Background: Asymptotic Freeness and Large Deviations

We start by recalling the basic asymptotic freeness result.

> **Proposition:** Let $\left(A_1^N, \dots, A_r^N\right)$ be an independent r-tuple of GUE $N \times N$ matrices. Then, the family $A_1^N, \dots, A_r^N$ converge in distribution to r independent semicircular elements, $s_1, \dots, s_r \in B(\mathcal{F}(\mathbb{C}^r))$, in the sense that for all $m \geq 1$ and all $1 \leq i_1, \dots, i_m \leq r$, we have
>
> $$\lim_{N\to\infty} E\left[\operatorname{tr}\left(A_{i_1}^N \cdots A_{i_m}^N\right)\right] = \varphi(s_{i_1} \cdots s_{i_m}),$$
>
> where φ is the vacuum state, $\varphi(T) = \langle T\Omega, \Omega\rangle$.

In fact, this collection is *almost surely* asymptotically free, in the following sense. Suppose we have two random matrices A^N and B^N defined on probability spaces (X_N, μ_N). Define

$$X \coloneqq \prod_{N\in\mathbb{N}} X_N$$
$$\mu \coloneqq \prod_{N\in\mathbb{N}} \mu_N,$$

where the latter is the product measure on X. The matrices A^N and B^N are said to be almost surely asymptotically free if there exists a noncommutative probability space (A, φ) and $a, b \in A$, and for almost all $x = (x_N)_N \in X$, we have $A^N(x_N), B^N(x_N) \in (\mathbb{M}_N, \operatorname{tr})$ converge in distribution to a, b.

Now, from here, we may ask a seemingly simple question: as N grows large, how likely are we to encounter other distributions? To make this sense more precise, we consider a random $N \times N$ self-adjoint matrix A, and let

$$\mu_A = \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_1}$$

be its empirical spectral distribution. This is a random probability measure on $\mathbb{R}$, and as $N \to \infty$, the semicircle law gives that μ_A converges weakly to the semicircle distribution; this can be strengthened to almost sure convergence by an application of the argument for asymptotic freeness. The question then becomes, how quickly does the deviation between μ_A and any other probability distribution ν decrease as N increases? This is where the theory of large deviations starts to take shape.

Much of this exposition related to the classical notions of entropy will be centered around results discussed in [MS17, Ch. 7].

Large Deviations for Random Variables

We start with one of the classical examples of convergence of random variables to introduce large deviations. Consider a sequence of independent and identically distributed real-valued random variables $(X_i)_i$ with common distribution μ. Set

$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$
$$m = E[X_1]$$
$$v = E\left[X_1^2\right] - m^2.$$

Then, we have that if $E\left[X_1^2\right] < \infty$, the central limit theorem says that $S_n \approx m + \frac{\sigma}{\sqrt{n}} N(0,1)$.

If μ is the standard Gaussian distribution, then this gives that S_n is distributed as $N(0, 1/n)$; we then get that

$$P(S_n \in [x, x+dx]) \approx \sqrt{\frac{n}{2\pi}} e^{-nx^2}\, dx.$$

Asymptotically, this gives that the probability that S_n is near the value $x \in \mathbb{R}$ decays exponentially in n determined by a rate function $I(x) = x^2/2$.

We will now generalize this result. In particular, if we let μ be any distribution discussed above (rather than simply the normal distribution), then we will find a rate function $I(x)$ such that

$$e^{-nI(x)} \sim P(S_n > x)$$

whenever $x > m$, and whenever $x < m$

$$e^{-nI(x)} \sim P(S_n < x).$$

For a given distribution μ, we can compute the rate function by using a family of basic manipulations. If $x > m$, then for all $\lambda \geq 0$, we may use Markov's inequality to obtain

$$\begin{aligned} P(S_n > x) &= P(nS_n > nx) \\ &= P\Big(e^{\lambda(nS_n - nx)} \geq 1\Big) \\ &\leq E\Big[e^{\lambda(nS_n - nx)}\Big] \\ &= e^{-\lambda nx} E\Big[e^{\lambda(X_1 + \cdots + X_n)}\Big] \\ &= \big(e^{-\lambda x} E\big[e^{\lambda X}\big]\big)^n, \end{aligned}$$

where X is identically distributed to each of the X_i, and we use the fact that the X_i are independent. We may then define

$$\Lambda(\lambda) = \ln E\big[e^{\lambda X}\big] \tag{$*$}$$

to be an extended real-valued function, but we only consider μ for which $\Lambda(\lambda)$ is finite for all λ in an open neighborhood of 0. The equation $(*)$ is known as the cumulant generating function for μ.

This gives the inequality

$$P(S_n > x) \leq e^{-n(\lambda x - \Lambda(\lambda))}.$$

Since ln is a concave function, Jensen's inequality gives

$$\begin{aligned} \Lambda(\lambda) &\geq E\big[\ln\big(e^{\lambda X}\big)\big] \\ &= E[\lambda X] \\ &= \lambda m. \end{aligned}$$

In particular, for any $\lambda < 0$ and $x > m$, we have $-n(\lambda x - \Lambda(\lambda)) \geq 0$, meaning this equation is valid for all λ. In particular, we have

$$P(S_n > x) \leq \inf_{\lambda \in \mathbb{R}} e^{-n\lambda x - \Lambda(\lambda)}.$$

Now, we observe that Λ is convex. This follows from Hölder's inequality

$$E\Big[e^{(1-t)\lambda_1 x + t\lambda_2 x}\Big] \leq E\big[e^{\lambda_1 x}\big]^{1-t} E\big[e^{\lambda_2 x}\big]^t$$

so that

$$\Lambda((1-t)\lambda_1 + t\lambda_2) \leq (1-t)\Lambda(\lambda_1) + t\Lambda(\lambda_2).$$

Defining the *Legendre transform* of Λ by

$$\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}} (\lambda x - \Lambda(\lambda)),$$

we find that this is a convex function of x, as it is a supremum of a family of convex functions of x.

Now, since $\Lambda(0) = 0$, it follows that $\Lambda^*(x) \geq 0$, and has $\Lambda^*(m) = 0$. In particular, this gives

$$P(S_n > x) \leq e^{-n\Lambda^*(x)}$$

whenever $x > m$.

It can also be shown that $e^{-n\Lambda^*(x)}$ is an asymptotic lower bound, in that

$$\liminf_{n\to\infty} \frac{1}{n} \ln P(x - \delta < S_n < x + \delta) \geq -\Lambda^*(x)$$

for all $x \in \mathbb{R}$ and all $\delta > 0$. The method for doing so is outlined in [MS17, Ch. 7, Section 2], and results in Cramér's theorem for real-valued random vectors.

Theorem (Cramér's Theorem)**:** Let $X_1, X_2, \dots$ be a sequence of independent and identically distributed random vectors in $\mathbb{R}^d$ with common distribution μ. Define

$$\begin{aligned} \Lambda(\lambda) &= \ln E\left[e^{\langle \lambda, X_i \rangle}\right] \\ \Lambda^*(x) &= \sup_{\Lambda \in \mathbb{R}^d} (\langle \lambda, x \rangle - \Lambda(\lambda)), \end{aligned}$$

and assume that $\Lambda(\lambda) < \infty$ for all λ. Set $S_n = \frac{1}{n}(X_1 + \cdots + X_n)$. Then, the distribution μ_{S_n} satisfies has

- $x \mapsto \Lambda^*(x)$ is convex;
- $\left\{x \in \mathbb{R}^d \mid \Lambda^*(x) \leq \alpha\right\}$ is compact for all $\alpha \in \mathbb{R}$;
- for any closed $F \subseteq \mathbb{R}^d$,
$$\limsup_{n\to\infty} \frac{1}{n} \ln P(S_n \in F) \leq -\inf_{x \in F} \Lambda^*(x),$$
- and for any open $G \subseteq \mathbb{R}^d$,
$$\liminf_{n\to\infty} \frac{1}{n} \ln P(S_n \in G) \geq -\inf_{x \in G} \Lambda^*(x).$$

This theorem defines precisely the large deviation principle that the partial sums satisfy — namely, it is the Legendre transform of the cumulant-generating function.

Large Deviations for the Empirical Distribution

Now, our next step is to develop an analogous large deviation principle for the empirical distribution of the random variables $X_1, X_2, \dots$. This will give us the idea of classical entropy.

We start by considering the case of (independent and identically distributed) random variables $X_i \colon \Omega \to A$ taking values in a finite set $\{a_1, \dots, a_d\}$, and define $p_k \coloneqq P(X_i = a_k)$. We expect that, as $n \to \infty$, the empirical distribution of the X_i should converge to the probability measure $(p_1, \dots, p_d)$ on A.

Toward this end, let $Y_i \colon \Omega \to \mathbb{R}^d$ be defined by

$$Y_i \coloneqq \left(\chi_{\{a_1\}}(X_i), \dots, \chi_{\{a_d\}}(X_i)\right).$$

We observe that p_k is the probability that Y_i will have 1 in position k and 0 elsewhere, and that $\frac{1}{n}(Y_1 + \cdots + Y_n)$ gives the relative frequency of $a_1, \dots, a_d$ — i.e., this has the same information as the empirical distribution of $X_1, \dots, X_n$.

Any probability measure on A is a d-tuple of positive real numbers satisfying $q_1 + \cdots + q_d = 1$. By Cramér's theorem and our discussion above, we have

$$\begin{aligned} P\left(\frac{1}{n}(\delta_{X_1} + \cdots + \delta_{X_n}) \approx (q_1, \ldots, q_d)\right) &= P\left(\frac{1}{n}(Y_1 + \cdots + Y_n) \approx (q_1, \ldots, q_d)\right) \\ &\sim e^{-n\Lambda^*(q_1,\ldots,q_d)}. \end{aligned}$$

Applying our definitions for Λ and Λ^*, we have

$$\begin{aligned} \Lambda(\lambda_1, \ldots, \lambda_d) &= \ln\bigl(p_1 e^{\lambda_1} + \cdots + p_d e^{\lambda_d}\bigr) \\ \Lambda^*(q_1, \ldots, q_d) &= \sup_{(\lambda_1,\ldots,\lambda_d)} \bigl(\lambda_1 q_1 + \cdots + \lambda_d q_d - \ln\bigl(p_1 e^{\lambda_1} + \cdots + p_d e^{\lambda_d}\bigr)\bigr). \end{aligned}$$

To compute the supremum over all tuples, we find that the partial derivatives with respect to each λ_i are given by

$$q_i - \frac{1}{p_1 e^{\lambda_1} + \cdots + p_d e^{\lambda_d}} p_i e^{\lambda_i},$$

so by concavity, we get that the maximum value occurs when

$$\lambda_i = \ln\left(\frac{q_i}{p_i}\right) + \Lambda(\lambda_1, \ldots, \lambda_d).$$

We thus get

$$\Lambda^*(q_1, \ldots, q_d) = \sum_{i=1}^{d} q_i \ln\left(\frac{q_i}{p_i}\right).$$

The quantity on the right is the relative Shannon entropy $H((q_1, \ldots, q_d)|(p_1, \ldots, p_d))$, which is strictly positive except when $q_1 = p_1, \ldots, q_d = p_d$.

In particular, we have that $(p_1, \ldots, p_d)$ admits a large deviation principle with rate function given by the relative Shannon entropy. That this holds for any distribution is known as Sanov's theorem.

Theorem (Sanov's Theorem)**:** Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed real random variables with common distribution μ, and let

$$\nu_n \coloneqq \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$$

be the empirical distribution. Then, the family $\{\nu_n\}_{n\geq 1}$ satisfies a large deviation principle given by the rate function

$$I(\nu) = \begin{cases} \int p \ln(p)\, d\mu & d\nu = p\, d\mu \\ +\infty & \text{else.} \end{cases}$$

One-Dimensional Free Entropy: A Heuristic Approach

The next logical step after defining a large deviation principle for a sequence of random variables is to define such a quantity for single free random variables.

In [Voi93], Voiculescu used a heuristic method based on random matrix theory to establish the value $\Sigma(x)$ for some random variable x in a C^*-probability space $(\mathcal{A}, \varphi)$ distributed according to a compactly supported measure ν on $\mathbb{R}$ (i.e., its spectral measure with respect to the state φ).

It can be established (as in [MS17, Exercise 1.8]) that if X is a $N \times N$ GUE matrix, then the density inside the real vector space $\mathbb{M}_N(\mathbb{C})_{\text{s.a.}}$ is given by

$$dP_N(X) = K e^{-N^2/2\,\mathrm{tr}\left(X^2\right)}\, dm,$$

where m is the Lebesgue measure on $\mathbb{R}^{N^2}$ and tr denotes the normalized trace. It can be shown, as in [AGZ10, Theorem 2.5.2], that the joint distribution of the eigenvalues $\lambda_1(X) \leq \cdots \leq \lambda_N(X)$ is absolutely continuous with respect to the Lebesgue measure and has density given by

$$dQ_N(\lambda_1, \dots, \lambda_N) = \frac{N^{N^2/2}}{(2\pi)^{N/2} \prod_{j=1}^{N} j!} e^{-\left(N \sum_{i=1}^{N} \lambda_i^2\right)/2} \prod_{i<j} (\lambda_i - \lambda_j)^2 \prod_{i=1}^{N} d\lambda_i$$

Heuristically, letting μ_A denote the empirical spectral distribution on $\mathbb{M}_N(\mathbb{C})_{\text{s.a.}}$ given by

$$\mu_A = \frac{1}{N}\left(\delta_{\lambda_1(A)} + \cdots + \delta_{\lambda_N(A)}\right),$$

we may consider a large deviation principle with $P_N(\mu_A \approx \nu) \sim e^{-N^2 I(\nu)}$, and write it out as

$$\begin{aligned} P_N(\mu_A \approx \nu) &= Q_N\left(\frac{1}{N}(\delta_{\lambda_1} + \cdots + \delta_{\lambda_N})\right) \\ &= \frac{N^{N^2/2}}{(2\pi)^{N/2} \prod_{j=1}^{N} j!} \int_E e^{-\left(N \sum_{i=1}^{N} \lambda_i^2\right)/2} \prod_{i<j} (\lambda_i - \lambda_j)^2 \prod_{i=1}^{N} d\lambda_i, \end{aligned}$$

where E is the given set in the Q_N. Whenever $\frac{1}{N}\left(\delta_{\lambda_1(A)} + \cdots + \delta_{\lambda_N(A)}\right) \approx \nu$, we have that

$$-\frac{N}{2} \sum_{i=1^N} \lambda_i^2 = -\frac{N^2}{2}\left(\frac{1}{N}\sum_{i=1}^{N} \lambda_i^2\right)$$

is a Riemann sum for the integral of t^2 with respect to ν. Furthermore, we have

$$\prod_{i<j} (\lambda_i - \lambda_j)^2 = e^{\sum_{i \neq j} \ln|\lambda_i - \lambda_j|}$$

The sum inside the argument of the exponential is a Riemann sum for the quantity $N^2 \iint \ln|s-t| d\nu(s) d\nu(t)$. All in all, we get the following large deviation principle, which was proven rigorously in [BAG97] for the general case of Gaussian random matrices with Dyson index β.

Theorem: Let

$$I(\nu) = -\iint \ln|s-t|\, d\nu(s)\nu(t) + \frac{1}{2}\int t^2\, d\nu(t) - \frac{3}{4}.$$

Then, the following hold.

(i) The function $I\colon \mathcal{M} \to [0, \infty]$ is a well-defined convex function with compact level sets on the space of probability measures on $\mathbb{R}$ that has minimum value 0 at Wigner's semicircle law.

(ii) The empirical spectral distribution satisfies a large deviation principle with respect to Q_N with rate function I. That is,

$$\liminf_{N \to \infty} \frac{1}{N^2} \ln Q_N\left(\frac{1}{N}(\delta_{\lambda_1} + \cdots + \delta_{\lambda_N}) \in G\right) \geq -\inf_{\nu \in G} I(\nu)$$

for any open $G \subseteq \mathcal{M}$, and

$$\limsup_{N \to \infty} \frac{1}{N^2} \ln Q_N\left(\frac{1}{N}(\delta_{\lambda_1} + \cdots + \delta_{\lambda_N}) \in F\right) \leq -\inf_{\nu \in F} I(\nu)$$

for any closed $F \subseteq \mathcal{M}$.

Microstates Free Entropy

Voiculescu's original definition of free entropy in the case with more than one free random variable can be viewed as a generalization of the microstates approach to entropy for a classical discrete random variable, discussed in the survey [Voi01]. For the sections following, I will assume a certain level of proficiency with the theory of von Neumann algebras, as this is the setting where Voiculescu developed the theories around free entropy in [Voi94].

Consider a discrete random variable with output values in $\{1, \dots, n\}$ assigned with probabilities $p_1, \dots, p_n$. The microstates of this discrete random variable for a fixed N are then the set

$$\{f \mid f\colon \{1, \dots, N\} \to \{1, \dots, n\}\},$$

and for a fixed ε, the microstates that approximate this distribution are those f with

$$\left| \frac{\left|f^{-1}(\{j\})\right|}{N} - p_j \right| < \varepsilon,$$

where $|\cdot|$ denotes the (necessarily finite) cardinality. We denote the collection of such f by $\Gamma(p_1, \dots, p_n; \varepsilon, N)$. One may find the Shannon entropy by evaluating the limit

$$\lim_{\varepsilon \to 0} \lim_{N \to \infty} \frac{1}{N} \ln|\Gamma(p_1, \dots, p_n; \varepsilon, N)|$$

Understanding the Microstates Free Entropy

Instead of a classical random variable, we will let (M, τ) be a tracial von Neumann algebra,[I] and let $(x_1, \dots, x_n)$ be a tuple of self-adjoint free random variables in M.

For a fixed R, we let the set of microstates $\Gamma_R(X_1, \dots, X_n; m, k, \varepsilon)$ admit three degrees of approximation: m denotes the level of approximation of mixed moments, k denotes the size of the (self-adjoint) approximation matrices that have operator norm at most R, and ε denotes the closeness of the approximation. Put into symbols, we select all n-tuples $(A_1, \dots, A_n) \in (\mathbb{M}_k(\mathbb{C})_{\text{s.a.}})^n$ with each $\|A_j\| < R$ satisfying

$$\left|\tau\left(x_{i_1} \cdots x_{i_p}\right) - \operatorname{tr}_k\left(A_{i_1} \cdots A_{i_p}\right)\right| < \varepsilon$$

for all $1 \leq p \leq m$ and all multi-indices $\mathbf{i}\colon \{1, \dots, p\} \to \{1, \dots, n\}$. Here, tr_k is the normalized trace.

As in the case of entropy for classical random variables, free entropy emerges from a certain large deviation principle. Specifically, recall that the distribution of a family of noncommutative random variables is given by the collection of mixed moments with respect to the trace,

$$\Delta(x_1, \dots, x_n) = \left\{\tau\left(x_{i_1} \cdots x_{i_p}\right) \;\middle|\; p \in \mathbb{N}, \mathbf{i}\colon \{1, \dots, p\} \to \{1, \dots, n\}\right\}.$$

If we have an n-tuple of independent GUE matrices $(A_1, \dots, A_n) \in (\mathbb{M}_k(\mathbb{C})_{\text{s.a.}})^n$, we know that as $k \to \infty$, that there is almost sure convergence in distribution to a family $(s_1, \dots, s_n)$ of free semicirculars in M. The large deviations will then be given by

$$P_N(\Delta(A_1, \dots, A_n) \approx \Delta(x_1, \dots, x_n)) \sim e^{-k^2 I(x_1, \dots, x_n)},$$

where I is the free entropy.

To compute this value, we let λ denote the Lebesgue measure on $(\mathbb{M}_k(\mathbb{C})_{\text{s.a.}})^n \cong \mathbb{R}^{nk^2}$, and define

$$\chi(x_1, \dots, x_n) = \sup_{R > 0} \inf_{m \in \mathbb{N}} \inf_{\varepsilon > 0} \limsup_{k \to \infty} \left(\frac{1}{k^2} \ln \lambda(\Gamma_R(x_1, \dots, x_n; m, k, \varepsilon)) + \frac{n}{2} \ln k\right).$$

It turns out that the value of R has minor influence, and we only need choose a fixed R greater than the norm of each x_i.

[I] A von Neumann algebra equipped with a faithful (injective on positive elements), normal (w^*-continuous), tracial ($\tau(xy) = \tau(yx)$) state ($\tau(1) = 1$) $\tau\colon M \to \mathbb{C}$.

Proposition ([Voi94, Proposition 2.2])**:** Let $C^2 = \tau\left(x_1^2 + \cdots + x_n^2\right)$. Then,

$$\chi(x_1, \dots, x_n) \leq 2^{-1} n \ln\left(2\pi e n^{-1} C^2\right).$$

Proof. We will instead prove a slightly different inequality. Define

$$\chi_R(x_1, \dots, x_n; m, k, \varepsilon) = \ln(\lambda(\Gamma_R(x_1, \dots, x_n; m, k, \varepsilon))),$$

and we will show that

$$\chi_R(x_1, \dots, x_n; m, k, \varepsilon) \leq 2^{-1} n k^2 \left(\ln\left(2\pi e n^{-1}\left(C^2 + n\varepsilon\right)\right) - \ln k\right),$$

assuming that $m \geq 2$. Applying the operations $\limsup_{k\to\infty}$, $\inf_{\varepsilon>0}$, $\inf_{m\in\mathbb{N}}$, and $\sup_{R>0}$ in succession will give us the free entropy.

For this, we use the p-dimensional Shannon entropy inequality

$$-\int f \ln(f)\, d\lambda_p \leq 2^{-1} p \ln\left(2\pi e p^{-1} a^2\right),$$

where f is some probability density function on $\mathbb{R}^p$ and

$$a^2 = \int \left(x_1^2 + \cdots + x_p^2\right) f\, d\lambda_p,$$

where λ_p denotes the Lebesgue measure on $\mathbb{R}^p$.

We apply this to the special case of the Lebesgue measure on $\left(\mathbb{M}_k(\mathbb{C})_{\text{s.a.}}\right)^n$, which is induced by the Hilbert–Schmidt metric

$$\langle (A_1, \dots, A_n), (B_1, \dots, B_n) \rangle = \sum_{i=1}^{n} \operatorname{tr}(A_i B_i),$$

and take the indicator function

$$f(A_1, \dots, A_n) = \begin{cases} 0 & (A_1, \dots, A_n) \notin \Gamma_R(x_1, \dots, x_n; m, k, \varepsilon) \\ (\lambda(\Gamma_R(x_1, \dots, x_n; m, k, \varepsilon)))^{-1} & (A_1, \dots, A_n) \in \Gamma_R(x_1, \dots, x_n; m, k, \varepsilon), \end{cases}$$

giving

$$\chi_R(x_1, \dots, x_n; m, k, \varepsilon) \leq 2^{-1} n k^2 \left(\ln\left(2\pi e n^{-1} k^{-2} a^2\right) - \ln k\right).$$

Finally, by the definition of the microstate space, we have

$$\left| \frac{1}{k} \int \operatorname{Tr}\left(\sum_{j=1}^{n} A_j^2\right) f\, d\lambda - \tau\left(\sum_{j=1}^{n} x_j^2\right) \right| < n\varepsilon,$$

meaning that $a^2 \leq C^2 + n\varepsilon$. □

Applications: Structural Properties of Free Group Factors

References

[Voi93] Dan Voiculescu. “The analogues of entropy and of Fisher’s information measure in free probability theory. I”. In: *Comm. Math. Phys.* 155.1 (1993), pp. 71–92. ISSN: 0010-3616,1432-0916. URL: http://projecteuclid.org/euclid.cmp/1104253200.

[Voi94] Dan Voiculescu. "The analogues of entropy and of Fisher's information measure in free probability theory. II". In: *Invent. Math.* 118.3 (1994), pp. 411–440. ISSN: 0020-9910,1432-1297. DOI: 10.1007/BF01231539. URL: https://doi.org/10.1007/BF01231539.

[BAG97] G. Ben Arous and A. Guionnet. "Large deviations for Wigner's law and Voiculescu's non-commutative entropy". In: *Probab. Theory Related Fields* 108.4 (1997), pp. 517–542. ISSN: 0178-8051,1432-2064. DOI: 10.1007/s004400050119. URL: https://doi.org/10.1007/s004400050119.

[Voi01] Dan Voiculescu. *Free Entropy*. 2001. arXiv: math/0103168 [math.OA]. URL: https://arxiv.org/abs/math/0103168.

[AGZ10] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An Introduction to Random Matrices*. Vol. 118. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2010, pp. xiv+492. ISBN: 978-0-521-19452-5.

[MS17] James A. Mingo and Roland Speicher. *Free Probability and Random Matrices*. Vol. 35. Fields Institute Monographs. Springer, New York; Fields Institute for Research in Mathematical Sciences, Toronto, ON, 2017, pp. xiv+336. ISBN: 978-1-4939-6941-8; 978-1-4939-6942-5. DOI: 10.1007/978-1-4939-6942-5. URL: https://doi.org/10.1007/978-1-4939-6942-5.