BREUILLARD-VARJU'S INEQUALITY BETWEEN ENTROPY AND MAHLER MEASURE

LECTURE NOTES - OBERWOLFACH OCT. 2017

OR LANDESBERG

These notes follow Emamanuel Breuillard and Peter P.Varju's paper titled "Entropy of Bernoulli Convolutions and Uniform Exponential Growth for Linear Groups" [1].

1. Introduction and Statement of Theorems

A *Bernoulli convolution* with parameter $\lambda \in (0,1)$ is the distribution μ_λ of the following random power series $\sum_{k=0}^{\infty} \pm\lambda^k$ where the signs $\pm$ are chosen independently with equal probability. The distribution of finite sum $\sum_{0\leqslant k\leqslant n-1} \pm\lambda^k$ will be denoted by $\mu_\lambda^{(n)}$. The *random walk entropy* h_λ of μ_λ is defined to be:

$$h_\lambda := \lim_{n\to\infty} \frac{H(\mu_\lambda^{(n)})}{n}$$

where $H(\mu_\lambda^{(n)})$ is the Shannon entropy of the finitely supported measure $\mu_\lambda^{(n)}$.

Let $\pi_\lambda(x) = a_r \cdot \Pi_{i=0}^{i=r}(x-\lambda_i) = a_r x^r + a_{r-1}x^{r-1} + ... + a_0$ be the minimal polynomial in $\mathbb{Z}[x]$ of an algebraic number $\lambda \in \overline{\mathbb{Q}}$, with $\lambda_1, ..., \lambda_r$ its Galois conjugates (including $\lambda_1 = \lambda$).

Definition 1.1. The Mahler measure of λ is define to be $M_\lambda := |a_r| \cdot \Pi_{|\lambda_i>1|}|\lambda_i|$.

This value is used as a height function for measuring the "complexity" of algebraic numbers and polynomials of integer coefficients (there are clearly only finitely many such numbers/poly. with bounded Mahler measure and bounded degree).

The main result we would like to discuss today is the following, connecting Mahler measure with the entropy of a Bernoulli convolution with parameter λ:

Theorem 1.1. *There exists a positive constant $c > 0$ such that given any algebraic number λ:*

$$c \cdot \min(1, \log\lambda) \leqslant h_\lambda \leqslant \min(1, \log\lambda).$$

This constant can be taken to be $c = 0.44$. log here and throughout will be taken to be in base 2.

A special case of Hochman's theorem on Bernoulli convolutions connects the random walk entropy h_λ with **algebraic** parameter $\lambda \in (\frac{1}{2}, 1)$ to the Hausdorff

[1] Preprint arXiv:1510.04043v2 [math.CA] 1 Jun 2016

dimension of the measure μ_λ:

$$\dim \mu_\lambda = \min\left(1, \frac{h_\lambda}{\log \lambda^{-1}}\right)$$

Hence theorem 1.1 provides an easily testable condition implying $\dim \mu_\lambda$ has full dimension.

Corollary 1.2. *If λ is a real algebraic number such that:*

$$\min(M_\lambda, 2)^{-0.44} \leqslant \lambda \leqslant 1$$

then $\dim \mu_\lambda = 1$.

It is a famous conjecture by Lehmer that the Mahler measure of all algebraic numbers is uniformly bounded away from 1 whenever λ is not 0 or a root of unity (it is a result of Kronecker showing these are the only cases when $M_\lambda = 1$).

Corollary 1.3. *If the Lehmer conjecture holds, then there exists an $\varepsilon > 0$ such that for every real algebraic $1 - \varepsilon < \lambda < 1$ the dimension of the Bernoulli convolution* $\dim \mu_\lambda = 1$.

proof of Corollary. If the Lehmer conjecture is true then there exists a $\delta > 0$ for which $\log M_\lambda > \delta$ for all $\lambda \in (\frac{1}{2}, 1)$. Hence for all λ large enough $\log \lambda^{-1} \leqslant c \cdot \delta$, where c is the uniform constant appearing in the statement of theorem 1.1. Therefore:

$$\dim \mu_\lambda = \min\left(1, \frac{h_\lambda}{log\lambda^{-1}}\right) \geqslant \min\left(1, \frac{c \cdot \min\left(1, logM_\lambda\right)}{c \cdot \delta}\right) = 1$$

□

In a following paper by Breuillard and Varju showed:

Theorem. *The following inclusion holds:*

$$\left\{\lambda \in \left(\frac{1}{2}, 1\right) : \dim \mu_\lambda < 1\right\} \subseteq \overline{\left\{\lambda \in \overline{\mathbb{Q}} \cap \left(\frac{1}{2}, 1\right) : \dim \mu_\lambda < 1\right\}}$$

where $\overline{\mathbb{Q}}$ is the set of algebraic numbers and $\overline{\{\cdot\}}$ denotes the topological closure in $\mathbb{R}$.

From this theorem the assumption of algebraicity in the above corollary can be removed, giving:

Corollary 1.4. *If the Lehmer conjecture holds then there exists an $\varepsilon > 0$ for which all $1 - \varepsilon < \lambda < 1$ admit $dim\mu_\lambda = 1$.*

We will focus in these notes in the proof of Theorem 1.1 and we begin with the (easier) proof of the upper bound.

2. Proof of the upper bound in theorem 1.1

As a consequence of Jensen's inequality we know $H(\mu_\lambda^{(n)}) \leqslant \log|supp\ \mu_\lambda^{(n)}|$. Hence h_λ is bounded above by the rate of exponential growth of the n-th sums:

$$\rho_\lambda = \lim_{n\to\infty} \frac{\log|supp\ \mu_\lambda^{(n)}|}{n}.$$

The limit exists by subadditivity. Note that $supp\ \mu_\lambda = \left\{\sum_{k=0}^{k=n-1} \pm\lambda^k\right\}$ and has cardinality 2^n whenever λ is not a root of a polynomial with coefficients in {-1,0,1} and actually $H(\mu_\lambda^{(n)}) = n$ in that case. Therefore Hochman's result directly implies that $\dim\mu_\lambda = 1$ whenever λ is algebraic and not a root of a polynomial with coefficients in {-1,0,1}. Hence our proof will focus on bounding ρ_λ in the case where λ is a root of such a polynomial.

Claim. *A root λ of a polynomial with coefficients in {-1,0,1} is an algebraic unit, i.e. both λ and λ^{-1} are algebraic integers (roots of monic polynomials of integer coefficients).*

Proof. Let p be a polynomial with coefficients in {-1,0,1} and let λ be some root of p. p can be factorized in $\mathbb{Z}[x]$ into irreducible components. Since all components have integer coefficients and p's leading coeffieint is ±1 we deduce λ is an algebraic integer (root of a monic irreducible polynomial with integer coefficients). The same reasoning shows λ's minimal polynomial has constant coefficient ±1 meaning $\lambda\cdot\lambda_2\cdots\lambda_r = \pm1$ and hence $\lambda^{-1} = \pm\lambda_2\cdots\lambda_r$ is a product of algebraic integers and thus itself an algebraic integer. □

A useful property of algebraic units is the fact that:

$$\Pi_{|\lambda_i|<1}|\lambda_i| = \frac{1}{M_\lambda}$$

as $a_r = 1$ and the product of all Galois conjugates equals ±1.

We will prove the following lemma which implies the upper bound in the theorem:

Lemma 2.1. *Let λ be an algebraic unit and denote by s the number of Galois conjugates of λ on the unit circle. Then $|supp\ \mu_\lambda^{(\ell)}| \leqslant C\ell^s M_\lambda^\ell$, where C is a constant depending only on λ. In particular:*

$$\rho_\lambda \leqslant \min\left(1, \log M_\lambda\right).$$

Proof. Denote by $\sigma_1, ..., \sigma_n : \mathbb{Q}(\lambda) \to \mathbb{R}$ the real Galois embeddings for which $|\sigma_i(\lambda)| > 1$, by $\tau_1, ..., \tau_m : \mathbb{Q}(\lambda) \to \mathbb{C}$ the complex Galois embeddings with $|\tau_j(\lambda)| \geqslant 1$ (taking only one of each complex conjugate pair of embeddings), and denote by $\rho_1, ..., \rho_o : \mathbb{Q}(\lambda) \to \mathbb{C}$ the complex Galois embedding for which $|\rho_k(\lambda)| < 1$ (here taking both from a pair of complex conjugate embeddings). Consider the set $A = supp\ \mu_\lambda^{(\ell)} = \left\{\sum_{k=0}^{\ell-1} \pm\lambda^k\right\}$ and note that its elements are algebraic integers (as $\mathcal{O}_\mathbb{Q}$ is a ring), hence for any two distinct $x, y \in A$ we have:

$$\Pi_{i,j,k}|\sigma_i(x-y)||\tau_j(x-y)|^2|\rho_k(x-y)| \geqslant 1$$

as the product inside $|\cdot|$ is both an algebraic integer and a rational (as it is fixed by the Galois group associated with λ over $\mathbb{Q}$). For any $1 \leqslant k \leqslant o$ we have

$$|\rho_k(x-y)| \leqslant |\rho_k(x)| + |\rho_k(y)| \leqslant 2\sum_{i=0}^{\ell-1} |\rho_k(\lambda)|^i \leqslant \frac{2}{1-|\rho_k(\lambda)|}$$

Hence there is a constant c_0 dependent on λ for which:

$$\Pi_{i,j,k}|\sigma_i(x-y)||\tau_j(x-y)|^2 \geqslant c_0.$$

Define the map $S : \mathbb{Q}(\lambda) \to \mathbb{R}^{n+2m}$ by:

$$S(x) = (\sigma_1(x), ..., \sigma_n(x), \mathrm{Re}(\tau_1(x)), \mathrm{Im}(\tau_1(x)), ..., \mathrm{Re}(\tau_m(x)), \mathrm{Im}(\tau_m(x)))$$

By the inequality above we deduce there exists a constant c_1 depending on λ for which any two distinct $x, y \in A$ admit $\|S(x-y)\| > c_1$. Consider the following set:

$$\begin{aligned}\Omega = \{(x_1, ..., x_{n+2m}) \in \mathbb{R}^{n+2m} :& \\ |x_i| \leqslant \frac{|\sigma_i(\lambda)|^\ell - 1}{|\sigma_i(\lambda)| - 1} + c_1&, \text{ and} \\ |x_{n+2j-1}|, |x_{2j}| \leqslant \frac{|\tau_j(\lambda)|^\ell - 1}{|\tau_j(\lambda)| - 1} + c_1 &\text{ if } |\tau_j(\lambda)| > 1, \text{ and} \\ |x_{n+2j-1}|, |x_{2j}| \leqslant l + c_1 &\text{ if } |\tau_j(\lambda)| = 1\}.\end{aligned}$$

It can be easily seen that the balls of radius $\frac{1}{2}c_1$ around the points of $S(A)$ are disjoint and contained inside Ω. On the other hand the volume of Ω is bounded by $C\ell^s M_\lambda^\ell$ giving the required bound on $|A|$. □

3. Differential and Gaussian averaged entropy

Denote $F(x) = -x\log x$, recall F is concave and admits $F(xy) = xF(y) + F(x)y$. Let X be a random variable in $\mathbb{R}^d$ with absolutely continuous distribution with respect to Lebesgue measure and with density f. We define the *differential entropy* of X to be:

$$H(X) := \int F(f(x))dx = -\int f(x)\log f(x)dx.^2$$

This is well defined whenever $f\log f \in L^1(\mathbb{R}^d)$.

Two useful properties of differential entropy:

- Given a linear map $A \in GL_d(\mathbb{R})$ and a RV X with finite differential entropy, the change of variables formula yields $H(AX) = H(X) + \log|\det A|$. In particular we see that the differential entropy may receive negative values.
- Given a continuous RV Y (with density f)and a discrete RV X both with finite entropy (differential and Shannon respectively), then: $H(X+Y) \leqslant H(X) + H(Y)$.

[2]We use the same notation for Shannon and differential entropy.

Proof. The density of $X+Y$ is $\mathbb{E}[f(y-X)] = \sum_i p_i f(y-x_i)$, hence:

$$\begin{aligned} H(X+Y) &= \int F(\sum_i p_i f(y-x_i))dy \\ &\leqslant \int \sum_i F(p_i f(y-x_i))dy \\ &= \int \sum_i F(p_i) f(y-x_i)dy + \int \sum_i p_i F(f(y-x_i))dy \\ &= \sum_i F(p_i) + \int F(f(y))dy = H(X) + H(Y) \end{aligned}$$

where the inequality id due to the subadditivity of F. □

We denote by G_A the centered Gaussian random variable in $\mathbb{R}^d$ with co-variance matrix $AA^t \in GL_d(\mathbb{R})$ and density:

$$g_A(x) = \frac{1}{(2\pi)^{\frac{d}{2}} |\det A|} \cdot e^{-\frac{1}{2}\|A^{-1}x\|^2}$$

Its differential entropy is $H(G_A) = \frac{d}{2}\log 2e\pi + \log|\det A|$ (which is the maximal entropy of all distributions with co-varianca matrix AA^t).

Given a bounded random variable in $\mathbb{R}^d$ and a matrix $B \in \mathrm{GL}_d(\mathbb{R})$ we define the following quantity:

$$H(X;B) := H(X+G_B) - H(G_B)$$

where G_B is a centered Gaussian RV with co-variance matrix BB^t independent of X. This may be thought of as the Gaussian-averaged entropy of X at scale B. Informally this quantity measure the how much information is needed to describe the law of X up to scale $B(\Delta)$ (Δ being the unit ball in $\mathbb{R}^d$). We also define $H(X;B_1|B_2) := H(X;B_1) - H(X;B_2)$.

Define the following partial order on $\mathrm{GL}_d(\mathbb{R})$ by $B_1 \preceq B_2$ whenever $B_2B_2^t - B_1B_1^t$ is a non-negative semi-definite matrix (or equivalently whenever $\forall x \in \mathbb{R}^d \quad \|B_1^t x\|_2 \leqslant \|B_2^t x\|_2$). Some basic properties of this quantity:

Lemma 3.1. *Let $B_1, B_2 \in \mathrm{GL}_d(\mathbb{R})$ with $B_1 \preceq B_2$. Assume that X, Y are two bounded independent random variables taking values in $\mathbb{R}^d$. Then:*

(1) $H(X;B_1) \geqslant 0$,
(2) $H(X;B_1) + H(Y;B_1) \geqslant H(X+Y;B_1)$,
(3) $H(X;B_1) \geqslant H(X;B_2)$,
(4) $H(X+Y;B_1|B_2) \geqslant H(X;B_1|B_2)$.

These properties are a direct consequence of the following:

Theorem 3.2 (Madiman's Submodularity Inequality)**.** *Assume X, Y, Z are three independent $\mathbb{R}^d$-valued random variables such that the distributions of Y, $X+Y$, $Y+Z$, $X+Y+Z$ are absolutely continuous with respect to Lebesgue measure and have finite differential entropy. Then:*

$$H(X+Y+Z) + H(Y) \leqslant H(X+Y) + H(Y+Z)$$

proof of lemma. Item (1) follows from the concavity of F:

$$H(X+G_{B_1}) = \int F(\mathbb{E}[g_{B_1}(x-X)])dx \geqslant \int \mathbb{E}[F(g_{B_1}(x-X))]dx = H(G_{B_1}).$$

Item (2) follows from the submodularity inequality applied to the three independent RVs $X' = X,\ Y' = G_{B_1},\ Z' = Y$:

$$H(X+G_{B_1}+Y) + H(G_{B_1}) \leqslant H(X+G_{B_1}) + H(Y+G_{B_1})$$

subtracting $2H(G_{B_1})$ from both sides gives the required inequality. For item (3), notice that $B_1 \preceq B_2$ implies there exists an $M \in \mathrm{GL}_d(\mathbb{R})$ such that $B_2B_2^t = B_1B_1^t + MM^t$. In particular if G_{B_1} and G_M are the respective and independent RVs then $G_{B_2} \approx G_{B_1} + G_M$. Hence by submodularity:

$$\begin{aligned} H(X+G_{B_2}) + H(G_{B_1}) &= H(X+G_{B_1}+G_M) + H(G_{B_1}) \\ &\leqslant H(X+G_{B_1}) + H(G_{B_1}+G_M) \\ &= H(X+G_{B_1}) + H(G_{B_2}) \\ &\iff H(X;B_2) \leqslant H(X;B_1). \end{aligned}$$

Item (4) is implied by taking $X' = Y,\ Y' = X + G_{B_1},\ Z' = G_M$:

$$\begin{aligned} H(X+Y+G_{B_2}) + H(X+G_{B_1}) &= H(Y+X+G_{B_1}+G_M) + H(G_{B_1}) \\ &\leqslant H(X+Y+G_{B_1}) + H(X+G_{B_1}+G_M) \\ &= H(X+Y+G_{B_1}) + H(X+G_{B_2}) \\ &\iff H(X+Y;B_1|B_2) \geqslant H(X;B_1|B_2). \end{aligned}$$

□

4. Proof of the lower bound in theorem 1.1

Before we begin with proof, a few words and notations on a special case of a random walk in $\mathrm{GL}_d(\mathbb{R})$ - "Bernoulli convolutions for matrices". Given a matrix $A \in \mathrm{GL}_d(\mathbb{R})$ with spectral radius strictly less than 1, define the random sum $X_A = \sum_{n=0}^{\infty} \pm A^n$ where the signs are chosen independently with equal probability. Notice that the sum almost always absolutely converges (by Gelfand $\rho(A) = \lim_{n\to\infty} \|A^n\|^{\frac{1}{n}}$). We denote the finite random sum of ℓ elements to be $X_A^{(\ell)} = \sum_{n=0}^{\ell-1}$. Let μ_A be the distribution of X_A and $\mu_A^{(\ell)}$ be the distribution of $X_A^{(\ell)}$. Given a vector $x \in \mathbb{R}^d$ we denote the projection $X_{A,x} = X_A x$ and $X_{A,x}^{(\ell)} = X_A^{(\ell)} x$ and their corresponding projected distributions $\mu_{A,x},\ \mu_{A,x}^{(\ell)}$. Notice that these measures enjoy a self-similarity property:

$$\mu_A = \mu_A^{(\ell)} * A^\ell \mu_A$$

where $A^\ell \mu_A$ is the push-forward of μ_A by the linear map $A^\ell \in \mathrm{GL}_d(\mathbb{R})$. Similarly, $\mu_{A,x} = \mu_{A,x}^{(\ell)} * A^\ell \mu_{A,x}$.

proof of the lower bound in theorem 1.1. Given $\lambda \in (\frac{1}{2}, 1)$ an algebraic unit, denote by $\lambda_1, ...\lambda_d$ the Galois conjugates of λ with modulus< 1. Take $A \in \mathrm{GL}_d(\mathbb{R})$ to be a real-valued matrix whose eigenvalues coincide with $\{\lambda_1, ..., \lambda_d\}$ (such a matrix always exists as the complex λ_i's come in complex conjugate

pairs). Notice that, by design, the spectral radius of A is strictly less than 1. Assume in addition that the operator norm of A is strictly less than 1 (w.r.t the Euclidean norm on $\mathbb{R}^d$).

The first simple but important observation is that the random walk of $\sum \pm\lambda^n$ on $\mathbb{R}$ may be "lifted" to $\sum \pm A^n$ on $\mathrm{GL}_d(\mathbb{R})$ in the following sense: If for some $a_1, ..., a_m, b_1, ..., b_k \in \mathbb{Q}$ λ admits the following identity $\sum_{i=1}^m a_i\lambda^i = \sum_{j=1}^m b_j\lambda^j$, then so does A, meaning $\sum_{i=1}^m a_i A^i x = \sum_{j=1}^m b_j A^j x$ for any $x \in \mathbb{R}^d$ (readily seen by expressing x via a basis of eigenvectors). Clearly the implication in the other direction holds as well. Hence:

$$H(\mu_\lambda^{(\ell)}) = H(\mu_A^{(\ell)}) \geqslant H(\mu_{A,x}^{(\ell)})$$

As a consequence we have:

$$h_\lambda = \lim_{\ell\to\infty} \frac{1}{\ell} H(\mu_\lambda^{(\ell)}) \geqslant \lim_{\ell\to\infty} \frac{1}{\ell} H(\mu_{A,x}^{(\ell)})$$

for any $x \in \mathbb{R}^d$. We shall now bound $H(\mu_{A,x}^{(\ell)})$ from below.

First we shall show that $H(\mu_{A,x}^{(\ell)}) \geqslant H(\mu_{A,x}; A^\ell|Id)$ which might be expected as both roughly measure the information of $\mu_{A,x}$ at scale $A^\ell(\Delta)$:

$$\begin{aligned} H(\mu_{A,x}; A^\ell|Id) &= H(\mu_{A,x}; A^\ell) - H(\mu_{A,x}; Id) \\ &= H(\mu_{A,x}^{(\ell)} * A^\ell\mu_{A,x}; A^\ell) - H(\mu_{A,x}; Id) \\ &\leqslant H(\mu_{A,x}^{(\ell)}; A^\ell) + H(A^\ell\mu_{A,x}; A^\ell) - H(\mu_{A,x}; Id) \\ &= H(\mu_{A,x}^{(\ell)}; A^\ell) \leqslant H(\mu_{A,x}^{(\ell)}) \end{aligned}$$

where the first inequality follows from property (2) of the lemma and the second inequality follows from the fact that $A^\ell G_{Id} \approx G_{A^\ell}$ and the change of variables formula for the differential entropy. In addition we have by definition:

$$H(\mu_{A,x}; A^\ell|Id) = \sum_{i=1}^{\ell} H(\mu_{A,x}; A^i|A^{i-1})$$

Our assumption that $\|A\|_{op} < 1$ implies $A^i \preceq A^{i-1}$ for all i. Hence by property (4) of the lemma we may deduce:

$$\begin{aligned} H(\mu_{A,x}; A^i|A^{i-1}) &= H(\mu_{A,x}^{(i-1)} * A^{i-1}\mu_{A,x}; A^i|A^{i-1}) \\ &\geqslant H(A^{i-1}\mu_{A,x}; A^i|A^{i-1}) = H(\mu_{A,x}; A|Id) \end{aligned}$$

In conclusion we recieve:

$$\begin{aligned} h_\lambda &= \lim_{\ell\to\infty} \frac{1}{\ell} H(\mu_A^{(\ell)}) \geqslant \lim_{\ell\to\infty} \frac{1}{\ell} H(\mu_{A,x}^{(\ell)}) \\ &\geqslant \lim_{\ell\to\infty} \frac{1}{\ell} H(\mu_A; A^\ell|Id) \geqslant H(\mu_{A,x}; A|Id) \end{aligned}$$

Using property (4) of the lemma again (for $A \preceq Id$) gives:

$$H(\mu_{A,x}; A|Id) = H(\mu_{A,x}^{(1)} * A\mu_{A,x}; A|Id) \geqslant H(\mu_{A,x}^{(1)}; A|Id) = H(\pm x; A|Id)$$

A result from linear algebra shows there exist two orthogonal matrices $u, v \in O_d(\mathbb{R})$ such that $D = uAv$ where $D = diag(1, 1, ..., \Pi\lambda_i)$. Note that

$\Pi\lambda_i = \frac{1}{M_\lambda}$. Exploiting the rotational symmetry of the normalized Gaussian distribution we conclude:

$$H(\pm x; A|Id) = H(\pm ux; D|Id)$$

Taking $x = te_d = (0, 0, ..., 0, t)$ one receives:

$$\begin{aligned} H(\pm x; D|Id) &= H(\pm te_d + DG_{Id}) - H(DG_{Id}) - (H(\pm te_d + G_{Id}) - H(G_{Id})) \\ &= H(\pm tD^{-1}e_d + G_{Id}) - H(G_{Id}) - (H(\pm te_d + G_{Id}) - H(G_{Id})) \\ &= H(\pm tM_\lambda e_d + G_{Id}) - H(\pm te_d + G_{Id}) \\ &= H(\pm tM_\lambda + G) - H(\pm t + G) \end{aligned}$$

where G is the standard one-dimensional Gaussian distribution. The second equality comes from the change of variables formula and the last one from $F(xy) = xF(y) + yF(x)$ which allows to integrate-out all the first $d - 1$ coordinates. An additional calculation shows there exists a uniform constant $c > 0$ with $H(\pm tM_\lambda + G) - H(\pm t + G) \geqslant c \min(1, \log M_\lambda)$ as required. $\square$