Chapter 7
Free Entropy χ: The Microstates Approach via Large Deviations

An important concept in classical probability theory is Shannon's notion of entropy. Having developed the analogy between free and classical probability theory, one hopes to find that a notion of *free entropy* exists in counterpart to the Shannon entropy. In fact there is a useful notion of free entropy. However, the development of this new concept is at present far from complete. The current state of affairs is that there are two distinct approaches to free entropy. These should give isomorphic theories, but at present we only know that they coincide in a limited number of situations.

The first approach to a theory of free entropy is via *microstates*. This is rooted in the concept of large deviations. The second approach is *microstates free*. This draws its inspiration from the statistical approach to classical entropy via the notion of Fisher information. The unification problem in free probability theory is to prove that these two theories of free entropy are consistent. We will in this chapter only talk about the first approach via microstates; the next chapter will address the microstates free approach.

7.1 Motivation

Let us return to the connection between random matrix theory and free probability theory which we have been developing. We know that a p-tuple $(A_N^{(1)}, \ldots, A_N^{(p)})$ of $N \times N$ matrices chosen independently at random with respect to the GUE density (compare Exercise 1.8), $P_N(A) = \text{const} \cdot \exp(-N\text{Tr}(A^2)/2)$, on the space of $N \times N$ Hermitian matrices converges almost surely (in moments with respect to the normalized trace) to a freely independent family $(s_1, \ldots, s_p)$ of semi-circular elements lying in a non-commutative probability space; see Theorem 4.4. The von Neumann algebra generated by p freely independent semi-circulars is the von Neumann algebra $L(\mathbb{F}_p)$ of the free group on p generators.

J.A. Mingo, R. Speicher, *Free Probability and Random Matrices*,
Fields Institute Monographs 35, DOI 10.1007/978-1-4939-6942-5_7

We ask now the following question: How likely is it to observe other distributions/operators for large N?

Let us consider the case $p = 1$ more closely. For a random Hermitian matrix $A = A^*$ (distribution as above) with real random eigenvalues $\lambda_1 \leq \cdots \leq \lambda_N$, denote by

$$\mu_A = \frac{1}{N}(\delta_{\lambda_1} + \cdots + \delta_{\lambda_N}) \tag{7.1}$$

the eigenvalue distribution of A (also known as the *empirical eigenvalue distribution*), which is a random measure on $\mathbb{R}$. Wigner's semi-circle law states that as $N \to \infty$, $P_N(\mu_A \approx \mu_W) \to 1$, where μ_W is the (non-random) semi-circular distribution and $\mu_A \approx \mu_W$ means that the measures are close in a sense that can be made precise. We are now interested in the deviations from this. What is the rate of decay of the probability $P_N(\mu_A \approx \nu)$, where ν is some measure (not necessarily the semi-circle)? We expect that

$$P_N(\mu_A \approx \nu) \sim e^{-N^2 I(\nu)} \tag{7.2}$$

for some *rate function* I vanishing at μ_W. By analogy with the classical theory of large deviations, I should correspond to a suitable notion of free entropy.

We used in the above the notion "≈" for meaning "being close" and "∼" for "behaves asymptotically (in N) like"; here they should just be taken on an intuitive level, later, in the actual theorems they will be made more precise.

In the next two sections, we will recall some of the basic facts of the classical theory of large deviations and, in particular, Sanov's theorem; this standard material can be found, for example, in the book [64]. In Section 7.4 we will come back to the random matrix question.

7.2 Large deviation theory and Cramér's theorem

Consider a real-valued random variable X with distribution μ. Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables with the same distribution as X, and put $S_n = (X_1 + \cdots + X_n)/n$. Let $m = E[X]$ and $\sigma^2 = \text{var}(X) = E[X^2] - m^2$. Then the law of large numbers asserts that $S_n \to m$, if $\mathrm{E}[|X|] < \infty$; while if $\mathrm{E}[X^2] < \infty$, the central limit theorem tells us that for large n

$$S_n \approx m + \frac{\sigma}{\sqrt{n}} N(0,1). \tag{7.3}$$

For example, if $\mu = N(0,1)$ is Gaussian, then $m = 0$ and S_n has the Gaussian distribution $N(0, 1/n)$, and hence

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

Thus the probability that S_n is near the value x decays exponentially in n at a rate determined by x, namely the *rate function* $I(x) = x^2/2$. Note that the convex function $I(x)$ has a global minimum at $x = 0$, the minimum value there being 0, which corresponds to the fact that S_n approaches the mean 0 in probability.

This behaviour is described in general by the following theorem of Cramér. Let X, μ, $\{X_i\}_i$, and S_n be as above. There exists a function $I(x)$, the rate function, such that

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

How does one calculate the rate function I for a given distribution μ? We shall let X be a random variable with the same distribution as the X_i's. For arbitrary $x > m$, one has for all $\lambda \geq 0$

$$\begin{aligned} P(S_n > x) &= P(nS_n > nx) \\ &= P(e^{\lambda(nS_n - nx)} \geq 1) \\ &\leq E[e^{\lambda(nS_n - nx)}] \qquad \text{(by Markov's inequality)} \\ &= e^{-\lambda nx} E[e^{\lambda(X_1 + \cdots + X_n)}] \\ &= (e^{-\lambda x} E[e^{\lambda X}])^n. \end{aligned}$$

Here we are allowing that $\mathrm{E}[e^{\lambda X}] = +\infty$. Now put

$$\Lambda(\lambda) := \log E[e^{\lambda X}], \tag{7.4}$$

the *cumulant generating series* of μ; c.f. Section 1.1. We consider Λ to be an extended real-valued function but here only consider μ for which $\Lambda(\lambda)$ is finite for all real λ in some open set containing 0; however, Cramér's theorem (Theorem 1) holds without this assumption. With this assumption Λ has a power series expansion with radius of convergence $\lambda_0 > 0$, and in particular all moments exist.

Exercise 1. Suppose that X is a real random variable and there is $\lambda_0 > 0$ so that for all $|\lambda| \leq \lambda_0$ we have $E(e^{\lambda X}) < \infty$. Then X has moments of all orders, and the function $\lambda \mapsto E(e^{\lambda X})$ has a power series expansion with a radius of convergence of at least λ_0.

Then the inequality above reads

$$P(S_n > x) \leq e^{-\lambda nx + n\Lambda(\lambda)} = e^{-n(\lambda x - \Lambda(\lambda))}, \tag{7.5}$$

which is valid for all $0 \le \lambda$. By Jensen's inequality we have, for all $\lambda \in \mathbb{R}$,

$$\Lambda(\lambda) = \log E[e^{\lambda X}] \ge E[\log e^{\lambda X}] = \lambda m. \tag{7.6}$$

This implies that for $\lambda < 0$ and $x > m$ we have $-n(\lambda x - \Lambda(\lambda)) \ge 0$, and so equation (7.5) is valid for all λ. Thus

$$P(S_n > x) \le \inf_\lambda e^{-n(\lambda x - \Lambda(\lambda))} = \exp\left(-n \sup_\lambda (\lambda x - \Lambda(\lambda))\right).$$

The function $\lambda \mapsto \Lambda(\lambda)$ is convex, and the *Legendre transform* of Λ defined by

$$\Lambda^*(x) := \sup_\lambda (\lambda x - \Lambda(\lambda)) \tag{7.7}$$

is also a convex function of x, as it is the supremum of a family of convex functions of x.

Exercise 2. Show that $(E(Xe^{\lambda X}))^2 \le E(e^{\lambda X})E(Xe^{\lambda X})$. Show that $\lambda \mapsto \Lambda(\lambda)$ is convex.

Note that $\Lambda(0) = \log 1 = 0$; thus, $\Lambda^*(x) \ge (0x - \Lambda(0)) = 0$ is non-negative, and hence equation (7.6) implies that $\Lambda^*(m) = 0$.

Thus, we have proved that, for $x > m$,

$$P(S_n > x) \le e^{-n\Lambda^*(x)}, \tag{7.8}$$

where Λ^* is the Legendre transform of the cumulant generating function Λ. In the same way, one proves the same estimate for $P(S_n < x)$ for $x < m$. This gives Λ^* as a candidate for the rate function. Moreover we have by Exercise 3 that $\lim_n \log[P(S_n > x)]^{1/n}$ exists and by Equation (7.8) this limit is less than $\exp(-\Lambda^*(x))$. If we assume that neither $P(X > x)$ nor $P(X < x)$ is 0, $\exp(-\Lambda^*(x))$ will be the limit. In general we have

$$-\inf_{y > x} \Lambda^*(y) \le \liminf_n \frac{1}{n} \log P(S_n > x) \le \limsup_n \frac{1}{n} \log P(S_n \ge x) \le -\inf_{y \ge x} \Lambda^*(y).$$

Exercise 3. Let $a_n = \log P(S_n > a)$. Show that

(i) for all m, n: $a_{m+n} \ge a_m + a_n$;
(ii) for all m

$$\liminf_{n\to\infty} \frac{a_n}{n} \ge \frac{a_m}{m};$$

(iii) $\lim_{n\to\infty} a_n/n$ exists.

However, in preparation for the vector-valued version, we will show that $\exp(-n\Lambda^*(x))$ is asymptotically a lower bound; more precisely, we need to verify that

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

for all x and all $\delta > 0$. By replacing X_i by $X_i - x$, we can reduce this to the case $x = 0$, namely, showing that

$$-\Lambda^*(0) \leq \liminf_{n\to\infty} \frac{1}{n} \log P(-\delta < S_n < \delta). \tag{7.9}$$

Note that $-\Lambda^*(0) = \inf_\lambda \Lambda(\lambda)$. The idea of the proof of (7.9) is then to perturb the distribution μ to $\tilde{\mu}$ such that $x = 0$ is the mean of $\tilde{\mu}$. Let us only consider the case where Λ has a global minimum at some point η. This will always be the case if μ has compact support and both $P(X > 0)$ and $P(X < 0)$ are not 0. The general case can be reduced to this by a truncation argument. With this reduction $\Lambda(\lambda)$ is finite for all λ, and thus Λ has an infinite radius of convergence (c.f. Exercise 1), and thus Λ is differentiable. So we have $\Lambda'(\eta) = 0$. Now let $\tilde{\mu}$ be the measure on $\mathbb{R}$ such that

$$d\tilde{\mu}(x) = e^{\eta x - \Lambda(\eta)} d\mu(x). \tag{7.10}$$

Note that

$$\int_{\mathbb{R}} d\tilde{\mu}(x) = e^{-\Lambda(\eta)} \int_{\mathbb{R}} e^{\eta x} d\mu(x) = e^{-\Lambda(\eta)} E[e^{\eta X}] = e^{-\Lambda(\eta)} e^{\Lambda(\eta)} = 1,$$

which verifies that $\tilde{\mu}$ is a probability measure. Consider now i.i.d. random variables $\{\tilde{X}_i\}_i$ with distribution $\tilde{\mu}$, and put $\tilde{S}_n = (\tilde{X}_1 + \cdots + \tilde{X}_n)/n$. Let $\tilde{X}$ have the distribution $\tilde{\mu}$. We have

$$\begin{aligned} E[\tilde{X}] = \int_{\mathbb{R}} x\, d\tilde{\mu}(x) &= e^{-\Lambda(\eta)} \int_{\mathbb{R}} x e^{\eta x} d\mu(x) = e^{-\Lambda(\eta)} \frac{d}{d\lambda} \int_{\mathbb{R}} e^{\lambda x} d\mu(x)\big|_{\lambda=\eta} \\ &= e^{-\Lambda(\eta)} \frac{d}{d\lambda} e^{\Lambda(\lambda)}\big|_{\lambda=\eta} = e^{-\Lambda(\eta)} \Lambda'(\eta) e^{\Lambda(\eta)} = \Lambda'(\eta) = 0. \end{aligned}$$

Now, for all $\epsilon > 0$, we have $\exp(\eta \sum x_i) \leq \exp(n\epsilon|\eta|)$ whenever $|\sum x_i| \leq n\epsilon$ and so

$$\begin{aligned} P(-\epsilon < S_n < \epsilon) &= \int_{|\sum_{i=1}^n x_i| < n\epsilon} d\mu(x_1) \cdots d\mu(x_n) \\ &\geq e^{-n\epsilon|\eta|} \int_{|\sum_{i=1}^n x_i| < n\epsilon} e^{\eta \sum x_i} d\mu(x_1) \cdots d\mu(x_n) \end{aligned}$$

$$= e^{-n\epsilon|\eta|} e^{n\Lambda(\eta)} \int_{|\sum_{i=1}^n x_i| < n\epsilon} d\tilde{\mu}(x_1)\cdots d\tilde{\mu}(x_n)$$
$$= e^{-n\epsilon|\eta|} e^{n\Lambda(\eta)} P(-\epsilon < \tilde{S}_n < \epsilon).$$

By the weak law of large numbers, $\tilde{S}_n \to E[\tilde{X}_i] = 0$ in probability, i.e. we have $\lim_{n\to\infty} P(-\epsilon < \tilde{S}_n < \epsilon) = 1$ for all $\epsilon > 0$. Thus for all $0 < \epsilon < \delta$

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{n} \log P(-\delta < S_n < \delta) &\geq \liminf_{n\to\infty} \frac{1}{n} \log P(-\epsilon < S_n < \epsilon)\\ &\geq \Lambda(\eta) - \epsilon|\eta|, \qquad \text{for all } \epsilon > 0\\ &\geq \Lambda(\eta)\\ &= \inf \Lambda(\lambda)\\ &= -\Lambda^*(0).\end{aligned}$$

This sketches the proof of Cramér's theorem for $\mathbb{R}$. The higher-dimensional form of Cramér's theorem can be proved in a similar way.

Theorem 1 (Cramér's Theorem for $\mathbb{R}^d$). *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random vectors, i.e. independent $\mathbb{R}^d$-valued random variables with common distribution μ (a probability measure on $\mathbb{R}^d$). Put*

$$\Lambda(\lambda) := \mathrm{E}[e^{\langle \lambda, X_i \rangle}], \ \lambda \in \mathbb{R}^d, \tag{7.11}$$

and

$$\Lambda^*(x) := \sup_{\lambda \in \mathbb{R}^d} \{\langle \lambda, x\rangle - \Lambda(\lambda)\}. \tag{7.12}$$

Assume that $\Lambda(\lambda) < \infty$ for all $\lambda \in \mathbb{R}^d$, and put $S_n := (X_1 + \cdots + X_n)/n$.

Then the distribution μ_{S_n} of the random variable S_n satisfies a large deviation principle with rate function Λ^**, i.e.*

- *$x \mapsto \Lambda^*(x)$ is lower semicontinuous (actually convex)*
- *Λ^* is* good*, i.e. $\{x \in \mathbb{R}^d : \Lambda^*(x) \leq \alpha\}$ is compact for all $\alpha \in \mathbb{R}$*
- *For any closed set $F \subset \mathbb{R}^d$,*

$$\limsup_{n\to\infty} \frac{1}{n} \log P(S_n \in F) \leq -\inf_{x\in F} \Lambda^*(x) \tag{7.13}$$

- *For any open set $G \subset \mathbb{R}^d$,*

$$\liminf_{n\to\infty} \frac{1}{n} \log P(S_n \in G) \geq -\inf_{x\in G} \Lambda^*(x). \tag{7.14}$$

7.3 Sanov's theorem and entropy

We have seen Cramér's theorem for $\mathbb{R}^d$; in an informal way, it says $P(S_n \approx x) \sim \exp(-n\Lambda^*(x))$. Actually, we are interested not in S_n, but in the empirical distribution $(\delta_{X_1} + \cdots + \delta_{X_n})/n$.

Let us consider this in the special case of random variables $X_i : \Omega \to A$, taking values in a finite alphabet $A = \{a_1, \ldots, a_d\}$, with $p_k := P(X_i = a_k)$. As $n \to \infty$, the empirical distribution of the X_i's should converge to the "most likely" probability measure $(p_1, \ldots, p_d)$ on A.

Now define the vector of indicator functions $Y_i : \Omega \to \mathbb{R}^d$ by

$$Y_i := (1_{\{a_1\}}(X_i), \ldots, 1_{\{a_d\}}(X_i)), \tag{7.15}$$

so that in particular p_k is equal to the probability that Y_i will have a 1 in the k-th spot and 0's elsewhere. Then the averaged sum $(Y_1 + \cdots + Y_n)/n$ gives the relative frequency of $a_1, \ldots, a_d$, i.e. it contains the same information as the empirical distribution of $(X_1, \ldots, X_n)$.

A probability measure on A is given by a d-tuple $(q_1, \ldots, q_d)$ of positive real numbers satisfying $q_1 + \cdots + q_d = 1$. By Cramér's theorem,

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

Here

$$\Lambda(\lambda_1, \ldots, \lambda_d) = \log E[e^{\langle \lambda, Y_i \rangle}] = \log(p_1 e^{\lambda_1} + \cdots + p_d e^{\lambda_d}).$$

Thus the Legendre transform is given by

$$\Lambda^*(q_1, \ldots, q_d) = \sup_{(\lambda_1, \ldots, \lambda_d)} \{\lambda_1 q_1 + \cdots + \lambda_d q_d - \Lambda(\lambda_1, \ldots, \lambda_d)\}.$$

We compute the supremum over all tuples $(\lambda_1, \ldots, \lambda_d)$ by finding the partial derivative $\partial/\partial\lambda_i$ of $\lambda_1 q_1 + \cdots + \lambda_d q_d - \Lambda(\lambda_1, \ldots, \lambda_d)$ to be

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

By concavity the maximum occurs when

$$\lambda_i = \log \frac{q_i}{p_i} + \log(p_1 e^{\lambda_1} + \cdots + p_d e^{\lambda_d}) = \log \frac{q_i}{p_i} + \Lambda(\lambda_1, \ldots, \lambda_d),$$

and we compute

$$\begin{aligned}
&\Lambda^*(q_1,\dots,q_d)\\
&= q_1\log\frac{q_1}{p_1}+\cdots+q_d\log\frac{q_d}{p_d}+(q_1+\cdots+q_d)\Lambda(\lambda_1,\dots,\lambda_d)-\Lambda(\lambda_1,\dots,\lambda_d)\\
&= q_1\log\frac{q_1}{p_1}+\cdots+q_d\log\frac{q_d}{p_d}.
\end{aligned}$$

The latter quantity is Shannon's relative entropy, $H((q_1,\dots,q_d)|(p_1,\dots,p_d))$, of $(q_1,\dots,q_d)$ with respect to $(p_1,\dots,p_d)$. Note that $H((q_1,\dots,q_d)|(p_1,\dots,p_d))\geq 0$, with equality holding if and only if $q_1=p_1,\dots,q_d=p_d$.

Thus $(p_1,\dots,p_d)$ is the most likely realization, with other realizations exponentially unlikely; their unlikelihood is measured by the rate function Λ^*; and this rate function is indeed Shannon's relative entropy. This is the statement of Sanov's theorem. We have proved it here for a finite alphabet; it also holds for continuous distributions.

Theorem 2 (Sanov's Theorem). *Let $X_1,X_2,\dots$ be i.i.d. real-valued random variables with common distribution μ, and let*

$$\nu_n=\frac{1}{n}(\delta_{X_1}+\cdots+\delta_{X_n}) \tag{7.16}$$

be the empirical distribution of $X_1,\dots,X_n$, which is a random probability measure on $\mathbb{R}$. Then $\{\nu_n\}_n$ satisfies a large deviation principle with rate function $I(\nu)=S(\nu,\mu)$ (called the relative entropy*) given by*

$$I(\nu)=\begin{cases}\int p(t)\log p(t)d\mu(t), & \text{if } d\nu=p\,d\mu\\ +\infty, & \text{otherwise.}\end{cases} \tag{7.17}$$

Concretely, this means the following. Consider the set $\mathcal{M}$ of probability measures on $\mathbb{R}$ with the weak topology (which is a metrizable topology, e.g. by the Lévy metric). Then for closed F and open G in $\mathcal{M}$, we have

$$\limsup_{n\to\infty}\frac{1}{n}\log P(\nu_n\in F)\leq-\inf_{\nu\in F}S(\nu,\mu) \tag{7.18}$$

$$\liminf_{n\to\infty}\frac{1}{n}\log P(\nu_n\in G)\geq-\inf_{\nu\in G}S(\nu,\mu). \tag{7.19}$$

7.4 Back to random matrices and one-dimensional free entropy

Consider again the space $\mathcal{H}_N$ of Hermitian matrices equipped with the probability measure P_N having density

$$dP_N(A)=\text{const}\cdot e^{-\frac{N}{2}\text{Tr}(A^2)}dA. \tag{7.20}$$

We let $\mathbb{R}^N_{\geq} = \{(x_1, \ldots, x_N) \in \mathbb{R}^N \mid x_1 \leq \cdots \leq x_N\}$. For a self-adjoint matrix A, we write the eigenvalues of A as $\lambda_1(A) \leq \cdots \leq \lambda_N(A)$. The joint eigenvalue distribution $\tilde{P}_N$ on $\mathbb{R}^N_{\geq}$ is defined by

$$\tilde{P}_N(B) := P_N\{A \in \mathcal{H}_N \mid (\lambda_1(A), \ldots, \lambda_N(A)) \in B\}. \tag{7.21}$$

The permutation group S_N acts on $\mathbb{R}^N$ by permuting the coordinates, with $\mathbb{R}^N_{\geq}$ as a fundamental domain (ignoring sets of measure 0). So we can use this action to transport $\tilde{P}_N$ around $\mathbb{R}^N$ to get a probability measure on $\mathbb{R}^N$.

One knows (e.g. see [7, Thm. 2.5.2]) that $\tilde{P}_N$ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^N$ and has density

$$d\tilde{P}_N(\lambda_1, \ldots, \lambda_N) = C_N \cdot e^{-\frac{N}{2}\sum_{i=1}^N \lambda_i^2} \prod_{i<j} (\lambda_i - \lambda_j)^2 \prod_{i=1}^N d\lambda_i, \tag{7.22}$$

where

$$C_N = \frac{N^{N^2/2}}{(2\pi)^{N/2} \prod_{j=1}^N j!}. \tag{7.23}$$

We want to establish a large deviation principle for the empirical eigenvalue distribution $\mu_A = (\delta_{\lambda_1(A)} + \cdots + \delta_{\lambda_N(A)})/N$ of a random matrix in $\mathcal{H}_N$.

One can argue heuristically as follows for the expected form of the rate function. We have

$$\begin{aligned} P_N\{\mu_A \approx \nu\} &= \tilde{P}_N\left\{\frac{1}{N}(\delta_{\lambda_1} + \cdots + \delta_{\lambda_N}) \approx \nu\right\} \\ &= C_N \cdot \int_{\{\frac{1}{N}(\delta_{\lambda_1}+\cdots+\delta_{\lambda_N})\approx\nu\}} e^{-\frac{N}{2}\sum \lambda_i^2} \prod_{i<j}(\lambda_i - \lambda_j)^2 \prod_{i=1}^N d\lambda_i. \end{aligned}$$

Now for $(\delta_{\lambda_1(A)} + \cdots + \delta_{\lambda_N(A)})/N \approx \nu$,

$$-\frac{N}{2}\sum_{i=1}^N \lambda_i^2 = -\frac{N^2}{2}\frac{1}{N}\sum_{i=1}^N \lambda_i^2$$

is a Riemann sum for the integral $\int t^2 d\nu(t)$. Moreover

$$\prod_{i<j}(\lambda_i - \lambda_j)^2 = \exp\left(\sum_{i<j} \log|\lambda_i - \lambda_j|^2\right) = \exp\left(\sum_{i\neq j} \log|\lambda_i - \lambda_j|\right)$$

is a Riemann sum for $N^2 \int\int \log|s-t| d\nu(s)d\nu(t)$.

Hence, heuristically, we expect that $P_N(\mu_A \approx \nu) \sim \exp(-N^2 I(\nu))$, with

$$I(\nu) = -\int\int \log|s-t|d\nu(s)d\nu(t) + \frac{1}{2}\int t^2 d\nu(t) - \lim_{N\to\infty} \frac{1}{N^2}\log C_N. \qquad (7.24)$$

The value of the limit can be explicitly computed as $3/4$. Note that by writing

$$s^2 + t^2 - 4\log|s-t| = s^2 + t^2 - 2\log(s^2+t^2) + 4\log\frac{\sqrt{s^2+t^2}}{|s-t|}$$

and using the inequalities

$$t - 2\log t \geq 2 - 2\log 2 \quad \text{for } t > 0 \qquad \text{and} \qquad 2(s^2+t^2) \geq (s-t)^2$$

we have for $s \neq t$ that $s^2 + t^2 - 4\log|s-t| \geq 2 - 4\log 2$. This shows that if ν has a finite second moment, the integral $\int\int(s^2 + t^2 - 4\log|s-t|)\,d\nu(s)d\nu(t)$ is always defined as an extended real number, possibly $+\infty$, in which case we set $I(\nu) = +\infty$, otherwise $I(\nu)$ is finite and is given by (7.24).

Voiculescu was thus motivated to use the integral $\int\int \log|s-t|\,d\mu_x(s)d\mu_x(t)$ to define in [181] the free entropy $\chi(x)$ for one self-adjoint variable x with distribution μ_x; see equation (7.30).

The large deviation argument was then made rigorous in the following theorem of Ben Arous and Guionnet [26].

Theorem 3. *Put*

$$I(\nu) = -\int\int \log|s-t|d\nu(s)d\nu(t) + \frac{1}{2}\int t^2 d\nu(t) - \frac{3}{4}. \qquad (7.25)$$

Then,

(*i*) $I : \mathcal{M} \to [0,\infty]$ *is a well-defined, convex, good function on the space,* $\mathcal{M}$, *of probability measures on* $\mathbb{R}$. *It has unique minimum value of* 0 *which occurs at the Wigner semi-circle distribution* μ_W *with variance* 1.

(*ii*) *The empirical eigenvalue distribution satisfies a large deviation principle with respect to* $\tilde{P}_N$ *with rate function* I*: we have for any open set* G *in* $\mathcal{M}$

$$\liminf_{N\to\infty} \frac{1}{N^2}\log\tilde{P}_N\left(\frac{\delta_{\lambda_1}+\cdots+\delta_{\lambda_N}}{N} \in G\right) \geq -\inf_{\nu\in G} I(\nu), \qquad (7.26)$$

and for any closed set F *in* $\mathcal{M}$

$$\limsup_{N\to\infty} \frac{1}{N^2}\log\tilde{P}_N\left(\frac{\delta_{\lambda_1}+\cdots+\delta_{\lambda_N}}{N} \in F\right) \leq -\inf_{\nu\in F} I(\nu). \qquad (7.27)$$

Exercise 4. The above theorem includes in particular the statement that for a Wigner semi-circle distribution μ_W with variance 1, we have

$$-\iint \log|s-t|\, d\mu_W(s) d\mu_W(t) = \frac{1}{4}. \tag{7.28}$$

Prove this directly!

Exercise 5.

(i) Let μ be a probability measure with support in $[-2, 2]$. Show that we have

$$\int_{\mathbb{R}} \int_{\mathbb{R}} \log|s-t| d\mu(s) d\mu(t) = -\sum_{n=1}^{\infty} \frac{1}{2n} \left(\int_{\mathbb{R}} C_n(t) d\mu(t) \right)^2,$$

where C_n are the Chebyshev polynomials of the first kind.

(ii) Use part (i) to give another derivation of (7.28).

7.5 Definition of multivariate free entropy

Let (M, τ) be a tracial W^*-probability space and $x_1, \dots, x_n$ self-adjoint elements in M. Recall that by definition the joint distribution of the non-commutative random variables $x_1, \dots, x_n$ is the collection of all mixed moments

$$\mathrm{distr}(x_1, \dots, x_n) = \{\tau(x_{i_1} x_{i_2} \cdots x_{i_k}) \mid k \in \mathbb{N}, i_1, \dots, i_k \in \{1, \dots, n\}\}.$$

In this section we want to examine the probability that the distribution of $(x_1, \dots, x_n)$ occurs in Voiculescu's multivariable generalization of Wigner's semi-circle law.

Let $A_1, \dots, A_n$ be independent Gaussian random matrices: $A_1, \dots, A_n$ are chosen independently at random from the sample space $M_N(\mathbb{C})_{sa}$ of $N \times N$ self-adjoint matrices over $\mathbb{C}$, equipped with Gaussian probability measure having density proportional to $\exp(-\mathrm{Tr}(A^2)/2)$ with respect to Lebesgue measure on $M_N(\mathbb{C})_{sa}$. We know that as $N \to \infty$ we have almost sure convergence $(A_1, \dots, A_n) \xrightarrow{\mathrm{distr}} (s_1, \dots, s_n)$ with respect to the normalized trace, where $(s_1, \dots, s_n)$ is a free semi-circular family. Large deviations from this limit should be given by

$$P_N\{(A_1, \dots, A_n) \mid \mathrm{distr}(A_1, \dots, A_n) \approx \mathrm{distr}(x_1, \dots, x_n)\} \sim e^{-N^2 I(x_1, \dots, x_n)},$$

where $I(x_1, \dots, x_n)$ is the *free entropy* of $x_1, \dots, x_n$. The problem is that this has to be made more precise and that, in contrast to the one-dimensional case, there is no analytical formula to calculate this quantity.

We use the equation above as motivation to define free entropy as follows. This is essentially the definition of Voiculescu from [182]; the only difference is that he also included a cut-off parameter R and required in the definition of the "microstate set" Γ that $\|A_i\| \leq R$ for all $i = 1, \dots, n$. Later it was shown by Belinschi and Bercovici [20] that removing this cut-off condition gives the same quantity.

Definition 4. Given a tracial W^*-probability space (M, τ) and an n-tuple $(x_1, \dots, x_n)$ of self-adjoint elements in M, we define the *(microstates) free entropy* $\chi(x_1, \dots, x_n)$ of the variables $x_1, \dots, x_n$ as follows. First, we put

$$\begin{aligned}\Gamma&(x_1, \dots, x_n; N, r, \epsilon)\\ &:= \big\{(A_1, \dots, A_n) \in M_N(\mathbb{C})^n_{sa} \mid |\mathrm{tr}(A_{i_1} \cdots A_{i_k}) - \tau(x_{i_1} \cdots x_{i_k})| \leq \epsilon \\ &\qquad\qquad \text{for all } 1 \leq i_1, \dots, i_k \leq n, 1 \leq k \leq r\big\}.\end{aligned}$$

In words, $\Gamma(x_1, \dots, x_n; N, r, \epsilon)$, which we call the *set of microstates*, is the set of all n-tuples of $N \times N$ self-adjoint matrices which approximate the mixed moments of the self-adjoint elements $x_1, \dots, x_n$ of length at most r to within ϵ.

Let Λ denote Lebesgue measure on $M_N(\mathbb{C})^n_{sa} \simeq \mathbb{R}^{nN^2}$. Then we define

$$\chi(x_1, \dots, x_n; r, \epsilon) := \limsup_{N \to \infty} \left(\frac{1}{N^2} \log \big(\Lambda(\Gamma(x_1, \dots, x_n; N, r, \epsilon))\big) + \frac{n}{2} \log(N) \right),$$

and finally put

$$\chi(x_1, \dots, x_n) := \lim_{\substack{r \to \infty \\ \epsilon \to 0}} \chi(x_1, \dots, x_n; r, \epsilon). \tag{7.29}$$

It is an important open problem whether the lim sup in the definition above of $\chi(x_1, \dots, x_n; r, \epsilon)$ is actually a limit.

We want to elaborate on the meaning of Λ, the Lebesgue measure on $M_N(\mathbb{C})^n_{sa} \simeq \mathbb{R}^{nN^2}$, and the normalization constant $n \log(N)/2$. Let us consider the case $n = 1$. For a self-adjoint matrix $A = (a_{ij})_{i,j=1}^N \in M_N(\mathbb{C})_{sa}$, we identify the elements on the diagonal (which are real) and the real and imaginary part of the elements above the diagonal (which are the adjoints of the corresponding elements below the diagonals) with an $N + 2\frac{N(N-1)}{2} = N^2$ dimensional vector of real numbers. The actual choice of this mapping is determined by the fact that we want the Euclidean inner product in $\mathbb{R}^{N^2}$ to correspond on the side of the matrices to the form $(A, B) \mapsto \mathrm{Tr}(AB)$. Note that

$$\mathrm{Tr}(A^2) = \sum_{i,j=1}^N a_{ij} a_{ji} = \sum_{i=1}^N (\mathrm{Re} a_{ii})^2 + 2 \sum_{1 \leq i < j \leq N} \big((\mathrm{Re} a_{ij})^2 + (\mathrm{Im} a_{ij})^2\big).$$

This means that there is a difference of a factor $\sqrt{2}$ between the diagonal and the off-diagonal elements. (The same effect made its appearance in Chapter 1, Exercise 8, when we defined the GUE by assigning different values for the covariances for variables on and off the diagonal – in order to make this choice invariant under conjugation by unitary matrices.) So our specific choice of a map between $M_N(\mathbb{C})$ and $\mathbb{R}^{N^2}$ means that we map the set $\{A \in M_N(\mathbb{C})_{sa} \mid \mathrm{Tr}(A^2) \leq R^2\}$ to the ball $B_{N^2}(R)$ of radius R in N^2 real dimensions. The pull back under this map of the Lebesgue measure on $\mathbb{R}^{N^2}$ is what we call Λ, the Lebesgue measure on $M_N(\mathbb{C})_{sa}$. The situation for general n is given by taking products.

Note that a microstate $(A_1, \dots, A_n) \in \Gamma(x_1, \dots, x_n; N, r, \epsilon)$ satisfies for $r \geq 2$

$$\frac{1}{N}\mathrm{Tr}(A_1^2 + \cdots + A_n^2) \leq \tau(x_1^2 + \cdots + x_n^2) + n\epsilon =: c^2,$$

and thus the set of microstates $\Gamma(x_1, \dots, x_n; N, r, \epsilon)$ is contained in the ball $B_{nN^2}(\sqrt{N}c)$. The fact that the latter grows logarithmically like

$$\frac{1}{N^2}\log \Lambda\left(B_{nN^2}(\sqrt{N}c)\right) = \frac{1}{N^2}\log\frac{(\sqrt{N}c\sqrt{\pi})^{nN^2}}{\Gamma(1+nN^2/2)} \sim -\frac{n}{2}\log N,$$

is the reason for adding the term $n\log N/2$ in the definition of $\chi(x_1, \dots, x_n; r, \epsilon)$.

7.6 Some important properties of χ

The free entropy has the following properties:

(*i*) For $n = 1$, much more can be said than for general n. In particular, one can show that the lim sup in the definition of χ is indeed a limit and that we have the explicit formula

$$\chi(x) = \int\int \log|s-t|d\mu_x(s)d\mu_x(t) + \frac{1}{2}\log(2\pi) + \frac{3}{4}. \tag{7.30}$$

Thus the definition of χ reduces in this case to the quantity from the previous section. Our discussion before Theorem 3 shows then that $\chi(x) \in [-\infty, \infty)$. For $n \geq 2$, no formula of this sort is known.

When x is a semi-circular operator with variance 1, we know the value of the double integral by (7.28); hence, for a semi-circular operator s with variance 1, we have

$$\chi(s) = \frac{1}{2}(1 + \log(2\pi)). \tag{7.31}$$

(*ii*) χ is subadditive:

$$\chi(x_1, \dots, x_n) \leq \chi(x_1) + \cdots + \chi(x_n). \tag{7.32}$$

This is an easy consequence of the fact that

$$\Gamma(x_1, \ldots, x_n; N, r, \epsilon) \subset \prod_{i=1}^{n} \Gamma(x_i; N, r, \epsilon).$$

Thus, in particular, by using the corresponding property from (*i*), we always have $\chi(x_1, \ldots, x_n) \in [-\infty, \infty)$.

(*iii*) χ is upper semicontinuous: if $(x_1^{(m)}, \ldots, x_n^{(m)}) \overset{\text{distr}}{\longrightarrow} (x_1, \ldots, x_n)$ for $m \to \infty$, then

$$\chi(x_1, \ldots, x_n) \geq \limsup_{m \to \infty} \chi(x_1^{(m)}, \ldots, x_n^{(m)}). \tag{7.33}$$

This is because if, for arbitrary words of length k with $1 \leq k \leq r$, we have

$$|\tau(x_{i_1}^{(m)} \cdots x_{i_k}^{(m)}) - \tau(x_{i_1} \cdots x_{i_k})| < \frac{\epsilon}{2}$$

for sufficiently large m, then

$$\Gamma(x_1^{(m)}, \ldots, x_n^{(m)}; N, r, \frac{\epsilon}{2}) \subset \Gamma(x_1, \ldots, x_n; N, r, \epsilon).$$

(*iv*) If $x_1, \ldots, x_n$ are free, then $\chi(x_1, \ldots, x_n) = \chi(x_1) + \cdots + \chi(x_n)$.

(*v*) $\chi(x_1, \ldots, x_n)$, under the constraint $\sum \tau(x_i^2) = n$, has a unique maximum when $x_1, \ldots, x_n$ is a free semi-circular family $(s_1, \ldots, s_n)$ with $\tau(s_i^2) = 1$. In this case

$$\chi(s_1, \ldots, s_n) = \frac{n}{2}(1 + \log(2\pi)). \tag{7.34}$$

(*vi*) Consider $y_j = F_j(x_1, \ldots, x_n)$, for some "convergent" non-commutative power series F_j, such that the mapping $(x_1, \ldots, x_n) \mapsto (y_1, \ldots, y_n)$ can be inverted by some other power series. Then

$$\chi(y_1, \ldots, y_n) = \chi(x_1, \ldots, x_n) + n \log(|\det|\mathcal{J}(x_1, \ldots, x_n)), \tag{7.35}$$

where $\mathcal{J}$ is a non-commutative Jacobian and $|\det|$ is the Fuglede-Kadison determinant. (We will provide more information on the Fuglede-Kadison determinant in Chapter 11.)

With the exception of (*ii*) and (*iii*), the statements above are quite non-trivial; for the proofs we refer to the original papers of Voiculescu [182, 186].

Exercise 6. (*i*) For an n-tuple $(x_1, \ldots, x_n)$ of self-adjoint elements in M and an invertible real matrix $T = (t_{ij})_{i,j=1}^n \in M_n(\mathbb{R})$, we put $y_i := \sum_{j=1}^n t_{ij} x_j \in M$

($i = 1, \dots, n$). Part (*vi*) of the above says then (by taking into account the meaning of the Fuglede-Kadison determinant for matrices, see (11.4)) that

$$\chi(y_1, \dots, y_n) = \chi(x_1, \dots, x_n) + \log|\det T|. \tag{7.36}$$

Prove this directly from the definitions.

(*ii*) Show that $\chi(x_1, \dots, x_n) = -\infty$ if $x_1, \dots, x_n$ are linearly dependent.

7.7 Applications of free entropy to operator algebras

One hopes that χ can be used to construct invariants for von Neumann algebras. In particular, we define the *free entropy dimension* of the n-tuple $x_1, \dots, x_n$ by

$$\delta(x_1, \dots, x_n) = n + \limsup_{\epsilon \searrow 0} \frac{\chi(x_1 + \epsilon s_1, \dots, x_n + \epsilon s_n)}{|\log \epsilon|}, \tag{7.37}$$

where $s_1, \dots, s_n$ is a free semi-circular family, free from $\{x_1, \dots, x_n\}$.

One of the main problems in this context is to establish the validity (or falsehood) of the following implication (or some variant thereof): if $\mathrm{vN}(x_1, \dots, x_n) = \mathrm{vN}(y_1, \dots, y_n)$, does this imply that $\delta(x_1, \dots, x_n) = \delta(y_1, \dots, y_n)$?

In recent years there have been a number of results which allow one to infer some properties of a von Neumann algebra from knowledge of the free entropy dimension for some generators of this algebra. Similar statements can be made on the level of the free entropy. However, there the actual value of χ is not important; the main issue is to distinguish finite values of χ from the situation $\chi = -\infty$.

Let us note that in the case of free group factors $\mathcal{L}(\mathbb{F}_n) = \mathrm{vN}(s_1, \dots, s_n)$, we have of course for the canonical generators $\chi(s_1, \dots, s_n) > -\infty$ and $\delta(s_1, \dots, s_n) = n$. (For the latter one should notice that the sum of two free semi-circulars is just another semi-circular, where the variances add; hence the numerator in (7.37) stays bounded for $\epsilon \to 0$ in this case.)

We want now to give the idea how to use free entropy to get statements about a von Neumann algebra. For this, let P be some property that a von Neumann algebra M may or may not have. Assume that we can verify that "M has P" implies that $\chi(x_1, \dots, x_n) = -\infty$ for any generating set $\mathrm{vN}(x_1, \dots, x_n) = M$. Then a von Neumann algebra for which we have at least one generating set with finite free entropy cannot have this property P. In particular, $\mathcal{L}(\mathbb{F}_n)$ cannot have P.

Three such properties where this approach was successful are property Γ, the existence of a Cartan subalgebra, and the property of being prime.

Let us first recall the definition of property Γ. We will use here the usual non-commutative L^2-norm, $\|x\|_2 := \sqrt{\tau(x^* x)}$, for elements x in our tracial W^*-probability space (M, τ).

Definition 5. A bounded sequence $(t_k)_{k \geq 0}$ in (M, τ) is *central* if $\lim_{k \to \infty} \|[x, t_k]\|_2 = 0$ for all $x \in M$, where $[\cdot, \cdot]$ denotes the commutator of two elements, i.e. $[x, t_k] = x t_k - t_k x$. If $(t_k)_k$ is a central sequence and $\lim_{k \to \infty} \|t_k - \tau(t_k) 1\|_2 = 0$, then $(t_k)_k$ is said to be a *trivial central sequence*. (M, τ) has *property Γ* if there exists a non-trivial central sequence in M.

Note that elements from the centre of an algebra always give central sequences; hence if M does not have property Γ, then it is a factor.

Definition 6. 1) Given any von Neumann subalgebra N of a von Neumann algebra M, we let the normalizer of N be the von Neumann subalgebra of M generated by all the unitaries $u \in M$ which normalize N, i.e. $uNu^* = N$. A von Neumann subalgebra N of M is said to be maximal abelian if it is abelian and is not properly contained in any other abelian subalgebra. A maximal abelian subalgebra is a *Cartan subalgebra* of M if its normalizer generates M.
2) Finally we recall that a finite von Neumann algebra M is *prime* if it cannot be decomposed as $M = M_1 \overline{\otimes} M_2$ for II_1 factors M_1 and M_2. Here $\overline{\otimes}$ denotes the von Neumann tensor product of M_1 and M_2; see [170, Ch. IV].

The above-mentioned strategy is the basis of the proof of the following theorem:

Theorem 7. *Let M be a finite von Neumann algebra with trace τ generated by self-adjoint operators $x_1, \dots, x_n$, where $n \geq 2$. Assume that $\chi(x_1, \dots, x_n) > -\infty$, where the free entropy is calculated with respect to the trace τ. Then*

(*i*) *M does not have property Γ. In particular, M is a factor.*
(*ii*) *M does not have a Cartan subalgebra.*
(*iii*) *M is prime.*

Corollary 8. *All this applies in the case of the free group factor $\mathcal{L}(\mathbb{F}_n)$ for $2 \leq n < \infty$; thus,*

(*i*) *$\mathcal{L}(\mathbb{F}_n)$ does not have property Γ.*
(*ii*) *$\mathcal{L}(\mathbb{F}_n)$ does not have a Cartan subalgebra.*
(*iii*) *$\mathcal{L}(\mathbb{F}_n)$ is prime.*

Parts (*i*) and (*ii*) of the theorem above are due to Voiculescu [185]; part (*iii*) was proved by Liming Ge [76]. In particular, the absence of Cartan subalgebras for $\mathcal{L}(\mathbb{F}_n)$ was a spectacular result, as it falsified the conjecture, which had been open for decades, that every II_1 factor should possess a Cartan subalgebra. Such a conjecture was suggested by the fact that von Neumann algebras obtained from ergodic measurable relations always have Cartan subalgebras, and for a while there was the hope that all von Neumann algebras might arise in this way.

In order to give a more concrete idea of this approach, we will present the essential steps in the proof for part (*i*) (which is the simplest part of the theorem above) and say a few words about the proof of part (*iii*). However, one should note that the absence of property Γ for $\mathcal{L}(\mathbb{F}_n)$ is an old result of Murray and von Neumann which can be proved more directly without using free entropy. The following follows quite closely the exposition of Biane [36].

7.7.1 The proof of Theorem 7, part (*i*)

We now give the main arguments and estimates for the proof of part (*i*) of Theorem 7. So let $M = \mathrm{vN}(x_1, \dots, x_n)$ have property Γ; we must prove that this implies $\chi(x_1, \dots, x_n) = -\infty$.

Let $(t_k)_k$ be a non-trivial central sequence in M. Then its real and imaginary parts are also central sequences (at least one of them non-trivial), and, by applying functional calculus to this sequence, we may replace the t_k's with a non-trivial central sequence of orthogonal projections $(p_k)_k$, and assume the existence of a real number θ in the open interval $(0, 1/2)$ such that $\theta < \tau(p_k) < 1 - \theta$ for all k and $\lim_{k\to\infty} \|[x, p_k]\|_2 = 0$ for all $x \in M$.

We then prove the following key lemma.

Lemma 9. *Let (M, τ) be a tracial W^*-probability space generated by self-adjoint elements $x_1, \dots, x_n$ satisfying $\tau(x_i^2) \leq 1$. Let $0 < \theta < \frac{1}{2}$ be a constant and $p \in M$ a projection such that $\theta < \tau(p) < 1 - \theta$. If there is $\omega > 0$ such that $\|[p, x_i]\|_2 < \omega$ for $1 \leq i \leq n$, then there exist positive constants C_1, C_2 depending only on n and θ such that $\chi(x_1, \dots, x_n) \leq C_1 + C_2 \log \omega$.*

Assuming this is proved, choose $p = p_k$. We can take $\omega_k \to 0$ as $k \to \infty$. Thus we get $\chi(x_1, \dots, x_n) \leq C_1 + C_2 \log \omega$ for all $\omega > 0$, implying $\chi(x_1, \dots, x_n) = -\infty$. (Note that we can achieve the assumption $\tau(x_i^2) \leq 1$ by rescaling our generators.) It remains to prove the lemma.

Proof: Take $(A_1, \dots, A_n) \in \Gamma(x_1, \dots, x_n; N, r, \epsilon)$ for N, r sufficiently large and ϵ sufficiently small. As p can be approximated by polynomials in $x_1, \dots, x_n$ and by an application of the functional calculus, we find a projection matrix $Q \in M_N(\mathbb{C})$ whose range is a subspace of dimension $q = \lfloor N\tau(p) \rfloor$ and such that we have (where the $\|\cdot\|_2$-norm is now with respect to tr in $M_N(\mathbb{C})$) $\|[A_i, Q]\|_2 < 2\omega$ for all $i = 1, \dots, n$. This Q is of the form

$$Q = U \begin{pmatrix} I_q & 0 \\ 0 & 0_{N-q} \end{pmatrix} U^*$$

for some $U \in \mathcal{U}(N)/\mathcal{U}(q) \times \mathcal{U}(N-q)$. Write

$$U^* A_i U = \begin{pmatrix} B_i & C_i^* \\ C_i & D_i \end{pmatrix}.$$

Then $\|[A_i, Q]\|_2 \leq 2\omega$ implies the same for the conjugated matrices, i.e.

$$\sqrt{\frac{2}{N}\mathrm{Tr}(C_i C_i^*)} = \left\| \begin{pmatrix} 0 & -C_i^* \\ C_i & 0 \end{pmatrix} \right\|_2 = \left\| \left[\begin{pmatrix} B_i & C_i^* \\ C_i & D_i \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right] \right\|_2 = \|[A_i, Q]\|_2 < 2\omega,$$

and thus we have for all $i = 1, \ldots, n$

$$\mathrm{Tr}(C_i C_i^*) < \frac{N}{2}(2\omega)^2 = 2N\omega^2.$$

Furthermore, $\tau(x_i^2) \leq 1$ implies that $\mathrm{tr}(A_i^2) \leq 1+\epsilon$ and hence $\mathrm{Tr}(A_i^2) \leq (1+\epsilon)N \leq 2N$, since we can take $\epsilon \leq 1$. Thus, in particular, we also have $\mathrm{Tr}(B_i^2) \leq 2N$ and $\mathrm{Tr}(D_i^2) \leq 2N$.

Denote now by $B_p(R)$ the ball of radius R in $\mathbb{R}^p$ centred at the origin and consider the map which sends our matrices $A_i \in M_N(\mathbb{C})$ to the Euclidean space $\mathbb{R}^{N^2}$. Then the latter conditions mean that each B_i is contained in a ball $B_{q^2}(\sqrt{2N})$ and that each D_i is contained in a ball $B_{(N-q)^2}(\sqrt{2N})$. For the rectangular $q \times (N-q)$ matrix $C_i \in M_{q,N-q}(\mathbb{C}) \simeq \mathbb{R}^{2q(N-q)}$, the condition $\mathrm{Tr}(CC^*) \leq 2N\omega^2$ means that C is contained in a ball $B_{2q(N-q)}(\sqrt{4N}\omega)$. (Here we get an extra factor $\sqrt{2}$, because all elements from C_i correspond to upper triangular elements from A_i.)

Thus, the estimates above show that we can cover $\Gamma(x_1, \ldots, x_n; N, r, \epsilon)$ by a union of products of balls:

$$\Gamma(x_1, \ldots, x_n; N, r, \epsilon) \subseteq \bigcup_{\substack{U \in \\ \mathcal{U}(N)/\mathcal{U}(q)\times\mathcal{U}(N-q)}} \left[U\left(B_{q^2}(\sqrt{2N}) \times B_{2q(N-q)}(\omega\sqrt{4N}) \times B_{(N-q)^2}(\sqrt{2N})\right)U^*\right]^n.$$

This does not give directly an estimate for the volume of our set Γ, as we have here a covering by infinitely many sets. However, we can reduce this to a finite cover by approximating the U's which appear by elements from a finite δ-net.

By a result of Szarek [169], for any $\delta > 0$, there exists a δ-net $(U_s)_{s\in S}$ in the Grassmannian $\mathcal{U}(N)/\mathcal{U}(q) \times \mathcal{U}(N-q)$ with $|S| \leq (C\delta^{-1})^{N^2-q^2-(N-q)^2}$ with C a universal constant.

For $(A_1, \ldots, A_n)$, Q, and U as above, we have that there exists $s \in S$ such that $\|U - U_s\| \leq \delta$ implies $\|[U_s^* A_i U_s, U^* Q U]\|_2 \leq 2\omega + 8\delta$. Repeating the arguments above for $U_s^* A_i U_s$ instead of $U^* A_i U$ (where we have to replace 2ω by $2\omega + 8\delta$), we get

$$\Gamma(x_1, \ldots, x_n; N, r, \epsilon) \subseteq \bigcup_{s\in S}\left[U_s\left(B_{q^2}(\sqrt{2N}) \times B_{2q(N-q)}\big((\omega+4\delta)\sqrt{4N}\big) \times B_{(N-q)^2}(\sqrt{2N})\right)U_s^*\right]^n, \tag{7.38}$$

and hence

$$\Lambda(\Gamma(x_1,\dots,x_n;N,r,\epsilon)) \le (C\delta^{-1})^{N^2-q^2-(N-q)^2} \times \Big[\Lambda\Big(B_{q^2}(\sqrt{2N})\Big)\,\Lambda\Big(B_{2q(N-q)}\big((\omega+4\delta)\sqrt{4N}\big)\Big)\,\Lambda\Big(B_{(N-q)^2}(\sqrt{2N})\Big)\Big]^n.$$

By using the explicit form of the Lebesgue measure of $B_p(R)$ as

$$\Lambda(B_p(R)) = \frac{R^p\pi^{p/2}}{\Gamma(1+\frac{p}{2})},$$

this simplifies to the bound

$$(C\delta^{-1})^{2q(N-q)}\left[\frac{(2N\pi)^{N^2/2}[\sqrt{2}(\omega+4\delta)]^{2q(N-q)}}{\Gamma(1+q^2/2)\Gamma(1+q(N-q))\Gamma(1+(N-q)^2/2)}\right]^n.$$

Thus

$$\frac{1}{N^2}\log\Lambda(\Gamma(x_1,\dots,x_n;N,r,\epsilon))+\frac{n}{2}\log N \le \tilde{C}_1+\tilde{C}_2\big(\log\delta^{-1}+n\log(\omega+4\delta)\big),$$

for positive constants $\tilde{C}_1,\tilde{C}_2$ depending only on n and θ. Taking now $\delta=\omega$ gives the claimed estimate with $C_1 := \tilde{C}_1 + n\log 5$ and $C_2 := (n-1)\tilde{C}_2$. □

One should note that our estimates work for all n. However, in order to have C_2 strictly positive, we need $n > 1$. For $n = 1$ we only get an estimate against a constant C_1, which is not very useful. This corresponds to the fact that for each i the smallness of the off-diagonal block C_i of U^*A_iU in some basis U is not very surprising; however, if we have the smallness of all such blocks $C_1,\dots,C_n$ of $U^*A_1U,\dots,U^*A_nU$ for a common U, then this is a much stronger constraint.

7.7.2 The proof of Theorem 7, part (*iii*)

The proof of part (*iii*) proceeds in a similar, though technically more complicated, fashion. Let us assume that our II_1 factor $M = \mathrm{vN}(x_1,\dots,x_n)$ has a Cartan subalgebra N. We have to show that this implies $\chi(x_1,\dots,x_n)=-\infty$.

First one has to rewrite the property of having a Cartan subalgebra in a more algebraic way, encoding a kind of "smallness". Voiculescu showed the following. For each $\epsilon > 0$, there exist a finite-dimensional C^*-subalgebra N_0 of N; $k(j)\in\mathbb{N}$ for all $1\le j\le n$; orthogonal projections $p_j^{(i)}, q_j^{(i)} \in N_0$ and elements $x_j^{(i)}\in M$ for all $j = 1, \dots, n$ and $1 \leq i \leq k(j)$ such that the following holds: $x_j^{(i)} = p_j^{(i)} x_j^{(i)} q_j^{(i)}$ for all $j = 1, \dots, n$ and $1 \leq i \leq k(j)$,

$$\|x_j - \sum_{1 \leq i \leq k(j)} (x_j^{(i)} + x_j^{(i)*})\|_2 < \epsilon \qquad \text{for all } j = 1, \dots, n, \tag{7.39}$$

and

$$\sum_{1 \leq j \leq n} \sum_{1 \leq i \leq k(j)} \tau(p_j^{(i)}) \tau(q_j^{(i)}) < \epsilon.$$

Consider now a microstate $(A_1, \dots, A_n) \in \Gamma(x_1, \dots, x_n; N, r, \epsilon)$. Since polynomials in the generators $x_1, \dots, x_n$ approximate the given projections $p_j^{(i)}, q_j^{(i)} \in N_0 \subset M$, the same polynomials in the matrices $A_1, \dots, A_n$ will approximate versions of these projections in finite matrices. Thus we find a unitary matrix such that $(UA_1U^*, \dots, UA_nU^*)$ is of a special form with respect to fixed matrix versions of the projections. This gives some constraints on the volume of possible microstates. Again, in order to get rid of the freedom of conjugating by an arbitrary unitary matrix, one covers the unitary $N \times N$ matrices by a δ-net S and gets so in the end a similar bound as in (7.38). Invoking from [169] the result that one can choose a δ-net with $|S| < (C/\delta)^{N^2}$ leads finally to an estimate for $\chi(x_1, \dots, x_n)$ as in Lemma 9. The bound in this estimate goes to $-\infty$ for $\epsilon \to 0$, which proves that $\chi(x_1, \dots, x_n) = -\infty$. □