UNIVERSALITY IN RANDOM MOMENT PROBLEMS

HOLGER DETTE*, DOMINIK TOMECKI†, AND MARTIN VENKER‡

Abstract. Let $\mathcal{M}_n(E)$ denote the set of vectors of the first n moments of probability measures on $E \subset \mathbb{R}$ with existing moments. The investigation of such moment spaces in high dimension has found considerable interest in the recent literature. For instance, it has been shown that a uniformly distributed moment sequence in $\mathcal{M}_n([0,1])$ converges in the large n limit to the moment sequence of the arcsine distribution. In this article we provide a unifying viewpoint by identifying classes of more general distributions on $\mathcal{M}_n(E)$ for $E = [a,b]$, $E = \mathbb{R}_+$ and $E = \mathbb{R}$, respectively, and discuss universality problems within these classes. In particular, we demonstrate that the moment sequence of the arcsine distribution is not universal for E being a compact interval. On the other hand, on the moment spaces $\mathcal{M}_n(\mathbb{R}_+)$ and $\mathcal{M}_n(\mathbb{R})$ the random moment sequences governed by our distributions exhibit for $n \to \infty$ a universal behaviour: The first k moments of such a random vector converge almost surely to the first k moments of the Marchenko-Pastur distribution (half line) and Wigner's semi-circle distribution (real line). Moreover, the fluctuations around the limit sequences are Gaussian. We also obtain moderate and large deviations principles and discuss relations of our findings with free probability.

1. Introduction

Let $\mathcal{P}(E)$ denote the set of probability measures on an (possibly infinite) interval $E \subset \mathbb{R}$ with finite moments of all orders. For a measure $\mu \in \mathcal{P}(E)$ denote by $m_j(\mu) = \int_E x^j d\mu(j)$ its j-th moment and define

$$\mathcal{M}_n(E) := \big\{(m_1(\mu), \dots, m_n(\mu)) : \mu \in \mathcal{P}(E)\big\}$$

as the set of moment sequences up to order n, generated by $\mathcal{P}(E)$. The set $\mathcal{M}_n(E)$ is convex and has been the subject of many studies beginning with Karlin and Shapeley (1953), Karlin and Studden (1966) and Krein and Nudelman (1977). In these classical works, geometric aspects of moment spaces were studied. While the even more classical moment problems deal with all *possible* moment sequences, a probabilistic investigation rather asks how a *typical* moment sequence looks like. This was initiated in Chang et al. (1993), where a uniform distribution on $\mathcal{M}_n([0,1])$ was considered. There it was shown that the first k moments of such a random vector $(m_1^{(n)}, \dots, m_n^{(n)})$ in $\mathcal{M}_n([0,1])$ obey a law of large numbers, when n tends to infinity (but k is fixed), that is

$$(m_1^{(n)}, \dots, m_k^{(n)}) \xrightarrow{d} (m_1^*, \dots, m_k^*), \quad n \to \infty, \tag{1.1}$$

$\xrightarrow{d}$ denoting convergence in distribution. Here $m_j^{(n)}$ is the j-th component of the random moment vector $(m_1^{(n)}, \dots, m_n^{(n)})$ and m_j^* is the j-th moment of the arcsine distribution (on the interval $[0,1]$). They also derived the central limit theorem

$$\sqrt{n}\big((m_1^{(n)}, \dots, m_k^{(n)}) - (m_1^*, \dots, m_k^*)\big) \xrightarrow{d} \mathcal{N}(0, \Sigma_k), \quad n \to \infty \tag{1.2}$$

2010 *Mathematics Subject Classification.* 60F05, 30E05, 60B20.

Key words and phrases. Random moment sequences, universality, CLT, large deviations principles, Stieltjes transform, free probability.

with the covariance matrix $\Sigma_k = (m^*_{i+j} - m^*_i m^*_j)^k_{i,j=1}$. Gamboa and Lozada-Chang (2004) investigated corresponding large deviations principles, while Lozada-Chang (2005) studied similar problems for moment spaces corresponding to more general functions defined on a bounded set.

More recently, Dette and Nagel (2012) defined special probability distributions on the non-compact moment spaces $\mathcal{M}_n([0,\infty))$ and $\mathcal{M}_{2n-1}(\mathbb{R})$. They could establish results analogous to (1.2) with the moments of the arcsine distribution replaced by those of the Marchenko-Pastur distribution (on $[0,\infty)$) and of the semicircle distribution (on $\mathbb{R}$), respectively.

In this article, we are going to investigate this surprising occurrence of the three distributions arcsine, Marchenko-Pastur and semicircle distribution in more detail. We are particularly interested in a possible universality of these distributions, as in random matrix theory the latter two appear naturally for large classes of random matrices with independent entries (see e.g. Bai and Silverstein (2010) and references therein). The arcsine measure also appears as a universal distribution of zeros of orthogonal polynomials with respect to weight functions on compact intervals (see Stahl and Totik (1992)). Especially for unbounded moment spaces a clarification of universality seems desirable, as there is no uniform measure and thus the consideration of a particular probability measure needs justification. In other words, we are asking for how typical the moment sequences of arcsine, semicircle and Marchenko-Pastur distribution are.

The paper will be organized as follows. In Section 2 we review some basic facts about moment spaces and introduce general classes of distributions on the moment spaces under consideration. They keep two key features of the uniform distribution on $\mathcal{M}_n([a,b])$ and can be used to interpolate between distributions on compact and non-compact moment spaces. For these distributions we derive laws of large numbers of the type (1.1). In particular, we show that for moment spaces $\mathcal{M}_n([a,b])$ corresponding to compact intervals there is no universality of the arcsine distribution. Instead, the arising measures are known as free binomial distributions, i.e. the analogues of the binomial distribution in free probability theory. On the other hand, for the moment spaces $\mathcal{M}_n([0,\infty))$ and $\mathcal{M}_n(\mathbb{R})$ the first k moments of a random vector always converge to the first k moments of Marchenko-Pastur and semicircle distributions, respectively. The occurrence of both distributions will be explained in terms of free Poissonian and free central limit theorems for the free binomial distribution. In Section 3 we consider central limit theorems of the form (1.2) and investigate moderate and large deviations principles for random moment sequences. All proofs are postponed to Section 4. Our results provide an extensive description of the distributional properties of random moment sequences and a unifying view on several findings in the recent literature.

2. Laws of Large Numbers

To motivate the class of distributions considered in this paper, we remark first that a real valued sequence $(m_i)_{i\in\mathbb{N}_0}$ is a sequence of moments corresponding to a Borel measure on the real line if and only if all Hankel matrices $(m_{i+j})^n_{i,j=0}$ are positive semi-definite (see Hamburger (1920)). Similar characterizations exist for measures supported on the half line $[0,\infty)$ and compact intervals, and the corresponding sequences are called Stieltjes and Hausdorff moment sequences (see Dette and Studden (1997)). Due to restrictions and relations of this type, the components of a random moment vector in $\mathcal{M}_n(E)$ are generically not independent coordinates. Moreover, for a compact interval E the moment space $\mathcal{M}_n(E)$ is a rather small set. For instance, it is known that the volume of $\mathcal{M}_n([0,1])$ is of order $\mathcal{O}(2^{-n^2})$ (see Karlin and Shapeley (1953)), as for a given moment sequence $(m_1,\ldots,m_{n-1}) \in \mathcal{M}_{n-1}([0,1])$, the possible range of the n-th moment m_n is very small.

For these reasons, we will consider different sets of coordinates that scale with the possible range of values. Although there are infinitely many choices of such coordinates, some are particularly natural and have found considerable attention in the literature. To be precise, assume that $(m_1, \ldots, m_{j-1}) \in \mathcal{M}_{j-1}([a,b])$ is a given vector of moments up to the order $j-1$. Then, because of convexity of $\mathcal{M}_j([a,b])$, the set of possible values m_j

$$\big\{ m_j(\mu) \big|\ \mu \in \mathcal{P}([a,b]);\ m_i(\mu) = m_i \ \text{ for all } i = 1, \ldots, j-1 \big\}$$

is a compact interval, say $[m_j^-, m_j^+]$. Following Dette and Studden (1997), we define for $m_j^+ \neq m_j^-$ and a given j-th moment m_j the j-th canonical moment p_j via

$$p_j := \frac{m_j - m_j^-}{m_j^+ - m_j^-}.$$

The canonical moments are left undefined if $m_j^- = m_j^+$ (in this case the vector $(m_1, \ldots, m_{j-1})$ is a boundary point of the set $\mathcal{M}_{j-1}([a,b])$ - see Karlin and Studden (1966)). Clearly, $p_j \in [0,1]$, and p_j gives the relative position of m_j in the available section of the set $\mathcal{M}_j([a,b])$. It is also worthwhile to mention that canonical moments are invariant under linear transformations of the measure (see Dette and Studden (1997), p. 13). The correspondence map

$$\varphi_n^{[a,b]} : \vec{p}_n = (p_1, \ldots, p_n) \mapsto \vec{m}_n = (m_1, \ldots, m_n) \tag{2.1}$$

between the canonical and ordinary moments is one-to-one from $(0,1)^n$ onto $\mathrm{Int}(\mathcal{M}_n([a,b]))$ (Int denoting the interior) and many classical quantities of the measure, especially of its associated orthogonal polynomials and the continued fraction expansion of its Stieltjes transform, have expressions in terms of the canonical moments (see Dette and Studden (1997) for more details). Canonical moments were introduced in a series of papers by Skibinsky (1967, 1968, 1969) and are closely related to the Verblunsky coefficients, which were investigated much earlier by Verblunsky (1935, 1936) for measures on the unit circle.

In case of the uniform distribution on $\mathcal{M}_n([0,1])$, as studied in Chang et al. (1993), the canonical moments have two important properties. After a change of variables by (2.1), the uniform distribution on $\mathcal{M}_n([0,1])$ has a density w.r.t. the Lebesgue measure on $(0,1)^n$ proportional to

$$\prod_{j=1}^{n} (p_j(1-p_j))^{n-j} = \exp\Big[\sum_{j=1}^{n} (n-j) \log(p_j(1-p_j)) \Big]. \tag{2.2}$$

Thus, the canonical moments are independent and for $n \gg j$ nearly identically distributed. To investigate a possible universality of the arcsine distribution, we will now define a class of distributions respecting these two properties. However, we will generalize the situation by allowing for different distributions of even and odd canonical moments. This takes into account the different roles that even and odd moments play. While even moments are always positive and give some rough information about the size of the support of the measure, odd moments give information about location of the support and the symmetry of the measure. In canonical moments, symmetry around the center of $[a,b]$ can be characterized easily as the property that all odd canonical moments are $1/2$ (see Skibinsky (1969)).

Let $V_1, V_2 : [0,1] \to \mathbb{R}$ be continuous functions. Define the probability measure $\mathbb{P}_{n,[a,b],V_{1,2}}$ on $\mathcal{M}_n([a,b])$ by $\mathbb{P}_{n,[a,b],V_{1,2}}(\partial \mathcal{M}_n([a,b])) = 0$ and on $\mathrm{Int}(\mathcal{M}_n([a,b]))$ via the density

$$P_{n,[a,b],V_{1,2}}(m_1, \ldots, m_n) := \frac{1}{Z_{n,[a,b],V_{1,2}}} \exp\Big[-n \sum_{j=1}^{\lfloor \frac{n+1}{2} \rfloor} V_1(p_{2j-1}) - n \sum_{j=1}^{\lfloor \frac{n}{2} \rfloor} V_2(p_{2j}) \Big] \tag{2.3}$$

w.r.t. the n-dimensional Lebesgue measure, where $p_j = p_j(m_1, \dots, m_j)$ is the j-th canonical moment of the sequence $(m_1, \dots, m_n) \in \mathrm{Int}(\mathcal{M}_n([a,b]))$ defined by (2.1) ($j = 1, \dots, n$) and $Z_{n,[a,b],V_{1,2}}$ is the normalization constant. By $\lfloor x \rfloor$ we denote the largest natural number smaller or equal to x. Note that the case $V_1(x) = V_2(x) \equiv 0$ and $[a,b] = [0,1]$ has been considered in Chang et al. (1993). The factors n in the exponent in (2.3) are asymptotically equivalent to the factor $n - j$ in (2.2). It follows from (2.2) that under $\mathbb{P}_{n,[a,b],V_{1,2}}$ the odd, respectively even, canonical moments are nearly i.i.d..

Let us now formulate our first result for random moment sequences on measures supported on the interval $[a,b]$. Here and later on, we will tacitly assume that the random variables $(m_j^{(n)})_{j,n\geq 1}$ are defined on the same probability space.

Theorem 2.1.

(1) *Let $a < b$ and $V_1, V_2 \in C^2((0,1))$ be continuous at 0 and 1. Assume that the functions*

$$W_1(p) := V_1(p) - \log(p(1-p)) \qquad \text{and} \qquad W_2(p) := V_2(p) - \log(p(1-p))$$

each have a unique minimizer $p_1^* \in (0,1)$ and $p_2^* \in (0,1)$, respectively. Let $m^{(n)} = (m_1^{(n)}, \dots, m_n^{(n)})$ be drawn from $\mathbb{P}_{n,[a,b],V_{1,2}}$ and abbreviate $q_i^* := 1 - p_i^*, i = 1,2$. Then we have for each $k \geq 1$ as $n \to \infty$

$$(m_1^{(n)}, \dots, m_k^{(n)}) \to (m_1^*, \dots, m_k^*)$$

almost surely and in L^1, where $m_1^*, \dots, m_k^*$ are the first k moments of a probability measure $\mu_{p_1^*,p_2^*} = \mu^{ac}_{p_1^*,p_2^*} + \mu^{d}_{p_1^*,p_2^*}$. Setting

$$l_\pm := a + (b-a)\left(\sqrt{p_1^* q_2^*} \pm \sqrt{p_2^* q_1^*}\right)^2,$$

the measures $\mu^{ac}_{p_1^*,p_2^*}$ and $\mu^{d}_{p_1^*,p_2^*}$ are given by

$$\mu^{ac}_{p_1^*,p_2^*}(dx) = \frac{\sqrt{(x-l_-)(l_+-x)}}{2\pi p_2^*(x-a)(b-x)} 1_{[l_-,l_+]}(x)dx,$$
$$\mu^{d}_{p_1^*,p_2^*} = \left(1 - \frac{p_1^*}{p_2^*}\right)_+ \delta_a + \left(\frac{p_1^* + p_2^* - 1}{p_2^*}\right)_+ \delta_b.$$

Here $(y)_+$ denotes the positive part of $y \in \mathbb{R}$ and δ_y is the Dirac measure at the point y.

(2) *If p_1^*, p_2^* are such that $\mu_{p_1^*,p_2^*}$ does not have atoms, then $\mu_{p_1^*,p_2^*}$ is the equilibrium measure on the interval $[a,b]$ to the external field*

$$Q(t) := -\left(\frac{p_1^*}{p_2^*} - 1\right)\log(t-a) - \left(\frac{1 - p_1^* - p_2^*}{p_2^*}\right)\log(b-t),$$

i.e. $\mu_{p_1^*,p_2^*}$ is the unique Borel probability measure on the interval $[a,b]$ minimizing the functional

$$\mu \mapsto \int_a^b Q(t)d\mu(t) - \int_a^b \int_a^b \log|t-s| d\mu(t)d\mu(s). \tag{2.4}$$

Remark 2.2.

(1) If $p_1^* = p_2^* = 1/2$, the measure $\mu_{p_1^*,p_2^*}$ in Theorem 2.1 is the arcsine distribution on the interval $[a,b]$. Note that this does not imply $V_1 = V_2 \equiv 0$. However, we see that for $p_1^* \neq 1/2$ or $p_2^* \neq 1/2$, the limiting measure (the measure having the limiting moments) is not the arcsine measure or an affine rescaling of it. We conclude that the moments of the arcsine measure are not universal within the class of random moment sequences in $\mathcal{M}_n([a,b])$ with nearly i.i.d. canonical moments. On the other hand, there is still some universality as the limiting measure only depends on V_1, V_2 via the parameters p_1^* and p_2^*.
(2) Since for probability measures supported on a fixed compact set convergence of moments is equivalent to convergence in distribution, the convergence result of Theorem 2.1 can be restated as follows: Let $\mu_n \in \mathcal{P}([a,b])$ be a random probability measure with first n moments $(m_1^{(n)}, \ldots, m_n^{(n)})$ which are $\mathbb{P}_{n,[a,b],V_{1,2}}$-distributed. Then μ_n converges a.s. (and in expectation) weakly to $\mu_{p_1^*,p_2^*}$ as $n \to \infty$.

The measure $\mu_{p_1^*,p_2^*}$ is known in the literature under (at least) two different names. In the context of probability theory on graphs, it is called Kesten-McKay measure (see Kesten (1959); McKay (1981)). It has also been studied in the context of orthogonal polynomials (see Cohen and Trenholme (1984); Saitoh and Yoshida (2001); Castro and Grünbaum (2013)). In free probability, it is called free binomial distribution (see Nica and Speicher (2006)). It will turn out useful to explain this naming in more detail.

Free probability is a variant of non-commutative probability theory initiated by Voiculescu (see Nica and Speicher (2006) or Chapter 22 by Speicher in Akemann et al. (2011) for an introduction and references) that has found its applications in particular in random matrix theory. For our purposes it suffices to know that free probability theory uses a different notion of independence, called freeness, that manifests itself in a different convolution of probability measures. A constructive approach to this convolution uses random matrices: Let $H_{1,n}, H_{2,n}$ be deterministic diagonal $n \times n$ matrices with diagonal entries $h_{1,n}(ii)$ and $h_{2,n}(ii)$, respectively. Assume that the empirical measures of the diagonal entries, i.e. the eigenvalues, converge for $n \to \infty$ weakly to probability measures of bounded support μ_1 and μ_2, respectively, that is

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{h_{j,n}(ii)} = \mu_j, \quad j = 1, 2, \qquad \text{weakly.}$$

Now let for each n a Haar distributed random unitary $n \times n$ matrix U_n be given on a common probability space. The Haar probability measure on the unitary group $\mathcal{U}_n$ is the unique Borel probability measure that is invariant under left (and right) multiplication with any group element. Letting $x_1, \ldots, x_n$ denote the n real random eigenvalues of the Hermitian random matrix $H_{1,n} + U_n H_{2,n} U_n^*$, the empirical measure of the x_i's converges for $n \to \infty$ almost surely in distribution to a non-random limit. This limit is called the free (additive) convolution of $\mu_1 \boxplus \mu_2$, in symbols

$$\mu_1 \boxplus \mu_2 := \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i} \qquad \text{a.s. weakly.}$$

In analogy to classical probability, the free binomial distribution with parameters $n \in \mathbb{N}$ and $p \in [0,1]$ is then the n-fold free convolution of the Bernoulli distribution $\mu = (1-p)\delta_0 + p\delta_1$ with itself. It seems convenient to extend the name to convolutions of measures $\mu = (1-p)\delta_c + p\delta_d$ with itself, $c, d \in \mathbb{R}$. Moreover, even fractional convolution numbers are possible using an analytic approach to the free convolution via the so-called R-transform (see (Akemann et al., 2011, Chapter 22)). It seems difficult to give a direct interpretation of the occurence of the free binomial distribution in the context of random moments. For instance it is not hard to verify that for $\mu = \frac{1}{2}\delta_c + \frac{1}{2}\delta_d$ the free convolution $\mu \boxplus \mu$ is the arcsine measure with support $[c+d-\sqrt{c^2+d^2}, c+d+\sqrt{c^2+d^2}]$, but in general the measure $\mu_{p_1^*,p_2^*}$ is not just a two-fold convolution of a Bernoulli measure with itself.

However, free probability indicates that universal limiting measures may be expected if random moment problems are considered for the moment spaces $\mathcal{M}_n(\mathbb{R}_+)$ with $\mathbb{R}_+ := [0,\infty)$ and $\mathcal{M}_n(\mathbb{R})$. Indeed, analogous to classical probability, there are free analogs of Poisson limit theorem and central limit theorem for the free binomial distribution (Akemann et al., 2011, Chapter 22). Typically, they are considered for $\mu = (1-p_m)\delta_0 + p_m\delta_1$ and show weak convergence of the rescaled n-th convolution power $\mu^{\boxplus m}$ to the free Poisson (Marchenko-Pastur distribution) or the free Gaussian law (semicircle distribution), as $m \to \infty$ and p_m converges to a zero or non-zero number, respectively.

The following corollary can be seen as a variant of these limit theorems. The proof is straightforward and will be omitted.

Corollary 2.3. *Let for each $m \in \mathbb{N}$ $a_m < b_m$ and $p_{1,m}^*, p_{2,m}^* \in (0,1)$ be given.*

(1) *Assume that, as $m \to \infty$,*

$$\begin{aligned} &a_m \to 0,\ b_m \to \infty, \quad p_{1,m}^*, p_{2,m}^* \to 0 \quad \text{such that} \\ &p_{i,m}^* b_m \to z_i^*, i = 1,2, \end{aligned}$$

for some constants $z_1^*, z_2^* > 0$. Then the measure $\mu_{p_{1,m}^*,p_{2,m}^*}$ defined in Theorem 2.1 on the interval $[a_m, b_m]$ converges in the large m limit weakly to the measure μ_{MP,z_1^*,z_2^*}, where with $l_\pm := (\sqrt{z_1^*} \pm \sqrt{z_2^*})^2$

$$\mu_{MP,z_1^*,z_2^*}(dx) = \left(1 - \frac{z_1^*}{z_2^*}\right)_+ \delta_0 + \frac{1}{2\pi z_2^*} \frac{\sqrt{(x-l_-)(l_+-x)}}{x} 1_{[l_-,l_+]}(x)dx. \tag{2.5}$$

The density of the absolutely continuous part of $\mu_{p_{1,m}^*,p_{2,m}^*}(x)$ converges pointwise to the density of the absolutely continuous part of μ_{MP,z_1^*,z_2^*} and uniformly within compact subsets of (l_-, l_+). Moreover, the moments of $\mu_{p_{1,m}^*,p_{2,m}^*}$ converge to the moments of μ_{MP,z_1^*,z_2^*}.

(2) *Assume that, as $m \to \infty$,*

$$\begin{aligned} &a_m \to -\infty,\ b_m \to \infty, \\ &p_{2,m}^* |a_m| b_m \to \beta^*, \quad a_m + (b_m - a_m)p_{1,m}^* \to \alpha^* \end{aligned}$$

for constants $\alpha^* \in \mathbb{R}, \beta^* > 0$. Then the measure $\mu_{p_{1,m}^*,p_{2,m}^*}$ defined in Theorem 2.1 on the interval $[a_m, b_m]$ converges weakly in the large m limit to the measure $\mu_{SC,\alpha^*,\beta^*}$, where with $l_\pm := \alpha^* \pm 2\sqrt{\beta^*}$

$$\mu_{SC,\alpha^*,\beta^*}(dx) = \frac{1}{2\pi\beta^*}\sqrt{(x-l_-)(l_+-x)}1_{[l_-,l_+]}(x)dx. \tag{2.6}$$

The density of the absolutely continuous part of $\mu_{p_{1,m}^*,p_{2,m}^*}(x)$ converges pointwise to the density of $\mu_{SC,\alpha^*,\beta^*}$ and uniformly within compact subsets of (l_-, l_+). Moreover, the moments of $\mu_{p_{1,m}^*,p_{2,m}^*}$ converge to the moments of $\mu_{SC,\alpha^*,\beta^*}$.

Remark 2.4.

(1) The measure μ_{MP,z_1^*,z_2^*} is called Marchenko-Pastur distribution (see Hiai and Petz (2000) or Nica and Speicher (2006)). For $z_1^* \geq z_2^*$ (absolutely continuous case) it is the equilibrium measure on $\mathbb{R}_+$ (in the sense of (2.4)) to the field

$$Q(t) = \frac{t}{z_2^*} - \frac{z_1^* - z_2^*}{z_2^*} \log t.$$

Besides its role in free probability theory as the free analog of the Poisson distribution it is particularly well-known for its universality in random matrix theory. More precisely, let X denote an $m \times n$ random matrix with real i.i.d. entries having mean 0 and variance $\sigma^2 > 0$. Assume that as $m, n \to \infty$ we have $m/n \to \lambda \in (0, \infty)$. Then the empirical distribution of the eigenvalues of the sample covariance matrix XX^T/n converges a.s. and in expectation weakly to μ_{MP,z_1,z_2}, where $z_1 := \sigma^2(1+\sqrt{\lambda})/(1+\sqrt{\lambda})^2$ and $z_2 := \lambda z_1$. For this result and generalizations we refer to Bai and Silverstein (2010) and references therein.

(2) The measure $\mu_{SC,\alpha^*,\beta^*}$ is called semicircle distribution. It is the equilibrium measure to the field

$$Q(t) = \frac{t^2}{2\beta^*} - \frac{\alpha^* t}{\beta^*}.$$

In free probability, it plays the role of the Gaussian distribution. In random matrix theory it is the universal limit of so-called Wigner matrices: Let X be an $n \times n$ random matrix with real i.i.d. mean 0 and variance $\sigma^2 > 0$ entries on and above the diagonal and the entries below the diagonal are chosen such that X is symmetric. Then the empirical distribution of the eigenvalues of $X/\sqrt{n}$ converges a.s. and in expectation weakly to $\mu_{SC,\alpha,\beta}$ as $n \to \infty$, where $\alpha = 0$ and $\beta = \sigma^2$, see e.g. Bai and Silverstein (2010).

The universality in these random matrix statements lies in the fact that the limiting distribution is always the same regardless of the distribution of the matrix entries.

(3) The measures $\mu_{p_1^*,p_2^*}, \mu_{MP,z_1^*,z_2^*}$ and $\mu_{SC,\alpha^*,\beta^*}$ all belong to the so-called free Meixner class. It consists of the free analogues of the six classical Meixner class distributions which are Gaussian, Poisson, gamma, binomial, negative binomial and hyperbolic secant distribution. The distributions of the free Meixner class enjoy some interesting characterizing properties, for instance having a generating function of resolvent type for the corresponding orthogonal polynomials (see Anshelevich (2007) for details) in analogy to the generating functions of the classical Meixner class being of exponential type (see Meixner (1934)).

Let us now turn to infinite moment spaces, starting with $\mathcal{M}_n(\mathbb{R}_+)$ (recall $\mathbb{R}_+ = [0, \infty)$). Following Dette and Nagel (2012), we may define the canonical moments $z_1, \dots, z_n$ of a moment sequence $m_1, \dots, m_n$ in the interior of $M_n(\mathbb{R}_+)$ as

$$z_k := \frac{m_k - m_k^-}{m_{k-1} - m_{k-1}^-}, \quad k = 1, \dots, n,$$

$m_0^- = 0, m_0 = 1$. Here one uses that given $m_1, \dots, m_{k-1}$, the section of possible values of m_k for given moments $(m_1, \dots, m_{k-1}) \in \mathrm{Int}(\mathcal{M}_{k-1}(\mathbb{R}_+))$ is an interval of the form $[m_k^-, \infty)$ (see Karlin and Studden (1966), Chapter V). Clearly, $z_k \in \mathbb{R}_+$. The correspondence

$$\varphi_n^{\mathbb{R}_+} : \vec{z}_n = (z_1, \dots, z_n) \mapsto \vec{m}_n = (m_1, \dots, m_n) \tag{2.7}$$

between canonical and ordinary moments is one-to-one from $(0,\infty)^n$ onto $\mathrm{Int}(\mathcal{M}_n(\mathbb{R}_+))$ (for all $n \in \mathbb{N}$). The Jacobian of this transformation is readily computed as

$$\left|\prod_{k=1}^{n} \frac{\partial m_k}{\partial z_k}\right| = \prod_{k=1}^{n} (m_{k-1} - m_{k-1}^{-}) = \prod_{k=2}^{n} z_1 z_2 \dots z_{k-1} = \prod_{k=1}^{n} z_k^{n-k} \ . \tag{2.8}$$

To define a probability measure on $\mathrm{Int}(\mathcal{M}_n(\mathbb{R}_+))$, consider continuous functions $V_1, V_2 : \mathbb{R}_+ \to \mathbb{R}$, such that for some $\varepsilon > 0$ and all z large enough the inequality

$$\frac{V_i(z)}{\log z} \geq 2 + \varepsilon, \ \ i = 1, 2 \tag{2.9}$$

holds. Then define a probability measure $\mathbb{P}_{n,\mathbb{R}_+,V_{1,2}}$ on $\mathcal{M}_n(\mathbb{R}_+)$ by $\mathbb{P}_{n,\mathbb{R}_+,V_{1,2}}(\partial\mathcal{M}_n(\mathbb{R}_+)) = 0$ and on $\mathrm{Int}(\mathcal{M}_n(\mathbb{R}_+))$ via the density

$$P_{n,\mathbb{R}_+,V_{1,2}}(m_1, \dots, m_n) := \frac{1}{Z_{n,\mathbb{R}_+,V_{1,2}}} \exp\Big[- n \sum_{j=1}^{\lfloor \frac{n+1}{2} \rfloor} V_1(z_j) - n \sum_{j=1}^{\lfloor \frac{n}{2} \rfloor} V_2(z_j) \Big], \tag{2.10}$$

where $Z_{n,\mathbb{R}_+,V_{1,2}}$ is the normalizing constant such that $P_{n,\mathbb{R}_+,V_{1,2}}$ is a probability density with respect to the Lebesgue measure on $\mathrm{Int}(\mathcal{M}_n(\mathbb{R}_+))$. This is possible due to (2.8) and (2.9). Because of (2.8), the canonical moments $z_1, z_2, \dots, z_k$ are independent under $\mathbb{P}_{n,\mathbb{R}_+,V_{1,2}}$ and for large n and fixed k nearly identically distributed.

Note that Dette and Nagel (2012) considered the special case of (2.10) with $V_1(t) = V_2(t) = t - \frac{c}{n}\log t$ and showed that under this measure the (ordinary) moments converge to those of the Marchenko-Pastur distribution. Here we will show that the moments of the Marchenko-Pastur distribution are in fact universal for all generic functions V_1, V_2.

Theorem 2.5. *Let $V_1, V_2 \in C^2((0,\infty))$ be continuous at 0, satisfy (2.9) and assume that*

$$W_1(z) := V_1(z) - \log z \qquad \text{and} \qquad W_2(z) := V_2(z) - \log z$$

each have a unique minimizer $z_1^* \in (0,\infty)$ and $z_2^* \in (0,\infty)$, respectively. Let the vector $m^{(n)} = (m_1^{(n)}, \dots, m_n^{(n)})$ be drawn from $\mathbb{P}_{n,\mathbb{R}_+,V_{1,2}}$. Then we have for any $k \geq 1$ as $n \to \infty$

$$(m_1^{(n)}, \dots, m_k^{(n)}) \to (m_1^*, \dots, m_k^*)$$

almost surely and in L^1, where $m_1^*, \dots, m_k^*$ are the first k moments of the Marchenko-Pastur distribution μ_{MP,z_1^*,z_2^*} defined in (2.5), that is

$$m_j^* = \sum_{i=0}^{\lfloor \frac{j-1}{2} \rfloor} \binom{j-1}{i} (z_1^*)^{i+1} (z_2^*)^i (z_1^* + z_2^*)^{j-1-i} \frac{1}{i+1} \binom{2i}{i}.$$

Next, we consider the moment space corresponding to measures supported on $\mathbb{R}$. We will use the recurrence coefficients of the corresponding orthogonal polynomials as a coordinate system. To be precise, note that for any measure $\mu \in \mathcal{P}(\mathbb{R})$ there is a sequence of monic polynomials $P_0(x), P_1(x), \dots$ with $\deg P_j = j$ that is orthogonal in $L^2(\mu)$. If μ is supported on finitely many points, the sequence is finite. In any case, $P_j(x)$ depends on the measure μ via its moment sequence $(m_1, \dots, m_{2j-1})$ only. The orthogonal polynomials satisfy a three-term recurrence relation of the form

$$P_{j+1}(x) = (x - \alpha_{j+1}) P_j(x) - \beta_j P_{j-1}(x), \qquad j = 1, \dots \tag{2.11}$$
$$P_0(x) = 1, \quad P_1(x) = x - \alpha_1$$

with recurrence coefficients $\alpha_1, \alpha_2, \dots \in \mathbb{R}$ and $\beta_1, \beta_2, \dots > 0$. For more details regarding orthogonal polynomials we refer to Chihara (1978). The mapping

$$\varphi_{2n-1}^{\mathbb{R}} : (\alpha_1, \beta_1, \alpha_2, \dots, \beta_{n-1}, \alpha_n) \mapsto \vec{m}_{2n-1} = (m_1, \dots, m_{2n-1}) \tag{2.12}$$

is one-to-one from $(\mathbb{R} \times (0, \infty))^{n-1} \times \mathbb{R}$ onto $\mathrm{Int}(\mathcal{M}_{2n-1}(\mathbb{R}))$ (for all $n \in \mathbb{N}$). Moreover, as observed by Dette and Nagel (2012), $(\alpha_1, \beta_1, \alpha_2, \dots, \beta_{n-1}, \alpha_n)$ constitutes a system of independent coordinates on the moment space $\mathcal{M}_{2n-1}(\mathbb{R})$. The corresponding Jacobian is given by

$$\det D\varphi_{2n-1}^{\mathbb{R}} = \prod_{j=1}^{n-1} \beta_j^{2n-2j-1}.$$

Similarly, we may define a map for moment spaces of even order.

Lemma 2.6. *There is a bijection*

$$\begin{aligned} \varphi_{2n}^{\mathbb{R}} \ &: \ (\mathbb{R} \times (0, \infty))^n \to \mathrm{Int}(\mathcal{M}_{2n}(\mathbb{R})), \\ (\alpha_1, \beta_1, \alpha_2, \dots, \alpha_n, \beta_n) &\mapsto (m_1, \dots, m_{2n}) \end{aligned} \tag{2.13}$$

between the recursion coefficients of the orthogonal polynomials and the corresponding moments. The Jacobian of $\varphi_{2n}^{\mathbb{R}}$ is

$$\det D\varphi_{2n}^{\mathbb{R}} = \prod_{j=1}^{n-1} \beta_j^{2n-2j}.$$

The values β_j have a simple interpretation in terms of moments, as

$$\beta_j = \frac{m_{2j} - m_{2j}^-}{m_{2j-2} - m_{2j-2}^-}, \qquad j = 1, \dots, n,$$

is the ratio of two consecutive even moments. The coefficients α_j give information about symmetry of the measure, e.g. for μ symmetric around 0, one has $\alpha_j = 0$ for all j. Taking into account these two different roles, we will again consider two continuous functions $V_1 : \mathbb{R} \to \mathbb{R}$ and $V_2 : \mathbb{R}_+ \to \mathbb{R}$ such that for some $\varepsilon > 0$ and $|\alpha|, \beta$ large enough

$$\frac{V_1(\alpha)}{\log|\alpha|} \geq 1 + \varepsilon, \qquad \frac{V_2(\beta)}{\log \beta} \geq 3 + \varepsilon. \tag{2.14}$$

With these notations we define the probability measure $\mathbb{P}_{n,\mathbb{R},V_{1,2}}$ on $\mathcal{M}_n(\mathbb{R})$ by $\mathbb{P}_{n,\mathbb{R},V_{1,2}}(\partial\mathcal{M}_n(\mathbb{R})) = 0$ and on $\mathrm{Int}(\mathcal{M}_n(\mathbb{R}))$ via the density

$$P_{n,\mathbb{R},V_{1,2}}(m_1, \dots, m_n) := \frac{1}{Z_{n,\mathbb{R},V_{1,2}}} \exp\Big[- n \sum_{j=1}^{\lfloor \frac{n+1}{2} \rfloor} V_1(\alpha_j) - n \sum_{j=1}^{\lfloor \frac{n}{2} \rfloor} V_2(\beta_j) \Big],$$

and obtain the following universal law of large numbers.

Theorem 2.7. *Let $V_1 \in C^2(\mathbb{R}), V_2 \in C^2((0, \infty))$ be continuous at 0 and satisfy (2.14). Furthermore, assume that*

$$W_1(\alpha) := V_1(\alpha) \qquad \textit{and} \qquad W_2(\beta) := V_2(\beta) - 2 \log \beta$$

each have unique minimizers $\alpha^* \in \mathbb{R}$ and $\beta^* \in (0, \infty)$, respectively. Let $m^{(n)} = (m_1^{(n)}, \dots, m_n^{(n)})$ be drawn from $\mathbb{P}_{n,\mathbb{R},V_{1,2}}$. Then for any $k \geq 1$ as $n \to \infty$

$$(m_1^{(n)}, \dots, m_k^{(n)}) \to (m_1^*, \dots, m_k^*)$$

almost surely and in L^1, where $m_1^*, \dots, m_k^*$ are the first k moments of the semicircle distribution $\mu_{SC,\alpha^*,\beta^*}$ defined in (2.6), that is

$$m_j^* = \sum_{i=0}^{\lfloor j/2 \rfloor} \binom{j}{2i} (\beta^*)^i (\alpha^*)^{j-2i} \frac{1}{i+1} \binom{2i}{i}. \tag{2.15}$$

We finish this section with some concluding remarks concerning the class of models we consider. We study random moment sequences with independent and nearly identically distributed canonical moments or recurrence coefficients, respectively. Dropping either of the two properties will in general result in non-universal limiting sequences even on unbounded intervals, if there is any limit at all. Nevertheless, other related models have been used for successful studies of random matrix models. More precisely, so-called Gaussian beta ensembles admit tridiagonal matrix models, see Dumitriu and Edelman (2002). More recently, Krishnapur et al. (2016) have used tridiagonal matrix models for studying non-Gaussian beta ensembles. They consider $\exp(-n \operatorname{Tr} Q(T)) \det(D\varphi_n^{\mathbb{R}})$ as density on the space of recursion coefficients, where T is the symmetric tridiagonal matrix (truncated Jacobi operator) with the α_j's on the main diagonal and β_j's on the neighboring diagonals, Q is a strictly convex polynomial and Tr denotes the trace. It is not hard to see from the results in Krishnapur et al. (2016) that the limiting moments corresponding to this model are those of the equilibrium measure to Q (see (2.4)), only for Q quadratic (this case is the one studied in Dumitriu and Edelman (2002)) the moments of the semicircle appear.

The connection between certain random matrix ensembles and canonical moments/recursion coefficients has also been used in Gamboa et al. (2016) and Gamboa et al. (2017) for deriving so-called sum rules for free binomial, semicircle and Marchenko-Pastur distribution.

3. Asymptotic Normality, Moderate and Large Deviations

In this section, we examine the fluctuations of the random moment sequences around their non-random limits. We state the central limit theorem and moderate and large deviations results. For the uniform distribution on the moment space $\mathcal{M}_n([0,1])$, results of this type were obtained by Chang et al. (1993) and Gamboa and Lozada-Chang (2004), respectively. The following theorem shows that the fluctuations of random moment vectors around their limits are Gaussian. We will adopt a short notation that allows us to state the three cases $E = [a,b]$, $E = \mathbb{R}_+$, $E = \mathbb{R}$ simultaneously. Note that the functions W_1, W_2 as well as the limiting moments m_j^* differ, depending on E.

Theorem 3.1. *In the situation of Theorem 2.1, Theorem 2.5 or Theorem 2.7, assume that $W_i''(y_i^*) \neq 0$ for $i = 1, 2$, where*

$$y_i^* := \begin{cases} p_i^* & , \text{ if } E = [a,b], \\ z_i^* & , \text{ if } E = \mathbb{R}_+, \\ \alpha^* & , \text{ if } E = \mathbb{R},\ i = 1, \\ \beta^* & , \text{ if } E = \mathbb{R},\ i = 2. \end{cases}$$

Then in any of the three cases $E = [a,b]$, $E = \mathbb{R}_+$, $E = \mathbb{R}$, for any $k \geq 1$ as $n \to \infty$

$$\sqrt{n}\big((m_1^{(n)}, \dots, m_k^{(n)}) - (m_1^*, \dots, m_k^*)\big) \xrightarrow{d} \mathcal{N}(0, \Sigma_k),$$

where the matrix Σ_k is given by

$$\Sigma_k = (D\varphi_k^E(\vec{y}^*))^t \operatorname{diag}(W_1''(y_1^*), W_2''(y_2^*), W_1''(y_1^*), \dots)^{-1} (D\varphi_k^E(\vec{y}^*)).$$

Here, the maps φ_k^E have been defined in (2.1), (2.7) and (2.12), (2.13), the diagonal matrix is of size $k \times k$ and $\vec{y}^* = (y_1^*, y_2^*, y_1^*, \dots) \in \mathbb{R}^k$.

In the case $E = \mathbb{R}_+$ and $z_1^* = z_2^*$, we have

$$(D\varphi_k^{\mathbb{R}_+}(\vec{y}^*))_{i,j} = (z_1^*)^{i-1} \left(\binom{2i}{i-j} - \binom{2i}{i-j-1} \right).$$

Theorem 3.1 shows that in all considered cases the $1/\sqrt{n}$-fluctuations of $m_1^{(n)}, \dots, m_k^{(n)}$ around $m_1^*, \dots, m_k^*$ are Gaussian. We will now study larger fluctuations. The appropriate tool for describing the exponentially small probabilities associated to these fluctuations is the large deviations principle. Recall that a sequence of random vectors $(X_n)_n$ with values in a Polish space $\mathcal{X}$ is said to satisfy a large deviations principle with speed $(b_n)_n, \lim_{n\to\infty} b_n = \infty$, and good rate function I, if $I : \mathcal{X} \to [0, \infty]$ is lower semi-continuous, has compact level sets $\{x \in \mathcal{X} : I(x) \le K\}, K \ge 0$ and for any open set $O \subset \mathcal{X}$ and closed set $U \subset \mathcal{X}$

$$\liminf_{n\to\infty} \frac{1}{b_n} \log P(X_n \in O) \ge -\inf_{x\in O} I(x), \tag{3.1}$$

$$\limsup_{n\to\infty} \frac{1}{b_n} \log P(X_n \in U) \le -\inf_{x\in U} I(x), \tag{3.2}$$

cf. (Dembo and Zeitouni, 2010, p. 6). The next theorem is a result on moderate deviations. It shows that on scales up to $o(1)$ the exponential leading order asymptotics are still given by the Gaussian distributions from Theorem 3.1, in particular they are universal.

Theorem 3.2. *Let the conditions of Theorem 3.1 be satisfied. Then for any of the three cases $E = [a, b]$, $E = \mathbb{R}_+$, $E = \mathbb{R}$, for any real-valued sequence $(a_n)_n$ with $\lim_{n\to\infty} a_n = \infty$ and $a_n = o(\sqrt{n})$, the sequence of random variables*

$$a_n\big((m_1^{(n)}, \dots, m_k^{(n)}) - (m_1^*, \dots, m_k^*)\big)$$

satisfies a large deviations principle on $\mathbb{R}^k$ with speed $b_n = \frac{n}{a_n^2}$ and good rate function

$$I(x) := \frac{1}{2} \| \operatorname{diag}(W_1''(y_1^*), W_2''(y_2^*), W_1''(y_1^*), \dots)^{1/2} D\varphi_k^E(\vec{y}^*)^{-1} x\|_2^2.$$

The next result shows that for fluctuations of order 1 a new, non-universal rate function arises.

Theorem 3.3. *Let the conditions of Theorem 2.1, Theorem 2.5 or Theorem 2.7 be satisfied. Then in each of the three cases, the sequence $(m_1^{(n)}, \dots, m_k^{(n)})_n$ satisfies a large deviations principle on $\mathcal{M}_k(E)$ with speed n and good rate function $I(m) := \infty$ for $m \in \partial\mathcal{M}_k(E)$ and for $m \in \operatorname{Int}(\mathcal{M}_k(E))$*

$$I(m) := \sum_{j=1}^{\lfloor \frac{k+1}{2} \rfloor} \left\{ W_1(y_{2j-1}) - W_1(y_1^*) \right\} + \sum_{j=1}^{\lfloor \frac{k}{2} \rfloor} \left\{ W_2(y_{2j}) - W_2(y_2^*) \right\}.$$

Here $y_i^*, i = 1, 2$ are as in Theorem 3.1 and $y_j, j = 1, \dots, k$ are defined similarly as p_j ($E = [a, b]$), z_j ($E = \mathbb{R}_+$) or for $E = \mathbb{R}$ as $\alpha_{\frac{j+1}{2}}$ (j odd) and $\beta_{j/2}$ (j even).

We remark in passing that the case $E = [0, 1]$, $V_1 = V_2 \equiv 0$ is Theorem 2.6 in Gamboa and Lozada-Chang (2004).

4. Proofs

Proof of Lemma 2.6. For each vector of moments $(m_1, \dots, m_{2n})$ in the interior of the moment space $\mathcal{M}_{2n}(\mathbb{R})$, we can find a probability measure μ with infinite support and the first $2n$ moments given by $m_1, \dots, m_{2n}$. It is easy to see that the following relationship holds between the monic orthogonal polynomials P_k corresponding to μ and their recursion coefficients α_i, β_i,

$$\int x^k P_k(x)\, d\mu(x) = \beta_1 \cdots \beta_k \tag{4.1}$$

$$\int x^{k+1} P_k(x)\, d\mu(x) = \beta_1 \cdots \beta_k(\alpha_1 + \cdots + \alpha_{k+1}). \tag{4.2}$$

From this we can immediately see that $\beta_1, \dots, \beta_k$ only depend on the moments $m_1, \dots, m_{2k}$, while $\alpha_1, \dots, \alpha_k$ only depend on the moments $m_1, \dots, m_{2k-1}$. On the other hand, we may determine each moment m_{2k} from $\beta_1, \dots, \beta_k, \alpha_1, \dots, \alpha_k$ and each moment m_{2k-1} from $\beta_1, \dots, \beta_{k-1}, \alpha_1, \dots, \alpha_k$. Therefore the mapping $\varphi_{2n}^{\mathbb{R}}$ in (2.12) is a well-defined bijection between $(\alpha_1, \beta_1, \dots, \alpha_n, \beta_n)$ and $(m_1, \dots, m_{2n})$. The corresponding Jacobian matrix $D\varphi_{2n}^{\mathbb{R}}$ is a lower triagonal matrix with determinant given by

$$\det D\varphi_{2n}^{\mathbb{R}} = \prod_{k=1}^{n} \left(\frac{\partial m_{2k-1}}{\alpha_k} \cdot \frac{\partial m_{2k}}{\beta_k} \right).$$

In order to calculate these derivatives, note that since the P_k are monic orthogonal polynomials we have

$$\int x^k P_{k-1}(x)\, d\mu(x) = m_{2k-1} + \sum_{i=0}^{2k-2} \lambda_i m_i$$

for some real numbers λ_i (that may depend on k). Since $m_1, \dots, m_{2k-2}$ only depend on $\beta_1, \dots, \beta_{k-1}, \alpha_1, \dots, \alpha_{k-1}$, we get with (4.2)

$$\frac{\partial m_{2k-1}}{\partial \alpha_k} = \frac{\partial \int x^k P_{k-1}(x)\, d\mu(x)}{\partial \alpha_k} = \beta_1 \cdots \beta_{k-1}.$$

A similar argument using (4.1) shows

$$\frac{\partial m_{2k}}{\partial \beta_k} = \beta_1 \cdots \beta_{k-1},$$

which leads to

$$\det D\varphi_{2n}^{\mathbb{R}} = \prod_{k=1}^{n} \prod_{j=1}^{k-1} \beta_j^2 = \prod_{j=1}^{n-1} \prod_{k=j+1}^{n} \beta_j^2 = \prod_{j=1}^{n-1} \beta_j^{2n-2j}.$$

□

We will now prove the large deviations principles, as they play an important role in the proofs of Theorems 2.1, 2.5 and 2.7.

Proof of Theorem 3.3. For the sake of brevity we restrict ourselves to the case $E = [a, b]$, the remaining cases can be proved analogously. To this extent, we will show that each $p_{2i-1}^{(n)}$ satisfies a large deviations principle on $[0, 1]$ with good rate function

$$I_1(p) := W_1(p) - W_1(p_1^*),\ p \in (0, 1), \quad I_1(p) := \infty,\ p \in \{0, 1\}, \tag{4.3}$$

where $W_1(p) = V_1(p) - \log(p(1-p))$. Analogously, the $p_{2i}^{(n)}$ satisfy a large deviations principle on $[0, 1]$ with good rate function $I_2(p) := W_2(p) - W_2(p_2^*)$ on the interval $(0, 1)$ and ∞ elsewhere. The assertion then follows from the independence of the p_i's and the contraction principle. Note that $\varphi_k^{[a,b]}$ is bijective and thus the rate function does not change when passing from canonical to ordinary moments.

For the upper bound (3.2), let $U \subset [0,1]$ be a closed set. If $U \subset \{0,1\}$, (3.2) is trivially true by definition of $\mathbb{P}_{n,[a,b],V_{1,2}}$ and thus we may assume $U \cap (0,1) \neq \emptyset$. Then, setting $W^U := \inf_{x\in U} W_1(x)$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_U e^{-nV_1(x)+(n-i)\log(x(1-x))}\, dx \leq \limsup_{n\to\infty} \frac{1}{n} \log \int_0^1 e^{-iV_1(x)-(n-i)W^U}\, dx = -W^U.$$

For the lower bound (3.1), let $O \subset [0,1]$ be an open set and define $W^O := \inf_{x\in O} W_1(x)$. Let $\varepsilon > 0$ be arbitrary. By continuity of W on the interval $(0,1)$ and openness of O we know that $O \cap \{W_1 < W^O + \varepsilon\}$ is a nonempty open set. This yields

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{n} \log \int_O e^{-nV_1(x)+(n-i)\log(x(1-x))}\, dx \\
\geq\ &\liminf_{n\to\infty} \frac{1}{n} \log \int\limits_{O\cap\{W_1<W^O+\varepsilon\}} e^{-nV_1(x)+(n-i)\log(x(1-x))}\, dx \\
\geq\ &\liminf_{n\to\infty} \frac{1}{n} \log \int\limits_{O\cap\{W_1<W^O+\varepsilon\}} e^{-iV_1(x)-(n-i)(W^O+\varepsilon)}\, dx = -W^O - \varepsilon.
\end{aligned}$$

Now let $\varepsilon \to 0$, then the assertion finally follows from the choice $U = O = [0,1]$ which shows that the normalization constant of the density satisfies

$$\lim_{n\to\infty} \frac{1}{n} \log \int_0^1 e^{-nV_1(x)+(n-i)\log(x(1-x))}\, dx = -\inf_{y\in(0,1)} W_1(y).$$

$\square$

Next, we will prove the results on laws of large numbers in Section 2. It follows from Theorem 3.3 and the Borel-Cantelli lemma that in all three cases $(m_1^{(n)}, \dots, m_k^{(n)}) \to (m_1^*, \dots, m_k^*)$ almost surely as $n \to \infty$, where m_j^* are determined by $p_i^*, z_i^*, i = 1,2$ or α^*, β^*, respectively. The convergence in L^1 follows for $E = [a,b]$ immediately by the boundedness of the moments. For unbounded E, it suffices to see that the $m_j^{(n)}$'s are uniformly integrable thanks to the exponential decay from the large deviations principle. It remains to identify the corresponding measures to the moment sequences $(m_1^*, m_2^*, \dots)$. The general technique to do this is to consider the Jacobi operator associated to the recurrence coefficients of the orthogonal polynomials and derive an equation for the Stieltjes transform of the desired measure via a continued fraction expansion. We start with the simplest case of Theorem 2.7, where we explain the strategy in detail.

We will make use of the following lemma.

Lemma 4.1. *Let μ be a Borel probability measure on $\mathbb{R}$ that is determined by its moments (i.e. the Hamburger moment problem to the moments of μ is determinate). Let $\alpha_1, \beta_1, \alpha_2, \beta_2 \dots$ denote the recurrence coefficients of the monic orthogonal polynomials to the measure μ (see (2.11)). If μ is supported on N points, we set $\beta_j := 0$ for $j \geq N$. Then the Stieltjes transform of μ,*

$$\Phi(z) := \int \frac{d\mu(x)}{z-x},$$

defined for $z \in \mathbb{C}^+ := \{z \in \mathbb{C} : \Im z > 0\}$, has the continued fraction expansion

$$\Phi(z) = \frac{1 \rfloor}{\lceil z-\alpha_1} - \frac{\beta_1 \rfloor}{\lceil z-\alpha_2} - \frac{\beta_2 \rfloor}{\lceil z-\alpha_3} - \dots .$$

Here the convergents

$$\frac{1}{|z-\alpha_1} - \frac{\beta_1}{|z-\alpha_2} - \cdots - \frac{\beta_l}{|z-\alpha_{l+1}}$$

converge locally uniformly in $\mathbb{C}^+$ *as* $l \to \infty$.

Although the connection between continued fractions, Stieltjes transforms and orthogonal polynomials is classical and this result should be well-known, we did not manage to find this lemma in the literature. For measures with compact support, it is called Markov's theorem. We will give an elementary derivation.

Proof of Lemma 4.1. Let μ be a measure whose support consists of precisely N distinct points. Then the monic orthogonal polynomials $P_1, \ldots, P_N$ up to order N with respect to μ and the corresponding recursion coefficients $\alpha_1, \beta_1, \alpha_2, \beta_2, \ldots, \beta_{N-1}, \alpha_N$ are well-defined. Moreover, if μ has masses $\omega_1, \ldots, \omega_N$ at the points $t_1, \ldots, t_N$ and m_j denotes the j-th moment of μ, the monic orthogonal polynomial P_N is proportional to the polynomial

$$\begin{aligned}\tilde{P}_N(t) &= \det \begin{pmatrix} 1 & m_1 & \ldots & m_{N-1} & 1 \\ m_1 & m_2 & \ldots & m_N & t \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ m_N & m_{N+1} & \ldots & m_{2N-1} & t^N \end{pmatrix} \\ &= \sum_{i_0=1}^{N} \cdots \sum_{i_{N-1}=1}^{N} \omega_{i_0} \ldots \omega_{i_{N-1}} t_{i_1}^1 t_{i_2}^2 \ldots t_{i_{N-1}}^{N-1} \det \begin{pmatrix} 1 & 1 & \ldots & 1 & 1 \\ t_{i_0} & t_{i_1} & \ldots & t_{i_{N-1}} & t \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ t_{i_0}^N & t_{i_1}^N & \ldots & t_{i_{N-1}}^N & t^N \end{pmatrix} .\end{aligned}$$

Now the determinant in the last line vanishes whenever two indices i_j and i_k coincide. If all indices are different, the determinant is equal (up to a sign) to the polynomial $\ell(t) = \prod_{i=1}^{N}(t-t_i)$. Consequently, the polynomials $\tilde{P}_N$ and P_N are also proportional to $\ell(t)$ and therefore vanish precisely at the the support points $t_1, \ldots t_N$ of the measure μ.

We now define for $z \in \mathbb{C}^+$ the continued fraction

$$f_j(z) := \frac{1}{|z-\alpha_1} - \frac{\beta_1}{|z-\alpha_2} - \frac{\beta_2}{|z-\alpha_3} - \cdots - \frac{\beta_{j-1}}{|z-\alpha_j}, \quad j = 1, \ldots, N.$$

Writing $f_j(z)$ as a single fraction $\frac{A_j(z)}{B_j(z)}$, we see that $A_j(z)$ and $B_j(z), j = 1, \ldots, m$ satisfy the recursions $A_0(z) := 0, B_0(z) := 1,\ A_1(z) := 1, B_1(z) := z - \alpha_1$ and

$$\begin{aligned} A_j(z) &= (z-\alpha_j)A_{j-1}(z) - \beta_{j-1}A_{j-2}(z), \\ B_j(z) &= (z-\alpha_j)B_{j-1}(z) - \beta_{j-1}B_{j-2}(z) \end{aligned}$$

for $2 \le j \le N$. Clearly, B_j is a polynomial in z of degree j with leading coefficient 1 and as it satisfies the same recursion as the orthogonal polynomials P_j, we conclude $B_j = P_j$ for $0 \le j \le N$. Furthermore, note that the sequence of functions

$$Q_j(z) := \int \frac{P_j(z) - P_j(t)}{z-t}\, d\mu(t)$$

satisfies the same recursion as A_j, from which we can conclude $Q_j = A_j$ for $0 \le j \le N$. As the roots of P_N are precisely the support points of the measure μ we obtain

$$f_N(z) = \frac{A_N(z)}{B_N(z)} = \frac{1}{P_N(z)} \int \frac{P_N(z)}{z-t}\, d\mu(t) = \int \frac{1}{z-t}\, d\mu(t),$$

which concludes the proof for a measure μ with finite support.

If μ has infinite support, all recursion coefficients β_j are strictly positive. Let N be an arbitrary natural number. There is a unique measure μ_N supported on N points such that the corresponding monic orthogonal polynomials have the recursion coefficients $\alpha_1, \beta_1, \ldots, \beta_{N-1}, \alpha_N$. By the arguments above, the Stieltjes transform of μ_N has the form

$$f_N(z) = \frac{1 |}{| z-\alpha_1} - \frac{\beta_1 |}{| z-\alpha_2} - \frac{\beta_2 |}{| z-\alpha_3} - \cdots - \frac{\beta_{N-1} |}{| z-\alpha_N}.$$

Since the recursion coefficients up to order N determine the moments of μ_N up to order $2N-1$, we know that $m_j(\mu_N) = m_j(\mu)$ for $1 \leq j \leq 2N-1$. Letting $N \to \infty$ thus shows $\lim_{N\to\infty} m_j(\mu_N) = m_j(\mu)$ for all j. Since the measure μ is uniquely determined by its moments, this implies the weak convergence $\mu_N \xrightarrow{w} \mu$. For any fixed $z \in \mathbb{C}^+$, the function $t \mapsto \frac{1}{z-t}$ is a bounded continuous function. Therefore the Stieltjes transform of μ_N converges to the Stieltjes transform of μ, i.e.

$$\int \frac{1}{z-t}\, d\mu(t) = \lim_{N\to\infty} \int \frac{1}{z-t}\, d\mu_N(t) = \frac{1 |}{| z-\alpha_1} - \frac{\beta_1 |}{| z-\alpha_2} - \frac{\beta_2 |}{| z-\alpha_3} - \cdots.$$

As $z \mapsto \frac{1}{z-t}$ is analytic in $\mathbb{C}^+$ and uniformly bounded away from the real line, f_N is analytic in $\mathbb{C}^+$ and for any compact $K \subset \mathbb{C}^+$ we have $\sup_{N, z\in K} |f_N(z)| \leq M$ for some $M > 0$. It follows by Montel's theorem that the convergence is uniform on K.

□

Proof of Theorem 2.7. Let $\mu_{SC,\alpha^*,\beta^*}$ be the measure for which the recurrence coefficients of the associated monic orthogonal polynomials are $\alpha_j = \alpha^*$ and $\beta_j = \beta^*$ for all j. From (2.12) we know that $\mu_{SC,\alpha^*,\beta^*}$ has finite moments. By Carleman's criterion (in terms of recurrence coefficients, see (Shohat and Tamarkin, 1943, p. 59), the Hamburger moment problem for the moments of $\mu_{SC,\alpha^*,\beta^*}$ is determinate, if

$$\sum_{j=1}^{\infty} \frac{1}{\sqrt{\beta_j}} = \infty, \tag{4.4}$$

which is clearly the case here. Thus by Lemma 4.1 the Stieltjes transform

$$\Phi_{SC,\alpha^*,\beta^*}(z) := \int \frac{d\mu_{SC,\alpha^*,\beta^*}(x)}{z-x},$$

has the continued fraction expansion

$$\Phi_{SC,\alpha^*,\beta^*}(z) = \frac{1 |}{| z-\alpha^*} - \frac{\beta^* |}{| z-\alpha^*} - \cdots = \frac{1}{z-\alpha^* - \beta^* \Phi_{SC,\alpha^*,\beta^*}(z)}, \tag{4.5}$$

where the dots ... in (4.5) mean a continued repetition of the last fraction before the dots. Solving algebraically for $\Phi_{SC,\alpha^*,\beta^*}(z)$ yields the two solutions

$$\frac{z-\alpha^* \mp \sqrt{(z-\alpha^*)^2 - 4\beta^*}}{2\beta^*}.$$

Since any Stieltjes transform maps the upper half plane to the lower half plane, we get

$$\Phi_{SC,\alpha^*,\beta^*}(z) = \frac{z-\alpha^* - \sqrt{(z-\alpha^*)^2 - 4\beta^*}}{2\beta^*}, \tag{4.6}$$

where we define $\sqrt{(z-\alpha^*)^2-4\beta^*}$ for $z\in\mathbb{C}^+$ as the branch with positive imaginary part. Note that $\sqrt{(z-\alpha)^2-4\beta}$ admits a continuous extension from $\mathbb{C}^+$ to $\mathbb{R}$ via

$$\lim_{y\to 0+}\sqrt{(x+iy-\alpha)^2-4\beta}=\begin{cases}-\sqrt{(x-\alpha)^2-4\beta} & ,x<\alpha-2\sqrt{\beta}\\ i\sqrt{4\beta-(x-\alpha)^2} & ,x\in[\alpha-2\sqrt{\beta},\alpha+2\sqrt{\beta}]\\ \sqrt{(x-\alpha)^2-4\beta} & ,x>\alpha+2\sqrt{\beta}\end{cases}.$$

Thus $\Phi_{SC,\alpha^*,\beta^*}$ has a continuous extension from the upper half plane to the real line and μ_{α^*,β^*} has a density on $\mathbb{R}$ which is given by the Stieltjes inversion formula (see e.g. (Nica and Speicher, 2006, Remark 2.20))

$$\begin{aligned}\frac{\mu_{\alpha^*,\beta^*}(dx)}{dx}&=-\frac{1}{\pi}\lim_{y\to 0+}\Im\Phi_{SC,\alpha^*,\beta^*}(x+iy)\\&=\frac{1}{2\pi\beta^*}\sqrt{4\beta^*-(x-\alpha^*)^2}\mathbb{1}_{[\alpha^*-2\sqrt{\beta^*},\alpha^*+2\sqrt{\beta^*}]}(x).\end{aligned}\tag{4.7}$$

It is well-known that (see (Nica and Speicher, 2006, Corollary 2.14)) the j-th moment of the semicircle distribution $\mu_{SC,0,1}$ is $\frac{1}{j+1}\binom{2j}{j}$, (2.15) follows by a simple computation. □

Proof of Theorem 2.1. Let $\mu_{p_1^*,p_2^*}$ be the probability measure determined by having canonical odd moments p_1^* and canonical even moments p_2^*. For a probability measure on $[a,b]$ with canonical moments $p_1,p_2,p_3,\dots$ the recurrence coefficients of its monic orthogonal polynomials are given by (cf. (Dette and Studden, 1997, Corollary 2.3.4, eq. (1.4.6)))

$$\begin{aligned}\alpha_j&=a+(b-a)(q_{2j-3}p_{2j-2}+q_{2j-2}p_{2j-1}),\\ \beta_j&=(b-a)^2q_{2j-2}p_{2j-1}q_{2j-1}p_{2j},\, j=1,\dots.\end{aligned}$$

Here we set $p_{-1}=p_0=0$ and as usual $q_j:=1-p_j$. In our case $\alpha_1=a+(b-a)p_1^*$, $\beta_1=(b-a)^2p_1^*q_1^*p_2^*$, and for $j\geq 2$ we have $\alpha_j=a+(b-a)(p_1^*q_2^*+p_2^*q_1^*)$, $\beta_j=(b-a)^2p_1^*q_1^*p_2^*q_2^*$. Since $[a,b]$ is compact, the moment problem is determinate and hence Lemma 4.1 yields that the Stieltjes transform

$$\Phi_{p_1^*,p_2^*}(z):=\int\frac{d\mu_{p_1^*,p_2^*}(x)}{z-x}$$

has the continued fraction expansion

$$\begin{aligned}\Phi_{p_1^*,p_2^*}(z)=&\frac{1}{\big| z-a-(b-a)p_1^*}\Big|-\frac{(b-a)^2p_1^*q_1^*p_2^*}{\big| z-a-(b-a)(p_1^*q_2^*+p_2^*q_1^*)}\Big|\\&-\frac{(b-a)^2p_1^*q_1^*p_2^*q_2^*}{\big| z-a-(b-a)(p_1^*q_2^*+p_2^*q_1^*)}\Big|-\dots,\\ =&\frac{1}{\big| z-a-(b-a)p_1^*}\Big|-(b-a)^2p_1^*q_1^*p_2^*\Phi_{SC,\alpha,\beta}(z),\end{aligned}$$

where $\Phi_{SC,\alpha,\beta}$ is from (4.5) with $\alpha:=a+(b-a)(p_1^*q_2^*+p_2^*q_1^*)$, $\beta:=(b-a)^2p_1^*q_1^*p_2^*q_2^*$. Thus by (4.6)

$$\begin{aligned}\Phi_{p_1^*,p_2^*}(z)&=\frac{2q_2^*}{2q_2^*(z-a-(b-a)p_1^*)-(z-\alpha-\sqrt{(z-\alpha)^2-4\beta})}\\&=\frac{(1-2p_2^*)z+\alpha-2q_2^*(a+(b-a)p_1^*)-\sqrt{(z-\alpha)^2-4\beta}}{2p_2^*(z-a)(b-z)}.\end{aligned}$$

As atoms of $\mu_{p_1^*,p_2^*}$ are simple poles of the Stieltjes transform, atoms can only be at a or b. They can be identified using the formula

$$\mu_{p_1^*,p_2^*}(\{x\}) = -\lim_{y\to 0+} y\Im\Phi_{p_1^*,p_2^*}(x+iy). \tag{4.8}$$

Using this, we get after some algebra for $x = a$

$$\mu_{p_1^*,p_2^*}(\{a\}) = \frac{p_2^* - p_1^* + |p_2^* - p_1^*|}{2p_2^*} = \begin{cases} 0, & \text{if } p_1^* \geq p_2^* \\ 1 - \frac{p_1^*}{p_2^*}, & \text{if } p_1^* < p_2^* \end{cases}.$$

For $x = b$, we get similarly

$$\mu_{p_1^*,p_2^*}(\{b\}) = \frac{p_1^* + p_2^* - 1 + |1 - p_1^* - p_2^*|}{2p_2^*} = \begin{cases} 0, & \text{if } p_1^* + p_2^* \leq 1, \\ \frac{p_1^*+p_2^*-1}{p_2^*}, & \text{if } p_1^* + p_2^* > 1. \end{cases}$$

$\Phi_{p_1^*,p_2^*}(z)$ has a continuous extension to $\mathbb{R}\setminus\{a,b\}$. Thus the measure is absolutely continuous on $\mathbb{R}\setminus\{a,b\}$ and the density of the absolutely continuous part $\mu^{ac}_{p_1^*,p_2^*}$ can be computed using (4.7) as

$$\frac{\mu^{ac}_{p_1^*,p_2^*}(dx)}{dx} = \frac{\sqrt{4\beta - (\alpha - x)^2}}{2\pi p_2^*(x-a)(b-x)}$$

for $x \in [\alpha - 2\sqrt{\beta}, \alpha + 2\sqrt{\beta}]$, and 0 elsewhere. This proves (1), since $l_\pm = \alpha \pm 2\sqrt{\beta}$.

For (2) we use the well-known fact from potential theory (cf. e.g. (Saff and Totik, 1997, Theorem I.3.3)) that μ is the minimizing measure of (2.4) if and only if it satisfies the Euler-Lagrange equations

$$Q(t) - 2\int \log|t-s|d\mu(s) \begin{cases} = l & , \text{ if } t \in \operatorname{supp}(\mu), \\ \geq l & , \text{ if } t \notin \operatorname{supp}(\mu), \end{cases} \tag{4.9}$$

where l is a real constant. Differentiating, we get for $t \in \operatorname{supp}(\mu)$

$$Q'(t) = 2H_\mu(t), \tag{4.10}$$

where

$$H_\mu(t) := \int \frac{d\mu(s)}{t-s}$$

is the Hilbert transform of μ. Note that the integral is understood as a principal value integral. The Hilbert transform of an absolutely continuous measure can be obtained from its Stieltjes transform Φ_μ via (see e.g. (Hiai and Petz, 2000, p. 94))

$$H_\mu(t) = \lim_{y\to 0+} \Re\Phi_\mu(t+iy).$$

In our case this gives together with (4.10)

$$Q'(t) = \frac{(1-2p_2^*)t + \alpha - 2q_2^*(a + (b-a)p_1^*)}{p_2^*(t-a)(b-t)} = -\frac{p_1^* - p_2^*}{p_2^*(t-a)} + \frac{1 - p_1^* - p_2^*}{p_2^*(b-t)}.$$

Integration yields

$$Q(t) = -\left(\frac{p_1^*}{p_2^*} - 1\right)\log(t-a) - \left(\frac{1 - p_1^* - p_2^*}{p_2^*}\right)\log(b-t) \tag{4.11}$$

on the support. The integration constant does not matter here and thus is set to 0 for simplicity. We will consider Q defined by (4.11) as function $Q : [a,b] \to \mathbb{R}\cup\{+\infty\}$. By construction, Q

satisfies the equation of (4.9) on the support of $\mu_{p_1^*,p_2^*}$. For the inequality in (4.9), we compute the Hilbert transform $H_{\mu_{p_1^*,p_2^*}}$ outside of the support of $\mu_{p_1^*,p_2^*}$ as

$$H_{\mu_{p_1^*,p_2^*}}(t) = \begin{cases} \frac{Q'(t)}{2} + \frac{\sqrt{(t-\alpha)^2-4\beta}}{2p_2^*(t-a)(b-t)} \geq \frac{Q'(t)}{2} & , t \leq \alpha - 2\sqrt{\beta}, \\ \frac{Q'(t)}{2} - \frac{\sqrt{(t-\alpha)^2-4\beta}}{2p_2^*(t-a)(b-t)} \leq \frac{Q'(t)}{2} & , t \geq \alpha + 2\sqrt{\beta}. \end{cases}$$

Consequently, $Q(t) - 2\int \log|t-s|\, d\mu_{p_1^*,p_2^*}(s)$ is nonincreasing on $[a, l_-)$, constant on $[l_-, l_+]$ and nondecreasing on $(l_+, b]$. This implies the inequality in (4.9) and thus proves (2). □

Proof of Theorem 2.5. It is not difficult to see that the recurrence coefficients for the orthogonal polynomials to a probability measure μ on $\mathbb{R}_+$ with canonical moments $z_1, z_2, \ldots$ are given by

$$\alpha_j = z_{2j-2} + z_{2j-1},$$
$$\beta_j = z_{2j-1} z_{2j}, j \geq 1$$

with the convention $z_0 := 0$.

Let μ_{MP,z_1^*,z_2^*} be the probability measure on $\mathbb{R}_+$ with canonical moments $z_{2j-1} = z_1^*$ and $z_{2j} = z_2^*$, $j = 1, \ldots$. Then the recursion coefficients of the corresponding orthogonal polynomials are $\alpha_1 = z_1^*, \alpha_j = z_1^* + z_2^*, j \geq 2$ and $\beta_j = z_1^* z_2^*, j \geq 1$. The Stieltjes transform of μ_{MP,z_1^*,z_2^*} will be denoted by Φ_{MP,z_1^*,z_2^*}. By (4.4), the moment problem is determinate and thus Φ_{MP,z_1^*,z_2^*} admits the continued fraction expansion

$$\Phi_{MP,z_1^*,z_2^*}(z) = \frac{1}{\left| z - z_1^* \right.} - \frac{z_1^* z_2^*}{\left| z - (z_1^* + z_2^*) \right.} - \quad \ldots \quad = \frac{1}{z - z_1^* - z_1^* z_2^* \Phi_{SC,\alpha,\beta}(z)},$$

where $\Phi_{SC,\alpha,\beta}(z)$ is from (4.5) with $\alpha := (z_1^* + z_2^*)$ and $\beta = z_1^* z_2^*$. Using (4.6), this gives

$$\begin{aligned} \Phi_{MP,z_1^*,z_2^*}(z) &= \frac{2\beta}{2\beta(z - z_1^*) - z_1^* z_2^*(z - \alpha - \sqrt{(z-\alpha)^2 - 4\beta})} \\ &= \frac{z - z_1^* + z_2^* - \sqrt{(z-\alpha)^2 - 4\beta}}{2z_2^* z}. \end{aligned}$$

Clearly, μ_{MP,z_1^*,z_2^*} can have an atom only at 0. A computation using (4.8) gives

$$\mu_{MP,z_1^*,z_2^*}(\{0\}) = \frac{z_2^* - z_1^* - |z_1^* - z_2^*|}{2z_2^*} = \begin{cases} 0, & \text{if } z_2^* \geq z_1^*, \\ 1 - \frac{z_1^*}{z_2^*}, & \text{if } z_2^* < z_1^*. \end{cases}$$

The density of the absolutely continuous part can again be determined using (4.7) as

$$\frac{\mu_{MP,z_1^*,z_2^*}(dx)}{dx} = \frac{\sqrt{4\beta - (\alpha - x)^2}}{2\pi z_2^* x}$$

for $x \in [\alpha - 2\sqrt{\beta}, \alpha + 2\sqrt{\beta}]$, $x \neq 0$, and 0 elsewhere. Now the statement of the theorem follows noting $l_\pm = \alpha \pm 2\sqrt{\beta}$. □

Proof of Theorem 3.1. We will only prove the case $E = \mathbb{R}_+$, as the remaining parts are shown by similar arguments. Consider a moment vector under the distribution $\mathbb{P}_{n,\mathbb{R}_+,V_{1,2}}$ defined by the density (2.10). We will show that the canonical moments satisfy

$$\sqrt{n}(z_{2i-1}^{(n)} - z_1^*) \xrightarrow{d} \mathcal{N}(0, W_1''(z_1^*)^{-1})$$
$$\sqrt{n}(z_{2i}^{(n)} - z_2^*) \xrightarrow{d} \mathcal{N}(0, W_2''(z_2^*)^{-1})$$

as $n \to \infty$. The assertion of the theorem then follows from the independence of the $z_i^{(n)}$ and an application of the delta-method.

By Scheffé's Lemma, weak convergence of a sequence of measures can be proved by showing pointwise convergence of the corresponding densities. The density of $\sqrt{n}(z_{2i-1}^{(n)} - z_1^*)$ is given by

$$f_n(x) := \frac{g_n(x)}{c_n},$$

where

$$g_n(x) := \exp\left\{-n(W_1(z_1^* + \tfrac{x}{\sqrt{n}}) - W_1(z_1^*))\right\}(z_1^* + \tfrac{x}{\sqrt{n}})^{-(2i-1)} 1_{\left\{z_1^* + \frac{x}{\sqrt{n}} > 0\right\}}$$

and c_n is an appropriate normalization constant. By Taylor's theorem we obtain that

$$W_1(z_1^* + x/\sqrt{n}) = W_1(z_1^*) + \frac{x^2}{2n} W_1''(z_1^* + \lambda x/\sqrt{n})$$

holds for some $\lambda \in [0,1]$. From this we can easily conclude

$$g_n(x) \xrightarrow{n\to\infty} \exp(-W_1''(z_1^*)x^2/2)(z_1^*)^{-(2i-1)},$$

and it remains to prove the convergence of the normalization constant. By assumption $W_1''(z_1^*) \neq 0$ and since z_1^* is a minimizer of W_1, we have $W_1''(z_1^*) > 0$. Hence we may choose $0 < \varepsilon < z_1^*$ so small that the inequality $W_1''(x) > W_1''(z_1^*)/2$ is satisfied for all x with $|x - z_1^*| < \varepsilon$. This yields

$$\begin{aligned}
c_n &= \int_{-z_1^*\sqrt{n}}^{\infty} \exp\left\{-n(W_1(z_1^* + x/\sqrt{n}) - W_1(z_1^*))\right\}(z_1^* + x/\sqrt{n})^{-(2i-1)}\, dx \\
&= \int_{-\varepsilon\sqrt{n}}^{\varepsilon\sqrt{n}} \exp\left\{-n(W_1(z_1^* + x/\sqrt{n}) - W_1(z_1^*))\right\}(z_1^* + x/\sqrt{n})^{-(2i-1)}\, dx + o(1) \\
&\xrightarrow{n\to\infty} \int \exp\left\{-W_1''(z_1^*)x^2/2\right\}(z_1^*)^{-(2i-1)}\, dx = \sqrt{\frac{2\pi}{W_1''(z_1^*)}}(z_1^*)^{-(2i-1)}.
\end{aligned}$$

Here we have used the dominated convergence theorem with dominating function

$$g(x) := \exp\left\{-W_1''(z_1^*)x^2/4\right\}(z_1^* - \varepsilon)^{-(2i-1)}.$$

The $o(1)$ term stems from the fact that outside of $(-\varepsilon\sqrt{n}, \varepsilon\sqrt{n})$ the function $W_1(z_1^* + x/\sqrt{n}) - W(z_1^*)$ is bounded from below by some positive constant $K > 0$. The remaining integral can then be bounded by

$$\sqrt{n}\exp(-(n-(2i-1))K) \int_0^\infty \exp\left\{ -(2i-1)(V_1(x) - V_1(z_1^*) + \log(z_1^*(1-z_1^*)))\right\} dx = o(1).$$

Hence the density f_n converges pointwise to a centered normal distribution with variance $1/W_1''(z_1^*)$, which completes the proof of the first part of the theorem.

It remains to determine the entries of $D\varphi_k^{\mathbb{R}_+}$ in the case $z_1^* = z_2^*$. In order to do this, we will follow the arguments in Dette and Nagel (2012). Therein, a double sequence $g_{i,j}$ is defined by

$$g_{i,j} := \begin{cases} 1 & , \text{ if } i = 0, \\ 0 & , \text{ if } i \neq 0, i > j, \\ g_{i,j-1} + z_{j-i+1} g_{i-1,j} & , \text{ if } i \neq 0, i \leq j. \end{cases}$$

An induction argument over the sum $i+j$ shows that $g_{i,j}$ is a homogeneous polynomial of degree i in $z_1, z_2, \ldots$. Consequently, the partial derivative $\frac{dg_{i,j}}{dz_k}$ is a homogeneous polynomial of degree $i-1$. Following the arguments of Dette and Nagel (2012) we have $g_{k,k} = m_k$ with

$$\frac{dm_i}{dz_r}(1,1,\ldots) = \binom{2i}{i-r} - \binom{2i}{i-r-1}$$

and thus

$$\frac{dm_i}{dz_r}(z_1^*, z_1^*, \ldots) = (z_1^*)^{i-1}\frac{dm_i}{dz_r}(1,1,\ldots) = (z_1^*)^{i-1}\left(\binom{2i}{i-r} - \binom{2i}{i-r-1}\right).$$

□

Proof of Theorem 3.2. We will only prove the case $E = [a,b]$, the remaining cases are treated similarly. We will first show that each $a_n(p_{2j-1}^{(n)} - p_1^*)$ satisfies a large deviations principle with good rate function $J(x) := W_1''(p_1^*)x^2/2$ and speed b_n, where $(a_n)_n$ and $(b_n)_n$ are chosen as in Theorem 3.2. In order to see this, let $U \subset \mathbb{R}$ be an arbitrary closed set and $0 < \varepsilon < 1$ sufficiently small so that $W_1''(y) \geq M > 0$ holds for all $y \in (p^* - \varepsilon, p^* + \varepsilon)$ and some constant $M > 0$. Set $\gamma := \inf_{x \in U} |x|$, $R(p) := (p(1-p))^{-(2i-1)}$ and let I_1 be the function (4.3). Note that $I_1 \geq 0$ with unique zero p_1^* and $I_1'' = W_1''$. For (3.2) we show first

$$\limsup_{n\to\infty} \frac{1}{b_n} \log \int_U e^{-nI_1(x/a_n + p_1^*)} R(x/a_n + p_1^*)\, dx \leq -W_1''(p^*)\frac{\gamma^2}{2}.$$

The case $\gamma = \infty$ is trivial, since then $U = \emptyset$, so we may assume $\gamma < \infty$. We will first consider $U \cap \{|x| \geq \varepsilon a_n\}$. We get

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{b_n} \log \int_U 1_{\{|x|\geq \varepsilon a_n\}} e^{-nI_1(x/a_n + p_1^*)} R(x/a_n + p_1^*)\, dx \\
&\leq \limsup_{n\to\infty} \frac{1}{b_n} \log \int_{\mathbb{R}} 1_{\{|x|\geq \varepsilon a_n\}} e^{-(2i-1)V_1(x/a_n + p_1^*)} \exp\Big(-(n-(2i-1)) \inf_{|y-p_1^*|\geq\varepsilon} I_1(y)\Big)\, dx \\
&\leq \limsup_{n\to\infty} \frac{1}{b_n} \log \int_{\mathbb{R}} a_n e^{-(2i-1)V_1(t)} \exp\Big(-(n-(2i-1)) \inf_{|y-p_1^*|\geq\varepsilon} I_1(y)\Big)\, dt \\
&\leq \limsup_{n\to\infty} a_n^2\Big(\log a_n - (n-(2i-1)) \inf_{|y-p_1^*|\geq\varepsilon} I_1(y)\Big)/n = -\infty.
\end{aligned}$$

For the set $U \cap \{|x| < \varepsilon a_n\}$, note that by Taylor's theorem

$$\begin{aligned}
&\int_U 1_{\{|x|<\varepsilon a_n\}} e^{-nI_1(x/a_n + p_1^*)} R(x/a_n + p_1^*)\, dx \\
&\leq \sup_{|y-p_1^*|<\varepsilon} R(y) \int_U 1_{\{|x|<\varepsilon a_n\}} \exp\big(-nx^2/(2a_n^2) \inf_{|z-p_1^*|<\varepsilon} W_1''(z)\big)\, dx \\
&\leq \sup_{|y-p_1^*|<\varepsilon} R(y) \int_{\mathbb{R}} \exp\Big(-\big((1-\varepsilon)n\gamma^2/(2a_n^2) + \varepsilon n x^2/(2a_n)\big) \inf_{|z-p_1^*|<\varepsilon} W_1''(z)\Big)\, dx \\
&\leq \sup_{|y-p_1^*|<\varepsilon} R(y) \exp\big(-(1-\varepsilon)b_n\gamma^2/2 \inf_{|z-p_1^*|<\varepsilon} W_1''(z)\big)\sqrt{2\pi\Big/\Big(\varepsilon b_n \inf_{|z-p_1^*|<\varepsilon} W_1''(z)\Big)}.
\end{aligned}$$

Consequently,

$$\limsup_{n\to\infty} \frac{1}{b_n} \log \int_U 1_{\{|x|<\varepsilon a_n\}} e^{-nI_1(x/a_n + p_1^*)} R(x/a_n + p_1^*)\, dx \leq -(1-\varepsilon) \inf_{|z-p_1^*|<\varepsilon} W_1''(z)\frac{\gamma^2}{2}.$$

Using $\log(a+b) \leq \log 2 + \max\{\log a, \log b\}$, $a, b \geq 0$, we conclude

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{b_n} \log \int_U e^{-nI_1(x/a_n+p_1^*)} R(x/a_n+p_1^*)\, dx \\
\leq\ & \max\Big\{ \limsup_{n\to\infty} \frac{1}{b_n} \log \int_U 1_{\{|x|<\varepsilon a_n\}} e^{-nI_1(x/a_n+p_1^*)} R(x/a_n+p_1^*)\, dx, \\
&\qquad \limsup_{n\to\infty} \frac{1}{b_n} \log \int_U 1_{\{|x|\geq\varepsilon a_n\}} e^{-nI_1(x/a_n+p_1^*)} R(x/a_n+p_1^*)\, dx \Big\} + \frac{\log 2}{b_n} \\
\leq\ & -(1-\varepsilon) \inf_{|z-p_1^*|<\varepsilon} W_1''(z) \frac{\gamma^2}{2}.
\end{aligned}$$

Letting $\varepsilon \to 0$ now yields

$$\limsup_{n\to\infty} \frac{1}{b_n} \log \int_U e^{-nI_1(x/a_n+p_1^*)} R(x/a_n+p_1^*)\, dx \leq -W_1''(p_1^*) \frac{\gamma^2}{2}. \tag{4.12}$$

For the lower bound (3.1), let $O \subset \mathbb{R}$ be an arbitrary nonempty open set. Set again $\gamma := \inf_{x\in O} |x| < \infty$. By the definition of γ the set $O \cap \{|x| < \gamma + \varepsilon\}$ is a nonempty open set. Therefore by Taylor's theorem

$$\begin{aligned}
&\int_O e^{-nI_1(x/a_n+p_1^*)} R(x/a_n+p_1^*)\, dx \\
\geq\ & \int_O 1_{\{|x|<\gamma+\varepsilon\}} e^{-nI_1(x/a_n+p_1^*)} R(x/a_n+p_1^*)\, dx \\
\geq\ & \inf_{|y-p^*|<(\gamma+\varepsilon)/a_n} R(y) \lambda(O \cap \{|x| < \gamma+\varepsilon\}) \exp\Big(-n(\gamma+\varepsilon)^2/(2a_n) \sup_{|y-p_1^*|<(\gamma+\varepsilon)/a_n} W_1''(y) \Big),
\end{aligned}$$

where λ is the Lebesgue measure. This yields

$$\liminf_{n\to\infty} \frac{1}{b_n} \log \int_O e^{-nI_1(x/a_n+p_1^*)} R(x/a_n+p_1^*)\, dx \geq -W_1''(p_1^*) \frac{(\gamma+\varepsilon)^2}{2}.$$

Letting $\varepsilon \to 0$ we therefore get

$$\liminf_{n\to\infty} \frac{1}{b_n} \log \int_O e^{-nI_1(x/a_n+p_1^*)} R(x/a_n+p_1^*)\, dx \geq -W_1''(p_1^*) \frac{\gamma^2}{2}. \tag{4.13}$$

Note that the density of $a_n(p_{2i-1}^{(n)} - p_1^*)$ is

$$\frac{1}{c_n} e^{-nI_1(x/a_n+p_1^*)} R(x/a_n+p_1^*)\, dx,$$

where c_n is the normalization constant. Plugging $U = O = \mathbb{R}$ into (4.12) and (4.13) shows $\lim_{n\to\infty} \frac{1}{b_n} \log c_n = 0$. This proves the large deviations principle for $a_n(p_{2i-1}^{(n)} - p_1^*)$.

Analogously, $a_n(p_{2i} - p_2^*)$ satisfies a large deviation principle with speed b_n and good rate function $W_2''(p_2^*)x^2/2$. Since the canonical moments are independent, we can conclude that the vector

$$a_n\big((p_1^{(n)}, \ldots, p_k^{(n)}) - \vec{y}^*\big)$$

satisfies a large deviations principle with speed b_n and good rate function $\|Hx\|_2^2/2$, where the matrix H is given by $H = \operatorname{diag}(W_1''(p_1^*), W_2''(p_2^*), W_1''(p_1^*), \ldots)^{1/2} \in \mathbb{R}^{k\times k}$. Recall that $\vec{y}^* = (p_1^*, p_2^*, p_1^*, \ldots) \in (0,1)^k$.

In order to transfer this large deviations principle to the sequence of ordinary moments, we need to apply the delta-method for large deviations. As Theorem 3.1 in Gao and Zhao (2011) states, the sequence

$$a_n\big((m_1^{(n)},\ldots,m_k^{(n)})-(m_1^*,\ldots,m_k^*)\big)=a_n\big(\varphi_k^{[a,b]}(p_1^{(n)},\ldots,p_k^{(n)})-\varphi_k^{[a,b]}(\vec{y}^*)\big)$$

satisfies a large deviations principle with good rate function

$$I(x):=\inf\{\|Hy\|_2^2/2 \mid (D\varphi_k^{[a,b]}(\vec{y}^*))y=x\}=\|HD\varphi_k^{[a,b]}(\vec{y}^*)^{-1}x\|_2^2/2.$$

□

Acknowledgements. The authors would like to thank M. Stein who typed parts of this manuscript with considerable technical expertise. The work of H. Dette and D. Tomecki was supported by the Deutsche Forschungsgemeinschaft (DFG Research Unit 1735, DE 502/26-2, RTG 2131: High-dimensional Phenomena in Probability - Fluctuations and Discontinuity). The work of M. Venker was supported by the European Research Council under the European Unions Seventh Framework Programme (FP/2007/2013)/ ERC Grant Agreement n. 307074 as well as by CRC 701 "Spectral Structures and Topological Methods in Mathematics".

References

Akemann, G., Baik, J., and Di Francesco, P., editors (2011). *The Oxford handbook of random matrix theory.* Oxford University Press, Oxford.

Anshelevich, M. (2007). Free Meixner states. *Commun. Math. Phys.*, 276(3):863–899.

Bai, Z. and Silverstein, J. W. (2010). *Spectral analysis of large dimensional random matrices.* Springer Series in Statistics. Springer, New York, second edition.

Castro, M. M. and Grünbaum, F. A. (2013). On a seminal paper by Karlin and McGregor. *SIGMA Symmetry Integrability Geom. Methods Appl.*, 9:Paper 020, 11.

Chang, F. C., Kemperman, J. H. B., and Studden, W. J. (1993). A normal limit theorem for moment sequences. *Ann. Probab.*, 21:1295–1309.

Chihara, T. S. (1978). *An Introduction to Orthogonal Polynomials.* Gordon and Breach, New York.

Cohen, J. M. and Trenholme, A. R. (1984). Orthogonal polynomials with a constant recursion formula and an application to harmonic analysis. *J. Funct. Anal.*, 59(2):175–184.

Dembo, A. and Zeitouni, O. (2010). *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability.* Springer-Verlag, Berlin. Corrected reprint of the second (1998) edition.

Dette, H. and Nagel, J. (2012). Distributions on unbounded moment spaces and random moment sequences. *Ann. Probab.*, 40(6):2690–2704.

Dette, H. and Studden, W. J. (1997). *Canonical Moments with Applications in Statistics, Probability and Analysis.* Wiley and Sons, New York.

Dumitriu, I. and Edelman, A. (2002). Matrix models for beta ensembles. *J. Math. Phys.*, 43(11):5830–5847.

Gamboa, F. and Lozada-Chang, L. V. (2004). Large deviations for random power moment problem. *Ann. Probab.*, 32:2819–2837.

Gamboa, F., Nagel, J., and Rouault, A. (2016). Sum rules via large deviations. *J. Funct. Anal.*, 270(2):509–559.

Gamboa, F., Nagel, J., and Rouault, A. (2017). Sum rules and large deviations for spectral measures on the unit circle. *Random Matrices Theory Appl.*, 6(1):1750005, 49.

Gao, F. and Zhao, X. (2011). Delta method in large deviations and moderate deviations for estimators. *Ann. Statist.*, 39(2):1211–1240.

Hamburger, H. (1920). Über eine Erweiterung des Stieltjesschen Momentenproblems. *Math. Ann.*, 81:235–319.

Hiai, F. and Petz, D. (2000). *The Semicircle Law, Free Random Variables and Entropy.* American Mathematical Society, R.I.

Karlin, S. and Shapeley, L. S. (1953). Geometry of moment spaces. In *Amer. Math. Soc. Memoir No. 12.* American Mathematical Society, Providence, Rhode Island.

Karlin, S. and Studden, W. (1966). *Tchebycheff systems: with applications in analysis and statistics.* Interscience Publishers.

Kesten, H. (1959). Symmetric random walks on groups. *Trans. Amer. Math. Soc.*, 92:336–354.

Krein, M. G. and Nudelman, A. A. (1977). *The Markov Moment Problem and Extremal Problems.* American Mathematical Society., Providence, RI.

Krishnapur, M., Rider, B., and Virág, B. (2016). Universality of the stochastic Airy operator. *Comm. Pure Appl. Math.*, 69(1):145–199.

Lozada-Chang, L. V. (2005). Large deviations on moment spaces. *Electron. J. Probab.*, 10:662–690.

McKay, B. D. (1981). The expected eigenvalue distribution of a large regular graph. *Linear Algebra Appl.*, 40:203–216.

Meixner, J. (1934). Orthogonale Polynomsysteme mit einer besonderen Gestalt der erzeugenden Funktion. *J. London Math. Soc.*, S1-9(1):6.

Nica, A. and Speicher, R. (2006). *Lectures on the combinatorics of free probability.* London Mathematical Society Lecture Note Series 335, Cambridge.

Saff, E. and Totik, V. (1997). *Logarithmic potentials with external fields*, volume 316 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin. Appendix B by Thomas Bloom.

Saitoh, N. and Yoshida, H. (2001). The infinite divisibility and orthogonal polynomials with a constant recursion formula in free probability theory. *Probab. Math. Statist.*, 21(1, Acta Univ. Wratislav. No. 2298):159–170.

Shohat, J. A. and Tamarkin, J. D. (1943). *The Problem of Moments.* American Mathematical Society Mathematical surveys, vol. I. American Mathematical Society, New York.

Skibinsky, M. (1967). The range of the $(n+1)$-th moment for distributions on $[0;1]$. *J. Appl. Probability*, 4:543–552.

Skibinsky, M. (1968). Extreme nth moments for distributions on [0, 1] and the inverse of a moment space map. *J. Appl. Probability*, 5:693–701.

Skibinsky, M. (1969). Some striking properties of binomial and beta moments. *Ann. Math. Stat.*, 40:1753–1764.

Stahl, H. and Totik, V. (1992). *General orthogonal polynomials*, volume 43 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, Cambridge.

Verblunsky, S. (1935). On positive harmonic functions: A contribution to the algebra of Fourier series. *Proc. London Math. Soc.*, 38:125–157.

Verblunsky, S. (1936). On positive harmonic functions (second paper). *Proc. London Math. Soc.*, 40:290–320.

[*]Department of Mathematics, Ruhr University Bochum, 44780 Bochum, Germany
E-mail address: holger.dette@ruhr-uni-bochum.de

[†]Department of Mathematics, Ruhr University Bochum, 44780 Bochum, Germany
E-mail address: dominik.tomecki@ruhr-uni-bochum.de

[‡]Faculty of Mathematics, Bielefeld University, 33501 Bielefeld, Germany
E-mail address: mvenker@math.uni-bielefeld.de