CHAPTER I

SPECTRAL PROPERTIES OF LARGE WISHART MATRICES

Traditionally, methods of the multi-dimensional statistical analysis were first developed for normal populations. For these the majority of multivariate problems can be solved analytically. In particular, the exact distribution densities were found for the joint distribution of entries of sample covariance matrices and also for the set of their eigenvalues (see in [68]). A number of functionals involved in multivariate analysis were evaluated exactly. For the fundamentals of the theory of multivariate analysis one can see [4]. We recall here the basic notions and introduce definitions which will be used below.

We say that a scalar random value X has the standard normal distribution if its distribution function is

$$F(u) = P(X \le u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} \exp(-x^2/2)dx.$$

This fact can be written as $X \sim \mathbf{N}(0,1)$. The Fourier transform of the standard normal density gives its characteristic function

$$\chi(\theta) = \int \exp(iu\theta)dF(u) = \exp(-\theta^2/2).$$

Let $Y = (Y_1, \dots, Y_n)$ be a random vector with independent components distributed as $\mathbf{N}(0,1)$. We perform the linear transformation $X = \vec{\mu} + ODY$, where $\vec{\mu}$ is a displacement vector, O is a rotation matrix, and D is a diagonal matrix with the diagonal elements $\sigma_i^2 > 0, \quad i = 1, \dots, n$. Denote the expectation operator by $\mathbf{E}$, and for vectors ξ and η define cov $(\vec{\xi}, \vec{\eta}) = \mathbf{E}\ (\vec{\xi} - \mathbf{E}\ \vec{\xi})(\vec{\eta} - \mathbf{E}\ \vec{\eta})^T$. We have $\mathbf{E}\ X = \vec{\mu}$. The matrix $\Sigma = \mathrm{cov}\ (X, X)$ is called the covariance matrix. Obviously, $\sigma_i^2 > 0$ are eigenvalues of Σ, $i = 1, \dots, n$. The differential element of probability for the random value Y is $dP = (2\pi)^{-n/2}\exp(-\mathbf{y}^2/2)d\mathbf{y}$, where $\mathbf{y} = (y_1, \dots, y_n)$ is a vector of realizations of Y, $d\mathbf{y} = dy_1 \dots dy_n$. The Jacobian

$$\det\left(\frac{\partial \mathbf{y}}{\partial \mathbf{x}}\right) = \prod_{i=1}^{u} \sigma_i^{-2},$$

$\Sigma = OD^2O^T$, and $\mathbf{y}^2 = (\mathbf{x} - \vec{\mu})^T\Sigma^{-1}(\mathbf{x} - \vec{\mu})$. We can rewrite dP in the form $dP = f(\mathbf{x})d\mathbf{x}$, where $d\mathbf{x} = dx_1 \dots, dx_n$ and

$$f(\mathbf{x}) = (2\pi)^{-n/2}(\det\Sigma)^{-1/2}\ \exp\left((\mathbf{x} - \vec{\mu})^T\Sigma^{-1}(\mathbf{x} - \vec{\mu})/2\right). \quad (1)$$

A multi-dimensional distribution is called (non-degenerate) normal if its density is (1). This fact is written as $X \sim \mathbf{N}(\vec{\mu}, \Sigma)$. The characteristic function for $\mathbf{N}(\vec{\mu}, \Sigma)$ is

$$\chi(\theta) = \int \exp(i\mathbf{x}^T\theta)d\mathbf{x} = \exp(i\vec{\mu}^T\theta - \theta^T\Sigma\theta/2),$$

$\mathbf{x}, \theta \in \mathbb{R}^n$, $\chi(0) = 1$. This expression is convenient to calculate moments of normal variables by differentiating. Let $\vec{\mu} = 0$. The moments of odd orders vanish because of the symmetricity of normal distribution, and

$$\mathbf{E}\ X_iX_j = -\frac{\partial^2}{\partial\theta_i\partial\theta_j}\chi(\theta)\bigg|_{\theta=0} = \Sigma_{ij},$$
$$\mathbf{E}\ X_iX_jX_kX_l = \Sigma_{ij}\Sigma_{kl} + \Sigma_{ik}\Sigma_{jl} + \Sigma_{il}\Sigma_{jk},$$

$i, j, k, l = 1, \dots, n$.

The normal distribution is characterized by well known properties; we recall a number of them:

- each exponential density function whose argument is a positively definite quadratic form is the density of a non-degenerate normal law;
- the matrix Σ is diagonal if and only if the components of X are independent;
- if $X \sim \mathbf{N}(\vec{\mu}_1, \Sigma_1)$ and $Y \sim \mathbf{N}(\vec{\mu}_2, \Sigma_2)$ are two independent random vectors, then $X + Y \sim \mathbf{N}(\vec{\mu}_1 + \vec{\mu}_2, \Sigma_1 + \Sigma_2)$;
- by integrating $f(\mathbf{x})$ with respect to a subset of components of $\mathbf{x}$, we obtain another normal law density.

Wishart Distribution

We consider samples $\mathfrak{X} = \{\mathbf{x}_m,\ m = 1, \ldots, N\}$ of independent observation vectors $\mathbf{x}_m \in \mathbb{R}^n$ distributed as $\mathbf{N}(\vec{\mu}, \Sigma)$ each (the vectors $\mathbf{x}_m$ can be regarded as random values and, at the same time, as realizations of a random value; here and in the following we will not distinguish them in notations). To estimate $\vec{\mu}$ and Σ the 'natural' estimators are used

$$\bar{\mathbf{x}} = N^{-1} \sum_{m=1}^{N} \mathbf{x}_m \quad \text{and} \quad C = \frac{1}{N} \sum_{m=1}^{N} (\mathbf{x}_m - \bar{\mathbf{x}})(\mathbf{x}_m - \bar{\mathbf{x}})^T. \tag{2}$$

It is well known that $\bar{\mathbf{x}} \to \vec{\mu}$ and $C \to \Sigma$ in probability as $N \to \infty$.

In the following, let $\vec{\mu} = 0$.

LEMMA 1.1. *There exists an orthogonal transformation of vectors*

$$\mathbf{y}_k = \sum_{m=1}^{N} \Omega_{km} \mathbf{x}_m \tag{3}$$

such that the vectors $\mathbf{y}_N = \sqrt{N}\bar{\mathbf{x}}$ *and* $\mathbf{y}_k \sim N(0, \Sigma)$, $k = 1, \ldots, N-1$, *are independent, and the sample covariance matrix* (2) *is equal to*

$$C = \frac{1}{N} \sum_{m=1}^{N-1} \mathbf{y}_m \mathbf{y}_m^T.$$

Proof. Define the vector $\Omega_N = (\Omega_{N1}, \ldots, \Omega_{NN})$, where $\Omega_{Nm} = N^{-1/2}$ for each $m = 1, \ldots, N$. In the space $\mathbb{R}^{N-1}$ orthogonal to Ω_N, one can introduce an orthonormal basis $\{\Omega_k\}$ of vectors $\Omega_k = \{\Omega_{k1}, \ldots, \Omega_{kN}\}$, $k = 1, \ldots, N-1$. Consider the transformation (3). Since arguments $\mathbf{x}_m$ are normal and independent, the vectors $\mathbf{y}_k$ are also normally distributed, $\mathbf{E}\ \mathbf{y}_k = 0$ for each k and

$$\mathbf{E}\ \mathbf{y}_k \mathbf{y}_l^T = \sum_{m'} \sum_{m} \Omega_{km'} \Omega_{lm} \mathbf{E}\ \mathbf{x}_{m'} \mathbf{x}_m^T = \delta_{kl} \Sigma,$$

where δ_{kl} is the Kronecker delta, $k, l = 1, \ldots, N$. The covariance $\operatorname{cov}(\mathbf{y}_k, \mathbf{y}_l) = 0$ if $k \neq l$. It means, for normal distributions, that $\mathbf{y}_k$ and $\mathbf{y}_l$ are independent for $k \neq l$, $k, l = 1, \ldots, N$. In particular,

$\bar{\mathbf{x}}$ is independent on $\{\mathbf{y}_1, \dots, \mathbf{y}_{N-1}\}$. For $k = l$, we obtain that cov $(\mathbf{y}_k, \mathbf{y}_k) = \Sigma, \quad l = 1, \dots, N$. The first statement of Lemma 1.1 is proved.

Now, we find that

$$N^{-1} \sum_{k=1}^{N-1} \mathbf{y}_k \mathbf{y}_k^T = N^{-1} \sum_{m=1}^{N} \mathbf{x}_m \mathbf{x}_m^T - \bar{\mathbf{x}}\bar{\mathbf{x}}^T = C.$$

The proof of Lemma 1.1 is complete. □

Corollary. The random values $\bar{\mathbf{x}}$ and C are independent.

Let us find the distribution law for entries of the matrices C. We note that C is a symmetric matrix, and it suffices to find the distribution of $\{C_{ij}\}$ for $i \le j$. It is simpler to find first the characteristic function. This must depend, as well as the distribution function, on $n(n+1)/2$ parameters. Let Ξ be a symmetric matrix of parameters, $\Xi = \{\Xi_{ij}\}$; denote $\tilde{\Xi}_{ij} = \varepsilon_{ij}\Xi_{ij}$, where $\varepsilon_{ii} = 1$ and $\varepsilon_{ij} = 1/2$ if $i \ne j, \; i, j = 1, \dots, n$. By definition the characteristic function for the distribution of C is

$$\chi(\Xi) = \mathbf{E} \exp[i \text{ tr } (C\tilde{\Xi})]. \tag{4}$$

Lemma 1.2. *For* $\mathbf{N}(0, \Sigma)$, *the characteristic function* (4) *is*

$$\chi(\Xi) = \det \left(I - \frac{2i}{N} \Sigma^{-1} \tilde{\Xi} \right)^{-(N-1)/2} \tag{5}$$

Proof. Let us write down the differential element of probability of events concerned with the sample $\mathfrak{X}$:

$$dP = (2\pi)^{-nN/2} (\det \Sigma)^{-N/2} \exp \left(-\frac{1}{2} \sum_{m=1}^{N} \mathbf{x}_m^T \Sigma^{-1} \mathbf{x}_m \right) \prod_{m=1}^{N} d\mathbf{x}_m.$$

We examine that the expression under the sign of the summation equals N tr $(\Sigma^{-1} S) = N$ tr $(\Sigma^{-1} C) + N \bar{\mathbf{x}}^T \Sigma^{-1} \bar{\mathbf{x}}$. Changing variables, we notice that the Jacobian

$$\det \left(\frac{\partial \mathbf{x}_1 \dots \partial \mathbf{x}_N}{\partial \mathbf{y}_1 \dots \partial \mathbf{y}_N} \right) = 1.$$

We integrate with respect to $y_N = \sqrt{N}\bar{\mathbf{x}}$ and find that $\chi(\Xi)$ equals

$$c\ (\det \Sigma)^{-(N-1)/2} \int \exp\left[\operatorname{tr}\ \left(-\frac{N}{2}\Sigma^{-1}C + i\tilde{\Xi}C\right)\right] \prod_{m=1}^{N} d\mathbf{y}_m, \quad (6)$$

where c is a number. Since $\chi(0) = 1$, the integration in (6) can be carried out formally by the substitution of $\tilde{\Sigma}^{-1} = \Sigma^{-1} - 2iN^{-1}\tilde{\Xi}$ instead of Σ^{-1}. It follows that

$$\chi(\Xi) = \left(\frac{\det \tilde{\Sigma}}{\det \Sigma}\right)^{(N-1)/2} = \det\left(I - \frac{2i}{N}\ \Sigma\tilde{\Xi}\right)^{-(N-1)/2}.$$

Lemma 1.2 is proved. □

The probability element for the variable C can be written as

$$dP = f_W(C)dC, \quad \text{where } dC = dC_{11}dC_{12}dC_{22}dC_{23}\ldots dC_{nn}$$

(lower subscripts enumerate entries of the matrix C). Let us find the density $f_W(C)$. Denote $|\Sigma| = \det\ \Sigma,\quad |C| = \det\ C$.

THEOREM 1.1. *If $N > n+2$ then the probability density $f_W = f_W(C)$ is equal to*

$$c\ |\Sigma|^{-(N-1)/2}|C|^{(N-n-2)/2}\ \exp\left(-\frac{N}{2}\operatorname{tr}\ (\Sigma^{-1})C\right) \quad (7)$$

if the matrix C is positive definite and to 0 otherwise; c is a normalization factor.

Proof. It is easier to prove this theorem starting from (5). Since there is one-to-one correspondence between densities and characteristic functions it suffices to prove that (7) is the Fourier transform of (5). It turns out that the normalization constant c does not depend on Σ. To prove it we notice that the value of the normalization coefficient c can depend only on n, N and the eigenvalues $\{\lambda_i\}$ of Σ. Consider the coordinate system in which the matrix Σ is diagonal. Let us perform an extension of one of the axes, say, number 1: $\tilde{x}_1 = kx_1$ where k is the extension coefficient. The eigenvalue $\lambda_1 = \operatorname{cov}\ (x_1, x_1)$ is multiplied by a factor k^2, whereas other λ_i do not change. The element C_{11} is multiplied by k^2 and elements C_{1i} are multiplied by k, $i = 2, \ldots, n$. Consequently the quantities $\det \Sigma$ and $\det C$ are multiplied by k^2. The product $dC = \prod_{i \le j} dC_{ij}$ is multiplied by k^{n+1} , and the product $|\Sigma|^{-(N-1)/2} |C|^{(N-n-2)/2} dC$ does not change. The matrix $\Sigma^{-1}C$ also does not change. We conclude that the normalization constant c also does not change. We proved that c depends only on n and N.

Let us write the normalization condition for $f_W(C)$ with Σ^{-1} replaced by $\Sigma^{-1} - 2i\tilde{\Xi}/N$. We obtain the equation

$$\text{const} \int |C|^{(N-n-2)/2} \exp\left(-\frac{N}{2} \text{tr } (\Sigma^{-1}C) + i \text{ tr } (C\tilde{\Xi})\right) dC = 1,$$

where (and in the following) the integration region is confined to positive definite matrices C. This equation can be rewritten as

$$\det \left(I - \frac{2i}{N}\Sigma\tilde{\Xi}\right)^{(N-1)/2} \mathbf{E} \exp[i \text{ tr } (C\tilde{\Xi})] = 1,$$

where the expectation is calculated by integrating with respect to $f_W(C)dC$.

One can see that $f_W(C)$ is a density with the characteristic function (5). This proves our theorem. $\square$

The distribution with the density (7) is called the Wishart distribution. Matrices whose entries have the Wishart distribution are called Wishart matrices.

To calculate moments $\mathbf{E}\, C^k$ using the distribution (7), one can use the differentiation with respect to parameters. Combining the differentiation with matrix multiplication, we must take into account that the matrices are symmetrical. Given a symmetric matrix variable X, we define the matrix operator

$$\nabla = \{\nabla_{ij}\} = \{\varepsilon_{ij} \frac{\partial}{\partial X_{ij}}\},$$

where $\varepsilon_{ij} = 1$ and $\varepsilon_{ij} = 1/2$, $i \ne j$, $i, j = 1, \ldots, n$. If X is symmetric, then $\nabla\varphi(X)$ is also symmetric and

$$\nabla_{ij}\varphi(X) = 1/2 \left(\frac{\partial\varphi(X)}{\partial X_{ij}} + \frac{\partial\varphi(X)}{\partial X_{ji}}\right),$$

where partial derivatives are calculated by the variation of a single entry of the matrix X.

It is easy to examine that

$$
\begin{aligned}
&\nabla \text{ tr } (AX) = X, \quad \nabla \exp[\text{tr } (AX)] = A, \quad \nabla \text{ } (\text{tr } X^k) = kX^{k-1},\\
&\nabla \text{ } |X| = |X|X^{-1}, \quad \nabla_{ij} X_{kl} = 1/2 \text{ } (\delta_{ik}\delta_{jl} + \delta_{ij}\delta_{kl}),\\
&\nabla_{ij} X^{-1}_{kl} = -1/2 \text{ } (X^{-1}_{ik} X^{-1}_{jl} + X^{-1}_{il} X^{-1}_{jk}),
\end{aligned}
\tag{8}
$$

where $i, j, k, l = 1, \ldots, n$, A is a symmetric matrix of constants, $|X| = \det X$, δ stands for the Kronecker delta, and X^{-1} with subscripts denotes the corresponding entries of the matrix X^{-1}. We obtain the following recurrent relations for the differentiation of powers of matrices:

$$
\begin{aligned}
&\nabla X = 1/2 \text{ } (n+1)I, \quad \nabla X^{-1} = -1/2 \text{ } (X^{-2} + X^{-1} \text{tr } X^{-1}),\\
&\nabla X^k = 1/2 \left(kX^{k-1} + \sum_{i=0}^{k-1} X^i \text{ tr } X^{k-1-i} \right), \quad k = 1, 2, \ldots\\
&\nabla X^{-k} = -1/2 \left(kX^{-k-1} + \sum_{i=1}^{k} X^{-i} \text{ tr } X^{-k-1+i} \right), \quad k = 1, 2, \ldots .
\end{aligned}
\tag{9}
$$

Remark 1. If $N > n + 2$, then

$$
\mathbf{E} \text{ } C^k = |X|^{(N-1)/2} [(-2N^{-1}\nabla)^k \text{ } |X|^{-(N-1)/2}] \Big|_{X=\Sigma^{-1}},
$$

$k = 0, 1, 2, \ldots$. This relations can be readily shown if to notice that $|\Sigma|^{-(N-1)/2} f_W$ is proportional to $\exp\left(-(N/2) \text{ tr } \Sigma^{-1} C\right)$.

To calculate negative moments of C, a special technique can be used offered by L. Haff (1981). Note that if $N > n + 2$ the density $f_W = 0$ on the integration region boundary since the matrix $C = \{C_{ij}\}$ becomes degenerate and $\det C = 0$. Transforming the volume integral to the surface one by Stokes' theorem, we obtain

$$
\int \nabla(\varphi(C) f_W) dC = 0,
$$

where 0 stands for zero matrix, for any differentiable symmetric matrix function $\varphi(C)$ if the integrand is absolute integrable. For example, for $\varphi(C) = 1$ we have

$$
\nabla f_W = 1/2 \left(-N\Sigma^{-1} + (N - n - 2) C^{-1} \right) f_W \tag{10}
$$

and the relation $\mathbf{E}\ C^{-1} = N(N-n-2)^{-1}\Sigma^{-1}$ follows. Evaluating $\nabla_{ij}(C_{kl}^{-1} f_W)$, we obtain a system of three linear equations to determine $\mathbf{E}\ C_{ij}^{-1} C_{kl}^{-1}$. If $N > n+4$, then the solution is

$$\mathbf{E}\ C_{ij}^{-1} C_{kl}^{-1} = N^2\ \frac{(N-n-3)\Sigma_{ij}^{-1}\Sigma_{kl}^{-1} + \Sigma_{ik}^{-1}\Sigma_{jl}^{-1} + \Sigma_{il}^{-1}\Sigma_{jk}^{-1}}{(N-n-1)(N-n-2)(N-n-4)},$$

$i, j, k, l, = 1, \dots, n$

Limit Moments of Wishart Matrices

E. Wigner in 1958 and V. Marchenko and L. Pastur in 1967 discovered a characteristic form of limiting spectra for random matrices of increasing dimension (see Introduction). Limit formulas obtained by these authors are applicable to the Wishart matrices for a special case $\Sigma = I$. In this section we investigate spectral functions of the increasing Wishart matrices using special properties of normal distributions.

The Case of the Identity Covariance Matrices

The first attempt to study limiting spectra of increasing sample covariance matrices for normal variables was made in [5] and [6]. By a straightforward evaluation of moments of variables for $X \sim N(0, I)$ in the asymptotics $n \to \infty$, $N = N(n) \to \infty$, $n/N \to y$, where n is the dimension of X, and N is the sample size, the author of [6] proved the existence of limits

$$M_k = \operatorname*{plim}_{n\to\infty} n^{-1} \text{tr}\ C^k = \operatorname*{plim}_{n\to\infty} n^{-1} \sum_{i=1}^{n} \lambda_i^k$$

for normalized average moments λ_i of the matrices C. These limit moments proved to be polynomials with respect to y

$$M_k = M_k(y) = \sum_{j=0}^{k-1} B_k^j y_i,$$

where the coefficients B_k^j satisfy the recurrent relations

$$B_k^j = \frac{k(k-1)}{(j+1)j} B_{k-1}^{j-1}, \quad j = 1, \dots, k-1; \quad k = 2, 3, \dots,$$

$B_k^0 = 1$, $k = 1, 2, \dots$. In particular,

$$B_k^1 = k(k-1)/2, \quad k = 2, 3, \dots ,$$

$$M_1(y) = y, \quad M_2(y) = 1 + y, \quad M_3(y) = 1 + 3y + y^2,$$
$$M_4(y) = 1 + 6y + 6y^2 + y^3. \tag{11}$$

Let us study the limit spectra of C starting from these moments. First we notice that $M_k(y)$ satisfy the following equations.

Remark 2. For any $y \geq 0$ and $k = 1, 2, \dots$, there holds the equation

$$\frac{d^2}{dy^2}[y(M_k(y) - 1)] = (k-1)kM_{k-1}(y), \quad M_0(y) = 1,$$

with the boundary conditions $M_k(0) = 1$ and $M_k'(0) = k(k+1)/2$, $k = 1, 2, \dots$.

Remark 3. For each $t \geq 0$ there exists the characteristic function

$$\chi(t, y) = \sum_{k=0}^{\infty} M_k(y) \, \frac{(it)^k}{k!}. \tag{12}$$

The convergence of this series follows from the relations

$$M_k(y) \leq \theta^k \; M_k(1) = \theta^k \; k^{-1} \sum_{j=0}^{k-1} \binom{k}{j+1} \binom{k}{j} = \theta^k \; k^{-1} \binom{2k}{k-1}, \tag{13}$$

where θ=max(1,y) and the logarithm of the right hand side is $O(k)$ as $k \to \infty$.

Remark 4. For $y \geq 0$ and $t \geq 0$ the equation holds

$$y \frac{\partial^2 \chi(t, y)}{\partial y^2} + 2 \frac{\partial \chi(t, y)}{\partial y} = it^2 \frac{\partial \chi(t, y)}{\partial t}$$

with the boundary conditions $\chi(0, 1) = 1$ for $y \geq 0$, $\chi(t, 0) = \exp(it)$, and $\chi'(t, 0) = -1/2 \; t^2 \exp(it)$, where the prime denotes the differentiation with respect to y.

Notice the following.

THEOREM 1.2. *The characteristic function* (12) *constructed with the moments* $\{M_k\}$ *is*

$$\chi(t,y) = 1 + i\int_0^t \frac{J_1(2u\sqrt{y})}{u\sqrt{y}}\,\exp[i\,(1+y)u]du,$$

where $y > 0$, *and* $J_1(\cdot)$ *is the Bessel function of the first order.*

Proof. Consider the integral equation with respect to function $\psi(u,y)$

$$\chi(t,y) = 1 + i\int_0^t \psi(u,y)\,\exp[i\,(1+y)u]du.$$

We can see that the equation for $\chi(t,y)$ in Remark 4 is equivalent to the equation

$$y\frac{\partial^2\psi}{\partial y^2} + 2(1+iyu)\frac{\partial\psi}{\partial y} = iu^2\frac{\partial\psi}{\partial u} - u^2\psi \tag{14}$$

for $\psi = \psi(u,y)$, $y \geq 0$, $u \geq 0$, with the boundary conditions

$$\psi(u,0) = 1, \quad \psi'(u,0) = -u^2/2, \quad u \geq 0; \quad \psi(0,y) = 1, \quad y \geq 0,$$

where the prime denotes the partial derivative with respect to y. The solution to (14) is

$$\psi = \psi(u,y) = \sum_{k=0}^{\infty} \frac{(-)^k y^k u^{2k}}{k!(k+1)!}.$$

We examine that this function satisfies the equations

$$u\frac{\partial\psi}{\partial u} = 2y\frac{\partial\psi}{\partial y}, \quad \text{and} \quad y\frac{\partial^2\psi}{\partial y^2} + 2\frac{\partial\psi}{\partial y} + u^2\psi = 0.$$

These two equations imply (14). One can see that the boundary conditions for $\psi(u,y)$ are satisfied.

It remains to examine that the series for $\psi(u,y)$ coincides with the series for $J_1(2u\sqrt{y})/u\sqrt{y}$. This proves Theorem 1.2. □

Denote

$$f(z) = (2\pi)^{-1}\int_{-\infty}^{+\infty} \chi(t,y)\,\exp(-izt)dt.$$

THEOREM 1.3. *If* $0 < y < 1$, *then*

$$f(z) = \begin{cases} (2\pi zy)^{-1}(4y - (z-1-y)^2)^{1/2} & \text{if} \quad u_1 \le z \le u_2, \\ 0 \quad \text{for other } z, \end{cases} \tag{15}$$

where $u_1 = c_1(1-\sqrt{y})^2, \quad u_2 = c_2(1+\sqrt{y})^2$.

Proof. We note that $\chi(t, y) \to 0$ as $t \to \infty$. Integrating by parts, we find that

$$\begin{aligned} f(z) &= \pi^{-1}\mathrm{Re} \int_0^\infty \chi(t,y)\exp(-itz)dt \\ &= \pi^{-1}\mathrm{Re} \int_0^\infty \frac{J_1(2t\sqrt{y})}{zt\sqrt{y}}\exp[i\ (1+y-z)t]dt \\ &= (2\pi zy)^{-1}\mathrm{Re}\ \sqrt{4y-(z-1-y)^2}\,. \end{aligned}$$

The assertion of Theorem 1.3 follows. □

We conclude that the $M_k(y)$ are moments with respect to the measure $f(z)dz$ defined by (15).

Let us use (15) for the calculation of average moments of eigenvalues of C. Substituting $z = 1 + y - 2\sqrt{y}\cos\varphi$ we find that

$$M_k(y) = \int z^k f(z)dz = \pi^{-1}\int_0^{2\pi}(1+y-2\sqrt{y}\cos\varphi)^{k-1}\sin^2\varphi d\varphi\,.$$

For $k = 1, 2, 3, 4$, we obtain the moments (11). For negative k, we find

$$M_{-1}(y) = (1-y)^{-1}, \quad M_{-2}(y) = (1-y)^{-3},$$

$$M_{-3}(y) = (1+y)(1-y)^{-5}, \quad M_{-4}(y) = (1+3y+y^2)(1-y)^{-7}.$$

General Case: Arbitrary Covariance Matrices

Let us calculate the limits of $n^{-1}\mathrm{tr}\ C^k$ using the method of parametric differentiation. To obtain the convergence in the increasing dimension asymptotics, we assume the convergence for the empirical distribution function

$$F_{0n}(u) = n^{-1} \sum_{i=1}^{n} \mathrm{ind}(\lambda_i^0 \leq u) \to F_0(u) \tag{16}$$

of the eigenvalues λ_i^0 of Σ almost everywhere for $u \geq 0$. Note that, under assumption (16), the limits exist

$$\Lambda_k = \lim_{n\to\infty} n^{-1} \mathrm{tr}\ \Sigma^k, \quad k > 0.$$

THEOREM 1.4. *If $n \to \infty$, $N \to \infty$ so that $n/N \to y > 0$ and* (16) *holds, then*

1^o *the limits in the square mean exist* $\underset{n\to\infty}{\mathrm{l.i.m.}}\ n^{-1}\mathrm{tr}\ C^k = M_k$ *such that*

$$M_k = M_k(y) = (L^k)_{11}/y, \quad k = 1, 2,$$

where an infinite matrix L has the entries

$$L_{ij} = \begin{cases} 0 & \text{if} \quad j < i-1, \\ 1 & \text{if} \quad j = i-1, \\ y\Lambda_{j-i+1} & \text{if} \quad j > i-1, \end{cases}$$

$i, j = 1, 2, \ldots$, and $(L^k)_{11}$ is the first entry of the matrix L^k;

2^o $\mathrm{var}(n^{-1}\mathrm{tr}\ C^k) = O(n^{-1}N^{-1})$ *as $n \to \infty$,* $\quad k = 1, 2, \ldots$.

Proof. We use Remark 1. Note that, for $X = \Sigma^{-1}$ as $n \to \infty$,

$$(-2N^{-1}\nabla)X^{-k} = y \sum_{i=1}^{k} X^{-1} n^{-1}\mathrm{tr}\ X^{-k-1+i} + O(n^{-1})K_1,$$
$$(-2N^{-1}\nabla)|X|^{-(N-1)/2} = |X|^{-(N-1)/2}X^{-1} + O(N^{-1})K_2,$$
$$(-2N^{-1}\nabla)n^{-1}\mathrm{tr}\ X^{-k} = O(n^{-1}N^{-1})K,$$

where symmetric positively semidefinite matrices K_1, K_2, and K_3 have bounded spectral norms. For $k > 0$ the application of the operator $-2N^{-1}\nabla$ to $n^{-1}\mathrm{tr}\ X^{-k}$ decreases the order of magnitude, whereas the application of the same operator to the remainder terms does not decrease their order of magnitude. By Remark 1 as $n \to \infty$ we obtain

$$\mathbf{E}\, C^k = \sum_{i=1}^{k} a_k^i \Sigma^i + o(1)K, \quad i = 1, 2, \dots$$

where the matrices K have bounded norms, $a_k^k = 1$ for $k = 1, 2, \dots,$ and

$$a_k^i = a_{k-1}^{i-1} + y \sum_{j=i}^{k-1} a_{k-1}^j \Lambda_{j-i+1}, \quad i = 1, \dots, k-1.$$

From these equations it follows that limits of $\mathbf{E}\, n^{-1}\text{tr}\ C^k$ exist, and the expression of M_k in terms of entries of L holds.

Denote $d = |X|^{-(N-1)/2}$ and define the differential operator $T = \text{tr}\ (-2N^{-1}\nabla)$. We have $Td = td$, where τ is a sum of traces of powers of X^{-1}. If $X = \Sigma^{-1}$, then $\tau = O(n)$. By Remark 1 it follows that

$$\text{var}(n^{-1}\text{tr}\ C^k) = n^{-2}(d^{-1}T(\tau d) - \tau^2).$$

But we have $T\tau = O(n^{-1}N^{-1})$. This proves Theorem 1.4. □

In particular, we find $M_1 = \Lambda_1$, $M_2 = \Lambda_2 + y\Lambda_1^2$, $M_3 = \lambda_3 + 3y\Lambda_2\Lambda_1 + y^2\Lambda_1^3$.

Note that the moments M_k are non-decreasing functions of Λ_k, $k = 1, 2, \dots$. But $\Lambda_k \le \|\Sigma\|^k$ for $N(0, \Sigma)$ for any $k > 0$. This means that M_k are not larger than $\|\Sigma\|^k$ multiplied by the moments M_k calculated for $\mathbf{N}(0, I)$ in the previous section. By (13) those are not larger than c^k for an appropriate constant $c > 0$. Thus the moments M_k defined by Theorem 1.4 satisfy the Carleman condition, and one can indicate a distribution function $F(u)$ such that $M_k = \int u^k dF(u)$, $k = 1, 2, \dots$.

To evaluate other functionals depending on C, the Stokes's theorem can be applied. We present an example.

Remark 5. Under assumptions of Theorem 1.4 the limits exist

$$N_k = \lim_{n\to\infty} \mathbf{E}\, n^{-1}\text{tr}\ (\Sigma C^k), \quad k = 0, 1, \dots$$

such that

$$N_0 = M_1 \quad \text{and} \quad N_k = M_{k+1} - y \sum_{i=0}^{k-1} N_i M_{k-i}, \quad k = 1, 2, \dots. \tag{17}$$

Let us prove these recurrent relations. For each integer $k > 0$

$$\int \nabla_{ij}\ (\Sigma C^k) f_W dC = 0 \quad i, j = 1, \dots, n,$$

since $|C| = 0$ on the boundaries of the integration region for $N > n+2$ and the density $f_W = 0$. Applying rules (8) and using Remark 1, we obtain that $\Sigma^{-1}\mathbf{E}\, C^k$ equals

$$(N-n-2)N^{-1}\mathbf{E}\, C^{k-1} + kN^{-1}\mathbf{E}\, C^k + N^{-1}\sum_{i=0}^{k-1}\mathbf{E}\, C^i \mathrm{tr}\ C^{k-1-i}. \quad (18)$$

We multiply both parts of (18) by n^{-1} and calculate the traces. Passing to the limit as $n \to \infty$, $n/N \to y$, we note that the second summand of the right hand side of (18) vanishes. Keeping only leading terms we obtain (17).

In particular, we have $N_1 = M_2 - yM_1^2$, $N_2 = M_3 - 2yM_2M_1 + y^2M_1^3$, etc.

Limit Formula for the Resolvent of Wishart Matrices

Let us study the resolvent $H = H(t) = (I + tC)^{-1}$ of the Wishart matrices C, $t \geq 0$, under the increasing dimension asymptotics. Consider a sequence $\mathfrak{P} = \{\mathfrak{P}_n\}$ of problems

$$\mathfrak{P}_n = (\mathfrak{S}, N, \mathfrak{X}, C, H(t))_n, \quad n = 1, 2, \dots, \quad (19)$$

in which spectra of the Wishart covariance matrices C and the resolvent $H(t)$ are studied by samples $\mathfrak{X}$ of size N from populations $\mathfrak{S} = \mathbf{N}(0, \Sigma)$. Denote $h_n(t) = n^{-1}\mathrm{tr}\ H(t)$ and $s_n(t) = 1 - n/N + n/Nh_n(t)$.

LEMMA 1.3. *If the spectral norms* $\|\Sigma\|$ *and* $\|\Sigma^{-1}\|$ *are uniformly bounded in the sequence* (19) *and* $n + 2 < N$ *for each* n, *then*

$$H(t) - \mathbf{E}\ (I + ts_n(t)\Sigma)H(t) = \Omega, \quad (20)$$

where $\|\Omega\| = O(N^{-1})$.

Proof. To apply Stokes' theorem we write

$$\int \nabla(Hf_W)dC = 0, \quad (21)$$

where $H = H(t)$ and 0 stands for the zero matrix. Applying the differentiation rules (8), we obtain

$$\nabla H = -t/2\ (H^2 + H \text{ tr } H).$$

By (8), we have $\nabla f_W = [-1/2\ N\Sigma^{-1} + (N-n-2)/2\ C^{-1}]f_W$. From (21) it follows that the integral

$$\int [(N-n-2)N^{-1}C^{-1}H - \Sigma^{-1}H - tN^{-1}H^2 - tN^{-1}H \text{ tr } H]\ f_W dC$$

vanishes. Let us multiply this expression from the left by Σ. Substituting $C^{-1}H = C^{-1} - tH$ and $(N-n-2)N^{-1}\mathbf{E}\ C^{-1} = \Sigma^{-1}$, we find that

$$I - \int (I + ts_n(t)\Sigma)H\ f_W dC = \\ - \int [2tN^{-1}\Sigma H + tN^{-1}\Sigma H^2]\ f_W dC.$$

Since $\|H\|$ are bounded, the right hand side presents a magnitude of the order of N^{-1}. The lemma is proved. □

Remark 6. Under conditions of Lemma 1.3, var $h_n(t) \to 0$ for each $t \geq 0$.

Theorem 1.5. *Suppose a sequence of problems* (19) *is given in which the populations* $\mathfrak{S} = \mathbf{N}(0,\Sigma)$ *are such that the spectral norms* $\|\Sigma\|$ *and* $\|\Sigma^{-1}\|$ *are uniformly bounded in* $\mathfrak{P}$, (16) *holds, and* $n/N \to y > 0$.

Then, for each $t \geq 0$, $h_n(t) \to h(t)$ *in the square mean,*

$$h(t) = \lim_{n\to\infty} n^{-1}\text{tr } (I + ts(t)\Sigma)^{-1},$$
$$\mathbf{E}\ (I + tC)^{-1} = (I + ts(t)\Sigma)^{-1} + \Omega, \tag{22}$$

where $s(t) = 1 - y + yh(t)$ *and* $\|\Omega\| \to 0$.

Proof. Denote $H = (I + tC)^{-1}$, $\bar{h}_n = \mathbf{E}h_n(t)$, $\bar{s}_n = \mathbf{E}s_n(t)$. By Remark 6, we have $\|\mathbf{E}\ (s_n - \bar{s}_n)\Sigma H\| \to 0$, where $s_n = s_n(t)$. Multiplying the expressions in (20) by $(I + t\bar{s}_n\Sigma)^{-1}$, we obtain $\mathbf{H} =$

$(I + t\bar{s}_n\Sigma)^{-1} + \Omega$, where $\|\Omega\| \to 0$. We calculate traces and obtain that

$$\bar{h}_n = n^{-1}\mathrm{tr}\ (I + ts_n\Sigma)^{-1} + o(1) = \eta(t\bar{s}_n) + o(1).$$

as $n \to \infty$. From this equation, it can be readily seen that the sequence $\{\bar{h}_n\}$ converges to some $h = h(t)$ and this implies the first statement of the theorem. The second statement can be derived from the same equations. Theorem 1.5 is proved. □

Remark 7. The limit moments $M_k(y)$ can be calculated using the first equation in (22) by differentiating.