



Spectral asymptotic expansion of Wishart matrices with exploding moments

Nathan Noiry

Laboratoire Modal'X,
UPL, Université Paris Nanterre,
F92000 Nanterre France.
E-mail address: nathan.noiry@parisnanterre.fr

Abstract. We study random covariance matrices whose entries have exploding moments meaning that the ratio between their k -th moment and the k -th power of their standard deviation goes to infinity with the size of the matrix. We compute an asymptotic expansion of the limiting spectral measure in the critical regime when this measure is close to the Marčenko–Pastur distribution. Explicit computations are given in the two classical cases of Bernoulli and truncated heavy tailed entries.

1. Introduction

Let X_n be a real random matrix of size $n \times m$ with i.i.d. entries which are centered and with second moment M_2 . We define the Wishart matrix $W_n = \frac{1}{nM_2} X_n X_n^T$, where X_n^T is the transpose of X_n . The spectral measure of W_n is the random probability law:

$$\mu_{W_n} = \frac{1}{n} \sum_{\lambda \in \text{Spec}(W_n)} \delta_\lambda,$$

where $\text{Spec}(W_n)$ is the spectrum of W_n and δ_λ the Dirac at λ . Since W_n is a positive symmetric matrix, its eigenvalues are nonnegative reals. The work of [Marčenko and Pastur \(1967\)](#) implies that μ_{W_n} weakly converges to a probability law μ_α as $n, m \rightarrow +\infty$ and $m/n \rightarrow \alpha > 0$. The law μ_α is given by:

$$\mu_\alpha(dx) = \frac{\sqrt{(b-x)(x-a)}}{2\pi x} dx + \mathbf{1}_{\alpha < 1} (1-\alpha) \delta_0(dx),$$

where $a = (1 - \sqrt{\alpha})^2$ and $b = (1 + \sqrt{\alpha})^2$.

When the ratio between the k -th moment and the k -th power of the standard deviation of the entries goes to infinity with the size n , the limiting spectrum

Received by the editors May 24th, 2018; accepted June 26th, 2018.

2010 Mathematics Subject Classification. 05C80, 60B20.

Key words and phrases. Wishart matrices, Marčenko–Pastur distribution, Erdős–Rényi bipartite random graphs, heavy tailed random variables.

may not be μ_α . However, when this ratio is of order $n^{k/2-1}$, the spectral measure converges to a limiting distribution (see the work of [Benaych-Georges and Cabanal-Duvillard, 2012](#) and [Male, 2017](#)) that can still be close to the Marčenko-Pastur law.

In this paper, we focus on two models. When X_n has Bernoulli entries with parameter c/n , $c > 0$, the limiting probability law $\mu_{\alpha,c}$ depends only on α and c . When the entries of X_n are in the domain of attraction of a β -stable law, one can truncate them at the n -th lowest and largest quantiles times a parameter $B > 0$. The resulting spectral law $\mu_{\alpha,\beta,B}$ depends only on α, β and B . In each case, except for the existence, the limiting spectral law remains poorly understood.

The main concern of this paper is to obtain an asymptotic expansion of $\mu_{\alpha,c}$ and $\mu_{\alpha,\beta,B}$ as $c \rightarrow +\infty$ and $B \rightarrow 0$. We propose first order formulas in term of moments. In each case, the leading term is the Marčenko-Pastur law with parameter α . More interestingly, the order one perturbation term involves in each case a signed measure $\mu_\alpha^{(1)}$ with total mass 0 for which we are able to obtain an explicit expression:

$$\mu_\alpha^{(1)}(dx) = \frac{x^2 - 2(\alpha + 1)x + (\alpha^2 + 1)}{2\alpha\pi\sqrt{(b-x)(x-a)}} \mathbf{1}_{(a,b)}(x)dx. \quad (1.1)$$

These results are the content of Theorems [2.1](#) and [2.4](#). They suggest that somehow, $\mu_\alpha^{(1)}$ is a typical perturbative term when a sequence of measure converges to the Marčenko-Pastur law. In the Bernoulli case, $\mu_{\alpha,c}$ can be interpreted as some transform of the spectrum of a large bipartite Erdős-Rényi random graph with parameters n , $m = \alpha n$ and c/n . Therefore, our method provides a first order development of the limiting spectrum of sparse bipartite random graphs when c becomes large. This is the content of [Corollary 2.2](#).

The work of [Benaych-Georges and Cabanal-Duvillard \(2012, Theorem 3.2\)](#) provides a characterization of the limiting spectra in terms of moments. The combinatorics of the formula has the flavor of free probability and can be compared to the work of [Ryan \(1998\)](#) in the symmetric case. However, it does not lead to direct computations as we do here. Our proof is based on an alternative formula for the limiting moments obtained in [Proposition 3.1](#). It involves a certain class of walks on planar rooted trees more amenable to analysis and can be compared to the work of [Zakharevich \(2006\)](#) for Wigner matrices.

[Enriquez and Ménard \(2016\)](#) derived similar developments of moments for the limiting spectra of diluted random graphs. Their proof relies on a combinatorial analysis of the so-called local limit of the sequence of graphs. Although an analogous analysis could be done in the setting of bipartite random graphs, our method bypasses this argument and allows to study both bipartite random graphs *and* truncated heavy tailed covariance matrices as particular cases of a more general formula, namely the combinatorial expression obtained for the limiting moments. We also cover the *weighted* cases, studied in the setting of bipartite random graphs by [Vengerovsky \(2014\)](#), where he obtained recursive expressions for the limiting spectral moments, which cannot lead to our developments. This is the content of [Corollaries 2.3](#) and [2.5](#).

Let us finally mention that from a practical perspective, the study of large sparse covariance matrices, which corresponds to the Bernoulli setting, could be of interest for the statistics community. Interpreting the entries of X_n as grades given by individuals to some products, our study treats the case of independent grades.

Our techniques can be naturally extended to stochastic blocks models (e.g. two communities and two types of products).

Organization of the paper. In section 2, we state our main results concerning the first order asymptotic expansion of the limiting spectra of Bernoulli and truncated heavy-tailed covariance matrices. In section 3 we derive a new formula for the limiting moments of Wishart matrices with exploding moments and briefly explain how this leads to a new proof of the convergence of the spectral measure, given by [Benaych-Georges and Cabanal-Duvillard \(2012, Theorem 3.2\)](#). Section 4 is devoted to the proof of the results of section 2.

2. Main results

2.1. *Spectra of Erdős-Rényi Bipartite Random Graphs.* A bipartite Erdős-Rényi random graph with parameters n , m and p is a percolation with parameter p on the bipartite complete graph having parts of respective sizes n and m . Up to a relabeling of the vertices, its adjacency matrix can be written

$$A = A(n, m, p) = \begin{pmatrix} 0 & X_n^T \\ X_n & 0 \end{pmatrix},$$

where X_n is a $n \times m$ matrix with i.i.d. entries having Bernoulli law with parameter p . The associated empirical spectral measure is the random probability law

$$\nu_A = \frac{1}{n} \sum_{\lambda \in \text{Spec}(A)} \delta_\lambda$$

that put mass $1/n$ at each eigenvalue of A , counting multiplicities. Notice that

$$\text{Spec}(A) = \left\{ \pm \sqrt{\lambda_i(X_n X_n^T)} \right\}_{1 \leq i \leq n}$$

where $\lambda_1(X_n X_n^T), \dots, \lambda_n(X_n X_n^T)$ are the eigenvalues of $X_n X_n^T$. Let f be the bijection of \mathbf{R}_+ : $f(x) = \sqrt{x}$. For a measure ν on \mathbf{R}_+ , define $\text{Sym}(\nu)(\cdot) = (\nu(\cdot) + \nu(-\cdot))/2$ the symmetrized version of ν . Then

$$\nu_A = (\text{Sym} \circ f_*) \mu_{X_n X_n^T},$$

where $f_* \mu$ is the pushforward of a measure μ by f . This one-to-one correspondence between probability measures on \mathbf{R}_+ and symmetric probability measures on \mathbf{R} allows to directly work on the Wishart's setting and study $\mu_{X_n X_n^T}$.

We are interested in the diluted regime $m = \alpha n$, $\alpha > 0$ and $p = c/n$, $c > 0$. In that case, the object of interest is the renormalized adjacency matrix $\frac{1}{\sqrt{c}} A$, so we study the following sequence of Wishart matrices:

$$W_n = \frac{1}{c} X_n X_n^T.$$

In the diluted regime, a lot of entries of X_n are equal to zero, which prevents μ_{W_n} to converge to the Marčenko-Pastur law μ_α defined in the introduction. However, there still exists a limiting probability law $\mu_{\alpha, c}$ that only depends on α and c . This is a consequence of [Benaych-Georges and Cabanal-Duvillard \(2012, Theorem 3.2\)](#). To see this, it suffices to show that the centered version $W'_n =$

$c^{-1}(X_n - \mathbf{E}[X_n])(X_n - \mathbf{E}[X_n])^T$ has the same limiting spectra. Denoting respectively F and F' the cumulative distribution functions of μ_{W_n} and $\mu_{W'_n}$, a consequence of Lidskii's inequalities is that:

$$\|F - F'\|_\infty \leq \frac{\text{rk}(W_n - W'_n)}{n},$$

where rk is the rank operator. Therefore, μ_{W_n} has the same limit as $\mu_{W'_n}$.

Another enlightening approach due to [Bordenave and Lelarge \(2010\)](#) is directly concerned with $\nu_{\alpha,c} = (\text{Sym} \circ f_*)\mu_{\alpha,c}$. The two authors remark that the spectra of a graph is continuous with respect to the local convergence introduced in [Aldous and Lyons \(2007\)](#); [Benjamini and Schramm \(2001\)](#). In our setting, it can be shown that the limiting local law $\mathcal{L}_{\alpha,c}$ has support on unimodular random trees. More precisely,

$$\mathcal{L}_{\alpha,c} = \frac{1}{1+\alpha} \delta_{\mathbf{GW}(c,\alpha c)} + \frac{\alpha}{1+\alpha} \delta_{\mathbf{GW}(\alpha c,c)},$$

where $\mathbf{GW}(x,y)$ stands for a Galton-Watson tree where individuals in even (resp. odd) generations reproduce with Poisson law with parameter x (resp. y). The measure $\nu_{\alpha,c}$ remains largely misunderstood. It can be proved, using same arguments as in the work of [Salez \(2015\)](#), that its set of atoms is dense in \mathbf{R} . Besides, a consequence of the work of [Bordenave et al. \(2017\)](#) is that $\nu_{\alpha,c}$ possesses a continuous part if and only if $c > 1$. When $c \rightarrow +\infty$, $\nu_{\alpha,c}$ converges towards $(\text{Sym} \circ f_*)\mu_\alpha$. A natural question is then to describe how $\nu_{\alpha,c}$ differs from ν_α as c becomes large. Our result provides a characterization of the perturbation of order $1/c$ in terms of moments of the Wishart counterpart $\mu_{\alpha,c}$.

Theorem 2.1. *For all $k \geq 1$, as $c \rightarrow \infty$:*

$$\int_{\mathbf{R}} x^k d\mu_{\alpha,c}(x) = \int_{\mathbf{R}} x^k d\mu_\alpha(x) + \frac{1}{c} \int_{\mathbf{R}} x^k d\mu_\alpha^{(1)}(x) + o\left(\frac{1}{c}\right),$$

where $\mu_\alpha^{(1)}$ is defined in Equation (1.1).

It leads to an asymptotic development of the spectra of large bipartite Erdős-Rényi random graph at large intensity c :

Corollary 2.2. *For all $k \geq 1$, as $c \rightarrow \infty$:*

$$\int_{\mathbf{R}} x^k d\nu_{\alpha,c}(x) = \int_{\mathbf{R}} x^k d\nu_\alpha(x) + \frac{1}{c} \int_{\mathbf{R}} x^k d\nu_\alpha^{(1)}(x) + o\left(\frac{1}{c}\right),$$

where $\nu_\alpha^{(1)} = (\text{Sym} \circ f_*)\mu_\alpha^{(1)}$ is given by the density:

$$\frac{x^5 - 2(\alpha + 1)x^3 + (\alpha^2 + 1)x}{2\alpha\pi\sqrt{(b-x^2)(x^2-a)}} \mathbf{1}_{\sqrt{a} < |x| < \sqrt{b}}.$$

Finally, let us mention that our method also applies in the *weighted* case where we study $Y_n(i,j) = X_n(i,j) \times \xi_n(i,j)$ with $\xi_n(i,j)$'s i.i.d. with law that does not depend on n and having all moments finite. In that case, the perturbation term is just multiplied by $\mathbf{E}[\xi^4]/\mathbf{E}[\xi^2]$, where ξ has the same law as the $\xi_n(i,j)$'s.

Corollary 2.3. *In the weighted case, for all $k \geq 1$, as $c \rightarrow \infty$:*

$$\int_{\mathbf{R}} x^k d\nu_{\alpha,c}(x) = \int_{\mathbf{R}} x^k d\nu_\alpha(x) + \frac{1}{c} \cdot \frac{\mathbf{E}[\xi^4]}{\mathbf{E}[\xi^2]} \int_{\mathbf{R}} x^k d\nu_\alpha^{(1)}(x) + o\left(\frac{1}{c}\right).$$

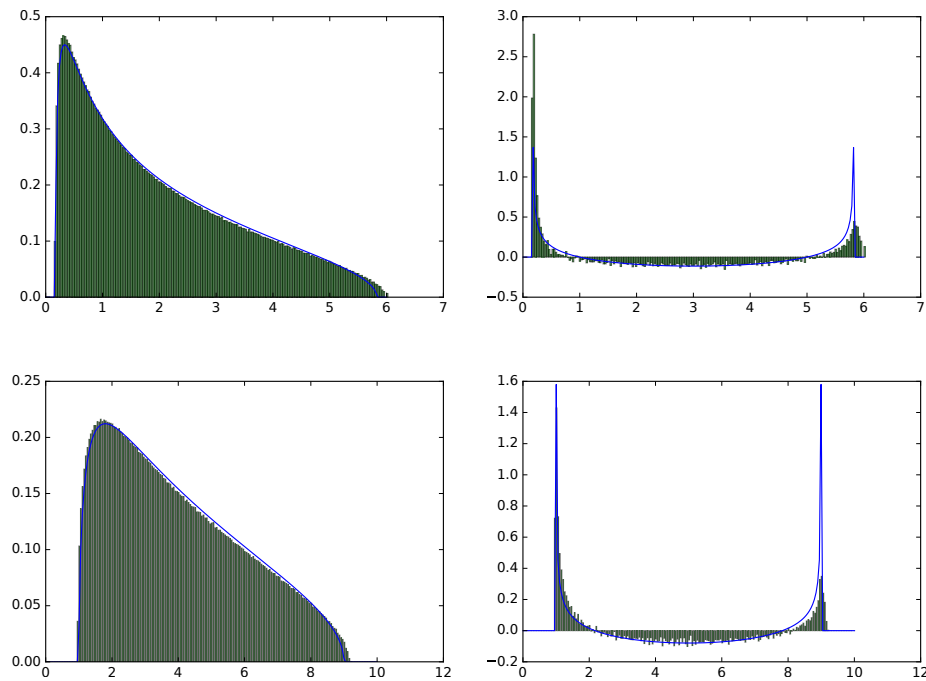


FIGURE 2.1. Numerical simulations for the spectrum of 100 Wishart matrices associated to random matrices of size $n \times \alpha n$ with i.i.d. entries with Bernoulli law of parameter c/n , with $c = 20$ and $n = 3000$. The theoretical densities of μ_α and $\mu_\alpha^{(1)}$ are drawn in blue. The top diagrams correspond to $\alpha = 2$ whereas the bottom diagrams correspond to $\alpha = 4$.

2.2. *Truncated heavy tailed random matrices.* For all $n \geq 1$, let X_n be a rectangular random matrix of size $n \times m$ having i.i.d. entries with heavy tailed law P which has cumulative distribution function F . As before, we suppose that the ratio m/n converges to $\alpha > 0$. We suppose that P is in the domain of attraction of a β -stable law, $\beta < 2$. By Breiman (1992, Theorem 9.34), it implies that there exist two reals $M^-, M^+ \geq 0$ such that $M^- + M^+ > 0$ and

$$\frac{F(-x)}{1 - F(x)} \xrightarrow{x \rightarrow +\infty} \frac{M^-}{M^+}, \quad (2.1)$$

We restrict our study to the case $M^+ > 0$: the following arguments easily adapt when $M^- > 0$ by considering $P(-\cdot)$.

Theorem 1.10 in Belinschi et al. (2009) ensures that the spectral measure of

$$n^{-\frac{2}{\beta}} X_n X_n^T$$

almost surely weakly converges towards a deterministic probability law $\mu_{\alpha, \beta}$ which only depends on α and β . We are here concerned with a truncated version. More precisely, define the quantiles q_n^- and q_n^+ by:

$$\begin{cases} F(q_n^-) & = 1/n \\ 1 - F(q_n^+) & = 1/n. \end{cases}$$

For all $B > 0$, we consider the sequence of random matrices $X_n^{(B)}$ of size $n \times m$ where

$$X_n^{(B)}(i, j) = X_n(i, j) \mathbf{1}_{Bq_n^- \leq X_n(i, j) \leq Bq_n^+} + Bq_n^- \mathbf{1}_{X_n(i, j) < Bq_n^-} + Bq_n^+ \mathbf{1}_{X_n(i, j) > Bq_n^+}.$$

Let $P_n^{(B)}$ be the law of the entries of $X_n^{(B)}$. This forms a sequence of probability measures that can be viewed as an approximation of P . The sequence of Wishart's matrices:

$$\frac{1}{nM_2(P_n^{(B)})} X_n^{(B)} X_n^{(B)T},$$

where $M_2(P_n^{(B)})$ is the second moment of $P_n^{(B)}$, satisfies the hypothesis of [Benaych-Georges and Cabanal-Duvillard \(2012, Theorem 3.2\)](#). Therefore, the associated sequence of spectral measures converges towards a probability law $\mu_{\alpha, \beta, B}$ which, as we will see, only depends on α, β and B .

The quantities q_n^- and q_n^+ corresponds to the lowest and largest n -th quantile of P . Therefore, our choice of law P_n can be interpreted as a truncation of the largest entries in each rows of X_n . If one had chosen a smaller order of truncation, one would have retrieved the Marčenko-Pastur regime. On the contrary, if one had chosen a larger order of truncation, the A_k 's defined in Equation (3.1) would have been all infinite, meaning that the truncation is not large enough to apply Theorem 3.2 of [Benaych-Georges and Cabanal-Duvillard \(2012\)](#). In this spirit, the parameter $B > 0$ can be seen as a finer adjustment of the truncation.

When $B \rightarrow 0$, we are able to obtain a first order expansion in terms of moments. Interestingly, it involves the signed measure $\mu_\alpha^{(1)}$ that also appears in the Bernoulli case.

Theorem 2.4. *For all $k \geq 1$, as $B \rightarrow 0$:*

$$\int_{\mathbf{R}} x^k d\mu_{\alpha, \beta, B}(x) = \int_{\mathbf{R}} x^k d\mu_\alpha(x) + B^\beta \cdot C(\beta, M^+, M^-) \int_{\mathbf{R}} x^k d\mu_\alpha^{(1)}(x) + o(B^\beta),$$

where $C(\beta, M^+, M^-) = \frac{(2-\beta)^2}{4-\beta} \cdot \frac{1}{1+(M^-/M^+)^{1/\beta}}$.

As in the Bernoulli case, our method directly applies to the weighted setting where we consider rectangular matrices $Y_n^{(B)}(i, j) = X_n^{(B)} \times \xi_n(i, j)$ with $\xi_n(i, j)$'s i.i.d. random variables, independent of $X_n^{(B)}$, whose law does not depend on n and has all moments finite. If ξ has the same law as this family, the corollary writes in the following way.

Corollary 2.5. *In the weighted case, for all $k \geq 1$, as $B \rightarrow 0$:*

$$\int_{\mathbf{R}} x^k d\mu_{\alpha, \beta, B}(x) = \int_{\mathbf{R}} x^k d\mu_\alpha(x) + B^\beta \cdot C(\beta, M^+, M^-) \frac{\mathbf{E}[\xi^4]}{\mathbf{E}[\xi^2]} \int_{\mathbf{R}} x^k d\mu_\alpha^{(1)}(x) + o(B^\beta).$$

3. Spectral moments of generalized Wishart's matrices

Our main results are obtained from a general formula that we derive for the limiting moments of size-dependent Wishart's matrices. More precisely, we consider the following setting. Let X_n be an $n \times m$ matrix having i.i.d. entries with centered law P_n which has k -th moment $M_k(P_n) < +\infty$. We make the following assumption:

$$\forall k \geq 2, \quad \frac{M_k(P_n)}{n^{k/2-1} M_2(P_n)^{k/2}} \xrightarrow{n \rightarrow +\infty} A_k \in [0, \infty). \tag{3.1}$$

In the regime $m/n \rightarrow \alpha > 0$, Benaych-Georges and Cabanal-Duvillard proved the convergence of the empirical spectral measures associated to the Wishart sequence:

$$W_n := \frac{1}{nM_2(P_n)} X_n X_n^T.$$

The limiting measure $\mu_{\mathcal{A}}$ only depends on $\mathcal{A} := (A_k)_{k \geq 2}$. They obtained a formula for the moments of $\mu_{\mathcal{A}}$ which has the flavor of free probability theory. We propose here an alternative formula which turns out to be more amenable to analysis for our purpose.

In order to properly state our result, we need to introduce the notion of word on a labeled graph. A labeled graph is a graph $G = (V, E)$ together with a labeling of the vertices, that is a one-to-one application from V to $\{1, \dots, |V|\}$. A relabeling of a labeled graph is a new choice of bijection between V and $\{1, \dots, |V|\}$. Note that there are $|V|!$ choices of labelings for a given graph $G = (V, E)$. A word of length $k \geq 1$ on a labeled graph G is a sequence of labels i_1, i_2, \dots, i_k such that $\{i_j, i_{j+1}\}$ is a pair of adjacent labels (that is the associated vertices are neighbours in G) for all $1 \leq j \leq k-1$. A word of length k is said to be closed if $i_1 = i_k$. Let $\mathbf{i} = i_1, \dots, i_k$ and $\mathbf{i}' = i'_1, \dots, i'_k$ be two words of length k on two labeled graphs G and G' having the same number of vertices. Then, \mathbf{i} and \mathbf{i}' are said to be equivalent if there exists a bijection σ of $\{1, \dots, |V|\}$ such that $\sigma(i_j) = i'_j$ for all $1 \leq j \leq k$. In words, \mathbf{i} and \mathbf{i}' are equivalents if there exists a relabeling of a G such that the word associated to \mathbf{i} is exactly \mathbf{i}' . One can check that this defines an equivalence relation on words on labeled graphs.

Recall that a planar rooted tree is a connected graph without loop embedded in the plane, with a distinguished vertex called the root. A vertex at odd (resp. even) distance from the root will be called an odd (resp. even) vertex. An edge with an odd (resp. even) origin vertex will be called an odd (resp. even) edge.

Proposition 3.1. *For all $k \geq 1$, the k -th moment of $\mu_{\mathcal{A}}$ is*

$$\int_{\mathbf{R}} x^k d\mu_{\mathcal{A}}(x) = \sum_{a=1}^k \sum_{l=1}^a \alpha^l \sum_{\substack{\mathbf{b}=(b_1, \dots, b_a) \\ b_1 \geq b_2 \geq \dots \geq b_a \geq 2 \\ b_1 + b_2 + \dots + b_a = 2k}} |\mathcal{W}_k(a, a+1, l, \mathbf{b})| \prod_{i=1}^a A_{b_i}. \quad (3.2)$$

where $\mathcal{W}_k(a, a+1, l, \mathbf{b})$ is a set of representatives of the equivalence classes of closed words on labeled rooted planar trees having “ a ” edges, of which l are odd edges, starting from the root and such that for all $1 \leq i \leq a$, one edge is browsed b_i times.

Remark 3.2. Let w be an element of $\mathcal{W}_k(a, a+1, l, \mathbf{b})$. Since it is a representative walk on a tree, starting and ending at the root, the multiplicity of each edge has to be even. In particular, \mathbf{b} must be a a -tuple of even integers summing to $2k$, and the sequence of odd parameters $(A_{2k})_{k \geq 1}$ characterizes the limiting law.

Proof: For all $k \geq 1$, denote by $M_k(\mu_{W_n})$ the k -th moment of μ_{W_n} . Its expected value $\mathbf{E}[M_k(\mu_{W_n})]$ is given by:

$$\frac{1}{n^{k+1} M_2(P_n)^k} \sum_{\substack{1 \leq i_1, \dots, i_k \leq n \\ 1 \leq j_1, \dots, j_k \leq m}} \mathbf{E}[X(i_1, j_1) X(i_2, j_2) \cdots X(i_k, j_k) X(i_1, j_k)]. \quad (3.3)$$

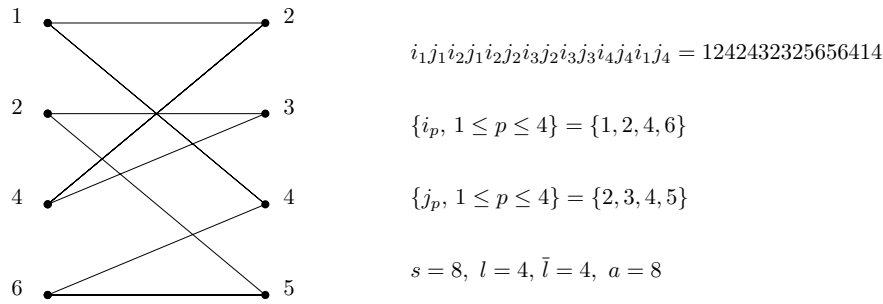


FIGURE 3.2. Example of a word (\mathbf{i}, \mathbf{j}) with its associated graph and quantities.

Denote (\mathbf{i}, \mathbf{j}) the generic word $i_1 j_1 i_2 \dots i_1 j_k$ appearing in (3.3). We define the bipartite graph $G = (V, E)$ associated to the word (\mathbf{i}, \mathbf{j}) by:

$$V = \{(i_r, \mathbf{i}), (j_r, \mathbf{j}); 1 \leq r \leq k\};$$

and

$$E = \{(i_r, \mathbf{i}), (j_r, \mathbf{j})\}, \{(i_{r+1}, \mathbf{i}), (j_r, \mathbf{j})\}; 1 \leq r \leq k\},$$

where we used the convention $k + 1 = 1$. The abstract symbols \mathbf{i} and \mathbf{j} are needed to obtain a bipartite graph since the i_r 's and j_r 's can have common values (see Figure 3.2 for an illustration). We will refer to (\mathbf{i}) - and (\mathbf{j}) -letters. In words, the vertices of G are the letters of the word (\mathbf{i}, \mathbf{j}) and two vertices are linked by an edge when they are consecutive in (\mathbf{i}, \mathbf{j}) . Denote by s the number of vertices, a the number of edges, l the number of \mathbf{j} -vertices and \bar{l} the number of \mathbf{i} -vertices in the word. Since G is connected, $s \leq a + 1$. Moreover, since P_n has zero mean, each edge must appear at least twice in the word to give a non-zero contribution in (3.3). As a consequence we obtain the bound $a \leq k$ because $i_1 j_1 \dots j_k$ possesses $2k$ edges counted with multiplicity.

Two words (\mathbf{i}, \mathbf{j}) and $(\mathbf{i}', \mathbf{j}')$ are said equivalent if one can find a permutation σ of $\{1, \dots, n\}$ and another one τ of $\{1, \dots, m\}$ such that

$$\forall p \in \{1, \dots, k\}, \quad \sigma(i_p) = i'_p \text{ and } \tau(j_p) = j'_p.$$

One can check that this is an equivalence relation on the words appearing in (3.3). Note that (\mathbf{i}, \mathbf{j}) has

$$C(s, l) = n(n - 1) \dots (n - l + s + 1) \times m(m - 1) \dots (m - l + 1) \sim \alpha^l n^s$$

equivalents. Fix $a \in \{1, \dots, k\}$, $1 \leq s \leq a + 1$ and $1 \leq l \leq a$. Let $\mathcal{B}_{a,k}$ be the set of a -tuples $\mathbf{b} = (b_1, \dots, b_a)$ of integers such that

- (1) $b_1 \geq b_2 \geq \dots \geq b_a \geq 2$;
- (2) $b_1 + \dots + b_a = 2k$.

For all $k \geq 1$ and $\mathbf{b} \in \mathcal{B}_{a,k}$, we introduce $\mathcal{W}_k(a, s, l, \mathbf{b})$ a set of representatives of the equivalence classes of words (\mathbf{i}, \mathbf{j}) such that the associated graph has a edges, s vertices of which l are \mathbf{j} -vertices and such that for all $1 \leq i \leq a$ there is an edge which has multiplicity b_i in (\mathbf{i}, \mathbf{j}) . We can rewrite (3.3) as:

$$\sum_{a=1}^k \sum_{s=1}^{a+1} \sum_{l=1}^s \frac{C(s, l)}{n^{a+1}} \sum_{\mathbf{b} \in \mathcal{B}_{a,k}} \sum_{(\mathbf{i}, \mathbf{j}) \in \mathcal{W}(a, s, l, \mathbf{b})} \prod_{1 \leq i \leq a} \frac{M_{b_i}(P_n)}{n^{b_i/2-1} M_2(P_n)^{b_i/2}}. \tag{3.4}$$

Since $C(s, l)n^{-a-1} \sim \alpha^l n^{s-a-1}$ when $n \rightarrow +\infty$ we deduce that when $s < a + 1$ the asymptotic contribution is zero. Hence a possible non-zero contribution arises only when $s = a + 1$. The Proposition is a consequence of (3.1). \square

Remark that our method could lead to an alternative proof of Benaych-Georges and Cabanal-Duvillard result. We only give a sketch of the remaining steps and do not enter the details here as it is not our main purpose. The variance of $M_k(\mu_{W_n})$ can be written

$$\frac{1}{n^{2(k+1)}M_2(P_n)^{2k}} \sum_{(\mathbf{i}, \mathbf{j}), (\mathbf{i}', \mathbf{j}')} \left(\mathbf{E}[P(\mathbf{i}, \mathbf{j})P(\mathbf{i}', \mathbf{j}')] - \mathbf{E}[P(\mathbf{i}, \mathbf{j})]\mathbf{E}[P(\mathbf{i}', \mathbf{j}')] \right),$$

where $P(\mathbf{i}, \mathbf{j})$ is the product $X(i_1, j_1)X(i_2, j_1) \cdots X(i_1, j_k)$. A combinatorial analysis of the contributions associated to words (\mathbf{i}, \mathbf{j}) and $(\mathbf{i}', \mathbf{j}')$ can lead to $\text{Var}(M_k(\mu_{W_n})) = O(1/n)$. Then, it can be shown that the limiting k -th moment does not grow faster than k^{ck} for a constant $c > 0$, which ensures that the limiting law is characterized by its moments.

4. Proof of the main results

The formula for the moments of the limiting law can be rewritten:

$$\sum_{l=1}^k \alpha^l |\mathcal{W}_k(k, k+1, l, (2, \dots, 2))| + A_4 \sum_{l=1}^{k-1} \alpha^l |\mathcal{W}_k(k, k+1, l, (4, 2, \dots, 2))| + \dots \quad (4.1)$$

We used that $A_2 = 1$ and that:

- a word of length $2k + 1$ on a planar rooted tree having k edges that starts and ends at the root and browses every edge *has to* browse every edge exactly 2 times;
- a word of length $2k + 1$ on a planar rooted tree having $k - 1$ edges that starts and ends at the root and browses every edge *has to* browse every edge 2 times except for one that is browsed 4 times.

The (numerous) remaining terms (represented by the “...”) involve the A_k 's for $k > 4$.

The first term of (4.1) is known to be the k -th moment of the Marčenko-Pastur law with parameter α . We will show that the second term

$$\sum_{l=1}^{k-1} \alpha^l |\mathcal{W}_k(k, k+1, l, (4, 2, \dots, 2))|$$

is the k -th moment of $\mu_\alpha^{(1)}$. Finally, we will see that A_4 is equal to $1/c$ in the Bernoulli case and to $B^\beta \cdot C(\beta, M^+, M^-)$ in the truncated heavy-tailed case, by identifying the asymptotic coefficients A_k 's in both settings.

4.1. *Identification of the signed measure $\mu_\alpha^{(1)}$.* Although this is a known result, we briefly prove that

$$\int_{\mathbf{R}} x^k d\mu_\alpha(x) = \sum_{l=1}^k \alpha^l |\mathcal{W}_k(k, k+1, l, (2, \dots, 2))|$$

in order to introduce some notations.

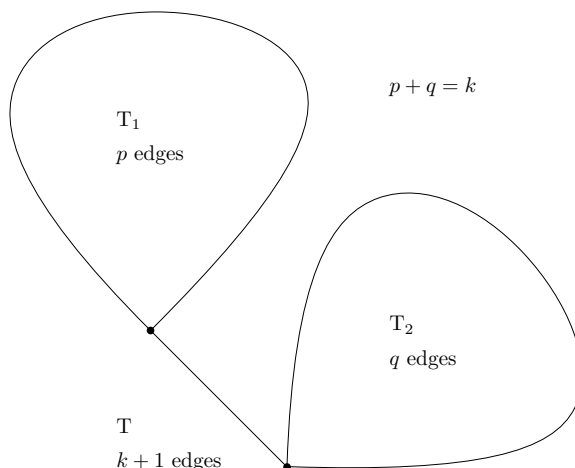


FIGURE 4.3. Decomposition of a planar tree.

Recall that $\mathcal{W}_k(k, k + 1, l, (2, \dots, 2))$ is a set of representatives of closed words starting at the root, of length $2k + 1$ on labeled planar rooted trees having k edges, l of these being odd edges. Therefore,

$$\sum_{l=1}^k \alpha^l |\mathcal{W}_k(k, k + 1, l, (2, \dots, 2))| = \sum_{T \in \mathcal{T}_k} \alpha^{l(T)},$$

where \mathcal{T}_k is the set of planar rooted trees having k edges. For convenience, we introduce

$$a_k := \sum_{T \in \mathcal{T}_k} \alpha^{l(T)} \quad \text{and} \quad b_k := \sum_{T \in \mathcal{T}_k} \alpha^{\bar{l}(T)},$$

where $\bar{l}(T)$ is the number of even edges of a given tree $T \in \mathcal{T}_k$.

It turns out that the a_k 's are the moments of μ_α . To obtain the term of order $1/c$ we will need to compute the generating series of the a_k 's and b_k 's.

Let T be a planar tree having $k + 1$ edges. Let T_1 be the tree induced by the first child of the root and T_2 the connected component of the root after removing the edge between the root and its first child (see Figure 4.3). Denoting p (resp. q) the number of edges of T_1 (resp. T_2), we have $p + q = k$. It is straightforward to obtain the relations $l(T) = 1 + \bar{l}(T_1) + l(T_2)$ and $\bar{l}(T) = l(T_1) + \bar{l}(T_2)$. Therefore

$$\begin{cases} a_{k+1} = \alpha \sum_{p+q=k} a_p b_q \\ b_{k+1} = \sum_{p+q=k} a_p b_q. \end{cases}$$

Denoting $A(z) = \sum_{k \geq 0} a_k z^k$ and $B(z) = \sum_{k \geq 0} b_k z^k$ the generating functions of the a_k 's and the b_k 's we obtain the functional relations:

$$\begin{cases} A = 1 + \alpha z AB \\ B = 1 + z AB. \end{cases} \tag{4.2}$$

It implies that $zA^2 + (\alpha z - z - 1)A + 1 = 0$. If we denote $S(z) := -z^{-1}A(z^{-1})$ the Stieltjes transform of the measure with moments a_k 's, then S satisfies the equation:

$$zS^2 - (\alpha - z - 1)S + 1 = 0. \tag{4.3}$$

The function S of the variable $z \in \mathbf{C}_+$ is the limit of the Stieltjes transform of the μ_{W_n} when $c \rightarrow +\infty$. The imaginary part of a Stieltjes transform is positive: this allows us to choose the right solution for equation (4.3). For a complex z , if we denote \sqrt{z} the square root having a positive imaginary part on the upper half plane:

$$S(z) = \frac{\alpha - z - 1 + \sqrt{(z-b)(z-a)}}{2z},$$

where $a = (1 - \sqrt{\alpha})^2$ and $b = (1 + \sqrt{\alpha})^2$. This is the Stieltjes transform of the Marčenko-Pastur law μ_α , as announced.

The second term involves $\mathcal{W}_k(k-1, k, l, (4, 2, \dots, 2))$ a set of equivalence classes of closed words of length $2k+1$ on labeled planar rooted tree having $k-1$ edges, starting at the root and such that each edge is browsed exactly two times except one which is browsed four times. Let us denote

$$a_k^{(1)} = \sum_{l=1}^{k-1} \alpha^l |\mathcal{W}_k(k-1, k, l, (4, 2, \dots, 2))|,$$

and

$$b_k^{(1)} = \sum_{\bar{l}=1}^{k-1} \alpha^{\bar{l}} |\mathcal{W}_k(k-1, k, \bar{l}, (4, 2, \dots, 2))|.$$

The associated generating series will be denoted $A^{(1)}$ and $B^{(1)}$. Notice that by definition $a_0^{(1)} = a_1^{(1)} = b_0^{(1)} = b_1^{(1)} = 0$. We are going to obtain a recursion linking the four generating series $A, B, A^{(1)}$ and $B^{(1)}$. The idea is to use a first generation decomposition of the planar rooted tree on which the words are written, and then to distinguish whether or not the quadruple edge is an edge of this generation. For all $k \geq 1$, we partition $\mathcal{W}_k(k-1, k, l, (4, 2, \dots, 2))$ into two parts:

$$\mathcal{W}_k^{(0)}(k-1, k, l, (4, 2, \dots, 2)) \bigsqcup \mathcal{W}_k^{(1)}(k-1, k, l, (4, 2, \dots, 2)),$$

where $\mathcal{W}_k^{(0)}(k-1, k, l, (4, 2, \dots, 2))$ is the set of representative belonging to $\mathcal{W}_k(k-1, k, l, (4, 2, \dots, 2))$ such that the quadruple edge is not a first generation edge, and $\mathcal{W}_k^{(1)}(k-1, k, l, (4, 2, \dots, 2))$ is the set of representatives belonging to $\mathcal{W}_k(k-1, k, l, (4, 2, \dots, 2))$ such that the quadruple edge is a first generation edge. The associated quantities will be denoted $a_k^{(1,0)}, a_k^{(1,1)}, A^{(1,0)}, \dots$. For example:

$$a_k^{(1,0)} = \sum_{l=1}^{k-1} \alpha^l |\mathcal{W}_k^{(0)}(k-1, k, l, (4, 2, \dots, 2))|.$$

A representative word $(\mathbf{i}, \mathbf{j}) \in \mathcal{W}_k(k-1, k, l, (4, 2, \dots, 2))$ can be written:

$$(\mathbf{i}, \mathbf{j}) = i_1 \mathbf{S}_1 \zeta \xi \mathbf{S}_2 \xi \zeta \mathbf{S}_3 \zeta \xi \mathbf{S}_4 \xi \zeta \mathbf{S}_5 i_1,$$

where:

- (1) $i_1 \mathbf{S}_1 \zeta \mathbf{S}_5 i_1$ is the contour of a planar tree having p_1 edges;
- (2) $\xi \mathbf{S}_2 \xi$ is the contour of a planar tree having p_2 edges;
- (3) $\zeta \mathbf{S}_3 \zeta$ is the contour of a planar tree having p_3 edges;
- (4) $\xi \mathbf{S}_4 \xi$ is the contour of a planar tree having p_4 edges;
- (5) $\xi \mathbf{S}_2 \xi \mathbf{S}_4 \xi$ is the contour of a planar tree having $p_2 + p_4$ edges.

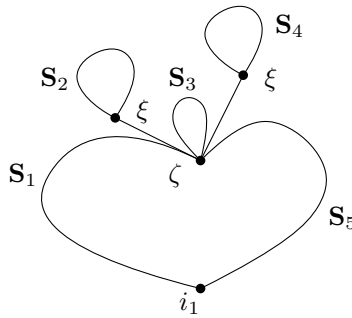


FIGURE 4.4. The writing (\mathbf{i}, \mathbf{j}) and its quadruple edge $\{\zeta, \xi\}$.

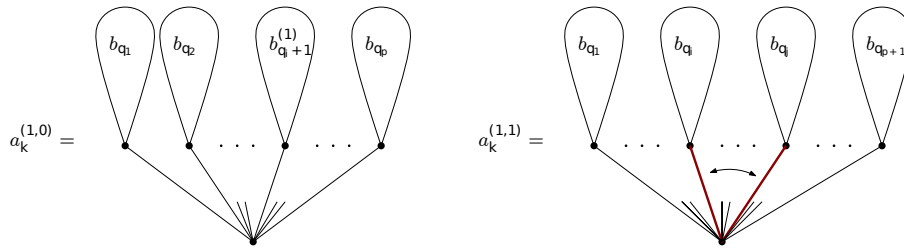


FIGURE 4.5. First edge decomposition of a word respectively in $\mathcal{W}_k^{(0)}(k - 1, k, l, (4, 2, \dots, 2))$ on the left and in $\mathcal{W}_k^{(1)}(k - 1, k, l, (4, 2, \dots, 2))$ on the right, where the quadruple edge is in red.

The above integers satisfy $p_1 + p_2 + p_3 + p_4 = k - 2$. See Figure 4.4 for an illustration. All of these conditions are sufficient to define a class of canonical representatives. Let T be the planar rooted tree on which a representative word (\mathbf{i}, \mathbf{j}) is written. Denote e_4 the quadruple edge, $T \setminus e_4$ the connected component of the root after removing e_4 and T^{e_4} the planar rooted tree formed by the descendants of e_4 . Then, the above conditions ensures that (\mathbf{i}, \mathbf{j}) is such that $T \setminus e_4$ and T^{e_4} are respectively browsed in lexicographic order.

Let $(\mathbf{i}, \mathbf{j}) \in \mathcal{W}_k^{(0)}(k - 1, k, l, (4, 2, \dots, 2))$. The underlying tree can have $p \in \{1, \dots, k - 2\}$ edges which are all browsed two times by (\mathbf{i}, \mathbf{j}) . One of the tree induced by the children of the root contains the quadruple edge, leading to p different choices. On another side, if $(\mathbf{i}, \mathbf{j}) \in \mathcal{W}_k^{(1)}(k - 1, k, l, (4, 2, \dots, 2))$ then the underlying tree can have $p \in \{1, \dots, k - 1\}$ edges out of which one is the quadruple edge. There are $\binom{p+1}{2}$ choices for the locations of the the visits of the quadruple edge. See Figure 4.5 for an illustration.

As a consequence, we get the following recursions:

$$a_k^{(1,0)} = \sum_{p=1}^{k-2} \alpha^p p \sum_{q_1 + \dots + q_p = k-p-1} b_{q_1+1}^{(1)} b_{q_2} \dots b_{q_p},$$

and

$$a_k^{(1,1)} = \sum_{p=1}^{k-2} \alpha^p \binom{p+1}{2} \sum_{q_1 + \dots + q_{p+1} = k-p-1} b_{q_1} b_{q_2} \dots b_{q_{p+1}}.$$

This yields

$$A^{(1,0)} = \frac{\alpha z B^{(1)}}{(1 - \alpha z B)^2} = \alpha z A^2 B^{(1)}$$

and

$$A^{(1,1)} = \frac{\alpha z^2 B^2}{(1 - \alpha z B)^3} = \alpha z^2 A^3 B^2,$$

where we used equation (4.2). The same arguments and computations give $B^{(1,0)} = z A^{(1)} B^2$ and $B^{(1,1)} = z^2 A^2 B^3$, to finally obtain

$$\begin{cases} A^{(1)} = \alpha z A^2 B^{(1)} + \alpha z^2 A^3 B^2 \\ B^{(1)} = z A^{(1)} B^2 + z^2 A^2 B^3. \end{cases}$$

We deduce, using equation (4.2), that $A^{(1)}$ is given by:

$$A^{(1)} = \frac{\alpha(zAB)^2}{1 - \alpha(zAB)^2}(zA^2B + A) = \frac{AB}{1 - \alpha(zAB)^2}\alpha(zAB)^2. \quad (4.4)$$

To obtain a more explicit formula for $A^{(1)}$, one can compute $\alpha(zAB)^2$ using first that $B = (A + \alpha - 1)/\alpha$ and then that $zA^2 = (1 - (\alpha - 1)z)A - 1$. After simplifications:

$$\begin{aligned} \alpha(zAB)^2 &= \frac{(1 - \alpha z - z)A + z - 1}{\alpha z} \\ &= \frac{(\alpha^2 + 1)z^2 - 2z(\alpha - 1) + 1 - (1 - \alpha z - z)\sqrt{\delta}}{2\alpha z^2}, \end{aligned} \quad (4.5)$$

since $A = (2z)^{-1}(1 - (\alpha - 1)z - \sqrt{\delta})$. Using $\sqrt{\delta} = -2zA - (\alpha - 1)z + 1$, one can then check that $\sqrt{\delta}AB = 1 - \alpha(zAB)^2$. From (4.5), we can finally rewrite (4.4) as

$$A^{(1)} = \frac{1}{\sqrt{\delta}} \frac{(\alpha^2 + 1)z^2 - 2z(\alpha + 1) + 1 - (1 - \alpha z - z)\sqrt{\delta}}{2\alpha z^2}.$$

Therefore, the function $S^{(1)}(z) = -\frac{1}{z}A^{(1)}(\frac{1}{z})$ is given by

$$S^{(1)}(z) = -\frac{z^2 - 2z(\alpha + 1) + (\alpha^2 + 1)}{2\alpha\sqrt{(z-b)(z-a)}} + \frac{z - \alpha - 1}{2\alpha}.$$

It corresponds to the Stieltjes transform of the measure $\mu_\alpha^{(1)}$ with density:

$$\frac{1}{\pi} \lim_{\varepsilon \rightarrow 0} \operatorname{Im}(S^{(1)}(x + i\varepsilon)) = \frac{x^2 - 2x(\alpha + 1) + (\alpha^2 + 1)}{2\alpha\pi\sqrt{(b-x)(x-a)}} \mathbf{1}_{(a,b)}.$$

4.2. *Asymptotic coefficients.* By Remark 3.2, it is sufficient to take l even.

The Bernoulli case. In this setting the computation is direct. As explained in Section 2.1, it suffices to consider the centered version of Bernoulli laws, which corresponds to take:

$$P_n = \frac{c}{n} \delta_{1-c/n} + \left(1 - \frac{c}{n}\right) \delta_{-c/n}.$$

Therefore, for all $k \geq 1$:

$$\frac{M_{2k}(P_n)}{n^{k-1} M_2(P_n)^k} \xrightarrow{n \rightarrow +\infty} c^{1-k}.$$

In particular $A_4 = 1/c$, which, combined with the above identification of $\mu_\alpha^{(1)}$, provides the proof of Theorem 2.1.

The truncated heavy-tailed case. Recall that P is in the domain of attraction of a β -stable law, which implies that its cumulative distribution function F satisfies Equation (2.1), where we supposed that $M^+ > 0$. In this regime, Breiman (1992, Theorem 9.34) provides

$$\frac{1 - F(\xi x)}{1 - F(x)} \xrightarrow{x \rightarrow +\infty} \xi^{-\beta},$$

meaning that $1 - F(\cdot)$ varies regularly with exponent $-\beta$. It implies that for all $k \geq 2$, the truncated moment function

$$U_k(x) := \int_0^x t^k dP(t)$$

varies regularly with exponent $k - \beta$. Theorem 2 of Feller (1971, VIII.9), known as Karamata’s estimate, yields the following asymptotic as $x \rightarrow +\infty$:

$$U_k(x) \sim \frac{\beta}{k - \beta} x^k (1 - F(x)).$$

The behavior of the left truncated moment $\bar{U}_k(x) = \int_{-x}^0 t^k dP(t)$ can be obtained in the same way. As $x \rightarrow +\infty$:

$$\bar{U}_k(x) \sim \frac{\beta}{k - \beta} (-x)^k F(-x) \mathbf{1}_{M^- > 0}.$$

Therefore, for all $k \geq 1$:

$$\begin{aligned} M_{2k} \left(P_n^{(B)} \right) &= U_{2k}(Bq_n^+) + \frac{(Bq_n^+)^{2k}}{n} + \bar{U}_{2k}(Bq_n^-) + \frac{(Bq_n^-)^{2k}}{n} \\ &\sim \frac{(Bq_n^+)^{2k}}{n} B^{-\beta} \left(1 + \frac{\beta}{2k - \beta} \right) + \frac{(Bq_n^-)^{2k}}{n} B^{-\beta} \left(1 + \frac{\beta}{2k - \beta} \right) \mathbf{1}_{M^- > 0} \\ &\sim \frac{(q_n^+)^{2k}}{n} \cdot B^{-\beta} \cdot \frac{2k}{2k - \beta} \left(1 + \left(\frac{M^-}{M^+} \right)^{1/\beta} \right) \end{aligned}$$

because on the event $M^- > 0$,

$$\frac{1 - F(q_n^+)}{1 - F \left(\left(\frac{M^+}{M^-} \right)^{1/\beta} q_n^- \right)} \xrightarrow{n \rightarrow +\infty} 1,$$

which yields $q_n^- \sim (M^-/M^+)^{1/\beta} q_n^+$. We finally obtain:

$$A_{2k} = B^{\beta(k-1)} \cdot \frac{2k}{2k - \beta} \left(\frac{2 - \beta}{2} \right)^k \left(1 + \left(\frac{M^-}{M^+} \right)^{1/\beta} \right)^{1-k}.$$

In particular, $A_4 = B^\beta \cdot C(\beta, M^+, M^-)$ which leads to Theorem 2.4.

Acknowledgements

The author would like to warmly thank his advisors Nathanaël Enriquez and Laurent Ménard for many helpful discussions and suggestions about this work. The author would also like to thank the anonymous referees for relevant comments on an earlier version of this work.

References

- D. Aldous and R. Lyons. Processes on unimodular random networks. *Electron. J. Probab.* **12**, no. 54, 1454–1508 (2007). [MR2354165](#).
- S. Belinschi, A. Dembo and A. Guionnet. Spectral measure of heavy tailed band and covariance random matrices. *Comm. Math. Phys.* **289** (3), 1023–1055 (2009). [MR2511659](#).
- F. Benaych-Georges and T. Cabanal-Duvillard. Marčenko-Pastur theorem and Bercovici-Pata bijections for heavy-tailed or localized vectors. *ALEA Lat. Am. J. Probab. Math. Stat.* **9** (2), 685–715 (2012). [MR3069381](#).
- I. Benjamini and O. Schramm. Recurrence of distributional limits of finite planar graphs. *Electron. J. Probab.* **6**, no. 23, 13 (2001). [MR1873300](#).
- C. Bordenave and M. Lelarge. Resolvent of large random graphs. *Random Structures Algorithms* **37** (3), 332–352 (2010). [MR2724665](#).
- C. Bordenave, A. Sen and B. Virág. Mean quantum percolation. *J. Eur. Math. Soc. (JEMS)* **19** (12), 3679–3707 (2017). [MR3730511](#).
- L. Breiman. *Probability*, volume 7 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (1992). ISBN 0-89871-296-3. [MR1163370](#).
- N. Enriquez and L. Ménard. Spectra of large diluted but bushy random graphs. *Random Structures Algorithms* **49** (1), 160–184 (2016). [MR3521277](#).
- W. Feller. *An introduction to probability theory and its applications. Vol. II*. Second edition. John Wiley & Sons, Inc., New York-London-Sydney (1971). [MR0270403](#).
- C. Male. The limiting distributions of large heavy Wigner and arbitrary random matrices. *J. Funct. Anal.* **272** (1), 1–46 (2017). [MR3567500](#).
- V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)* **72** (114), 507–536 (1967). [MR0208649](#).
- Ø. Ryan. On the limit distributions of random matrices with independent or free entries. *Comm. Math. Phys.* **193** (3), 595–626 (1998). [MR1624843](#).
- J. Salez. Every totally real algebraic integer is a tree eigenvalue. *J. Combin. Theory Ser. B* **111**, 249–256 (2015). [MR3315609](#).
- V. Vengerovsky. Eigenvalue distribution of a large weighted bipartite random graph. *Zh. Mat. Fiz. Anal. Geom.* **10** (2), 240–255, 259, 261 (2014). [MR3236969](#).
- I. Zakharevich. A generalization of Wigner’s law. *Comm. Math. Phys.* **268** (2), 403–414 (2006). [MR2259200](#).