G_n(f_n)\}_{n\in\mathbb{N}}$ is uniformly tight, then*

$$M(\{f_n\}_{n\in\mathbb{N}}) := \sup_{n\in\mathbb{N}} |\mathbb{E}\left[G_n(\mathfrak{f}_n)\right]| \quad \textit{with} \quad \mathfrak{f}_n(\lambda) := f_n((\lambda \vee 0) \wedge (b(\gamma) + \delta)) \quad (3.17)$$

is finite. Furthermore, for $M > 2M(\{f_n\}_{n\in\mathbb{N}})$, $E_{3,n}(f_n, M) := \{|G_n(f_n)| > M\}$ satisfies

$$\mathbb{P}(E_{1,n}^c \cap E_{3,n}(f_n, M)) \leq 2 \exp(-M^2/(8L^2))).$$

Proposition 5 immediately follows from the Markov inequality for moment generating functions, while Proposition 6 is mainly based on Corollary 1.8 (b) of Guionnet and Zeitouni (2000).

To apply Proposition 6, assumptions (i) and (ii) need to be established for all sufficiently large n ; (i) is true when f_n' exists and is uniformly bounded on $[0, b(\gamma)+\delta]$ because $(f_n(x^2))' = 2xf_n'(x^2)$. For (ii), the following lemma provides a sufficient condition.

Lemma 7. *In the setting of Theorem 3, suppose there is an open neighborhood $\Omega \subset \mathbb{C}$ of $I(\gamma)$ such that (i) $\{f_n\}_{n\in\mathbb{N}}$ is analytic and locally bounded in Ω and (ii) $f_n \to f$ pointwise on $I(\gamma)$. Then*

$$G_n(f_n) - G_n(f) \xrightarrow{p} 0$$

as $n \to \infty$. In particular, $G_n(f_n)$ has the same limiting Gaussian distribution as $G_n(f)$.

The proof relies on and adapts parts of the proof of Bai and Silverstein (2004) Theorem 1.1, along with the Vitali-Porter and Weierstrass theorems(e.g. Schiff (2013, Ch. 1.4, 2.4)). This lemma is sufficient for the uniform tightness required for (ii) of Proposition 6, because of Slutsky's theorem and Prohorov's Theorem(e.g. Van der Vaart (2000) Theorem 2.4). Consequently, we obtain the following corollary.

Corollary 8. *For functions $\{f_n\}_{n\in\mathbb{N}}$, assume that for $n' \in \mathbb{N}$ (i) $\{f_n'\}_{n>n'}$ is uniformly bounded by L' on $[0, b(\gamma)+\delta]$, (ii) $\{f_n\}_{n>n'}$ is analytic and locally bounded in an open neighborhood $\Omega \subset \mathbb{C}$ of $[a(\gamma), (1+\sqrt{\gamma})^2]$ and (iii) $f_n \to f$ pointwise on $[a(\gamma), (1+\sqrt{\gamma})^2]$. Then $G_n(f_n) \xrightarrow{d} N(\mu(f), \sigma^2(f))$ and*

$$\mathbb{P}(E_{1,n}^c \cap E_{3,n}(f_n, M)) \leq 2\exp(-M^2/(32(b(\gamma)+\delta)L'^2))$$

for $M > 2M(\{f_n\}_{n>n'})$ and all $n > n'$.

Now it is easy to see that $\{g_n\}_{n>n'}$ satisfies sufficient conditions for Proposition 5 and Corollary 8 for $U_f = \delta^{-1}$, $n' = n_0(\delta)$ and $L' = \delta^{-2}$, from $|g_n(\lambda)| \leq (\rho_n - b(\gamma) - \delta)^{-1} < \delta^{-1}$ for all $\lambda \in [0, b(\gamma)+\delta]$ and $n > n_0(\delta)$. Hence, (3.16) gives

Corollary 9. *For any $\delta \in (0, \min(\eta(\ell,\gamma)/3, b(\gamma)/2))$ and $M > 0$,*

$$\mathbb{P}(E_{1,n}^c \cap \{n^{1/2}|\hat{\ell} - \rho_n| > M\}) = O(\exp(-c(\gamma,\ell,\delta)M))$$

for a constant $c(\gamma,\ell,\delta)$ depending only on γ,ℓ,δ.

Finally, we verify **Step 1** as follows : let $\delta \in (0, \min(\eta(\ell,\gamma)/3, \gamma/2))$ and take $\epsilon > 0$ such that $\epsilon^2 + 3\epsilon < \gamma^2/8$. Then, if $\max(|G_n(m_2)|, |G_n(m_1)|) \le n\epsilon$ for $n > n_0(\delta)$,

$$\mathsf{F}_n(m_2) - \mathsf{F}_n(m_1)^2 \ge \mathsf{F}_{\gamma_n}(m_2) - \epsilon - (\mathsf{F}_{\gamma_n}(m_1) + \epsilon)^2 = \gamma_n^2 - (\epsilon^2 + 3\epsilon) > (\gamma - \delta)^2 - \gamma^2/8 > \gamma^2/8$$

since $\mathsf{F}_{\gamma_n}(m_1) = 1, \mathsf{F}_{\gamma_n}(m_2) = 1 + \gamma_n^2$ from Yao, Zheng and Bai (2015, Proposition. 2.13), and $\delta > |\rho_n - \rho(\ell,\gamma)| = \ell|\gamma_n - \gamma|/(\ell - 1) \ge |\gamma_n - \gamma|$. Therefore, $E_{1,n}^c \cap \{|\hat{\ell} - \rho_n| \le \delta\} \cap E_{3,n}^c(m_1, n\epsilon) \cap E_{3,n}^c(m_2, n\epsilon) \subset E_{0,n}$ from (3.12), i.e. **Step 1** is established by Proposition 4, Proposition 6 and Corollary 9.

Last but not least, we have the following corollary for moments for the future use, from Corollary 8 and Theorem 2.20 of Van der Vaart (2000).

Corollary 10. *For functions $\{f_n\}_{n\in\mathbb{N}}$ and f satisfying the conditions for Corollary 8 and any sequence of measurable E_n such that $E_n \subset E_{1,n}^c$ and $\lim_{n\to\infty} \mathbb{P}(E_n) = 1$,*

$$\lim_{n\to\infty} \mathbb{E}\left[I(E_n)(G_n(f_n))^k\right] = \tau_k(f), \forall k \in \mathbb{N},$$

where $\tau_k(f)$ denotes the k^{th} moment of $N(\mu(f), \sigma^2(f))$. In particular, since $\{g_n\}_{n\in\mathbb{N}}, g$ and $\{E_{0,n}\}_{n\in\mathbb{N}}$ satisfy these sufficient conditions, $\lim_{n\to\infty} \mathbb{E}_n\left[(G_n(g_n))^k\right] = \tau_k(g)$ holds.

3.4. Edgeworth expansion for sums of independent random variables

A heuristic conversion between characteristic function and Edgeworth expansion is described in Hall (1992, pg. 48). Justification for the conversion is the main subject of Chapter

VI of Petrov (1975), and leads to his Theorem 7, which we state in modified form in Theorem 11 below. For us it yields an expression of $\mathbb{P}\left(M_n \leq x \mid \Lambda\right)$ up to the accuracy of $o(n^{-1/2})$.

For clarity, we first define relevant notations. Let $(X_{ni})_{n\in\mathbb{N},i\in\{1,\cdots n\}}$ be a triangular array of random variables with zero means and finite variances, and assume that $X_{n1},\cdots,X_{nn}$ are independent for all $n\in\mathbb{N}$. Furthermore,

- $\overline{V}_n := n^{-1}\sum_{i=1}^n \mathrm{Var}\left[X_{ni}\right]$ is positive for all sufficiently large n.
- $\bar{\chi}_{v,n}$ is the average v^{th} cumulant of $\overline{V}_n^{-1/2}X_{ni}$'s, for $v\in\mathbb{N}$.
- $C_n(t) := \mathbb{E}\left[\exp(it\overline{V}_n^{-1/2}\sum_{j=1}^n X_{ni})\right]$.
- For $v\in\mathbb{N}$,

$$Q_{vn}(x) := \sum_{w=1}^{v}\frac{1}{w!}\left(\sum_{*(w,v)}\prod_{k=1}^{w}\frac{\bar{\chi}_{j_k+2,n}}{(j_k+2)!}\right)(-1)^{v+2w}\frac{d^{v+2w}}{dx^{v+2w}}\Phi(x),$$

  where the summation $*(w,v)$ is over $\{(j_1,\cdots,j_w)\in\mathbb{N}^w \mid j_1+\cdots+j_w=v\}$.

One verifies that $Q_{vn}(x)$ is a product of $\phi(x)$ and a degree-$(3v-1)$ polynomial of x with coefficients being polynomials of $\bar{\chi}_{j,n}, j\in\{3,\cdots,v+2\}$. Further, Q_{vn} is even for odd v and odd for even v.

Theorem 11. *For fixed $k\geq 3$, $l\geq 0$ and for $(X_{ni})_{n\in\mathbb{N},i\in\{1,\cdots n\}}$, assume that there exist $r_1(k), r_2(n;k,\tau), r_3(n;k,l,\epsilon)$ satisfying the following regularity conditions :*

R1 *For all sufficiently large $n\in\mathbb{N}$,*

$$n^{-1}\overline{V}_n^{-k/2}\sum_{i=1}^n \mathbb{E}\left[|X_{ni}|^k\right] \leq r_1(k) < \infty.$$

R2 *For some $\tau\in(0,1/2)$,*

$$n^{-1}\overline{V}_n^{-k/2}\sum_{i=1}^n \mathbb{E}\left[I(\overline{V}_n^{-1/2}|X_{ni}|>n^\tau)|X_{ni}|^k\right] \leq r_2(n;k,\tau) = o(1).$$

R3 *A generalized Cramer's condition*

$$n^{(k+l-2)/2}\int_{|t|>\epsilon}|t|^{l-1}|C_n(t)|dt \le r_3(n;k,l,\epsilon) = o(1)$$

holds for some $\epsilon \in (0, 3/(4H_3))$ *and all* $n > n_3(k,l,\epsilon)$, *where* $H_3 := r_1(k)^{3/k} < \infty$ *is an upper bound of the average third absolute moments(by power mean inequality).*

Then, there exists $N = N(k,l,\tau,\epsilon,n_3)$ *such that for* $n > N$, *the inequality*

$$\left|\frac{d^l}{dx^l}\mathbb{P}(n^{-1/2}\overline{V}_n^{-1/2}\sum_{i=1}^n X_{ni} \le x) - \frac{d^l}{dx^l}(\Phi(x) + \sum_{v=1}^{k-2} n^{-v/2}Q_{vn}(x))\right| \le n^{-(k-2)/2}\delta(n)$$

holds for all $x \in \mathbb{R}$. *Here* $\delta(n) = o(1)$ *depends only on* $n, k, l, \tau, \epsilon, r_1(k), r_2(n;k,\tau)$ *and* $r_3(n;k,l,\epsilon)$.

Our reason for presenting this theorem along with the explicit dependence of the constants is that it provides a uniform bound on the (derivatives of) difference between the distribution function and corresponding Edgeworth expansion for all sufficiently large n. Also, we briefly comment on the regularity conditions : **R1** is about boundedness of $\bar{\chi}_{v,n}, v = 3, \cdots, k$, while **R2**, **R3** are related to tail behavior ; in particular, **R2** resembles the Lindeberg condition for the CLT.

Back to our problem, we state a special case of Theorem 11 when $k = 3$ and $l = 0$.

Corollary 12. *For* $(X_{ni})_{n\in\mathbb{N}, i\in\{1,\cdots n\}}$ *satisfying* ***R1***, ***R2*** *and* ***R3*** *for* $k = 3$ *and* $l = 0$,

$$\mathbb{P}(n^{-1/2}\overline{V}_n^{-1/2}\sum_{i=1}^n X_{ni} \le x) = \Phi(x) + n^{-1/2}\bar{\chi}_{3,n}(1-x^2)\phi(x)/6 + o(n^{-1/2}),$$

uniformly in $x \in \mathbb{R}$.

Now from (2.3) and (2.4), observe that conditioned on Λ, $S_n((1+n^{-1/2}x_nh_n)g_n)$ is a sum of independent random variables. That is, Corollary 12 is applicable for $X_{ni} = c_{ni}(z_i^2 - 1)$

where $c_{ni} := (1 + n^{-1/2}x_n h_n(\lambda_i))g_n(\lambda_i)$, so long as the corresponding regularity conditions **R1**, **R2** and **R3** hold. In the moments analysis below, we show that this is the case on $E_{0,n}$ with the *same* $r_1(k), r_2(n; k, \tau), r_3(n; k, l, \epsilon)$, and $n_3(k, l, \epsilon)$.

Moments analysis. Note that $(z_i^2 - 1)$ are mean zero i.i.d. with the characteristic function $\exp(-i\theta)(1-2i\theta)^{-1/2}$, and so the k^{th} cumulant is $\kappa_k = 2^{k-1}(k-1)!$ for $k \in \mathbb{N}$. In particular, adopting the notations above, we have

$$\overline{V}_n = 2n^{-1}\sum_{i=1}^n c_{ni}^2, \quad \bar{\chi}_{k,n} = \kappa_k \overline{V}_n^{-k/2} n^{-1}\sum_{i=1}^n c_{ni}^k, \quad |C_n(t)| = \prod_{i=1}^n (1 + 4\overline{V}_n^{-1} c_{ni}^2 t^2)^{-1/4}.$$

We will show that there exists a positive C such that

$$C \max_{i=1,\cdots,n} c_{ni}^2 \leq \overline{V}_n \tag{3.18}$$

for all $x \in \mathbb{R}$ on $E_{0,n}$, for all sufficiently large n. Note that c_{ni} depends on x. Let us assume (3.18) for now and verify that **R1**, **R2** and **R3** hold uniformly in $x \in \mathbb{R}$ on $E_{0,n}$. First,

$$n^{-1}\overline{V}_n^{-k/2}\sum_{j=1}^n \mathbb{E}\left[|X_{nj}|^k\right] = \overline{V}_n^{-k/2} n^{-1}\sum_{i=1}^n |c_{ni}|^k \mathbb{E}\left[|z_1^2 - 1|^k\right] \leq C^{-k/2}\mathbb{E}\left[|z_1^2-1|^k\right],$$

hence **R1** holds with $r_1(k) = C^{-k/2}\mathbb{E}\left[|z_1^2 - 1|^k\right]$ for all $k \in \mathbb{N}$. Now use the Markov inequalities and then **R1** to get

$$n^{-1}\overline{V}_n^{-k/2}\sum_{i=1}^n \mathbb{E}\left[I(\overline{V}_n^{-1/2}|X_{ni}| > n^\tau)|X_{ni}|^k\right] \leq n^{-\tau-1}\overline{V}_n^{-(k+1)/2}\sum_{i=1}^n \mathbb{E}\left[|X_{ni}|^{k+1}\right] \leq n^{-\tau} r_1(k+1),$$

which shows that **R2** holds with $r_2(n; k, \tau) = n^{-\tau} r_1(k+1)$ for any $\tau \in (0, 1/2)$ and $k \in \mathbb{N}$.

For any $m \in \{1, \cdots, n\}$, define $s_m := \sum_{1 \leq i_1 < \cdots < i_m \leq n} \prod_{j=1}^m c_{ni_j}^2$ and $n_m := n^m - n!/(n-m)!$. We then have

$$\begin{aligned}(n\overline{V}_n/2)^m = (\sum_{i=1}^n c_{ni}^2)^m &= \sum_{1 \leq i_1, \cdots, i_m \leq n} \prod_{j=1}^m c_{ni_j}^2 \\ &\leq n_m \max_{i=1,\cdots,n} c_{ni}^{2m} + m! s_m \leq C^{-m} n_m \overline{V}_n^m + m! s_m,\end{aligned}$$

so that $(2\overline{V}_n^{-1})^m s_m \geq (n^m - (2C^{-1})^m n_m)/m!$. Hence

$$\prod_{i=1}^{n}(1 + 4\overline{V}_n^{-1}c_{ni}^2 t^2) \geq (4\overline{V}_n^{-1}t^2)^m s_m \geq (2nt^2)^m(1 - (2C^{-1})^m n_m/n^m)/m!.$$

Now $\lim_{n\to\infty} n_m/n^m = 0$ for any fixed $m \in \mathbb{N}$, so, with $m = 4(k+l)$, it follows that $|C_n(t)| \leq 2(m!)^{1/4}(2nt^2)^{-(k+l)}$ for all $n > n_3(k,l,\epsilon)$. This implies **R3** with $r_3(n;k,l,\epsilon) = 2^{-(k+l-2)}(4(k+l)!)^{1/4}n^{-(k+l+2)/2}\epsilon^{-(2k+l)}/(2k+l)$ for any $\epsilon \in (0, 3/(4H_3))$ and $k \geq 3, l \geq 0$.

Proof of (3.18). Throughout the proof, $n > n_0(\delta)$ and $\Lambda \in E_{0,n}$ are assumed, so that $\lambda_i \in [0,\rho)$, $g_n(\lambda_i) = (\rho_n - \lambda_i)^{-1} \in [\rho_n^{-1}, \delta^{-1}]$ and $|h_n(\lambda_i)| = |r_n - \lambda_i g_n(\lambda_i)| \leq \max(r_n, \rho\delta^{-1})$. Consequently,

$$|c_{ni}| = |1 + n^{-1/2}x_n h_n(\lambda_i)|g_n(\lambda_i) \leq \delta^{-1}(1 + \max(r_n, \rho\delta^{-1})|n^{-1/2}x_n|),$$

so that $\max_{i=1,\cdots,n} c_{ni}^2 \leq C_1(1 + C_2|n^{-1/2}x_n|)^2$ for positive constants C_1, C_2 independent of n and x. Therefore, it suffices to show that there exists a positive ϵ such that

$$\epsilon(1 + C_2|n^{-1/2}x_n|)^2 \leq \overline{V}_n/2, \tag{3.19}$$

for all $x_n \in \mathbb{R}$. Let $v_k = \mathsf{F}_n(g_n^2 h_n^k)$ for $k = 0, 1, 2$, and then write $\overline{V}_n/2 = v_2(n^{-1/2}x_n)^2 + 2v_1(n^{-1/2}x_n) + v_0$. Hence (3.19) is equivalent to

$$2(C_2\epsilon - v_1\text{sign}(x_n))|n^{-1/2}x_n| \leq (v_2 - \epsilon C_2^2)(n^{-1/2}x_n)^2 + (v_0 - \epsilon)$$

for all $x_n \in \mathbb{R}$. In view of the AM-GM inequality and its equality condition, this is equivalent to $0 \leq (v_0 - \epsilon), (v_2 - \epsilon C_2^2)$ and $(v_1 + C_2\epsilon)^2 \leq (v_2 - C_2^2\epsilon)(v_0 - \epsilon)$. But then the first and the third inequalities yield the second, so the desired condition is

$$\epsilon \in (0, \min(v_0, (v_2 v_0 - v_1^2)(v_0 C_2^2 + 2v_1 C_2 + v_2)^{-1}))).$$

This is true when

$$v_2 v_0 - v_1^2 \geq C_4 \tag{3.20}$$

for a positive C_4, because $v_0 \geq 1$, $v_0 C_2^2 + 2v_1 C_2 + v_2 = v_0(C_2 + v_1/v_0)^2 + (v_0 v_2 - v_1^2)/v_0$ is positive when (3.20) holds, and bounded above on $E_{0,n}$. Finally, since $(\sum a_i^2)(\sum b_i^2) - (\sum a_i b_i)^2 = \sum_{i<j}(a_i b_j - a_j b_i)^2$ and $h_n(\lambda') - h_n(\lambda) = \lambda g_n(\lambda) - \lambda' g_n(\lambda') = \rho_n g_n(\lambda) g_n(\lambda')(\lambda - \lambda')$, we have

$$\begin{aligned} v_2 v_0 - v_1^2 &= \mathsf{F}_n(g_n^2 h_n^2)\mathsf{F}_n(g_n^2) - \mathsf{F}_n(g_n^2 h_n)^2 = n^{-2} \sum_{1\leq i<j\leq n} (g_n(\lambda_i) g_n(\lambda_j))^2 \left(h_n(\lambda_i) - h_n(\lambda_j)\right)^2 \\ &= \rho_n^2 n^{-2} \sum_{1\leq i<j\leq n} (g_n(\lambda_i) g_n(\lambda_j))^4 (\lambda_i - \lambda_j)^2 \\ &\geq \rho_n^{-6} n^{-2} \sum_{1\leq i<j\leq n} (\lambda_i - \lambda_j)^2 = \rho_n^{-6}(\mathsf{F}_n(m_2) - \mathsf{F}_n(m_1)^2) \geq (\rho+\gamma)^{-6}\gamma^2/8, \end{aligned}$$

so we have shown (3.18), and consequently the claim. □

First order Edgeworth expansion for M_n. From Corollary 12 and (3.14), we have

$$\mathbb{E}_n \left[\mathbb{P}\left(M_n \leq 2\sigma_n^{-1} x \mid \Lambda \right) - (\Phi(y_n) + n^{-1/2} \overline{V}_n^{-3/2} \bar{\kappa}_{3,n} (1 - y_n^2)\phi(y_n)/6) \right] = o(n^{-1/2})$$

uniformly in $x \in \mathbb{R}$, where $y_n := \overline{V}_n^{-1/2}(2\sigma_n^{-1} x - n^{-1/2} G_n(g_n))$ and $\bar{\kappa}_{3,n} = 8n^{-1} \sum_{i=1}^n c_{ni}^3$. It then suffices to show that

$$\begin{aligned} &\mathbb{E}_n \left[\Phi(y_n) + n^{-1/2} \overline{V}_n^{-3/2} \bar{\kappa}_{3,n} (1 - y_n^2)\phi(y_n)/6 \right] \\ = \;& \Phi(x) + n^{-1/2}(\kappa_{2,n}^{-3/2} \kappa_{3,n}(1 - x^2)/6 - \kappa_{2,n}^{-1/2} \mu(g_n))\phi(x) + o(n^{-1/2}), \end{aligned} \tag{3.21}$$

uniformly in $x \in \mathbb{R}$. To this end, we introduce the following notation.

Definition 1. For $\alpha > 0$ and a polynomial $p_n(t) = \sum_{i=0}^k c_{ni} t^i$ with random coefficients c_{ni}'s, p_n is $PO(n^{-\alpha}; E_{0,n})$ if $\mathbb{E}_n [|c_{ni}|] = O(n^{-\alpha}), i = 0, \cdots, k$.

With this definition, we will show that

$$\overline{V}_n - \kappa_{2,n} = PO(n^{-1}; E_{0,n}), \quad \bar{\kappa}_{3,n} - \kappa_{3,n} = PO(n^{-1/2}; E_{0,n}), \tag{3.22}$$

when both are treated as polynomials of $x_n = \rho_n^{-1}\sigma_n x$. To prove the first part, observe that

$$\begin{aligned}\overline{V}_n - \kappa_{2,n} &= 2(v_2(n^{-1/2}x_n)^2 + 2v_1(n^{-1/2}x_n) + v_0 - \mathsf{F}_{\gamma_n}(g_n^2))\\ &= 2n^{-1}(v_2x_n^2 + 2n^{-1/2}G_n(g_n^2h_n)x_n + G_n(g_n^2)),\end{aligned}$$

where the second equality uses $v_1 - n^{-1}G_n(g_n^2h_n) = \mathsf{F}_{\gamma_n}(g_n^2h_n) = r_n\mathsf{F}_{\gamma_n}(g_n^2) - \mathsf{F}_{\gamma_n}(m_1g_n^3) = 0$, from (3.8). Also, it is clear that $g_n^2h_n$ and g_n^2 satisfy the sufficient conditions for Corollary 8, hence Corollary 10 implies that $(\overline{V}_n - \kappa_{2,n})$ is $PO(n^{-1}; E_{0,n})$. The second part of (3.22) can be proved in a similar yet simpler way ; namely,

$$\bar{\kappa}_{3,n} - \kappa_{3,n} = 8n^{-1/2}(n^{-1}u_3x_n^3 + 3n^{-1/2}u_2x_n^2 + 3u_1x_n + n^{-1/2}G_n(g_n^3)),$$

where $u_k = \mathsf{F}_n(g_n^3h_n^k), k = 1, 2, 3$. These are also absolutely bounded on $E_{0,n}$.

To exploit (3.22), we introduce a trivial inequality and its consequence as follows.

Proposition 13. *For any univariate polynomial p(with deterministic coefficients) and a positive s, there exists a constant $C(p, s)$ such that $|p(t)\exp(-st^2)| \le C(p, s)$ for all $t \in \mathbb{R}$.*

Corollary 14. *If p_n is $PO(n^{-\alpha}; E_{0,n})$ for some $\alpha > 0$, then for any positive s,*

$$\sup_{t\in\mathbb{R}} |\mathbb{E}_n\left[p_n(t)\exp(-st^2)\right]| = O(n^{-\alpha}).$$

Now we show

$$\mathbb{E}_n\left[\Phi(y_n) - \Phi(x) + n^{-1/2}\kappa_{2,n}^{-1/2}\mu(g_n)\phi(x)\right] = o(n^{-1/2}), \tag{3.23}$$

$$\mathbb{E}_n\left[\overline{V}_n^{-3/2}\bar{\kappa}_{3,n}(1 - y_n^2)\phi(y_n) - \kappa_{2,n}^{-3/2}\kappa_{3,n}(1 - x^2)\phi(x)\right] = O(n^{-1/2}) \tag{3.24}$$

uniformly in $x \in \mathbb{R}$, which implies (3.21) along with Proposition 13 and the tail bound on $E_{0,n}$. These are fairly easy to prove on any compact subset of $\mathbb{R}$, but for uniform convergence, the proof is more delicate, due to the dependence of $\overline{V}_n$ and $\bar{\kappa}_{3,n}$ on x. Although a wide interval of x would be practically meaningful, we prove uniform convergence here.

Proof of (3.23) *and* (3.24). Observe that on $E_{0,n}$, $\overline{V}_n$ and $\kappa_{2,n}$ are bounded below by a positive constant uniformly in $x \in \mathbb{R}$, in view of $\overline{V}_n \geq v_2^{-1}(v_0 v_2 - v_1^2)$ and (3.20). On the other hand, by the AM-GM inequality and (3.20), we have the upper bound

$$\overline{V}_n \leq 4(n^{-1} v_2 x_n^2 + v_0). \tag{3.25}$$

Now we can prove (3.23) as follows : let $\alpha_n = \overline{V}_n^{-1/2} \kappa_{2,n}^{1/2}$, then it suffices to show that

$$\mathbb{E}_n \left[\Phi(y_n) - \Phi(\alpha_n x) + n^{-1/2} \overline{V}_n^{-1/2} G_n(g_n) \phi(\alpha_n x) \right], \tag{3.26}$$

$$\mathbb{E}_n \left[\Phi(\alpha_n x) - \Phi(x) \right], \tag{3.27}$$

$$\mathbb{E}_n \left[n^{-1/2} \overline{V}_n^{-1/2} G_n(g_n)(\phi(\alpha_n x) - \phi(x)) \right], \quad \text{and} \tag{3.28}$$

$$\mathbb{E}_n \left[n^{-1/2} \kappa_{2,n}^{-1/2} (\alpha_n - 1) G_n(g_n) \phi(x) \right] \tag{3.29}$$

are $O(n^{-1})$ uniformly in $x \in \mathbb{R}$, because $\mathbb{E}_n [G_n(g_n) - \mu(g_n)] = o(1)$ from Corollary 10. From the second order Taylor expansion of $\Phi(y_n)$ centered at $\alpha_n x$ and using Proposition 13, (3.26) is $O(n^{-1} \mathbb{E}_n [G_n(g_n)^2])$, and hence $O(n^{-1})$ uniformly in $x \in \mathbb{R}$, by Corollary 10. Next, for (3.27) and (3.28), we consider two cases :

(case 1) $x^2 \leq n$: This assumption implies that $\overline{V}_n$ is bounded above by a positive constant on $E_{0,n}$, by (3.25). Therefore, on $E_{0,n}$, α_n is bounded below by a positive α_0, and thus $\exp(-st^2) \leq \exp(-s\beta_0^2 x^2)$ for all t between x and $\alpha_n x$ and for all positive s, where $\beta_0 = \min(\alpha_0, 1)$. Using this fact, $|t| \exp(-t^2/2) \leq \exp(-t^2/4)$, and the first order Taylor expansions of $\Phi(\alpha_n x)$ and $\phi(\alpha_n x)$ centered at x, it follows that (3.27), (3.28) are

$$O(\mathbb{E}_n \left[|(\alpha_n - 1)x| \exp(-\beta_0^2 x^2/2) \right]), \quad O(n^{-1/2} \mathbb{E}_n \left[|G_n(g_n)(\alpha_n - 1)x| \exp(-\beta_0^2 x^2/4) \right]),$$

respectively. These are $O(n^{-1})$ uniformly in $x \in [-\sqrt{n}, \sqrt{n}]$, because of

$$\alpha_n - 1 = \overline{V}_n^{-1/2} (\kappa_{2,n} - \overline{V}_n)(\overline{V}_n^{1/2} + \kappa_{2,n}^{1/2})^{-1} = PO(n^{-1}; E_{0,n}), \tag{3.30}$$

Corollary 14 and the Cauchy-Schwarz inequality(for the second case).

(case 2) $x^2 > n$: In this case we have $\overline{V}_n = O(n^{-1}x^2)$ on $E_{0,n}$ from (3.25). Then $|\alpha_n x|^{-1} = O(n^{-1/2})$ on $E_{0,n}$ uniformly in $x \in [-\sqrt{n}, \sqrt{n}]^c$, and hence from $0 < 1 - \Phi(|t|) \leq \phi(|t|)/|t| = O(|t|^{-2})$, we conclude that $1 - \Phi(|x|)$, $1 - \Phi(|\alpha_n x|)$, $\phi(x)$, $\phi(\alpha_n x)$ are all $O(n^{-1})$ uniformly in $x \in [-\sqrt{n}, \sqrt{n}]^c$, and so the same is true for (3.27), (3.28).

Combining these cases gives the desired result for (3.27) and (3.28). Furthermore, (3.29) immediately follows from (3.30), Corollary 14 and the Cauchy-Schwarz inequality.

In a similar manner to the proof of (3.23) just given, we can decompose the RHS of (3.24) into

$$\mathbb{E}_n \left[\overline{V}_n^{-3/2}\bar{\kappa}_{3,n}((1 - z_n^2)\phi(z_n) - (1 - (\alpha_n x)^2)\phi(\alpha_n x))\right], \tag{3.31}$$

$$\mathbb{E}_n \left[\overline{V}_n^{-3/2}\bar{\kappa}_{3,n}((1 - (\alpha_n x)^2)\phi(\alpha_n x) - (1 - x^2)\phi(x))\right], \tag{3.32}$$

$$\mathbb{E}_n \left[\overline{V}_n^{-3/2}(\bar{\kappa}_{3,n} - \kappa_{3,n})(1 - x^2)\phi(x)\right], \tag{3.33}$$

$$\mathbb{E}_n \left[\kappa_{2,n}^{-3/2}(\alpha_n^3 - 1)\kappa_{3,n}(1 - x^2)\phi(x)\right], \tag{3.34}$$

which are to be shown to be $O(n^{-1/2})$ uniformly in $x \in \mathbb{R}$. From (3.19), $\overline{V}_n^{-3/2}|\bar{\kappa}_{3,n}|$ is bounded above uniformly in on $E_{0,n}$, which leads to the desired result for (3.31) and (3.32) by the same methods as for (3.26) and (3.28), with small changes in details ; the first order Taylor expansion suffices for (3.31), and case 2 for (3.32) requires $0 < (t^2 - 1)\phi(t) \leq 8t^{-2}$ if $t^2 > 1$. Finally, (3.22) and (3.30) give the desired properties for (3.33) and (3.34), respectively. $\square$

3.5. Delta method for Edgeworth expansion

In this section, we prove that $\delta_n x_n$ is ignorable in the sense of **Step 4**. The decomposition given in (3.13) is inspired by the discussion in Hall (1992, Chap. 2.7). The delta method is briefly introduced there as follows : for two statistics U_n and U_n' whose limiting distributions are $N(0,1)$, if $\Delta_n := U_n - U_n'$ is of order $O_p(n^{-j/2})$ for $j \in \mathbb{N}$, then "generally", $\mathbb{P}\left(U_n \leq x\right) - \mathbb{P}\left(U_n' \leq x\right)$ is of order $O(n^{-j/2})$. Therefore, if the $(j-1)^{th}$ order Edgeworth expansion for U_n is easy to calculate, so is for U_n'. However, neither sufficient conditions nor a rigorous proof for this method is given there. Furthermore, Δ_n is linear in x in our case. Hence, we prove a version of the delta method for Edgeworth expansion in our context.

Proposition 15. *Suppose that U_n admits the first order Edgeworth expansion*

$$\mathbb{P}_n\left(U_n \leq x\right) = \Phi(x) + n^{-1/2} p_1(x)\phi(x) + o(n^{-1/2})$$

uniformly in $x \in \mathbb{R}$, for a polynomial p_1. Also, assume that random variables J_n do not depend on x, and satisfy $\mathbb{P}_n(|J_n| > n^{-1/2}\epsilon_n) = o(n^{-1/2})$ for a non-random sequence $\{\epsilon_n\}$ converging to 0. Then

$$\mathbb{P}_n\left(U_n + xJ_n \leq x\right) - \mathbb{P}_n\left(U_n \leq x\right) = o(n^{-1/2})$$

uniformly in $x \in \mathbb{R}$.

Proof. Note that

$$|\mathbb{P}_n\left(U_n + xJ_n \leq x\right) - \mathbb{P}_n\left(U_n \leq x\right)| \leq \mathbb{P}_n(|J_n| > n^{-1/2}\epsilon_n) + \mathbb{P}_n(|U_n - x| \leq |x| n^{-1/2}\epsilon_n),$$

hence from the assumption $\mathbb{P}_n(|J_n| > n^{-1/2}\epsilon_n) = o(n^{-1/2})$ it suffices to show that

$$\mathbb{P}_n(|U_n - x| \leq |x| n^{-1/2}\epsilon_n) = o(n^{-1/2})$$

uniformly in $x \in \mathbb{R}$. This follows from the uniform convergence assumption on the first order Edgeworth expansion of U_n, and the following inequalities : for $y \in [-1/2, 1/2]$, by

Proposition 13,

$$|\Phi(x(1+y))-\Phi(x))| \le |xy| \max_{z\in[-1/2,1/2]} \phi(x(1+z)) \le |xy|\phi(x/2) = O(|y|),$$

$$\begin{aligned}|p_1(x(1+y))\phi(x(1+y))-p_1(x)\phi(x)| &\le |xy| \max_{z\in[-1/2,1/2]} |p_2(x(1+z))|\phi(x(1+z))\\ &\le |xy||p_2|(|3x/2|)\phi(x/2) = O(|y|).\end{aligned}$$

Here p_2 is the polynomial satisfying $\frac{d}{dx}(p_1(x)\phi(x)) = p_2(x)\phi(x)$, and $|p_2|$ is the polynomial with coefficients being the absolute values of coefficients of p_2. □

Finally, we prove (3.15) using this proposition with $U_n = \sigma_n M_n/2$, $J_n = \rho_n^{-1}\sigma_n^2\delta_n/2$ and $\epsilon_n \asymp n^{-\zeta}$ for any $\zeta \in (0,1/2)$. Recall the definition of δ_n from (3.10) : $\delta_n = n^{-1}G_n(\lambda g_n^2) - (\nu_n - n^{-1/2}r_n S_n(g_n))$. As $\mathbb{P}_n(|n^{-1}G_n(m_1 g_n^2)| > n^{-1/2-\zeta}) = o(n^{-1/2})$ by Proposition 6, we only need to consider $(\nu_n - n^{-1/2}r_n S_n(g_n))$. Observe that from (3.3) and (3.4),

$$(\hat{\ell}-\rho_n)(1+\ell n^{-1}z'\Lambda R^2(\rho_n)z) = n^{-1/2}\ell\rho_n[S_n(g_n)+n^{-1/2}G_n(g_n)] + (\hat{\ell}-\rho_n)\nu_n.$$

Multiply both sides by $-n^{-1}z'\Lambda R(\hat{\ell})R^2(\rho_n)z$ to yield

$$(1+\ell n^{-1}z'\Lambda R^2(\rho_n)z)\nu_n = -n^{-1/2}\ell\rho_n[S_n(g_n)+n^{-1/2}G_n(g_n)]\cdot n^{-1}z'\Lambda R(\hat{\ell})R^2(\rho_n)z + \nu_n^2,$$

because of (3.5). Consequently, on $E_{0,n}$ we have

$$\begin{aligned}|\nu_n - n^{-1/2}r_nS_n(g_n)| &\le (1+\ell n^{-1}z'\Lambda R^2(\rho_n)z)|\nu_n - n^{-1/2}r_nS_n(g_n)|\\ &\le n^{-1/2}\ell\rho_n|S_n(g_n)|\cdot|n^{-1}z'\Lambda R(\hat{\ell})R^2(\rho_n)z + (1+\ell n^{-1}z'\Lambda R^2(\rho_n)z)\tfrac{r_n}{\ell\rho_n}|\\ &\quad + n^{-1}\ell\rho_n|G_n(g_n)|\cdot|n^{-1}z'\Lambda R(\hat{\ell})R^2(\rho_n)z| + \nu_n^2.\end{aligned}$$

Furthermore, the following holds from (3.8), the resolvent identity and (3.2)

$$
\begin{aligned}
&n^{-1}z'\Lambda R(\hat{\ell})R^2(\rho_n)z + (1+\ell n^{-1}z'\Lambda R^2(\rho_n)z)\tfrac{r_n}{\ell\rho_n}\\
=&n^{-1}z'\Lambda R(\hat{\ell})R^2(\rho_n)z + (1+\ell n^{-1}z'\Lambda R^2(\rho_n)z)\mathsf{F}_{\gamma_n}(m_1 g_n^3)/(1+\ell\mathsf{F}_{\gamma_n}(m_1 g_n^2))\\
=&(\hat{\ell}-\rho_n)n^{-1}z'\Lambda R(\hat{\ell})R^3(\rho_n)z + n^{-1}z'\Lambda R^3(\rho_n)z + \mathsf{F}_{\gamma_n}(m_1 g_n^3) + (n^{-1}z'\Lambda R^2(\rho_n)z - \mathsf{F}_{\gamma_n}(m_1 g_n^2))\tfrac{r_n}{\rho_n}\\
=&(\hat{\ell}-\rho_n)n^{-1}z'\Lambda R(\hat{\ell})R^3(\rho_n)z + n^{-1/2}S_n(m_1 g_n^2(\tfrac{r_n}{\rho_n}-g_n)) + n^{-1}G_n(m_1 g_n^2(\tfrac{r_n}{\rho_n}-g_n)).
\end{aligned}
$$

Now considering that $\ell, \rho_n, r_n, \|\Lambda\|_\infty, \|R(\hat{\ell})\|_\infty, \|R(\rho_n)\|_\infty$ are absolutely bounded on $E_{0,n}$ for $n > n_0(\delta)$, and $\nu_n = -(\hat{\ell}-\rho_n)n^{-1}z'\Lambda R(\hat{\ell})R^2(\rho_n)z$ (3.5), it suffices to show that

$$
\mathbb{P}_n(|S_n(g_n)| > n^{1/4-\zeta/2}), \quad \mathbb{P}_n(|\hat{\ell}-\rho_n| > n^{-1/4-\zeta/2}), \quad \mathbb{P}_n(n^{-1}z'z > 2), \quad \mathbb{P}_n(|G_n(g_n)| > n^{1/2-\zeta})
$$

are of probability $o(n^{-1/2})$ for any $\zeta \in (0, 1/2)$. Each such bound can be easily deduced from Proposition 5, Proposition 6 and Corollary 9.

4. Discussion

This study clearly leaves some natural questions for further research. We considered a single supercritical spike; extension to a finite number of separated simple supercritical eigenvalues is presumably straightforward. Less immediately clear is the situation with a supercritical eigenvalue of multiplicity $K > 1$, as the limiting distribution for the associated K eigenvalues is $GOE(K)$ rather than ordinary Gaussian.

A common use of Edgeworth approximations is to improve the coverage properties of confidence intervals based on Gaussian limit theory. In ongoing work, we are exploring such improvements for one- and two-sided intervals for ℓ.

Development of a second order Edgeworth approximation (kurtosis correction) would appear to require a first order or skewness correction for certain linear statistics in the Bai-Silverstein central limit theorem, which is not yet available.

We assumed that the observations x_j were Gaussian and that assumption is used in an important way to create the i.i.d. variates $z = (z_i) = U'Z_1$, independent of the noise eigenvalues Λ, as input to the conditional Edgeworth expansion. Thus extension of the results to non Gaussian x_j is an open issue for future work.

Acknowledgements

We thank Zhou Fan for providing the main idea of the proofs of Proposition 6 and Corollary 12, and referees for helpful comments. Supported by NIH R01 EB001988, NSF DMS 1407813 and a Samsung scholarship.

References

Bai, Z. D. and Silverstein, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32(1A)**, 553-605.

Bai, Z. D. and Yao, J. (2005). On the convergence of the spectral empirical process of Wigner matrices. *Bernoulli*, **11(6)**, 1059-1092.

Baik, J., Arous, G. B., and Péché, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33(5)**, 1643-1697.

Benaych-Georges, F., Guionnet, A., and Maida, M. (2011). Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices. *Electron. J. Probab.*, **16(60)**, 1621-1662.

Guionnet, A. and Zeitouni, O. (2000). Concentration of the spectral measure for large matrices. *Electron. Comm. Probab.* **5**, 119-136.

Hall, P. (1992). *The bootstrap and Edgeworth expansion.* Springer Science & Business Media

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*

29, 295-327.

Muirhead, R. J. and Chikuse, Y. (1975). Asymptotic expansions for the joint and marginal distributions of the latent roots of the covariance matrix. *Ann. Statist.* **3**, 1011-1017.

Paul, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica* **17**, 1617-1642.

Petrov, V. V. (1975). *Sums of independent random variables.* Vol.**82.** Springer Science & Business Media

Schiff, J. L. (2013). *Normal families.* Springer Science & Business Media.

Van der Vaart, A. W. (2000). *Asymptotic statistics.* Vol.**3.** Cambridge University Press.

Wainwright, M. (2015). *Basic tail and concentration bounds.* `http://www.stat.berkeley.edu/~mjwain/stat210b/Chap2_TailBounds_Jan22_2015.pdf`

Yao, J., Zheng, S. and Bai, Z. D. (2015). *Sample covariance matrices and high-dimensional data analysis.* Cambridge University Press.

Department of Statistics, Stanford University, Stanford, CA 94305, U.S.A.

E-mail: jeha@stanford.edu

Department of Statistics, Stanford University, Stanford, CA 94305, U.S.A.

E-mail: imj@stanford.edu