1

Introduction

1.1 Large-Dimensional Data and New Asymptotic Statistics

In a multivariate analysis problem, we are given a sample $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ of random observations of dimension p. Statistical methods, such as principal component analysis, have been developed since the beginning of the 20th century. When the observations are Gaussian, some nonasymptotic methods exist, such as Student's test, Fisher's test, or the analysis of variance. However, in most applications, observations are non-Gaussian, at least in part, so that nonasymptotic results become hard to obtain and statistical methods are built using limiting theorems on model statistics.

Most of these asymptotic results are derived under the assumption that the data dimension p is fixed while the sample size n tends to infinity (large sample theory). This theory had been adopted by most practitioners until very recently, when they were faced with a new challenge: the analysis of large dimensional data.

Large-dimensional data appear in various fields for different reasons. In finance, as a consequence of the generalisation of Internet and electronic commerce supported by the exponentially increasing power of computing, online data from markets around the world are accumulated on a giga-octet basis every day. In genetic experiments, such as micro-arrays, it becomes possible to record the expression of several thousand of genes from a single tissue. Table 1.1 displays some typical data dimensions and sample sizes. We can see from this table that the data dimension p is far from the "usual" situations where p is commonly less than 10. We refer to this new type of data as *large-dimensional data.*

It has been observed for a long time that several well-known methods in multivariate analysis become inefficient or even misleading when the data dimension p is not as small as, say, several tens. A seminal example was provided by Dempster in 1958, when he established the inefficiency of Hotelling's T^2 in such cases and provided a remedy (named a non-exact test). However, by that time, no statistician was able to discover the fundamental reasons for such a breakdown in the well-established methods.

To deal with such large-dimensional data, a new area in asymptotic statistics has been developed where the data dimension p is no longer fixed but tends to infinity *together* with the sample size n. We call this scheme *large-dimensional asymptotics*. For multivariate analysis, the problem thus turns out to be which one of the large sample scheme and the large-dimensional scheme is closer to reality. As Huber (1973) argued, some statisticians might say that five samples for each parameter on average is enough to use large sample asymptotic results. Now, suppose there are $p = 20$ parameters and we have a sample of size $n = 100$. We may consider the case as $p = 20$ being fixed and n tending to infinity

Table 1.1. *Examples of large-dimensional data*

	Data dimension p	Sample size n	$y = p/n$
Portfolio	~ 50	500	0.1
Climate survey	320	600	0.21
Speech analysis	$a \times 10^2$	$b \times 10^2$	~ 1
ORL face database	1440	320	4.5
Micro-arrays	1000	100	10

(large sample asymptotics), $p = 2\sqrt{n}$, or $p = 0.2n$ (large-dimensional asymptotics). So, we have at least three different options among which to choose for an asymptotic setup. A natural question then, is, which setup is the best choice among the three? Huber strongly suggested studying the situation of increasing dimension together with the sample size in linear regression analysis.

This situation occurs in many cases. In parameter estimation for a structured covariance matrix, simulation results show that parameter estimation becomes very poor when the number of parameters is more than four. Also, it is found that in linear regression analysis, if the covariates are random (or have measurement errors) and the number of covariates is larger than six, the behaviour of the estimates departs far from the theoretical values, unless the sample size is very large. In signal processing, when the number of signals is 2 or 3 and the number of sensors is more than 10, the traditional multivariate signal classification (music) approach provides very poor estimation of the number of signals, unless the sample size is larger than 1000. Paradoxically, if we use only half of the data set, namely, we use the data set collected by only five sensors, the signal number estimation is almost 100 percent correct if the sample size is larger than 200. Why would this paradox occur? Now, if the number of sensors (the dimension of data) is p, then one has to estimate p^2 parameters ($\frac{1}{2}p(p+1)$ real parts and $\frac{1}{2}p(p-1)$ imaginary parts of the covariance matrix). Therefore, when p increases, the number of parameters to be estimated increases proportionally to p^2, while the number ($2np$) of observations increases proportionally to p. This is the underlying reason for this paradox. This suggests that one has to revise the traditional MUSIC method if the sensor number is large.

An interesting problem was discussed by Bai and Saranadasa (1996), who theoretically proved that when testing the difference of means of two high-dimensional populations, the Dempster (1958) non-exact test is more powerful than Hotelling's T^2-test, even when the T^2-statistic is well defined. It is well known that statistical efficiency will be significantly reduced when the dimension of data or number of parameters becomes large. Thus, several techniques for dimension reduction were developed in multivariate statistical analysis. As an example, let us consider a problem in principal component analysis. If the data dimension is 10, one may select three principal components so that more than 80 percent of the information is reserved in the principal components. However, if the data dimension is 1000 and 300 principal components are selected, one would still have to face a large dimensional problem. If, again, three principal components only are selected, 90 percent or even more of the information carried in the original data set could be lost. Now, let us consider another example.

Example 1.1 Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ be a sample from p-dimensional Gaussian distribution $\mathcal{N}_p(0, \mathbf{I}_p)$ with mean zero and unit covariance matrix. The corresponding sample covariance matrix is

$$\mathbf{S}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^*.$$

An important statistic in multivariate analysis is

$$T_n = \log(\det \mathbf{S}_n) = \sum_{j=1}^{p} \log \lambda_{n,j},$$

where $\{\lambda_{n,j}\}_{1\le j\le p}$ are the eigenvalues of $\mathbf{S}_n$. When p is fixed, $\lambda_{n,j} \to 1$ almost surely as $n \to \infty$, and thus $T_n \to 0$. Furthermore, by taking a Taylor expansion of $\log(1 + x)$, one can show that

$$\sqrt{\frac{n}{p}}\, T_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, 2)$$

for any fixed p. This suggests the possibility that T_n remains asymptotically Gaussian for large p provided that $p = O(n)$. However, this is not the case. Let us see what happens when $p/n \to y \in (0, 1)$ as $n \to \infty$. Using results on the limiting spectral distribution of $\mathbf{S}_n$ (see Chapter 2), it is readily seen that almost surely,

$$\frac{1}{p} T_n \to \int_{a(y)}^{b(y)} \frac{\log x}{2\pi y x} [\{b(y) - x\}\{x - a(y)\}]^{1/2}\, dx = \frac{y-1}{y} \log(1 - y) - 1 \equiv d(y) < 0,$$

where $a(y) = (1 - \sqrt{y})^2$ and $b(y) = (1 + \sqrt{y})^2$ (details of this calculation of the integral are given in Example 2.11). This shows that almost surely,

$$\sqrt{\frac{n}{p}}\, T_n \simeq d(y)\sqrt{np} \to -\infty.$$

Thus, any test which assumes asymptotic normality of T_n will result in a serious error.

These examples show that the classical large sample limits are no longer suitable for dealing with large-dimensional data analysis. Statisticians must seek out new limiting theorems to deal with large-dimensional statistical problems. In this context, the theory of random matrices (RMT) proves to be a powerful tool for achieving this goal.

1.2 Random Matrix Theory

RMT traces back to the development of quantum mechanics in the 1940s and the early 1950s. In this field, the energy levels of a system are described by eigenvalues of a Hermitian operator $\mathbf{A}$ on a Hilbert space, called the Hamiltonian. To avoid working with an infinite-dimensional operator, it is common to approximate the system by discretisation, amounting to a truncation, keeping only the part of the Hilbert space that is important to the problem under consideration. Thus $\mathbf{A}$ becomes a finite but large-dimensional random linear operator, that is, a large-dimensional random matrix. Hence, the limiting behaviour of large-dimensional random matrices attracts special interest among experts in quantum mechanics,

and many limiting laws were discovered during that time. For a more detailed review on applications of RMT in quantum mechanics and other related areas in physics, the reader is referred to Mehta's (2004) *Random Matrices*.

Since the late 1950s, research on the limiting spectral properties of large-dimensional random matrices has attracted considerable interest among mathematicians, probabilists, and statisticians. One pioneering work is the semicircular law for a Gaussian (or Wigner) matrix (Wigner 1955; 1958). Wigner proved that the expected spectral distribution of a large-dimensional Wigner matrix tends to the semicircular law. This work was later generalised by Arnold (1967, 1971) and Grenander (1963) in various aspects. In another direction related to the class of Gaussian Wishart matrices, or more generally, the class of sample covariance matrices, breakthrough work was done by Marčenko and Pastur (1967) and Pastur (1972, 1973): the authors discovered the Marčenko-Pastur law under fairly general conditions. The asymptotic theory of spectral analysis of large-dimensional sample covariance matrices was later developed by many researchers, including Bai et al. (1986), Grenander and Silverstein (1977), Jonsson (1982), Wachter (1978), Yin (1986), and Yin and Krishnaiah (1983). Also, Bai et al. (1986, 1987), Silverstein (1985), Wachter (1980), Yin (1986), and Yin and Krishnaiah (1983) investigated the limiting spectral distribution of the multivariate Fisher matrix, or more generally, of products of random matrices (a random Fisher matrix is the product of a sample covariance matrix by the inverse of another independent sample covariance matrix). The early 1980s saw major contributions on the existence of limiting spectral distributions and their explicit forms for certain classes of random matrices. In particular, Bai and Yin (1988) proved that the spectral distribution of a sample covariance matrix (suitably normalised) tends to the semicircular law when the dimension is relatively smaller than the sample size. In recent years, research on RMT is turning toward the second-order limiting theorems, such as the central limit theorem for linear spectral statistics, the limiting distributions of spectral spacings, and extreme eigenvalues.

1.3 Eigenvalue Statistics of Large Sample Covariance Matrices

This book is about the theory of large sample covariance matrices and their applications to high-dimensional statistics. Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ be a sample of random observations of dimension p. The population covariance matrix is denoted by $\mathbf{\Sigma} = \mathrm{cov}(\mathbf{x}_i)$. The corresponding *sample covariance matrix* is defined as

$$\mathbf{S}_n = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^*, \tag{1.1}$$

where $\overline{\mathbf{x}} = n^{-1}\sum_i \mathbf{x}_i$ denotes the *sample mean*. Almost all statistical methods in multivariate analysis rely on this sample covariance matrix: principle component analysis, canonical correlation analysis, multivariate regressions, one-sample or two-sample hypothesis testing, factor analysis, and so on.

A striking fact in multivariate analysis of large-dimensional statistics is that many important statistics are functions of the eigenvalues of sample covariance matrices. The statistic T_n in Example 1.1 is of this type, and following is yet another example.

Example 1.2 Let the covariance matrix of a population have the form $\mathbf{\Sigma} = \mathbf{\Sigma}_q + \sigma^2\mathbf{I}$, where $\mathbf{\Sigma}$ is $p \times p$ and $\mathbf{\Sigma}_q$ has rank q ($q < p$). Suppose $\mathbf{S}_n$ is the sample covariance matrix based on a sample of size n drawn from the population. Denote the eigenvalues of $\mathbf{S}_n$ by $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$. Then the test statistic for the hypothesis H_0: $\text{rank}(\mathbf{\Sigma}_q) = q$ against H_1: $\text{rank}(\mathbf{\Sigma}_q) > q$ is given by

$$Q_n = \frac{1}{p-q}\sum_{j=q+1}^{p}\lambda_j^2 - \left(\frac{1}{p-q}\sum_{j=q+1}^{p}\lambda_j\right)^2.$$

In other words, the test statistic Q_n is the variance of the $p-q$ smallest eigenvalues of $\mathbf{S}_n$.

Therefore, understanding the asymptotic properties of eigenvalue statistics such as T_n and Q_n has paramount importance in data analysis when the dimension p is getting large with respect to the sample size. The spectral analysis of large-dimensional sample covariance matrices from RMT provides powerful tools for the study of such eigenvalue statistics. For instance, the Marčenko-Pastur law describes the global behaviour of the p eigenvalues of a sample covariance matrix so that pointwise limits of eigenvalue statistics are determined by integrals of appropriate functions with respect to the Marčenko-Pastur law (see Example 1.1 for the case of T_n). Moreover, fluctuations of these eigenvalue statistics are described by central limit theorems found in Bai and Silverstein (2004) and Zheng (2012). Similarly to the case of classical large sample theory, such CLTs constitute the cornerstones of statistical inference with large-dimensional data.

1.4 Organisation of the Book

The book has a quite simple structure. The first set of chapters presents in detail the core of fundamental results from RMT regarding sample covariance matrices and random Fisher matrices. These results are selected in such a way that they are applied and used in the subsequent chapters of the book. More specifically, Chapter 2 introduces the limiting spectral distributions of general sample covariance matrices, namely, the Marčenko-Pastur distributions, and the limiting spectral distributions of random Fisher matrices. Detailed examples of both limits are also provided. In Chapter 3, the two fundamental CLTs from Bai and Silverstein (2004) and Zheng (2012) are presented in detail. Simple application examples of these CLTs are given. We also introduce a *substitution principle* that deals with the effect in the CLTs induced by the use of adjusted sample sizes $n_i - 1$ in place of the (raw) sample sizes n_i in the definition of sample covariance matrices and Fisher matrices.

The remaining chapters collect large-dimensional statistical problems where the classical large sample methods fail and the new asymptotic methods from the RMT provide a valuable remedy. The problems run from the "simple" and classical two-sample test problem (Chapter 5) to the current and advanced topic of the Markowitz portfolio optimisation problem (Chapter 12). Topics from Chapters 4–9 are classical topics in multivariate analysis; they are here re-analysed under the large-dimensional scheme. The last three chapters cover three modern topics in large-dimensional statistics. Methods and results reported in those chapters have been so far available only in research papers.

A characteristic feature of the book is that Chapters 4–12 are quite independent each other so that they can be read in an arbitrary order once the material in Chapters 2 and 3 is understood. Notice, however, that dependence between some of these chapters might exist occasionally, but this remains very limited.

Finally, we have included an appendix to introduce the basics on contour integration. The reason is that in the CLTs developed in Chapter 3 for linear spectral statistics of sample covariance matrices and of random Fisher matrices, the mean and covariance functions of the limiting Gaussian distributions are expressed in terms of contour integrals, and explicit calculations of such contour integrals frequently appear in various chapters of this book. As such calculations are not always taught in non-mathematical curricula, it is hoped that the appendix will help the reader to follow some basic calculations in the use of CLTs.

Notes

On the interplay between random matrix theory and large-dimensional statistics, supplementary information can be found in the excellent introductory papers by Bai (2005), Johnstone (2007), and Johnstone and Titterington (2009). A more recent review is by Paul and Aue (2014).

2

Limiting Spectral Distributions

2.1 Introduction

Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ be a sample of random observations of dimension p. The *sample covariance matrix* is defined as

$$\mathbf{S}_n = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^* = \frac{1}{n-1}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^* - \frac{n}{n-1}\overline{\mathbf{x}}\overline{\mathbf{x}}^*, \tag{2.1}$$

where $\overline{\mathbf{x}} = n^{-1}\sum_i \mathbf{x}_i$ denotes the *sample mean*. Many of traditional multivariate statistics are functions of the eigenvalues $\{\lambda_k\}$ of the sample covariance matrix $\mathbf{S}_n$. In the most basic form, such statistics can be written as

$$T_n = \frac{1}{p}\sum_{k=1}^{p}\varphi(\lambda_k) \tag{2.2}$$

for some specific function φ. Such a statistic is called a *linear spectral statistic* of the sample covariance matrix $\mathbf{S}_n$. For example, the so-called *generalised variance* discussed later in Chapter 4 (see (4.1)) is

$$T_n = \frac{1}{p}\log|\mathbf{S}_n| = \frac{1}{p}\sum_{k=1}^{p}\log(\lambda_k).$$

So this particular T_n is a linear spectral statistic of the sample covariance matrix $\mathbf{S}_n$ with "test function" $\varphi(x) = \log(x)$.

In two-sample multivariate analysis with, say, an $\mathbf{x}$-sample and a $\mathbf{y}$-sample, interesting statistics will still be of the previous form in (2.2), where, however, the eigenvalues $\{\lambda_k\}$ will be those of the so-called *Fisher matrix* $\mathbf{F}_n$. Notice that each of the two examples has a corresponding sample covariance matrix, say, $\mathbf{S}_\mathbf{x}$ and $\mathbf{S}_\mathbf{y}$. The Fisher matrix associated with these samples is the quotient of the two sample matrices, namely, $\mathbf{F}_n = \mathbf{S}_\mathbf{x}\mathbf{S}_\mathbf{y}^{-1}$ (assuming the latter is invertible).

Linear spectral statistics of sample covariance matrices or Fisher matrices are at the heart of the new statistical tools developed in this book. In this chapter and the next, we introduce the theoretical backgrounds of these statistics. More specifically, this chapter deals with the first-order limits of such statistics, namely, to answer the question, when and how T_n should converge to some limiting value ℓ as both the dimension p and the sample size grow to infinity?

Clearly the question should relate to the "joint limit" of the p eigenvalues $\{\lambda_k\}$. The formal concepts to deal with the question are called the *empirical spectral distributions* and *limiting*

spectral distributions. This chapter, introduces these distributions for the sample covariance matrix $\mathbf{S}_n$ and the two-sample Fisher matrix $\mathbf{F}_n$.

2.2 Fundamental Tools

This section introduces some fundamental concepts and tools used throughout the book.

2.2.1 Empirical and Limiting Spectral Distributions

Let $\mathcal{M}_p(\mathbb{C})$ be the set of $p \times p$ matrices with complex-valued elements.

Definition 2.1 Let $\mathbf{A} \in \mathcal{M}_p(\mathbb{C})$ and $\{\lambda_j\}_{1\le j\le p}$; its *empirical spectral distribution* (ESD) is

$$F^{\mathbf{A}} = \frac{1}{p}\sum_{j=1}^{p} \delta_{\lambda_j},$$

where δ_a denotes the Dirac mass at a point a.

In general, the ESD $F^{\mathbf{A}}$ is a probability measure on $\mathbb{C}$; it has support in $\mathbb{R}$ (resp. on $\mathbb{R}_+$) if $\mathbf{A}$ is Hermitian (resp. nonnegative definite Hermitian). For example, the two-dimensional rotation

$$\mathbf{A} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

has eigenvalues $\pm i$ so that $F^{\mathbf{A}} = \frac{1}{2}(\delta_{\{i\}} + \delta_{\{-i\}})$ is a measure on $\mathbb{C}$, whereas the symmetry

$$\mathbf{B} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

has eigenvalues ± 1 so that $F^{\mathbf{B}} = \frac{1}{2}(\delta_{\{1\}} + \delta_{\{-1\}})$ has support on $\mathbb{R}$. In this book, we are mainly concerned with covariance matrices. Because these are Hermitian and nonnegative definite, the corresponding ESDs will have support on $\mathbb{R}_+$.

Definition 2.2 Let $\{\mathbf{A}_n\}_{n\ge 1}$ be a sequence from $\mathcal{M}_p(\mathbb{C})$. If the sequence of corresponding ESDs $\{F^{\mathbf{A}_n}\}_{n\ge 1}$ vaguely converges to a (possibly defective) measure F, we call F the *limiting spectral distribution* (LSD) of the sequence of matrices $\{\mathbf{A}_n\}$.

The preceding vague convergence means that for any continuous and compactly supported function φ, $F^{\mathbf{A}_n}(\varphi) \to F(\varphi)$ as $n \to \infty$. It is well known that if the LSD F is indeed non-defective, that is, $\int F(dx) = 1$, the vague convergence turns into the stronger (usual) weak convergence, that is, $F^{\mathbf{A}_n}(\varphi) \to F(\varphi)$, for any continuous and bounded function φ.

When dealing with a sequence of sample covariance matrices $\{\mathbf{S}_n\}$, their eigenvalues are random variables and the corresponding ESDs $\{F^{\mathbf{S}_n}\}$ are random probability measures on $\mathbb{R}_+$. A fundamental question in random matrix theory is whether the sequence $\{F^{\mathbf{S}_n}\}$ has a limit (in probability or almost surely).

2.2.2 Stieltjes Transform

The eigenvalues of a matrix are continuous functions of entries of the matrix. But these functions have no closed forms when the dimension of the matrix is larger than 4. So special methods are needed for their study. Three important methods are employed in this area: moment method, Stieltjes transform, and orthogonal polynomial decomposition of the exact density of eigenvalues. For the sake of our exposition, we concentrate on the Stieltjes transform method, which is indeed widely used in the literature of large-dimensional statistics.

We denote by Γ_μ the support of a finite measure μ on $\mathbb{R}$. Let

$$\mathbb{C}^+ := \{z \in \mathbb{C} \,:\, \Im(z) > 0\}$$

be the (open) upper half complex plan with positive imaginary part.

Definition 2.3 Let μ be a finite measure on the real line. Its Stieltjes transform (also called the Cauchy transform in the literature) is defined as

$$s_\mu(z) = \int \frac{1}{x - z}\mu(dx), \quad z \in \mathbb{C} \setminus \Gamma_\mu.$$

The results of this section are given without proofs; they can be found in textbooks such as Kreĭn and Nudel′man (1977).

Proposition 2.4 *The Stieltjes transform has the following properties:*

1. s_μ *is holomorphic on* $\mathbb{C} \setminus \Gamma_\mu$
2. $z \in \mathbb{C}^+$ *if and only if* $s_\mu(z) \in \mathbb{C}^+$
3. *If* $\Gamma_\mu \subset \mathbb{R}_+$ *and* $z \in \mathbb{C}^+$, *then* $zs_\mu(z) \in \mathbb{C}^+$
4. $|s_\mu(z)| \le \dfrac{\mu(1)}{dist(z, \Gamma_\mu) \vee |\Im(z)|}$

The next result is an inversion result.

Proposition 2.5 *The mass* $\mu(1)$ *can be recovered through the formula*

$$\mu(1) = \lim_{v \to \infty} -iv s_\mu(iv).$$

Moreover, for all continuous and compactly supported φ: $\mathbb{R} \to \mathbb{R}$,

$$\mu(\varphi) = \int_{\mathbb{R}} \varphi(x)\mu(dx) = \lim_{v \downarrow 0} \frac{1}{\pi} \int_{\mathbb{R}} \varphi(x) \Im s_\mu(x + iv)dx.$$

In particular, for two continuity points $a < b$ *of* μ,

$$\mu([a, b]) = \lim_{v \downarrow 0} \frac{1}{\pi} \int_a^b \Im s_\mu(x + iv)dx.$$

The next proposition characterises functions that are Stieltjes transforms of bounded measures on $\mathbb{R}$.

Proposition 2.6 *Assume that the following conditions hold for a complex valued function* $g(z)$:

1. g *is holomorphic on* $\mathbb{C}^+$

2. $g(z) \in \mathbb{C}^+$ *for all* $z \in \mathbb{C}^+$
3. $\limsup_{v\to\infty} |ivg(iv)| < \infty$

Then g *is the Stieltjes transform of a bounded measure on* $\mathbb{R}$.

Similar to the characterisation of the weak convergence of finite measures by the convergence of their Fourier transforms, Stieltjes transform characterises the vague convergence of finite measures. This a key tool for the study of the ESDs of random matrices.

Theorem 2.7 *A sequence* $\{\mu_n\}$ *of probability measures on* $\mathbb{R}$ *converges vaguely to some positive measure* μ *(possibly defective) if and only if their Stieltjes transforms* $\{s_{\mu_n}\}$ *converge to* s_μ *on* $\mathbb{C}^+$.

To get the weak convergence of $\{\mu_n\}$, one checks the vague convergence of the sequence using this theorem and then to ensure that the limiting measure μ is a probability measure, that is, to check $\mu(1) = 1$ through Proposition 2.5 or by some direct observation.

The Stieltjes transform and the RMT are closely related to each other. Indeed, the Stieltjes transform of the ESD $F^{\mathbf{A}}$ of a $n \times n$ Hermitian matrix $\mathbf{A}$ is by definition

$$s_{\mathbf{A}}(z) = \int \frac{1}{x - z} F^{\mathbf{A}}(dx) = \frac{1}{n} \operatorname{tr}(\mathbf{A} - z\mathbf{I})^{-1}, \tag{2.3}$$

which is the resolvent of the matrix $\mathbf{A}$ (up to the factor $1/n$). Using a formula for the trace of an inverse matrix (see Bai and Silverstein, 2010, Theorem A.4), we have

$$s_n(z) = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{a_{kk} - z - \boldsymbol{\alpha}_k^*(\mathbf{A}_k - z\mathbf{I})^{-1}\boldsymbol{\alpha}_k}, \tag{2.4}$$

where $\mathbf{A}_k$ is the $(n-1) \times (n-1)$ matrix obtained from $\mathbf{A}$ with the kth row and column removed and $\boldsymbol{\alpha}_k$ is the kth column vector of $\mathbf{A}$ with the kth element removed. If the denominator $a_{kk} - z - \boldsymbol{\alpha}_k^*(\mathbf{A}_k - z\mathbf{I})^{-1}\boldsymbol{\alpha}_k$ can be proved to be equal to $g(z, s_n(z)) + o(1)$ for some function g, then a LSD F exists, and its Stieltjes transform is the solution to the equation

$$s = 1/g(z, s).$$

Its applications are discussed in more detail later in the chapter.

2.3 Marčenko-Pastur Distributions

The *Marčenko-Pastur distribution* F_{y,σ^2} *(M-P law) with index* y *and scale parameter* σ has the density function

$$p_{y,\sigma^2}(x) = \begin{cases} \frac{1}{2\pi xy\sigma^2}\sqrt{(b-x)(x-a)}, & \text{if } a \le x \le b, \\ 0, & \text{otherwise,} \end{cases} \tag{2.5}$$

with an additional point mass of value $1 - 1/y$ at the origin if $y > 1$, where $a = \sigma^2(1 - \sqrt{y})^2$ and $b = \sigma^2(1 + \sqrt{y})^2$. Here, the constant y is the dimension to sample size ratio index and σ^2 the scale parameter. The distribution has mean σ^2 and variance $y\sigma^4$. The support interval has a length of $b - a = 4\sigma^2\sqrt{y}$.