The Annals of Probability
2019, Vol. 47, No. 4, 2582–2617
https://doi.org/10.1214/18-AOP1324
© Institute of Mathematical Statistics, 2019

ON THE DIMENSION OF BERNOULLI CONVOLUTIONS

BY EMMANUEL BREUILLARD[1] AND PÉTER P. VARJÚ[2]

University of Cambridge

The Bernoulli convolution with parameter $\lambda \in (0, 1)$ is the probability measure μ_λ that is the law of the random variable $\sum_{n\ge 0} \pm\lambda^n$, where the signs are independent unbiased coin tosses.

We prove that each parameter $\lambda \in (1/2, 1)$ with $\dim \mu_\lambda < 1$ can be approximated by algebraic parameters $\eta \in (1/2, 1)$ within an error of order $\exp(-\deg(\eta)^A)$ such that $\dim \mu_\eta < 1$, for any number A. As a corollary, we conclude that $\dim \mu_\lambda = 1$ for each of $\lambda = \ln 2, e^{-1/2}, \pi/4$. These are the first explicit examples of such transcendental parameters. Moreover, we show that Lehmer's conjecture implies the existence of a constant $a < 1$ such that $\dim \mu_\lambda = 1$ for all $\lambda \in (a, 1)$.

1. Introduction. Let $\lambda \in (0, 1)$ be a real number and let $\xi_0, \xi_1, \dots$ be a sequence of independent random variables with $\mathbb{P}(\xi_n = 1) = \mathbb{P}(\xi_n = -1) = 1/2$. We define the Bernoulli convolution μ_λ with parameter λ as the law of the random variable $\sum_{n=0}^{\infty} \xi_n \lambda^n$.

This paper is concerned with the problem of determining the set of parameters λ such that $\dim \mu_\lambda < 1$. It turns out (see [8], Theorem 2.8) that μ_λ is always exact dimensional, that is, there is a number $0 \le \alpha \le 1$ such that

$$\lim_{r\to 0} \frac{\log \mu_\lambda(x - r, x + r)}{\log r} = \alpha \tag{1.1}$$

for μ_λ-almost every x. We call α the (local) dimension of μ_λ and denote this number by $\dim \mu_\lambda$.

The main result of this paper is the following. We denote by $\mathcal{P}_d$ the set of polynomials of degree at most d all of whose coefficients are -1, 0 or 1. We write

$$E_{d,\alpha} = \{\eta \in (1/2, 1) : \dim \mu_\eta < \alpha \text{ and } P(\eta) = 0 \text{ for some } P \in \mathcal{P}_d\}.$$

THEOREM 1. *Let* $\lambda \in (1/2, 1)$ *be such that* $\dim \mu_\lambda < 1$.
Then for every $\varepsilon > 0$, *there is a number* $A > 0$ *such that the following holds. For every sufficiently large integer* d_0, *there is an integer*

$$d \in [d_0, \exp^{(5)}(\log^{(5)}(d_0) + A)]$$

Received October 2016; revised November 2017.
[1]Supported by ERC Grant no. 617129 "GeTeMo".
[2]Supported by the Royal Society.
MSC2010 subject classifications. 28A80, 42A85.
Key words and phrases. Bernoulli convolution, self-similar measure, dimension, entropy, convolution, transcendence measure, Lehmer's conjecture.

and there is $\eta \in E_{d,\dim \mu_\lambda+\varepsilon}$ *such that*

$$|\lambda - \eta| \le \exp\bigl(-d^{\log^{(3)} d}\bigr).$$

In this paper, the base of the log and exp functions are 2; however, in most places this normalization makes no difference. When we want to use the natural base, we use the notation ln and $e^{(\cdot)}$. We denote by $\log^{(a)}$ and $\exp^{(a)}$ the a-fold iteration of the log and exp functions.

Theorem 1 has a converse.

THEOREM 2. *Let* $\lambda \in (1/2, 1)$ *and let* $\alpha < 1$. *Suppose that there is a sequence* $\{\eta_n\}$ *such that* $\lim \eta_n = \lambda$ *and* $\liminf \dim \mu_{\eta_n} \le \alpha$ *for all* n. *Then* $\dim \mu_\lambda \le \alpha$.

This is an immediate consequence of the fact that the function $\lambda \mapsto \dim \mu_\lambda$ is lower semicontinuous. This was proved, for instance, by Hochman and Shmerkin in [12], Theorem 1.8, but this fact was already known to experts in the area, see the discussion in [12], Section 6. We also give a short proof based on our techniques in Section 2.4.

We formulate some corollaries.

COROLLARY 3. *We have*

$$\bigl\{\lambda \in (1/2, 1) : \dim \mu_\lambda < 1\bigr\} \subseteq \overline{\bigl\{\lambda \in \overline{\mathbb{Q}} \cap (1/2, 1) : \dim \mu_\lambda < 1\bigr\}},$$

where $\overline{\mathbb{Q}}$ *is the set of algebraic numbers and* $\overline{\{\cdot\}}$ *denotes the closure of the set with respect to the natural topology of real numbers.*

We note that the only known examples of parameters $\lambda \in (1/2, 1)$ such that $\dim \mu_\lambda < 1$ are the inverses of Pisot numbers (see [9], Theorem I.2 together with [8], Theorem 2.8 and [31], Theorem 4.4), that is algebraic integers all of whose Galois conjugates are inside the open unit disk. The set of Pisot numbers is closed (see [21]). If one were able to prove that there are no more algebraic parameters with the property $\dim \mu_\lambda < 1$, then this would follow also for transcendental parameters from our result.

The dimension of Bernoulli convolutions for algebraic parameters has been studied in the paper [3]. Recall that Lehmer's conjecture states that there is some numerical constant $\varepsilon_0 > 0$ such that the Mahler measure M_λ (the definition is recalled below in (1.6)) of every algebraic number λ is either 1 or at least $1 + \varepsilon_0$. It was proved in [3] that Lehmer's conjecture implies that there exists a number $a < 1$ such that $\dim \mu_\lambda = 1$ for all algebraic numbers $\lambda \in (a, 1)$. We can now drop the condition of algebraicity in that result thanks to Corollary 3 and we obtain the following.

COROLLARY 4. *If Lehmer's conjecture holds*, *then there is an absolute constant* $a < 1$ *such that* $\dim \mu_\lambda = 1$ *for all* $\lambda \in (a, 1)$.

We also have the following result.

COROLLARY 5. *Let* $\lambda \in (1/2, 1)$ *be a number such that*

$$|P(\lambda)| > \exp(-d^{\log^{(3)} d}) \tag{1.2}$$

for all $P \in \mathcal{P}_d$ *for all sufficiently large* d.
Then $\dim \mu_\lambda = 1$.

A simple calculation shows that $|P'(x)| < d(d+1)/2$ for all $x \in (0, 1)$ and $P \in \mathcal{P}_d$. If there is a number η that is a root of a polynomial $P \in \mathcal{P}_d$ such that

$$|\lambda - \eta| \le \frac{2}{d(d+1)} \exp(-d^{\log^{(3)} d}),$$

then $|P(\lambda)| \le \exp(-d^{\log^{(3)} d})$. We will see in the proof of Theorem 1 that the factor $2/d(d+1)$ is insignificant and that this slightly stronger approximation also holds in the setting of the theorem.

There is a large variety of explicit transcendental numbers, for which the estimate (1.2) has been established. In Sprindžuk's classification of numbers, all $\widetilde{S}$-numbers, all $\widetilde{T}$-numbers and those $\widetilde{U}$-numbers, for which $H_0 \ge 2$ satisfy (1.2). See [4], Chapter 8.1, for the notation.

In particular, we have $\dim \mu_\lambda = 1$ for each of

$$\lambda \in \{\ln 2, e^{-1/2}, \pi/4\};$$

see, for example [29], Figure 1, as well as for many Mahler numbers, see, for example, [32]. For further examples, we refer the reader to the references in [4], pages 189 and in [29, 32].

If one is interested in the smallest possible value that $\dim \mu_\lambda$ can take then it is enough to look at algebraic parameters thanks to the following result.

COROLLARY 6. *We have*

$$\min_{\lambda \in (1/2,1)} \dim \mu_\lambda = \inf_{\lambda \in (1/2,1) \cap \overline{\mathbb{Q}}} \dim \mu_\lambda.$$

Indeed, let $\dim \mu_{\lambda_0} = \min_{\lambda \in (1/2,1)} \dim \mu_\lambda$. By Theorem 1, for each $\varepsilon > 0$, there is an algebraic parameter $\eta \in (1/2, 1)$ such that $\dim \mu_\eta < \dim \mu_{\lambda_0} + \varepsilon$, and this proves the claim.

Hare and Sidorov [10] proved that $\dim \mu_\lambda \ge 0.81$ for all Pisot parameters $\lambda \in (1/2, 1)$. The authors of that paper explained to us in private communication that their result can be extended to arbitrary algebraic parameters in $(1/2, 1)$. Combined with Corollary 6, this gives 0.81 as an explicit uniform lower bound for the dimension of μ_λ for all parameters in $(1/2, 1)$.

1.1. *Background.* For thorough surveys on Bernoulli convolutions, we refer to [20] and [25]. For a discussion of the more recent developments, see [27].

Bernoulli convolutions originate in a paper of Jessen and Wintner [13] and they have been studied by Erdős in [6, 7]. If $\lambda < 1/2$, then $\operatorname{supp}\mu_\lambda$ is a Cantor set, and it is easily seen that $\dim\mu_\lambda = 1/\log\lambda^{-1}$. (Recall that log is base 2 in this paper.) If $\lambda = 1/2$, then μ_λ is the normalized Lebesgue measure restricted to the interval $[-2, 2]$.

It has been noticed by Erdős [6] that μ_λ may be singular with respect to the Lebesgue measure even if $\lambda > 1/2$. In particular, he showed that μ_λ is singular whenever $\lambda^{-1} \neq 2$ is a Pisot number. Moreover, Garsia [9], Theorem I.2 (together with [8], Theorem 2.8 and [31], Theorem 4.4) showed that $\dim\mu_\lambda < 1$ if $\lambda^{-1} \neq 2$ is a Pisot number.

The typical behaviour is absolute continuity for parameters in $(1/2, 1)$. Indeed, Erdős [7] showed that μ_λ is absolutely continuous for almost all $\lambda \in (a, 1)$, where $a < 1$ is an absolute constant. This has been extended by Solomyak [24] to almost all $\lambda \in (1/2, 1)$.

Very recently Hochman [11], Theorem 1.9, made a further breakthrough on this problem.

THEOREM 7 (Hochman). *Let* $\lambda \in (1/2, 1)$ *be such that* $\dim\mu_\lambda < 1$.

Then for every $A > 0$, *there is a number* d_0 *such that for all integers* $d > d_0$, *there is an algebraic number* η *that is a root of a polynomial in* $\mathcal{P}_d$ *such that*

$$|\lambda - \eta| \leq \exp(-Ad).$$

In comparison with Theorem 1, Hochman's result has the advantage that it provides an algebraic approximation of an exceptional parameter at each scale. On the other hand, Theorem 1 provides a smaller error and the information that the approximating parameter is also exceptional (i.e., $\dim\mu_\eta < 1$).

Theorem 7 also implies that the set of exceptional parameters

$$\{\lambda \in (1/2, 1) : \dim\mu_\lambda < 1\}$$

is of packing dimension 0. Building on this result, Shmerkin [22] proved that

$$\{\lambda \in (1/2, 1) : \mu_\lambda \text{ is singular}\}$$

is of Hausdorff dimension 0. We recall that a set of packing dimension 0 is also a set of Hausdorff dimension 0.

See also the very recent paper of Shmerkin [23], where he proves a stronger version of Hochman's result for the L^q-dimension of Bernoulli convolutions. He also concludes that outside an exceptional set of Hausdorff dimension 0 for the parameter, Bernoulli convolutions are absolutely continuous with a density in L^q for any $q < \infty$. Moreover, his methods can establish that the density has fractional derivatives.

Theorem 7 also implies a conditional result on $\dim \mu_\lambda$ for transcendental parameters. Hochman proved that $\dim \mu_\lambda = 1$ for all transcendental parameters $\lambda \in (1/2, 1)$ if the answer is affirmative to the following question posed by him [11], Question 1.10. Is there an absolute constant $C > 0$ such that

$$|\eta_1 - \eta_2| \geq \exp(-Cd) \tag{1.3}$$

holds for any two different numbers $\eta_1 \neq \eta_2$ that are roots of (not necessarily the same) polynomials in $\mathcal{P}_d$? However, such a bound is not yet available; the best known result in this direction is due to Mahler ([19], Theorem 2) who proved

$$|\eta_1 - \eta_2| \geq \exp(-Cd \log d), \tag{1.4}$$

where C is an absolute constant. (See Theorem 21 below for more details.)

The work of Hochman [11] also gives a formula for the dimension of μ_λ, if λ is an algebraic number. Denote by h_λ the entropy of the random walk on the semigroup generated by the transformations $x \mapsto \lambda \cdot x + 1$ and $x \mapsto \lambda \cdot x - 1$. More precisely, let

$$h_\lambda = \lim_{n\to\infty} \frac{1}{n} H\left(\sum_{i=0}^{n-1} \xi_i \lambda^i\right) = \inf \frac{1}{n} H\left(\sum_{i=0}^{n-1} \xi_i \lambda^i\right),$$

where $H(\cdot)$ denotes the Shannon entropy of a discrete random variable. With this notation Hochman's formula is

$$\dim \mu_\lambda = \min(-h_\lambda / \log \lambda, 1). \tag{1.5}$$

(See [3], Section 3.4, where the formula is derived in this form from Hochman's main result.)

The quantity h_λ has been studied in the paper [3]. It was proved there ([3], Theorem 5) that there is an absolute constant $c_0 > 0$ such that for any algebraic number, we have

$$c_0 \cdot \min(\log M_\lambda, 1) \leq h_\lambda \leq \min(\log M_\lambda, 1).$$

The log's in this formula, as well as those that appear in the definition of entropy, are base 2. Numerical calculations reported in that paper indicate that one can take $c_0 = 0.44$. This result combined with Hochman's formula implies that $\dim \mu_\lambda = 1$ provided λ is an algebraic number with $1 > \lambda > \min(2, M_\lambda)^{-c_0}$. Here, and everywhere in the paper, we denote by M_λ the Mahler measure of an algebraic number λ. That is, if $P(x) = a_d \prod (x - \lambda_j)$ is the minimal polynomial of λ in $\mathbb{Z}[x]$, then by definition,

$$M_\lambda = |a_d| \prod_{j:|\lambda_j|>1} |\lambda_j|. \tag{1.6}$$

1.2. *The strategy of the proof.* This section gives an informal account of the proof of Theorem 1. All the arguments presented here will be repeated in a rigorous fashion later in the paper. Therefore, we take a rather relaxed approach towards our estimates. In particular, we will write $\lessapprox$ to indicate an inequality that could be made valid by inserting suitable constants in appropriate places.

The proof of our results builds on the techniques introduced by Hochman in [11] using entropy estimates.

We work with the following notion of entropy. Let X be a bounded random variable and let $r > 0$ be a real number. We define

$$H(X; r) := \int_0^1 H(\lfloor X/r + t \rfloor)\,dt.$$

On the right hand side, $H(\cdot)$ denotes the Shannon entropy of a discrete random variable. In addition, we define the conditional entropies

$$H(X; r_1|r_2) := H(X; r_1) - H(X; r_2).$$

We will study the basic properties of these quantities in Section 2. In particular, $H(X; r)$ is a nonincreasing function of r. Furthermore $0 \le H(X; r) \le \log r^{-1} + O(1)$, where the implied constant depends only on $\operatorname{esssup}|X|$. By abuse of notation, we write $H(\mu; r_1|r_2) = H(X; r_1|r_2)$ and similar expressions if μ denotes the law of X.

These quantities differ from those used by Hochman in that they involve an averaging over a random translation. This averaging endows these quantities with some useful properties as we will see in Section 2.2, which often comes in handy. The idea of this averaging procedure originates in Wang's paper [30], Section 4.1.

We fix a number $\lambda \in (1/2, 1)$ until the end of the section. For a set $I \subset \mathbb{R}_{>0}$, we write μ_λ^I for the law of the random variable

$$\sum_{n\in\mathbb{Z}:\lambda^n\in I} \xi_n \lambda^n.$$

We note that in this notation $\mu_\lambda^{(0,1]} = \mu_\lambda$ and $\mu_\lambda^{(\lambda^n,1]}$ is the law of $\sum_{j=0}^{n-1} \xi_j \lambda^j$, the first n term truncation of the series defining Bernoulli convolutions.

We note that

$$\dim \mu_\lambda = \lim_{n\to\infty} \frac{H(\mu_\lambda; \lambda^n)}{n \log \lambda^{-1}},$$

(see Lemma 14) and that $H(\mu_\lambda^{(\lambda^n,1]}; \lambda^n) \approx H(\mu_\lambda; \lambda^n)$ up to additive constants independent of n. Hence,

$$H\big(\mu_\lambda^{(\lambda^n,1]}; \lambda^n\big) \approx n \log \lambda^{-1} \dim \mu_\lambda. \tag{1.7}$$

We now assume that $\dim \mu_\lambda < 1$ and we assume by contradiction that the algebraic approximations to λ claimed in Theorem 1 do not exist. In the first part of the proof given in Section 3, we search for integers n with the property that

$$H\big(\mu_\lambda^{(\lambda^n,1]}; r\big) \ge n \log \lambda^{-1} (\dim \mu_\lambda + \varepsilon) \tag{1.8}$$

for a suitable scale $r \approx n^{-Cn}$. Equation (1.8) is a small improvement over (1.7) when we replace λ^n with the smaller scale r.

If $\varepsilon > 0$ is small enough so that the right-hand side of (1.8) is $< n$, and if (1.8) fails, then there are pairs of choices of the signs in the sum

$$\sum_{j=0}^{n-1} \pm\lambda^j$$

that give the same value within an error of r. For each such pair, there corresponds a nonzero polynomial $P \in \mathcal{P}_{n-1}$ such that $|2P(\lambda)| < r$. In Section 3, we show that these polynomials must have a common root η and $|\lambda - \eta| < n^{-4n}$. Since this collection of polynomials is rich enough to cause the failure of (1.8), we obtain

$$H\big(\mu_\eta^{(\eta^n,1]}\big) \le n \log \lambda^{-1}(\dim \mu_\lambda + \varepsilon),$$

which yields $h_\eta \le \log \lambda^{-1}(\dim \mu_\lambda + \varepsilon)$. Plugging this into (1.5), we get $\dim \mu_\eta \le \dim \mu_\lambda + \varepsilon'$, where ε' is arbitrarily close to ε if n is sufficiently large. Hence, $\eta \in E_{n,\dim \mu_\lambda + \varepsilon'}$.

Then we choose another integer n' such that $|\lambda - \eta|$ is just slightly larger than $n'^{-4n'}$. If (1.8) fails again for n' and for a suitable r', then we can repeat the above argument to find another number $\eta' \in E_{n',\dim \mu_\lambda + \varepsilon}$ such that $|\lambda - \eta'| < n'^{-4n'}$. Then $|\eta - \eta'| < 2n'^{-4n'}$, and we can conclude $\eta = \eta'$ thanks to (1.4) (the result of Mahler on the separation between roots of polynomials in $\mathcal{P}_n$). However, we carefully chose n' to make sure that $|\lambda - \eta'| < n'^{-4n'} < |\lambda - \eta|$, hence we cannot have $\eta = \eta'$, which shows that (1.8) must hold for at least one of n or n'.

The way we exploited Mahler's bound (1.4) is reminiscent to Hochman's argument for showing $\dim \mu_\lambda = 1$ for all transcendental $\lambda \in (1/2, 1)$ assuming the stronger bound (1.3) discussed in the previous section.

We will use the (indirect) assumption on the lack of algebraic approximations to λ to control n' in terms of n. Indeed, if (1.8) fails for n, we get that it holds for n' with

$$n'^{4n'} \lessapprox |\lambda - \eta|^{-1} < \exp\big(n^{\log^{(3)} n}\big). \tag{1.9}$$

This will enable us to produce suitably many integers n in a given range such that (1.8) holds.

In the second part of the proof, which we discuss in Section 4, we use the identity

$$\mu_\lambda^{I_1 \dot\cup \cdots \dot\cup I_k} = \mu_\lambda^{I_1} * \cdots * \mu_\lambda^{I_k}$$

and argue that entropy increases under convolution to improve on the bound (1.8). We use the following result from [28], Theorem 3.

THEOREM 8. *For every* $0 < \alpha \le 1/2$, *there are numbers* $C, c > 0$ *such that the following holds. Let* μ, ν *be two compactly supported probability measures on* $\mathbb{R}$. *Let* $\sigma_2 < \sigma_1 < 0$ *and* $0 < \beta \le 1/2$ *be real numbers. Suppose that*

$$H(\mu; 2^{\sigma}|2^{\sigma+1}) < 1 - \alpha$$

for all $\sigma_2 < \sigma < \sigma_1$. *Suppose further that*

$$H(\nu; 2^{\sigma_2}|2^{\sigma_1}) > \beta(\sigma_1 - \sigma_2).$$

Then

$$H(\mu * \nu; 2^{\sigma_2}|2^{\sigma_1}) > H(\mu; 2^{\sigma_2}|2^{\sigma_1}) + c\beta(\log \beta^{-1})^{-1}(\sigma_1 - \sigma_2) - C.$$

We note that the supremum of the values $H(\mu; r|2r)$ may take over all probability measures μ is 1 (see (2.5) below and the comment following it). We will see (in Lemma 13) that the assumption $\dim \mu_\lambda < 1$ implies that there is a number $\alpha > 0$ such that $H(\mu_\lambda^I; r|2r) < 1 - \alpha$ for all $r > 0$ and for all $I \subset \mathbb{R}_{>0}$. This means that the hypothesis of Theorem 8 holds for $\mu = \mu_\lambda^I$ for all $I \subset \mathbb{R}_{>0}$ with an α depending only on λ.

We give a brief and informal explanation on how this result will be used. Suppose that (1.8) holds for some n and r. Now (1.7) implies

$$H(\mu_\lambda^{(\lambda^n,1]}; \lambda^n) < (\dim \mu_\lambda + \varepsilon/2)n \log \lambda^{-1},$$

if n is sufficiently large, so we can show that

$$H(\mu_\lambda^{(\lambda^n,1]}; r|\lambda^n) \ge \varepsilon' n,$$

for some ε' depending only on ε and λ.

For simplicity of exposition, we assume now that the stronger bound

$$H(\mu_\lambda^{(\lambda^n,1]}; r|r^{9/10}) \ge \varepsilon' n$$

holds. There is no way to justify this hypothesis; in the actual proof we need to consider a suitable decomposition of the scales between λ^n and r.

Using scaling properties of entropy, we can write

$$H(\mu_\lambda^{(\lambda^{n(j+1)},\lambda^{nj}]}; r\lambda^{jn}|r^{9/10}\lambda^{jn}) \ge \varepsilon' n.$$

We consider this inequality for $j = 0, 1, \ldots, N-1$ for some $N \approx (\log r^{-1})/n$ so that $r^{1/10} \le \lambda^{jn} \le 1$ for each j in the range. Hence

$$H(\mu_\lambda^{(\lambda^{n(j+1)},\lambda^{nj}]}; r^{11/10}|r^{9/10}) \ge \varepsilon' n,$$

because $[r^{11/10}, r^{9/10}] \supset [r\lambda^{jn}, r^{9/10}\lambda^{jn}]$.

We can now apply Theorem 8 $N \approx \log n$ times with

$$\beta \approx \frac{n}{\log r^{-1}} \approx \frac{1}{\log n},$$

and we obtain

$$H\big(\mu_\lambda^{(\lambda^{nN},1]}; r^{11/10}|r^{9/10}\big) \gtrapprox \frac{\log r^{-1}}{\log\log n}, \tag{1.10}$$

that is, the average entropy of a digit is at least $\approx (\log\log n)^{-1}$.

Then we will apply Theorem 8 again in a second stage. Let $n_1, n_2, \dots$ be a sequence of integers such that (1.8) and hence (1.10) holds. We apply Theorem 8 repeatedly again with $\beta_i \approx 1/\log^{(2)}(n_i)$, and find that the average entropy of a digit between suitable scales is at least

$$\approx \sum \frac{1}{\log^{(2)}(n_i)\log^{(3)}(n_i)}.$$

If n_i does not grow faster than $\exp^{(2)}(i\log^{(2)} i)$, then the above sum can be arbitrarily large contradicting the fact that the entropy of a digit cannot exceed 1. This contradiction ends the proof.

Note that using the argument that we presented in the beginning of this sketch, one can show that the lack of the algebraic approximations claimed in Theorem 1 implies that we can find a sequence n_i that satisfies our requirement (1.8) and also satisfies the growth condition

$$n_{i+1}^{n_{i+1}} \lessapprox \exp\big(n_i^{\log^{(3)} n_i}\big),$$

see (1.9). We can use this to prove $n_i \lessapprox \exp^{(2)}(i\log^{(2)} i)$ by induction.

1.3. *Notation.* We denote by the letters c, C and their indexed variants various constants that could in principle be computed explicitly following the proofs step by step. The value of these constants denoted by the same symbol may change between occurrences. We keep the convention that we denote by lower case letters the constants that are best thought of as "small" and by capital letters the ones that are "large".

We denote by log and exp the base 2 logarithm and exponential functions and write ln for the logarithm in base e. We denote by $\log^{(a)}$ and $\exp^{(a)}$ the a-fold iterates of the log and exp functions.

The letter λ denotes a number in $(0, 1)$. For a bounded set $I \subset \mathbb{R}_{>0}$, we denote by μ_λ^I the law of the random variable

$$\sum_{n\in\mathbb{Z}:\lambda^n\in I} \xi_n\lambda^n,$$

where ξ_n is a sequence of independent unbiased ±1 valued random variables. In particular, we write $\mu_\lambda = \mu_\lambda^{(0,1]}$.

We denote by $\mathcal{P}_d$ the set of polynomials of degree at most d with coefficients ±1 and 0.

1.4. *The organization of this paper.* We begin by discussing some basic properties of entropy in Section 2, which we will rely on throughout the paper. Section 3 contains the first part of the proof of the main result focusing on the initial entropy estimate (1.8) mentioned above. The proof of Theorem 1 is completed in Section 4, where we exploit Theorem 8 to improve on our initial entropy estimate.

2. Preliminaries on entropy. The purpose of this section is to provide some background material on entropy.

2.1. *Shannon and differential entropies.* If X is a discrete random variable, we write $H(X)$ for its Shannon entropy, that is

$$H(X) = \sum_{x \in \mathcal{X}} -\mathbb{P}(X = x) \log \mathbb{P}(X = x),$$

where $\mathcal{X}$ denotes the set of values X takes. We recall that the base of log is 2 throughout the paper. If X is an absolutely continuous random variable with density $f : \mathbb{R} \to \mathbb{R}_{\geq 0}$, we write $H(X)$ for its differential entropy, that is

$$H(X) = \int -f(x) \log f(x)\, dx.$$

This dual use for $H(\cdot)$ should cause no confusion, as the type of the random variable will always be clear from the context. If μ is a probability measure, we write $H(\mu) = H(X)$, where X is a random variable with law μ.

Shannon entropy is always nonnegative. Differential entropy on the other hand can take negative values. For example, if $a \in \mathbb{R}_{>0}$, and X is a random variable with finite differential entropy, then it follows from the change of variables formula that

$$H(aX) = H(X) + \log a, \tag{2.1}$$

which can take negative values when a varies. On the other hand, if X takes countably many values, the Shannon entropy of aX is the same as that of X. Note that both entropies are invariant under translation by a constant in $\mathbb{R}$.

We define $F(x) := -x \log(x)$ for $x > 0$ and recall that F is concave. From the concavity of F and Jensen's inequality, we see that for any discrete random variable X taking at most N different values,

$$H(X) \leq \log N. \tag{2.2}$$

Let X and Y be two discrete random variables. We define the conditional entropy of X relative to Y as

$$\begin{aligned} H(X|Y) &= \sum_{y \in \mathcal{Y}} \mathbb{P}(Y = y) H(X|Y = y) \\ &= \sum_{y \in \mathcal{Y}} \mathbb{P}(Y = y) \sum_{x \in \mathcal{X}} -\frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} \log \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}. \end{aligned}$$

We recall some well-known properties. We always have $0 \le H(X|Y) \le H(X)$, and $H(X|Y) = H(X)$ if and only if the two random variables are independent (see [5], Theorem 2.6.5).

We recall the following result from [17], Theorem I.

PROPOSITION 9 (Submodularity inequality). *Assume that X, Y, Z are three independent $\mathbb{R}$-valued random variables such that the distributions of Y, $X+Y$, $Y+Z$ and $X+Y+Z$ are absolutely continuous with respect to Lebesgue measure and have finite differential entropy. Then*

$$H(X+Y+Z) + H(Y) \le H(X+Y) + H(Y+Z). \tag{2.3}$$

This result goes back in some form at least to a paper by Kaimanovich and Vershik [14], Proposition 1.3. The version in that paper assumes that the laws of X, Y and Z are identical. The inequality was rediscovered by Madiman ([17], Theorem I) in the greater generality stated above. Then it was recast in the context of entropy analogues of sumset estimates from additive combinatorics by Tao [26] and Kontoyannis and Madiman [15]. And indeed Proposition 9 can be seen as an entropy analogue of the Plünnecke–Ruzsa inequality in additive combinatorics. For the proof of this exact formulation, see [3], Theorem 7.

2.2. *Entropy at a given scale*. We recall the notation

$$H(X;r) = \int_0^1 H(\lfloor X/r + t \rfloor)\, dt$$

and

$$H(X; r_1|r_2) = H(X; r_1) - H(X; r_2).$$

These quantities originate in the work of Wang [30], and they also play an important role in the paper [16], where a quantitative version of Bourgain's sum-product theorem is proved.

We continue by recording some useful facts about these notions. If N is an integer, then we have the following interpretation, which follows easily from the definition

$$H(X; N^{-1}r|r) = \int_0^1 H(\lfloor N(r^{-1}X + t)\rfloor | \lfloor r^{-1}X + t\rfloor). \tag{2.4}$$

Indeed, $\lfloor r^{-1}X + t\rfloor$ is a function of $\lfloor N(r^{-1}X + t)\rfloor$, hence

$$H(\lfloor N(r^{-1}X + t)\rfloor | \lfloor r^{-1}X + t\rfloor) = H(\lfloor N(r^{-1}X + t)\rfloor) - H(\lfloor r^{-1}X + t\rfloor).$$

Combining this interpretation with (2.2), we see that

$$H(\mu; r|2r) \le 1 \tag{2.5}$$

for any probability measure μ. This upper bound is best possible, as demonstrated by the uniform measures on long intervals.

It is immediate from the definitions that we have the scaling formulae

$$H(sX; sr) = H(X; r), \qquad H(sX; sr_1|sr_2) = H(X; r_1|r_2),$$

for any random variable X and real numbers $s, r, r_1, r_2 > 0$. In particular, we have

$$H(\mu_\lambda^{\lambda^k I}; \lambda^k r_1|\lambda^k r_2) = H(\mu_\lambda^I; r_1|r_2), \tag{2.6}$$

for any integer k, real numbers $r_1, r_2 > 0$ and $I \subset \mathbb{R}_{>0}$.

The next lemma gives an alternative definition for entropy at a given scale.

LEMMA 10 ([28], Lemma 5). *Let X be a bounded random variable in $\mathbb{R}$. Then*

$$H(X; r) = H(X + I_r) - H(I_r) = H(X + I_r) - \log(r),$$

where I_r is a uniform random variable in $[0, r]$ independent of X.

It follows from the definition that being an average of Shannon entropies $H(X; r)$ is always nonnegative. Similarly, we see from (2.4) that $H(X; r_1|r_2)$ is also nonnegative if r_2/r_1 is an integer. We will see below that this holds also for any $r_2 \geq r_1$.

The next lemma shows that conditional entropy between scales of integral ratio cannot decrease by taking convolution of measures.

LEMMA 11 ([28], Lemma 6). *Let X and Y be two bounded independent random variables in $\mathbb{R}$. Let $r_2 > r_1 > 0$ be two numbers such that $r_2/r_1 \in \mathbb{Z}$. Then*

$$H(X + Y; r_1|r_2) \geq H(X; r_1|r_2).$$

We record an instance of this lemma that is of particular importance to us. We have

$$H(\mu_\lambda^{I_1}; r_1|r_2) \geq H(\mu_\lambda^{I_2}; r_1|r_2) \tag{2.7}$$

for any $I_2 \subset I_1 \subset \mathbb{R}_{>0}$ provided the ratio of the scales r_2/r_1 is an integer. Unfortunately, this may fail if the ratio of the scales in nonintegral, but we always have the following. If $r_2/r_1 \geq 2$, we can find $r_1 \leq t_1 \leq t_2 \leq r_2$ such that t_2/t_1 is an integer and

$$H(\mu_\lambda^{I_2}; t_1|t_2) \geq H(\mu_\lambda^{I_2}; r_1|r_2)/2.$$

We combine this with (2.7) and (2.9) (see below) and get

$$H(\mu_\lambda^{I_1}; r_1|r_2) \geq H(\mu_\lambda^{I_1}; t_1|t_2) \geq H(\mu_\lambda^{I_2}; t_1|t_2) \geq H(\mu_\lambda^{I_2}; r_1|r_2)/2. \tag{2.8}$$

(It is possible to prove a variant of this with a small additive error term instead of the multiplicative constant; see [28], Lemma 9. However, for the purposes of this paper (2.8) is more convenient.)

We recall a result form [16] (see also [28], Lemma 8), which establishes that $H(X; r)$ is a monotone increasing and Lipschitz function of $-\log r$; in particular $H(X; r_1|r_2)$ is nonnegative for all $r_1 \leq r_2$.

LEMMA 12. *Let X be a bounded random variable in $\mathbb{R}$. Then for any $r_2 \geq r_1 > 0$ we have*

$$0 \leq H(X; r_1) - H(X; r_2) \leq 2(\log r_2 - \log r_1).$$

This lemma implies that $H(\mu; r_1|r_2) \geq 0$, whenever $r_2 \geq r_1$. Moreover, we have

$$H(\mu; r_1|r_2) \geq H(\mu; s_1|s_2) \tag{2.9}$$

provided $(s_1, s_2) \subset (r_1, r_2)$.

2.3. *Bernoulli convolutions of dimension less than* 1. The purpose of this section is to show that the entropy of a single digit for a Bernoulli convolution that is of dimension less than 1 is bounded away from 1. This implies that Theorem 8 applies to $\mu = \mu_\lambda^I$ for any I between any scales if $\dim \mu_\lambda < 1$ with an α depending only on λ.

LEMMA 13. *Let $\lambda \in (1/2, 1)$ be such that $\dim \mu_\lambda < 1$. Then there is a number $\alpha > 0$ such that*

$$H(\mu_\lambda; r|2r) < 1 - \alpha$$

for all $r > 0$.

Thanks to (2.6) and (2.7), the same conclusion holds for μ_λ^I for any $I \subset \mathbb{R}_{>0}$ in place of μ_λ.

We begin by recalling the relation between the dimension and the entropy of Bernoulli convolutions, which is folklore.

LEMMA 14. *Let $\lambda \in (0, 1)$. Then*

$$\dim \mu_\lambda = \lim_{r \to 0} \frac{H(\mu_\lambda; r)}{\log r^{-1}}.$$

PROOF. By [8], Theorem 2.8, μ_λ is exact dimensional. By [31], Theorem 4.4, the Rényi entropy dimension of an exact dimensional measure coincides with its local dimension (the number α in (1.1)).

Thus,

$$\dim \mu_\lambda = \lim_{r \to 0} \frac{H(\lfloor r^{-1} X \rfloor)}{\log r^{-1}},$$

where X is a random variable with law μ_λ. Moreover, the same formula holds for any translates of X, and the claim follows by dominated convergence. $\square$

We fix $\lambda \in (1/2, 1)$ such that $\dim \mu_\lambda < 1$. By Lemma 14, there are numbers N and $\alpha_0 > 0$ such that

$$H(\mu_\lambda; 2^{-n}) < n(1-\alpha_0) \tag{2.10}$$

for any $n > N$.

We assume to the contrary that there is a number r such that

$$H(\mu_\lambda; r|2r) \geq 1-\alpha, \tag{2.11}$$

where $\alpha > 0$ is a suitably small number depending only on α_0 to be specified later.

To contradict (2.10), we aim to produce more digits of high entropy. One source of these digits will be the scaling formula (2.6), which together with (2.7) implies

$$H(\mu_\lambda; \lambda^k r|2\lambda^k r) \geq H(\mu_\lambda^{(0,\lambda^k]}; \lambda^k r|2\lambda^k r) = H(\mu_\lambda; r|2r) \geq 1-\alpha. \tag{2.12}$$

The other source is the next lemma.

LEMMA 15. *Let μ be a compactly supported probability measure on $\mathbb{R}$ and let $r > 0$ be a number. Then*

$$1 - H(\mu; 2r|4r) \leq 4\big(1 - H(\mu; r|2r)\big).$$

PROOF. Write χ_s for the uniform probability measure on the interval $[0, s]$ and let $\eta_s = (\delta_0 + \delta_s)/2$, where δ_x denotes the unit mass supported at the point x. By Lemma 10, we have

$$\begin{aligned} H(\mu; r|2r) &= H(\mu * \chi_r) - H(\chi_r) - H(\mu * \chi_{2r}) + H(\chi_{2r}) \\ &= 1 - \big(H(\mu * \eta_r * \chi_r) - H(\mu * \chi_r)\big). \end{aligned} \tag{2.13}$$

Note that $\chi_{2r} = \chi_r * \eta_r$.

By submodularity (Proposition 9), we have

$$H(\mu * \eta_r * \eta_r * \chi_r) - H(\mu * \eta_r * \chi_r) \leq H(\mu * \eta_r * \chi_r) - H(\mu * \chi_r),$$

hence

$$H(\mu * \eta_r * \eta_r * \chi_r) - H(\mu * \chi_r) \leq 2\big(H(\mu * \eta_r * \chi_r) - H(\mu * \chi_r)\big).$$

We note the identity

$$\eta_r * \eta_r = \frac{\eta_{2r} + \delta_r}{2}.$$

By concavity of $F(x) = -x \log x$, we have

$$H(\mu * \eta_r * \eta_r * \chi_r) \geq H(\mu * \eta_{2r} * \chi_r)/2 + H(\mu * \delta_r * \chi_r)/2.$$

Thus,

$$H(\mu * \eta_{2r} * \chi_r) - H(\mu * \chi_r) \le 2\big(H(\mu * \eta_r * \eta_r * \chi_r) - H(\mu * \chi_r)\big)$$
$$\le 4\big(H(\mu * \eta_r * \chi_r) - H(\mu * \chi_r)\big).$$

We use submodularity again to write

$$\begin{aligned} H(\mu * \eta_{2r} * \chi_{2r}) - H(\mu * \chi_{2r}) &= H(\mu * \eta_{2r} * \eta_r * \chi_r) - H(\mu * \eta_r * \chi_r) \\ &\le H(\mu * \eta_{2r} * \chi_r) - H(\mu * \chi_r) \\ &\le 4\big(H(\mu * \eta_r * \chi_r) - H(\mu * \chi_r)\big). \end{aligned}$$

We combine this with (2.13) and conclude the lemma. □

PROOF OF LEMMA 13. We assume to the contrary that (2.10) and (2.11) hold and we fix two integers K, J.

Using Lemma 15 repeatedly, we find that

$$H\big(\mu_\lambda; 2^k r | 2^{k+1} r\big) \ge 1 - 4^k \alpha$$

holds for all $k \in \mathbb{Z}_{\ge 0}$. We sum these inequalities for $k = 0, \dots, K-1$, and arrive at

$$H\big(\mu_\lambda; r | 2^K r\big) \ge K - 4^K \alpha.$$

We choose an integer m such that $2^{-K-1} \le \lambda^m \le 2^{-K}$ and use (2.12) together with the above argument to conclude

$$H\big(\mu_\lambda; \lambda^{jm} r | 2^K \lambda^{jm} r\big) \ge K - 4^K \alpha$$

for any $j \in \mathbb{Z}_{\ge 0}$. We sum this for $j = 0, \dots, J-1$ and use (2.9) to get

$$H\big(\mu_\lambda; \lambda^{(J-1)m} r\big) \ge J\big(K - 4^K \alpha\big).$$

Since $\lambda^m \ge 2^{-K-1}$, we get

$$H\big(\mu_\lambda; 2^{-(J(K+1)+\lceil \log r^{-1} \rceil)}\big) \ge J\big(K - 4^K \alpha\big) \ge J(K+1)\big(1 - 4^K \alpha - (K+1)^{-1}\big).$$

We set the parameters. We take K to be large enough so that $(K+1)^{-1} < \alpha_0/3$. Then we take α small enough so that $4^K \alpha < \alpha_0/3$. Finally, we take J sufficiently large so that we get a contradiction to (2.10) for $n = J(K+1) + \lceil \log r^{-1} \rceil$. This proves the lemma. □

2.4. *Lower semicontinuity.* The purpose of this section is to establish the following result.

LEMMA 16. *The function $\lambda \mapsto \dim \mu_\lambda$ is lower semicontinuous.*

PROOF. By Lemma 14, we have

$$\dim \mu_\lambda = \lim_{n\to\infty} \frac{H(\mu_\lambda;\lambda^n|1)}{-n\log\lambda}.$$

For each n, the function

$$\lambda \mapsto \frac{H(\mu_\lambda;\lambda^n|1)}{-n\log\lambda}$$

is continuous. This follows from [28], Lemma 7 and Lemma 12.

We show that the sequence of functions

$$f_k(\lambda) = \frac{H(\mu_\lambda;\lambda^{2^k}|1)-2}{-2^k\log\lambda}$$

increases pointwise, and this completes the proof. We can write

$$H(\mu_\lambda;\lambda^{2^{k+1}}|1) = H(\mu_\lambda;\lambda^{2^{k+1}}|\lambda^{2^k}) + H(\mu_\lambda;\lambda^{2^k}|1).$$

Let $\lambda^{2^k}/2 \le r \le \lambda^{2^k}$ be such that $r/\lambda^{2^{k+1}} \in \mathbb{Z}_{>0}$. We apply (2.9), then Lemmata 11 and 12 and then (2.6) to the first term on the right-hand side and obtain

$$\begin{aligned} H(\mu_\lambda;\lambda^{2^{k+1}}|\lambda^{2^k}) &\ge H(\mu_\lambda;\lambda^{2^{k+1}}|r) \ge H(\mu_\lambda^{(0,\lambda^{2^k}]};\lambda^{2^{k+1}}|r) \\ &\ge H(\mu_\lambda^{(0,\lambda^{2^k}]};\lambda^{2^{k+1}}|\lambda^{2^k}) - 2 = H(\mu_\lambda;\lambda^{2^k}|1) - 2. \end{aligned}$$

Combining our estimates, we find

$$f_{k+1}(\lambda) = \frac{H(\mu_\lambda;\lambda^{2^{k+1}}|1)-2}{-2^{k+1}\log\lambda} \ge \frac{2(H(\mu_\lambda;\lambda^{2^k}|1)-2)}{-2^{k+1}\log\lambda} = f_k(\lambda),$$

as required. □

3. Initial bounds on entropy using Diophantine considerations.

The purpose of this section is to prove the following two results, which provide the initial lower bounds on the entropy of Bernoulli convolutions that we will bootstrap in the next section.

THEOREM 17. *For every $\varepsilon > 0$, there is a number $c > 0$ such that the following holds for all n large enough (depending only on ε). Let $0 < r < n^{-3n}$ and $0 < \lambda \le 1-\varepsilon$ be numbers. Suppose $H(\mu_\lambda^{(\lambda^n,1]};r) < n$.*

Then there is an algebraic number η that is a root of a polynomial in $\mathcal{P}_n$ such that

$$|\eta - \lambda| < r^c$$

and

$$h_\eta \le \frac{H(\mu_\lambda^{(\lambda^n,1]};r)}{n}.$$

Recall that

$$h_\eta = \lim_{n\to\infty} \frac{H(\mu_\eta^{(\eta^n,1]})}{n} = \inf \frac{H(\mu_\eta^{(\eta^n,1]})}{n}.$$

THEOREM 18. *For every $\varepsilon > 0$, there is a number $c > 0$ such that the following holds for all n large enough (depending only on ε). Let $0 < \lambda \le 1 - \varepsilon$ be a number. Suppose that there is an algebraic number η that is a root of a polynomial in $\mathcal{P}_n$ and $|\lambda - \eta| < n^{-4n}$.*

Then

$$H(\mu_\lambda^{(\lambda^n,1]}; r) = n$$

for all $r \le |\lambda - \eta|^{1/c}$.

REMARK 19. We note that the constant c in both theorems can be taken independent of ε, and in fact, arbitrarily close to 1, provided we assume $0 < r < n^{-Bn}$ in Theorem 17 and $|\lambda - \eta| < n^{-Bn}$ in Theorem 18 for some suitably large B depending on ε.

We outline the main idea behind the proofs of these theorems. If $H(\mu_\lambda^{(\lambda^n,1]}; r)$ is "small", then there are "many" choices of signs $a_i, b_i \in \{-1, 1\}$ such that

$$\frac{1}{2}\big|(a_0\lambda^0 + \cdots + a_{n-1}\lambda^{n-1}) - (b_0\lambda^0 + \cdots + b_{n-1}\lambda^{n-1})\big| < r/2.$$

Observe that the expression on the left hand side is (the absolute value of) a polynomial in λ of degree at most $n - 1$ with coefficients in $\{-1, 0, 1\}$.

In the next proposition, we consider a collection of such polynomials that take "small" values at λ and conclude that they have a common zero η near λ. To prove Theorem 17, we will use this to estimate the Shannon entropy of $\mu_\eta^{(\eta^n,1]}$ and conclude that h_η is small.

PROPOSITION 20. *For every $\varepsilon > 0$, there is a number $c > 0$ such that the following holds for all n large enough (depending only on ε). Let $A \subset \mathcal{P}_n$ be a set of polynomials and let $0 < r < n^{-3n}$ and $\lambda \in \mathbb{C}$ be numbers. Suppose $\varepsilon < |\lambda| < 1 - \varepsilon$ and $|P(\lambda)| \le r$ for all $P \in A$.*

Then there is a number $\eta \in \mathbb{C}$ such that $P(\eta) = 0$ for all $P \in A$ and

$$|\eta - \lambda| \le r^c.$$

This proposition will be proved using a Bézout identity expressing the greatest common divisor D of the elements of A as

$$D = Q_1 P_1 + \cdots + Q_m P_m,$$

where $P_i \in A$ and $Q_i \in \mathbb{Q}[x]$ whose degree and coefficients are controlled. We will then argue that D must be "small" at λ, hence it must have a zero near λ.

To deduce Theorem 18, we will exploit the fact that the roots of the polynomials in $\mathcal{P}_n$ repel each other. If λ can be approximated by a root η of a polynomial in $\mathcal{P}_n$ with "very small" error, then this approximation is unique. If we set the scale r smaller than $|\lambda - \eta|^{1/c}$ with the constant c from Theorem 17, then that theorem implies the claim.

The result that we use about the separation between roots of polynomials in $\mathcal{P}_n$ is the following one due to Mahler.

THEOREM 21 (Mahler). *Let $n \geq 9$. Let $\eta \neq \eta'$ be two algebraic numbers each of which is a root of a polynomial in $\mathcal{P}_n$. Then $|\eta - \eta'| > 2n^{-4n}$.*

PROOF. Let $P \in \mathbb{Z}[X]$ of degree d. By Mahler's result [19], Theorem 2, it follows that the distance between any two distinct roots of P is at least

$$\sqrt{3}d^{-(d+2)/2}M(P)^{-(d-1)},$$

where $M(P)$ is the Mahler measure of P.

If η and η' are Galois conjugates, then we take P to be their minimal polynomial. If they are not Galois conjugates, then we take P to be the product of their minimal polynomials.

In either case, the degree of P is at most $2n$, and its Mahler measure is at most the product of the Mahler measures of the polynomials in $\mathcal{P}_n$ whose roots η and η' are. By [2], Lemma 1.6.7, we have $M(P) \leq n + 1$. Therefore, we have

$$|\eta - \eta'| \geq \sqrt{3}(2n)^{-n-1}(n+1)^{-2n+1} > (2n)^{-3n} > 2n^{-4n},$$

provided $n \geq 9$. $\square$

Finally, we note that Theorem 21 offers an alternative way to prove a weaker version of Proposition 20. Indeed, one can argue that any $P \in A$ must have a zero near λ, because $P(\lambda)$ is "small". Then one may use Theorem 21 to conclude that these zeros must coincide.

However, our argument based on the Bézout identity has the advantage that it gives a similar result (with weaker approximation) even without the hypothesis $|\lambda| < 1 - \varepsilon$. We formulate this below in Proposition 25. Although that result is not required for the proof of Theorem 1, we find it of independent interest.

In addition, our approach based on the Bézout identity could be used to give an alternative proof of Theorem 21 with a worse constant, but we do not pursue this here.

The rest of this section is organized as follows. We formulate and prove the Bézout identity in Section 3.1. Section 3.2 is devoted to the proof of Proposition 20. Finally, we prove Theorems 17 and 18 in Section 3.3.

3.1. *The Bézout identity.* The purpose of this section is to prove the following result.

PROPOSITION 22. *Let $A \subset \mathcal{P}_n$ be a set of polynomials and let D be their greatest common divisor in $\mathbb{Z}[x]$. Then there is a number $m \le n+1$ and polynomials $P_1, \dots, P_m \in A$ and $Q_1, \dots, Q_m \in \mathbb{Q}[x]$ such that*

$$D = \sum_{j=1}^{m} Q_j P_j$$

and

$$\deg(Q_j) \le n-1, \qquad h(Q_j) \le 2^n (2n)!$$

for all j.

Here and everywhere below, $h(Q)$ denotes the naive height, the maximum of the numerators and denominators of the coefficients of Q. We begin with some preliminary observations.

LEMMA 23. *Let $D \in \mathbb{Z}[x]$ be a polynomial that divides a polynomial $P \in \mathcal{P}_n$ for some $n \in \mathbb{Z}_{\ge 1}$. Then $l^1(D) \le 2^n n$.*

Here, and everywhere below, $l^p(D)$ denotes the l^p norm of the vector formed from the coefficients of D.

PROOF. By [2], Lemma 1.6.7, we have

$$M(D) \le M(P) \le \big(\deg(P)+1\big)^{1/2} h(P) \le (n+1)^{1/2} \le n.$$

We also have (see [18], Equation (4))

$$l^1(D) \le 2^n M(D) \le 2^n n,$$

which was to be proved. □

LEMMA 24. *Let $n \in \mathbb{Z}_{>0}$ and let $v_1, \dots, v_N \in \{-1, 0, 1\}^n$ be vectors. Suppose that $w \in \mathbb{Z}^n$ is in the $\mathbb{Q}$-span of $v_1, \dots, v_N$. Then there are rational numbers $\lambda_1, \dots, \lambda_N$ such that*

$$w = \lambda_1 v_1 + \cdots + \lambda_N v_N, \tag{3.1}$$

at most n of the λ_i are nonzero and their numerators and denominators are bounded in absolute value by $\max(n!, l^1(w)(n-1)!)$.

PROOF. We select a nonzero minor of maximal rank from the matrix $[v_1, \dots, v_N]$ and then solve the equation using Cramer's rule.

The rank is at most n, hence the number of nonzero λ_i is indeed at most n. The nonzero λ_i are the ratio of two determinants of rank at most n. In the denominator all entries come from the entries of v_i, hence they are -1, 0 or 1. This determinant is clearly bounded by $n!$.

The entries of the numerator are similarly -1, 0 or 1 except for one column whose entries come from w. Expanding the determinant in that column we obtain the bound $l^1(w)(n-1)!$. □

PROOF OF PROPOSITION 22. By the Nullstellensatz or simply by the Euclidean algorithm, there are polynomials $P_1, \dots, P_m \in A$ and $Q_1, \dots, Q_m \in \mathbb{Q}[x]$ such that

$$D = \sum_{j=1}^{m} Q_j P_j. \tag{3.2}$$

We may assume that the polynomials P_j are linearly independent. Indeed, we could achieve this situation by expressing some of the polynomials P_j that appear in (3.2) by linear combinations of others. This yields $m \le n+1$.

We may also assume that $\deg(Q_j) < \deg(P_m) \le n$ for all $j < m$. Indeed, if this was false for some $j < m$, we can write $Q_j = Q'_j P_m + R_j$ and replace Q_j by R_j and Q_m by $Q_m + Q'_j P_j$. This substitution does not change the value of (3.2), since

$$\begin{aligned} R_j P_j + (Q_m + Q'_j P_j) P_m &= (Q_j - Q'_j P_m) P_j + (Q_m + Q'_j P_j) P_m \\ &= Q_j P_j + Q_m P_m. \end{aligned}$$

These substitutions can be executed simultaneously without affecting each other.

We observe that

$$\begin{aligned} \deg(Q_m P_m) &\le \max\bigl(\deg(D), \deg(Q_1 P_1), \dots, \deg(Q_{m-1} P_{m-1})\bigr) \\ &< \deg(P_m) + n, \end{aligned}$$

which in turn gives $\deg(Q_m) < n$.

We write

$$P_j = p_{j,n} x^n + \cdots + p_{j,0}, \qquad D = d_n x^n + \cdots + d_0,$$

where we allow $p_{j,n} = 0$ and $d_n = 0$. We consider the vectors

$$v_{j,k} = (\underbrace{0, \dots, 0}_{k}, p_{j,0}, \dots, p_{j,n}, \underbrace{0, \dots, 0}_{n-1-k}) \in \{-1, 0, 1\}^{2n}$$

for $j = 1, \dots, m$ and $k = 0, \dots, n-1$ and

$$w = (d_0, \dots, d_n, 0, \dots, 0) \in \mathbb{Z}^{2n}.$$

By (3.2), w is in the $\mathbb{Q}$-span of the vectors $v_{j,k}$. By Lemma 23, we have $l^1(w) \le 2^n n$. We apply Lemma 24 to find rational numbers $\lambda_{j,k}$ with numerators and denominators bounded by $n2^n(2n-1)!$ such that

$$w = \sum \lambda_{j,k} v_{j,k}.$$

We conclude the proof by replacing Q_j by $\sum \lambda_{j,k} x^k$. $\square$

3.2. *Small polynomials have a common root.* The purpose of this section is to prove Proposition 20 and its following variant.

PROPOSITION 25. *Let $n \in \mathbb{Z}$ be sufficiently large (larger than an absolute constant), let $A \subset \mathcal{P}_n$ be a set of polynomials and let $0 < r < (2n)^{-2n}$ and $\lambda \in \mathbb{C}$ be numbers. Suppose $|P(\lambda)| \le r$ for all $P \in A$.*

Then there is a number $\eta \in \mathbb{C}$ such that $P(\eta) = 0$ for all $P \in A$ and

$$|\eta - \lambda| < r^{1/n}(2n)^2.$$

We give a bound on the number of roots a polynomial in $\mathcal{P}_n$ may have away from the unit circle using Jensen's formula. This will be used in the proof of Proposition 20 to show that such a polynomial can take very small values only near its roots.

LEMMA 26. *There is a function $a(k) : \mathbb{Z}_{>0} \to (0, 1)$ such that $\lim_{k\to\infty} a(k) = 1$ and the following holds. Let $P \in \mathcal{P}_n$ be a nonzero polynomial for some $n \in \mathbb{Z}_{\ge 0}$. Then there are at most k nonzero roots of P of absolute value less than $a(k)$.*

This result is not new; see, for example, [1].

PROOF OF LEMMA 26. Without loss of generality, we may assume that $|P(0)| = 1$. Indeed, we may divide P by an appropriate power of x to obtain a new polynomial that has this property. We prove the lemma taking

$$a(k) = \frac{k}{k+1} \cdot \frac{1}{(k+1)^{1/k}}.$$

We denote by $z_1, \dots, z_K$ the roots of P of absolute value less than $a(k)$. We set $r = k/(k+1)$ and apply Jensen's formula on the disk of radius r:

$$\sum_{j=1}^{K} \log \frac{|r|}{|z_j|} \le \int_0^1 \log\bigl|P\bigl(re^{2\pi i t}\bigr)\bigr|\,dt.$$

We note that

$$\bigl|P(z)\bigr| \le 1 + |z| + |z|^2 + \cdots = \frac{1}{1-r} = k+1$$

for all z with $|z| = r$.

Thus

$$K \cdot \log\big((k+1)^{1/k}\big) \le \log(k+1),$$

which yields $K \le k$, as claimed. □

PROOF OF PROPOSITIONS 25 AND 20. We begin with Proposition 25. We denote by D the greatest common divisor of the polynomials in A. Note that the hypothesis (when A is nonempty) implies that $|\lambda| \le 2$. We use Proposition 22 and the fact that $|P(\lambda)| \le r$ for all $P \in A$. We get

$$\big|D(\lambda)\big| \le (n+1)2^{2n+1}(2n)! \cdot r < (2n)^{2n} \cdot r. \tag{3.3}$$

Since $|D(\lambda)| < 1$ and D has integer coefficients, D is not constant. We denote by $\eta_1, \dots, \eta_d$ the roots of D taking multiplicities into account. Then

$$\big|D(\lambda)\big| = \prod_{j=1}^{d} |\eta_j - \lambda|,$$

hence there is some j such that

$$|\eta_j - \lambda| \le \big|D(\lambda)\big|^{1/d} < r^{1/n}(2n)^2,$$

as claimed.

To prove Proposition 20, we apply Lemma 26 and find that there is a number k depending only on ε such that any polynomial in $\mathcal{P}_n$ has at most k nonzero roots of modulus at most $1 - \varepsilon/2$. Since D divides such a polynomial, the same bound holds for its roots.

We denote by $\eta_1, \dots, \eta_l$ the nonzero roots of D of modulus at most $1 - \varepsilon/2$. Then $l \le k$ and

$$\big|D(\lambda)\big| \ge (\varepsilon/2)^{\deg D - l} \prod_{j=1}^{l} |\eta_j - \lambda|.$$

Thus, there is some j such that

$$|\eta_j - \lambda|^l \le \big|D(\lambda)\big| \cdot (\varepsilon/2)^{-n}.$$

Since $r < n^{-3n}$, we have from (3.3),

$$\big|D(\lambda)\big| \cdot (\varepsilon/2)^{-n} \le (4n/\varepsilon)^{2n} \cdot r < r^{1/10},$$

if n is large enough. Hence

$$|\eta_j - \lambda| < r^{1/(10l)},$$

as required. □

REMARK 27. The constant c in Proposition 20 can be taken arbitrarily close to 1 if $r < n^{-Bn}$ for B suitably large.

Indeed, in the setting of the above proof, denote by η_j a root of D of minimal distance to λ among $\eta_1, \dots, \eta_l$. By Theorem 21, there is at most one root at distance at most n^{-4n} from λ, hence $|\lambda - \eta_i| \geq n^{-4n}$ for all $i \neq j$. From this, we obtain

$$|\lambda - \eta_j| \leq n^{4kn} (\varepsilon/2)^{-n} D(\lambda) \leq r^c,$$

where c can indeed be taken arbitrarily close to 1, provided r is as small as we assumed above.

We will see in the next section that the constant c in Theorems 17 and 18 are the same as in Proposition 20, hence this justifies the claims made in Remark 19.

3.3. *Completing the proofs.*

PROOF OF THEOREM 17. Let $\xi_0, \dots, \xi_{n-1}$ be a sequence of independent unbiased ± 1-valued random variables. Let $t \in \mathbb{R}$ be such that

$$H(\mu_\lambda^{(\lambda^n,1]}; r) \geq H\left(\left\lfloor r^{-1} \sum_{j=0}^{n-1} \xi_j \lambda^j + t \right\rfloor \right).$$

For each $a \in \mathbb{Z}$ let

$$\Omega_a = \left\{ (\omega_0, \dots, \omega_{n-1}) \in \{-1, 1\}^n : \left\lfloor r^{-1} \sum_{j=0}^{n-1} \omega_j \lambda^j + t \right\rfloor = a \right\}.$$

We note the identity

$$H\left(\left\lfloor r^{-1} \sum_{j=0}^{n-1} \xi_j \lambda^j + t \right\rfloor \right) = \sum_{a \in \mathbb{Z}} \frac{|\Omega_a|}{2^n} \log\left(\frac{2^n}{|\Omega_a|} \right).$$

In particular $|\Omega_a| > 1$ for at least one $a \in \mathbb{Z}$, because $H(\mu_\lambda^{(\lambda^n,1]}) < n$.

We consider the set of polynomials

$$A = \bigcup_{a \in \mathbb{Z}} \left\{ \sum_{j=0}^{n-1} \frac{\omega_j - \omega'_j}{2} x^j : \omega \neq \omega' \in \Omega_a \right\}.$$

Since $|\Omega_a| > 1$ for at least one $a \in \mathbb{Z}$, A is not empty. We observe that $P \in \mathcal{P}_n$ and $|P(\lambda)| \leq r$ for each $P \in A$. We apply Proposition 20 and find $\eta \in \mathbb{C}$ such that $|\eta - \lambda| \leq r^c$ and $P(\eta) = 0$ for all $P \in A$.

For any $a \in \mathbb{Z}$ and $\omega, \omega' \in \Omega_a$, we have

$$\sum_{j=0}^{n-1} \frac{\omega_j - \omega'_j}{2} \eta^j = 0,$$

hence

$$\sum_{j=0}^{n-1} \omega_j \eta^j = \sum_{j=0}^{n-1} \omega'_j \eta^j.$$

Thus,

$$H\left(\sum_{j=0}^{n-1} \xi_j \eta^j\right) \le \sum_{a \in \mathbb{Z}} \frac{|\Omega_a|}{2^n} \log\left(\frac{2^n}{|\Omega_a|}\right).$$

We combine our inequalities to obtain

$$H(\mu_\lambda^{(\lambda^n,1]}; r) \ge H\left(\sum_{j=0}^{n-1} \xi_j \eta^j\right) \ge n h_\eta.$$

Recall $h_n = \inf H(\mu_\lambda^{(\lambda^n,1]})/n$. □

PROOF OF THEOREM 18. The constant c is the same as in Theorem 17. We suppose to the contrary that $H(\mu_\lambda^{(\lambda^n,1]}; r) < n$ and apply that theorem.

We find an algebraic number η', which is a root of a polynomial in $\mathcal{P}_n$ such that $|\lambda - \eta'| < r^c \le |\lambda - \eta|$. In particular, $\eta' \ne \eta$. Moreover,

$$|\eta - \eta'| \le |\lambda - \eta'| + |\lambda - \eta| < 2n^{-4n},$$

which contradicts Theorem 21. □

4. Increasing entropy of convolutions. In this section, we apply Theorem 8 to improve on the entropy estimates that we obtained in the previous section. We begin with two preliminary results in the next two sections and conclude the proof of Theorem 1 in Section 4.3

4.1. *First stage of entropy increment.* The purpose of this section is the following proposition.

PROPOSITION 28. *Let* $1/2 < \lambda < 1$ *and* $n, K, \alpha > 0$ *be numbers, with* $K \ge 10$. *Suppose*

$$H(\mu_\lambda; r|2r) \le 1 - \alpha \qquad \text{for all } r > 0, \tag{4.1}$$

$$H(\mu_\lambda^{(\lambda^n,1]}; \lambda^{Kn}|\lambda^{10n}) \ge \alpha n. \tag{4.2}$$

Suppose further $n > C_0(\log K)^2$, *where* C_0 *is a suitably large number depending only on* α *and* λ.

Then, there are numbers $R_1, \dots, R_k$ *and* $a_1, \dots, a_k$ *such that*

$$\lambda^{-9n} \le R_i \le \lambda^{-Kn}, \qquad R_{i+1} \ge R_i^2, \qquad a_i \ge \frac{c}{\log K},$$

$$\sum_{i=1}^{k} a_i \ge c, \qquad H\big(\mu_\lambda^{(R_i r, R_i^2 r]}; r|Ar\big) \ge \frac{c a_i}{\log K} \log A$$

for each $i = 1, \dots, k$, *for any* $r \le \lambda^{2Kn}$ *and for any* $\max(\lambda^{-2}, 2) \le A \le \lambda^{-n}$, *where* c *is a constant that depends only on* α *and* λ.

In the proof of Theorem 1, we fix a parameter λ such that $\dim \mu_\lambda < 1$. By Lemma 13, this implies that (4.1) holds at all scales. Furthermore, we will show that (4.2) also holds for the appropriate choice of n and K. To this end, we will use the results of Section 3. In Section 4.2, we refine the conclusion of this proposition by further applications of Theorem 8.

We begin the proof of the proposition with a technical lemma. If we have a bound for the entropy of $\mu^{(a,b]}$ between some scales, then we can use the scaling identity (2.6) to obtain bounds for $\mu^{(ar,br]}$ between some other scales. We take this idea a step further in the next lemma, which will be used in the proof of Proposition 28 to construct measures, to which we can apply Theorem 8.

LEMMA 29. *Let* $a_1, a_2, b_1, b_2, r_1, r_2, s_1, s_2$ *be numbers such that the following holds*

$$0 \le a_i < b_i \le 1, \qquad 0 \le r_i < s_i \qquad \text{for } i = 1, 2,$$

$$\lambda^{-1} s_1/s_2 \le a_1/a_2,$$

$$\lambda r_1/r_2 \ge b_1/b_2,$$

$$\max(2, \lambda^{-2}) \le s_2/r_2 \le s_1/r_1.$$

Suppose

$$H\big(\mu_\lambda^{(a_1,b_1]}; r_1|s_1\big) \ge \beta \log(s_1/r_1)$$

for some $\beta \ge 0$. *Then*

$$H\big(\mu_\lambda^{(a_2,b_2]}; r_2|s_2\big) \ge \frac{\beta}{6} \log(s_2/r_2).$$

We comment on the inequalities imposed in the lemma, which may look unmotivated on first reading. They are designed to ensure that for any scaling factor t, the inclusion of scales $t[r_2, s_2] \subset [\lambda r_1, \lambda^{-1} s_1]$ implies $(a_1, b_1] \subset t(a_2, b_2]$.

PROOF. We choose a sequence of integers $k_1 > \cdots > k_N$ such that the intervals $[\lambda^{k_i} r_2, \lambda^{k_i} s_2]$ cover $[r_1, s_1]$, that is, we have

$$\lambda r_1 \le \lambda^{k_1} r_2 \le r_1, \qquad s_1 \le \lambda^{k_N} s_2 \le \lambda^{-1} s_1,$$

and

$$\lambda^{k_i} s_2 \geq \lambda^{k_{i+1}} r_2$$

holds for all $1 \leq i \leq N-1$.

We may choose the sequence in such a way that the overlaps between the intervals $[\lambda^{ki} r_2, \lambda^{ki} s_2]$ are minimal, so that

$$\lambda^{k_{i+1}} r_2 \geq \lambda^{k_i+1} s_2$$

for all $i \leq N-2$. If this is the case, we have

$$\lambda^{k_{i+1}} s_2 \geq \lambda^{k_i} s_2 \cdot (\lambda s_2/r_2)$$

and then

$$\lambda^{k_i} s_2 \geq \lambda^i r_1 \cdot (s_2/r_2)^i \geq r_1 \cdot (s_2/r_2)^{i/2}$$

follows for $1 \leq i \leq N-1$ by induction. Clearly, we may assume $\lambda^{k_{N-1}} s_2 < s_1$, since otherwise we would not need the interval $[\lambda^{k_N} r_2, \lambda^{k_N} s_2]$ to cover $[r_2, s_2]$. Hence, we may assume that

$$N \leq 2\frac{\log(s_1/r_1)}{\log(s_2/r_2)} + 1 \leq 3\frac{\log(s_1/r_1)}{\log(s_2/r_2)}.$$

Using (2.9), we write

$$\begin{aligned}
&\sum_{i=1}^{N} H\bigl(\mu_\lambda^{(a_1,b_1]}; \lambda^{k_i} r_2 | \lambda^{k_i} s_2\bigr) \\
&\quad \geq H\bigl(\mu_\lambda^{(a_1,b_1]}; r_1 | \lambda^{k_2} r_2\bigr) + \sum_{i=2}^{N-1} H\bigl(\mu_\lambda^{(a_1,b_1]}; \lambda^{k_i} r_2 | \lambda^{k_{i+1}} r_2\bigr) \\
&\qquad + H\bigl(\mu_\lambda^{(a_1,b_1]}; \lambda^{k_N} r_2 | s_1\bigr) \\
&\quad = H\bigl(\mu_\lambda^{(a_1,b_1]}; r_1 | s_1\bigr) \geq \beta \log(s_1/r_1).
\end{aligned}$$

Thus there is some i such that

$$H\bigl(\mu_\lambda^{(a_1,b_1]}; \lambda^{k_i} r_2 | \lambda^{k_i} s_2\bigr) \geq \frac{\beta}{N} \log(s_1/r_1) \geq \frac{\beta}{3} \log(s_2/r_2).$$

Using $\lambda^{k_i} r_2 \geq \lambda r_1$, $\lambda^{k_i} s_2 \leq \lambda^{-1} s_1$ and the assumptions in the statement of the lemma, we have

$$b_1/b_2 \leq \lambda r_1/r_2 \leq \lambda^{k_i} \leq \lambda^{-1} s_1/s_2 \leq a_1/a_2,$$

hence

$$\bigl(\lambda^{k_i} a_2, \lambda^{k_i} b_2\bigr] \supset (a_1, b_1].$$

Therefore, we can use (2.6) and (2.8) and write

$$H\big(\mu_\lambda^{(a_2,b_2]}; r_2|s_2\big) = H\big(\mu_\lambda^{(\lambda^{k_i}a_2,\lambda^{k_i}b_2]}; \lambda^{k_i}r_2|\lambda^{k_i}s_2\big)$$
$$\geq \frac{1}{2}H\big(\mu_\lambda^{(a_1,b_1]}; \lambda^{k_i}r_2|\lambda^{k_i}s_2\big) \geq \frac{\beta}{6}\log(s_2/r_2). \qquad \square$$

PROOF OF PROPOSITION 28. Write

$$J = \left\lceil \frac{\log(K/10)}{\log(11/10)} \right\rceil.$$

Then $\lambda^{Kn} > \lambda^{11n(11/10)^{J-1}}$, hence

$$H\big(\mu_\lambda^{(\lambda^n,1]}; \lambda^{11(11/10)^{J-1}n}|\lambda^{10n}\big) \geq \alpha n.$$

For each integer $0 \leq j < J$ define b_j by

$$H\big(\mu_\lambda^{(\lambda^n,1]}; \lambda^{11(11/10)^j n}|\lambda^{10(11/10)^j n}\big) = b_j n. \tag{4.3}$$

Then $b_0 + \cdots + b_{J-1} \geq \alpha$.

We fix a $j \in \{0, \dots, J-1\}$. Put $T = \lfloor (11/10)^j \rfloor$. We note the identity

$$\mu_\lambda^{(\lambda^{Tn},1]} = \mu_\lambda^{(\lambda^{Tn},\lambda^{(T-1)n}]} * \mu_\lambda^{(\lambda^{(T-1)n},\lambda^{(T-2)n}]} * \cdots * \mu_\lambda^{(\lambda^n,1]},$$

and set out to apply Theorem 8 and find a lower bound on the entropy of $\mu_\lambda^{(\lambda^{Tn},1]}$ between suitably chosen scales.

Scaling (4.3) (see (2.6)), we can write

$$H\big(\mu_\lambda^{(\lambda^{(t+1)n},\lambda^{tn}]}; \lambda^{(11(11/10)^j+t)n}|\lambda^{(10(11/10)^j+t)n}\big) = b_j n.$$

For $t = 0, \dots, T-1$, we have

$$\big[\lambda^{(11(11/10)^j+t)n}, \lambda^{(10(11/10)^j+t)n}\big] \subseteq \big[\lambda^{12(11/10)^j n}, \lambda^{10(11/10)^j n}\big],$$

hence we can use (2.9) to get

$$H\big(\mu_\lambda^{(\lambda^{(t+1)n},\lambda^{tn}]}; \lambda^{12(11/10)^j n}|\lambda^{10(11/10)^j n}\big) \geq b_j n.$$

We can now apply Theorem 8 repeatedly T times for $t = 0, \dots, T-1$ with

$$\mu = \mu_\lambda^{(\lambda^{tn},\lambda^{(t-1)n}]} * \cdots * \mu_\lambda^{(\lambda^n,1]} \quad \text{or} \quad \mu = \delta_0 \qquad \text{if } t = 0,$$
$$\nu = \mu_\lambda^{(\lambda^{(t+1)n},\lambda^{tn}]},$$
$$\beta = \frac{b_j}{2(11/10)^j \log\lambda^{-1}} \geq \frac{cb_j}{T},$$

where $c > 0$ is a constant depending only on λ. We obtain

$$H\big(\mu_\lambda^{(\lambda^{Tn},1]}; \lambda^{12(11/10)^j n}|\lambda^{10(11/10)^j n}\big) \geq T\left(\frac{cb_j}{T\log(T/b_j)}\log\big(\lambda^{-2(11/10)^j n}\big) - C\right).$$

Assume that j is such that $b_j \geq \alpha/2J$. Since $j \leq J-1$, the definitions of T and J yield $T \leq K$ and $b_j \geq c/\log K$ for some other constant depending only on α and λ, which we keep denoting by c by abuse of notation. Since we assumed $n > C_0(\log K)^2$ for any fixed number C_0 depending on α and λ, the term TC becomes negligible. Thus,

$$H\big(\mu_\lambda^{(\lambda^{Tn},1]}; \lambda^{12(11/10)^j n} | \lambda^{10(11/10)^j n}\big) \geq c \frac{b_j}{\log K} \log\big(\lambda^{-2(11/10)^j n}\big). \tag{4.4}$$

Now we combine (4.4) with Lemma 29. To that end, we need to choose a number Q_j in such a way that the following inequalities are satisfied:

$$\lambda^{-1} \frac{\lambda^{10(11/10)^j n}}{Ar} \leq \frac{\lambda^{Tn}}{Q_j r}, \tag{4.5}$$

$$\lambda \frac{\lambda^{12(11/10)^j n}}{r} \geq \frac{1}{Q_j^2 r}, \tag{4.6}$$

$$A \leq \lambda^{-2(11/10)^j n}. \tag{4.7}$$

Since (4.7) always holds when $A \leq \lambda^{-n}$ (which we assumed in the statement of the proposition), we need to consider only the first two conditions. We observe that the first condition is the most restrictive when A is as small as possible, hence we may assume $A = \lambda^{-2}$. Recall $T \leq (11/10)^j$. Hence, (4.5) and (4.6) hold if we choose Q_j to satisfy

$$Q_j \leq \lambda^{-9(11/10)^j n - 1},$$
$$Q_j^2 \geq \lambda^{-12(11/10)^j n - 1}.$$

So we can put $Q_j = \lambda^{-9(11/10)^j n}$ and satisfy these inequalities.

We can now apply Lemma 29 to (4.4) with

$$r_1 = \lambda^{12(11/10)^j n}, \qquad s_1 = \lambda^{10(11/10)^j n}, \qquad a_1 = \lambda^{Tn}, \qquad b_1 = 1,$$
$$r_2 = r, \qquad s_2 = Ar, \qquad a_2 = Q_j r, \qquad b_2 = Q_j^2 r$$

and write

$$H\big(\mu_\lambda^{(Q_j r, Q_j^2 r]}; r | Ar\big) \geq c \frac{b_j}{6 \log K} \log A.$$

We note $1 \leq (11/10)^j \leq K/10$. Hence,

$$\lambda^{-9n} \leq Q_j \leq \lambda^{-9Kn/10} < \lambda^{-Kn}.$$

Finally, we define $a_i = b_{j_i}$ and $R_i = Q_{j_i}$ for a suitably chosen sequence j_i. We first select those j such that $b_j > \alpha/(2J)$. This ensures that the above argument applies to all selected j and that $b_j \geq c/\log K$. We still have $\sum' b_j \geq \alpha/2$, where $\sum'$ indicates summation over those j that we selected. Second, we select j's from an arithmetic progression with common difference 10 such that the sum of the selected b_j's are maximal among the possible choices. Then we still have $\sum'' b_j \geq \alpha/20$, where $\sum''$ indicates summation over those j that we selected during the second cut. Moreover, this choice ensures that $Q_{j'} \geq Q_j^{(11/10)^{10}} > Q_j^2$ if $j' > j$ are two selected indices. Therefore, this subsequence satisfies all the requirements of the proposition. $\square$

4.2. *Second stage of entropy increment.* In the proof of Theorem 1, we will choose sequences of suitable parameters $\{n_j\}$ and $\{K_j\}$ such that the conditions of Proposition 28 hold. In this section, we consider such sequences and apply Theorem 8 again together with the conclusion of Proposition 28 to obtain even stronger entropy bounds. Since the entropy between the scales r and Ar cannot be larger than $\log A$, this will lead to a constraint showing that the sequence K_j has to grow very fast. In the proof of Theorem 1, this will lead to a contradiction with the hypothesis of that theorem.

PROPOSITION 30. *Let $1/2 < \lambda < 1$, $\alpha > 0$ be numbers, let $\{n_j\}_{j=1}^N$ be a sequence of positive integers, and let $\{K_j\}_{j=1}^N$ be a sequence of real numbers each ≥ 10.*

Suppose

$$n_{j+1} \geq K_j n_j \qquad \text{for all } j = 1, \dots, N-1, \tag{4.8}$$

$$H(\mu_\lambda; r|2r) \leq 1 - \alpha \qquad \text{for all } r > 0, \tag{4.9}$$

$$H\big(\mu_\lambda^{(\lambda^{n_j},1]}; \lambda^{K_j n_j} | \lambda^{10 n_j}\big) \geq \alpha n_j \qquad \text{for all } j = 1, \dots, N, \tag{4.10}$$

$$n_j \geq C_0 (\log K_j)^2, \tag{4.11}$$

where C_0 is a sufficiently large number depending only on λ and α. Suppose further that n_1 is sufficiently large so that $\lambda^{-n_1} \geq \max(2, \lambda^{-2})$.

Then

$$\sum_{j=1}^{N} \frac{1}{\log K_j \log\log K_j} < C\left(1 + \frac{\sum_{j=1}^N \log K_j}{n_1}\right), \tag{4.12}$$

where C is a constant that depends only on α and λ.

PROOF. Set $A = \lambda^{-n_1}$ and $r = \lambda^{2K_N n_N}$, so $A \geq \max(2, \lambda^{-2})$. We apply Proposition 28 with $n = n_j$ and $K = K_j$. We find numbers $R_{j,i} \in [\lambda^{-9n_j}, \lambda^{-K_j n_j}]$ and $a_{j,i} \geq c(\log K_j)^{-1}$ such that $\sum_i a_{j,i} \geq c$ for each j and

$$H\big(\mu_\lambda^{(R_{j,i} r, R_{j,i}^2 r]}; r|Ar\big) \geq \frac{c a_{j,i}}{\log K_j} \log A$$

for each j and i.

We observe that

$$R_{j+1,\bullet} \geq \lambda^{-9n_{j+1}} \geq \lambda^{-9K_j n_j} > R_{j,\bullet}^2$$

for $j = 1, \dots, N-1$. We also recall $R_{j,i+1} \geq R_{j,i}^2$ from Proposition 28. Thus, the intervals $(R_{j,i}r, R_{j,i}^2 r]$ are disjoint.

This means that we can write

$$\mu_\lambda = \nu * \mathop{\ast}_{j=1}^{N} \mathop{\ast}_{i} \mu_\lambda^{(R_{j,i}r, R_{j,i}^2 r]},$$

for some probability measure ν. We can then apply Theorem 8 repeatedly with

$$\nu = \mu_\lambda^{(R_{j,i}r, R_{j,i}^2 r]} \quad \text{and} \quad \beta = \frac{ca_{j,i}}{\log K_j}$$

for each j and i. Note $\log \beta^{-1} \leq C \log\log K_j$, since $a_{j,i} \geq c/(\log K_j)$, where C is a constant that depends only on λ and α. We obtain

$$H(\mu_\lambda; r|Ar) \geq \sum_{j=1}^{N} \sum_i \left(\frac{ca_{j,i}}{\log K_j \log\log K_j} \log A - C \right),$$

where $c, C > 0$ are some numbers that depend only on λ and α.

Since $\sum_i a_{j,i} \geq c$, for each j and the entropy between scales of ratio A cannot be larger than $2 \log A$ (see Lemma 12), we get

$$\sum_{j=1}^{N} \frac{c}{\log K_j \log\log K_j} \log A < 2 \log A + C \sum_{j=1}^{N} \log K_j.$$

This proves the claim upon dividing both sides by $c \log A$, since $\log A = \log(\lambda^{-1}) n_1$. □

4.3. *Proof of Theorem* 1. Let $1/2 < \lambda < 1$ be a number such that $\dim \mu_\lambda < 1$, and fix a small number $\varepsilon > 0$ such that $\dim \mu_\lambda + 4\varepsilon < 1$. We fix a large number A, whose value will be set at the end of the proof depending only on λ and ε. We assume to the contrary that there are arbitrarily large integers n_0 such that

$$|\eta - \lambda| > \exp\big(-\exp\big(\log n \log^{(3)} n\big)\big) \tag{4.13}$$

for all $\eta \in E_{n,\dim \mu_\lambda + 4\varepsilon}$ for all $n \in [n_0, \exp^{(5)}(\log^{(5)}(n_0) + A)]$. We show that this leads to a contradiction provided A is a sufficiently large number depending on λ and ε.

The assumption $\dim \mu_\lambda < 1$ implies that there is $\alpha > 0$ such that

$$H(\mu_\lambda; r|2r) < 1 - \alpha \tag{4.14}$$

for all r; see Lemma 13. In addition, we have

$$H(\mu_\lambda; r) \leq (\dim \mu_\lambda + \varepsilon) \log r^{-1} \tag{4.15}$$

by Lemma 14 for all sufficiently small r (depending on ε and λ). Moreover, (4.14) and (4.15) hold for the measure μ_λ^I in place of μ_λ for any $I \subset (0,1]$. Indeed,

$$H(\mu_\lambda; r) = \lim_{N\to\infty} H(\mu_\lambda; r|Nr),$$
$$H(\mu_\lambda^I; r) = \lim_{N\to\infty} H(\mu_\lambda^I; r|Nr),$$

so $H(\mu_\lambda^I; r) \le H(\mu_\lambda; r)$ follows from (2.7).

It follows from the work of Hochman ([11], Theorem 1.3) that[3]

$$H(\mu_\lambda^{(\lambda^n,1]}; \lambda^{10n}|\lambda^n) < \varepsilon \log(\lambda^{-1})n, \tag{4.16}$$

if n is large enough (depending on ε and λ).

We fix an integer n_0 such that (4.13) holds, and which is sufficiently large; we require, in particular, that (4.16) holds for all $n \ge n_0$, (4.15) holds for all $r < \lambda^{n_0}$ and $\lambda^{-n_0} \ge \max(2, \lambda^{-2})$. We define a sequence of integers $n_1, n_2, \dots, n_N$ by a recursive procedure. Suppose that n_j is already defined for some $j \ge 0$ and we choose the value of n_{j+1} as follows. We take

$$m = \left\lceil \frac{4n_j \log n_j}{c_0 \log \lambda^{-1}} \right\rceil,$$

where c_0 denotes the minimum of the constants c from Theorems 17 and 18 applied with $1-\lambda$ in the role of ε.

We consider two cases. First, suppose

$$H(\mu_\lambda^{(\lambda^m,1]}; m^{-4m/c_0}) \ge m \log(\lambda^{-1})(\dim \mu_\lambda + 3\varepsilon). \tag{4.17}$$

In this case, we have

$$\begin{aligned}
&H(\mu_\lambda^{(\lambda^m,1]}; m^{-4m/c_0}|\lambda^{10m}) \\
&\quad = H(\mu_\lambda^{(\lambda^m,1]}; m^{-4m/c_0}) - H(\mu_\lambda^{(\lambda^m,1]}; \lambda^m) - H(\mu_\lambda^{(\lambda^m,1]}; \lambda^{10m}|\lambda^m) \\
&\quad \ge m \log(\lambda^{-1})(\dim \mu_\lambda + 3\varepsilon) - m \log(\lambda^{-1})(\dim \mu_\lambda + \varepsilon) - \varepsilon m \log(\lambda^{-1}) \\
&\quad \ge \varepsilon m \log(\lambda^{-1}).
\end{aligned}$$

We used (4.17), (4.15) and (4.16). To estimate $H(\mu_\lambda^{(\lambda^m,1]}; \lambda^m)$, we used (4.15). In this case, we set $n_{j+1} = m$.

Second, suppose

$$H(\mu_\lambda^{(\lambda^m,1]}; m^{-4m/c_0}) < m \log(\lambda^{-1})(\dim \mu_\lambda + 3\varepsilon).$$

[3] We could avoid using Hochman's result here if we replaced the number 10 by $1+\varepsilon$. If we do this, then Propositions 28 and 30 and their proofs need to be adjusted accordingly, which would turn the calculations even more tedious.

We apply Theorem 17 and find that there is an algebraic number η that is a root of a polynomial in $\mathcal{P}_m$, $h_\eta < \log(\lambda^{-1})(\dim\mu_\lambda + 3\varepsilon)$ and $|\lambda - \eta| < m^{-4m}$. We assume as we may that n_0 is sufficiently large that this guarantees $h_\eta < \log\eta^{-1}(\dim\mu_\lambda + 4\varepsilon)$. We note $|\eta - \overline{\eta}| < 2m^{-4m}$, hence η is real by Theorem 21. By Hochman's formula (1.5) for the dimension of Bernoulli convolutions for algebraic parameters, we have $\dim\mu_\eta < \dim\mu_\lambda + 4\varepsilon$ and hence $\eta \in E_{m,\dim\mu_\lambda+4\varepsilon}$.

In this case, we set n_{j+1} to be the largest integer n such that $|\lambda - \eta| < n^{-4n}$. In particular, $n_{j+1} \geq m$. It follows from Theorem 18 applied with $n = n_{j+1}$ and $r = (n_{j+1}+1)^{-4(n_{j+1}+1)/c_0}$ that

$$H\big(\mu_\lambda^{(\lambda^{n_{j+1}},1]}; (n_{j+1}+1)^{-4(n_{j+1}+1)/c_0}\big) = n_{j+1}.$$

A calculation similar to what we did in the previous case yields

$$H\big(\mu_\lambda^{(\lambda^{n_{j+1}},1]}; n_{j+1}^{-4n_{j+1}/c_0}|\lambda^{10n_{j+1}}\big) \geq \big(1 - \log\lambda^{-1}\big)n_{j+1},$$

if n_0 is sufficiently large. (Recall $\dim\mu_\lambda + 4\varepsilon < 1$.)

We set

$$K_{j+1} = \frac{4\log(n_{j+1})}{c_0\log\lambda^{-1}}$$

and note that

$$H\big(\mu_\lambda^{(\lambda^{n_{j+1}},1]}; \lambda^{K_{j+1}n_{j+1}}|\lambda^{10n_{j+1}}\big) \geq \varepsilon\log(\lambda^{-1})n_{j+1}$$

holds in both cases (provided $\varepsilon\log\lambda^{-1} < 1 - \log\lambda^{-1}$, which we may assume).

The choice of K_{j+1} and m in the recursive definition ensures that $n_{j+2} \geq K_{j+1}n_{j+1}$. Moreover, $n_{j+1} \geq C_0(\log K_{j+1})^2$ also holds with an arbitrarily large constant C_0, provided n_0 is sufficiently large. This means that Proposition 30 is applicable to the sequences $\{n_j\}$ and $\{K_j\}$. We estimate how fast these sequences may grow. Let m and η be as in the definition of n_{j+1} above. Suppose that

$$m \in \big[n_0, \exp^{(5)}\big(\log^{(5)} n_0 + A\big)\big]. \tag{4.18}$$

(We will return to this condition at the end of the proof.) Then

$$|\lambda - \eta| > \exp\big(-\exp\big(\log m \log^{(3)} m\big)\big)$$

by the indirect assumption (4.13), and hence

$$\exp(n_{j+1}) < n_{j+1}^{4n_{j+1}} < |\lambda - \eta|^{-1} < \exp^{(2)}\big(\log m \log^{(3)} m\big),$$

which together with $m \leq Cn_j\log n_j$ (for some C depending only on λ) yields

$$n_{j+1} < \exp\big(2\log n_j \log^{(3)} n_j\big),$$

provided n_0 is sufficiently large.

CLAIM. *For each* $j \geq 0$, *we have*

$$n_j < \exp^{(2)}\big((2j + j_0)\log^{(2)}(2j + j_0)\big),$$

where $j_0 = \log^{(2)}(n_0)$.

PROOF. The claim is trivial for $j = 0$, and we prove the $j > 0$ case by induction. We suppose that the claim holds for some j and prove that it also holds for $j + 1$. We first note

$$\log^{(3)} n_j < 2\log(2j + j_0).$$

We can write

$$\begin{aligned} n_{j+1} &< \exp\big(2\log n_j \log^{(3)} n_j\big) \\ &< \exp\big(2\exp\big((2j + j_0)\log^{(2)}(2j + j_0)\big) \cdot 2\log(2j + j_0)\big) \\ &= \exp^{(2)}\big((2j + j_0)\log^{(2)}(2j + j_0) + \log^{(2)}(2j + j_0) + 2\big) \\ &< \exp^{(2)}\big(\big(2(j + 1) + j_0\big)\log^{(2)}\big(2(j + 1) + j_0\big)\big), \end{aligned}$$

where the last line holds, because we assumed that n_0 is large enough, so in particular, we have $\log^{(2)}(2j + j_0) > 2$. This proves the claim. □

Using the above claim, we note that for some positive C_λ depending on λ only,

$$\begin{aligned} \frac{1}{\log K_j \log^{(2)} K_j} &= \frac{1}{\log(C_\lambda \log n_j)\log^{(2)}(C_\lambda \log n_j)} \\ &\geq \frac{1}{2(2j + j_0)\log(2j + j_0)\log^{(2)}(2j + j_0)}, \end{aligned}$$

provided n_0 is large enough.

We can write

$$\begin{aligned} &\sum_{j=1}^{N} \frac{1}{\log K_j \log\log K_j} \\ &\qquad \geq \sum_{j=1}^{N} \frac{1}{2(2j + j_0)\log(2j + j_0)\log^{(2)}(2j + j_0)} \\ &\qquad \geq c\big(\log^{(3)}(N + j_0) - \log^{(3)} j_0\big), \end{aligned}$$

where c is an absolute constant.

We write $B = 2C/c$, where c is the above constant and C is the constant from Proposition 30 applied with the minimum of $\varepsilon \log \lambda^{-1}$ and α in the role of α. We put

$$N := \big\lfloor \exp^{(3)}\big(\log^{(3)}(j_0) + B\big)\big\rfloor.$$

Then

$$\sum_{j=1}^{N} \frac{1}{\log K_j \log\log K_j} \geq c\big(\log^{(3)} N - \log^{(3)} j_0\big) \geq cB \geq 2C. \tag{4.19}$$

On the other hand, we can write

$$\log K_j \leq 2(2j + j_0)\log^{(2)}(2j + j_0) \leq 6N \log^{(2)}(3N) < 10N^2$$

for $j \leq N$, if n_0 and hence j_0 is sufficiently large, and this yields

$$\sum_{j=1}^{N} \log K_j \leq 10N^3.$$

We note that

$$N \leq \exp^{(3)}\big(\log^{(2)}(j_0)\big) = \exp(j_0) = \log(n_0),$$

if n_0 and hence j_0 is sufficiently large. This and $n_1 > n_0$ implies

$$\frac{\sum_{j=1}^{N} \log K_j}{n_1} < 1,$$

provided n_0 is sufficiently large, and hence we have a contradiction with (4.19) and Proposition 30.

It remains to verify that condition (4.18) holds each time we used it. Clearly we always had $n_0 \leq m \leq n_N$. Since $N \geq j_0$, we have

$$\begin{aligned} 2N + j_0 &\leq 4N \leq \exp\big(2 + \exp^{(2)}\big(\log^{(3)}(j_0) + B\big)\big) \\ &\leq \exp^{(2)}\big(2 + \exp\big(\log^{(3)}(j_0) + B\big)\big) \leq \exp^{(3)}\big(\log^{(3)}(j_0) + B + 2\big). \end{aligned}$$

In addition,

$$\begin{aligned} &(2N + j_0)\log^{(2)}(2N + j_0) \\ &\quad\leq \exp^{(3)}\big(\log^{(3)}(j_0) + B + 2\big)\exp\big(\log^{(3)}(j_0) + B + 2\big) \\ &\quad\leq \exp\big(\exp^{(2)}\big(\log^{(3)}(j_0) + B + 2\big) + \log^{(3)}(j_0) + B + 2\big) \\ &\quad\leq \exp\big(2\exp^{(2)}\big(\log^{(3)}(j_0) + B + 2\big)\big) \\ &\quad\leq \exp^{(3)}\big(\log^{(3)}(j_0) + B + 3\big). \end{aligned}$$

Then we have

$$n_N \leq \exp^{(2)}\big((2N + j_0)\log^{(2)}(2N + j_0)\big) \leq \exp^{(5)}\big(\log^{(5)}(n_0) + B + 3\big).$$

This shows that (4.18) holds provided $A \geq B + 3$. This completes the proof of the theorem.

Acknowledgments. We are grateful to Yann Bugeaud, Kevin Hare, Mike Hochman, Nikita Sidorov and Evgeniy Zorin for helpful discussions. We are also grateful to Mike Hochman for pointing out the converse of Theorem 1. We thank the anonymous referee, and also Sébastien Gouëzel, Nicolas de Saxcé and Ariel Rapaport for a very careful reading of our manuscript and for numerous comments and suggestions that greatly improved the presentation of our paper.

REFERENCES

[1] BEAUCOUP, F., BORWEIN, P., BOYD, D. W. and PINNER, C. (1998). Multiple roots of $[-1, 1]$ power series. *J. Lond. Math. Soc.* (2) **57** 135–147. MR1624809

[2] BOMBIERI, E. and GUBLER, W. (2006). *Heights in Diophantine Geometry*. *New Mathematical Monographs* **4**. Cambridge Univ. Press, Cambridge. MR2216774

[3] BREUILLARD, E. and VARJÚ, P. P. (2019). Entropy of Bernoulli convolutions and uniform exponential growth for linear groups. *J. Anal. Math.* To appear. Available at arXiv:1510.04043v2.

[4] BUGEAUD, Y. (2004). *Approximation by Algebraic Numbers*. *Cambridge Tracts in Mathematics* **160**. Cambridge Univ. Press, Cambridge. MR2136100

[5] COVER, T. M. and THOMAS, J. A. (2006). *Elements of Information Theory*, 2nd ed. Wiley, Hoboken, NJ. MR2239987

[6] ERDÖS, P. (1939). On a family of symmetric Bernoulli convolutions. *Amer. J. Math.* **61** 974–976. MR0000311

[7] ERDÖS, P. (1940). On the smoothness properties of a family of Bernoulli convolutions. *Amer. J. Math.* **62** 180–186. MR0000858

[8] FENG, D.-J. and HU, H. (2009). Dimension theory of iterated function systems. *Comm. Pure Appl. Math.* **62** 1435–1500. MR2560042

[9] GARSIA, A. M. (1963). Entropy and singularity of infinite convolutions. *Pacific J. Math.* **13** 1159–1169. MR0156945

[10] HARE, K. G. and SIDOROV, N. (2010). A lower bound for Garsia's entropy for certain Bernoulli convolutions. *LMS J. Comput. Math.* **13** 130–143. MR2638985

[11] HOCHMAN, M. (2014). On self-similar sets with overlaps and inverse theorems for entropy. *Ann. of Math.* (2) **180** 773–822. MR3224722

[12] HOCHMAN, M. and SHMERKIN, P. (2012). Local entropy averages and projections of fractal measures. *Ann. of Math.* (2) **175** 1001–1059. MR2912701

[13] JESSEN, B. and WINTNER, A. (1935). Distribution functions and the Riemann zeta function. *Trans. Amer. Math. Soc.* **38** 48–88. MR1501802

[14] KAĬMANOVICH, V. A. and VERSHIK, A. M. (1983). Random walks on discrete groups: Boundary and entropy. *Ann. Probab.* **11** 457–490. MR0704539

[15] KONTOYIANNIS, I. and MADIMAN, M. (2014). Sumset and inverse sumset inequalities for differential entropy and mutual information. *IEEE Trans. Inform. Theory* **60** 4503–4514. MR3245338

[16] LINDENSTRAUSS, E. and VARJÚ, P. P. (2016). Work in progress.

[17] MADIMAN, M. (2008). On the entropy of sums. In *Information Theory Workshop*, 2008. *ITW '08. IEEE* 303–307.

[18] MAHLER, K. (1960). An application of Jensen's formula to polynomials. *Mathematika* **7** 98–100. MR0124467

[19] MAHLER, K. (1964). An inequality for the discriminant of a polynomial. *Michigan Math. J.* **11** 257–262. MR0166188

[20] PERES, Y., SCHLAG, W. and SOLOMYAK, B. (2000). Sixty years of Bernoulli convolutions. In *Fractal Geometry and Stochastics*, *II* (*Greifswald/Koserow*, 1998). *Progress in Probability* **46** 39–65. Birkhäuser, Basel. MR1785620

[21] SALEM, R. (1944). A remarkable class of algebraic integers. Proof of a conjecture of Vijayaraghavan. *Duke Math*. *J*. **11** 103–108. MR0010149

[22] SHMERKIN, P. (2014). On the exceptional set for absolute continuity of Bernoulli convolutions. *Geom*. *Funct*. *Anal*. **24** 946–958. MR3213835

[23] SHMERKIN, P. (2019). On Furstenberg's intersection conjecture, self-similar measures, and the L^q norms of convolutions. *Ann*. *of Math*. (2) **189** 319–391. MR3919361

[24] SOLOMYAK, B. (1995). On the random series $\sum \pm\lambda^n$ (an Erdős problem). *Ann*. *of Math*. (2) **142** 611–625. MR1356783

[25] SOLOMYAK, B. (2004). Notes on Bernoulli convolutions. In *Fractal Geometry and Applications*: *A Jubilee of Benoît Mandelbrot*. *Part* 1. *Proc*. *Sympos*. *Pure Math*. **72** 207–230. Amer. Math. Soc., Providence, RI. MR2112107

[26] TAO, T. (2010). Sumset and inverse sumset theory for Shannon entropy. *Combin*. *Probab*. *Comput*. **19** 603–639. MR2647496

[27] VARJÚ, P. P. Recent progress on Bernoulli convolutions. In *European Congress of Mathematics*: *Berlin*, *July* 18–22, 2016. Eur. Math. Soc., Zurich.

[28] VARJÚ, P. P. (2016). Absolute continuity of Bernoulli convolutions for algebraic parameters. *J*. *Amer*. *Math*. *Soc*. **32** 351–397. MR3904156

[29] WALDSCHMIDT, M. (1978). Transcendence measures for exponentials and logarithms. *J*. *Aust*. *Math*. *Soc*. *A* **25** 445–465. MR0508469

[30] WANG, Z. (2011). Quantitative density under higher rank Abelian algebraic toral actions. *Int*. *Math*. *Res*. *Not*. *IMRN* **16** 3744–3821. MR2824843

[31] YOUNG, L. S. (1982). Dimension, entropy and Lyapunov exponents. *Ergodic Theory Dynam*. *Systems* **2** 109–124. MR0684248

[32] ZORIN, E. (2013). Algebraic independence and normality of the values of Mahler's functions. Available at arXiv:1309.0105v2.

CENTRE FOR MATHEMATICAL SCIENCES
UNIVERSITY OF CAMBRIDGE
WILBERFORCE ROAD
CAMBRIDGE CB3 0WA
UNITED KINGDOM
E-MAIL: efjb2@cam.ac.uk
pv270@dpmms.cam.ac.uk