Background: Asymptotic Freeness and Large Deviations

We start by recalling the basic asymptotic freeness result discussed in class.

> **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

$$\begin{aligned} X &\coloneqq \prod_{N\in\mathbb{N}} X_N \\ \mu &\coloneqq \prod_{N\in\mathbb{N}} \mu_N, \end{aligned}$$

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

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

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

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

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

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;
- $\left\{x \in \mathbb{R}^d \;\middle|\; \Lambda^*(x) \leq \alpha\right\}$ 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.

Yet, the next step is to move from the sequence of partial sums to developing a large deviation principle for the empirical distribution itself. 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

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

One-Dimensional Free Entropy

Microstates Free Entropy

Applications: Structural Properties of Free Group Factors

References

[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.