

# CUT-OFF PHENOMENON FOR THE $AX+B$ MARKOV CHAIN OVER A FINITE FIELD

EMMANUEL BREUILLARD AND PÉTER P. VARJÚ

ABSTRACT. We study the Markov chain  $x_{n+1} = ax_n + b_n$  on a finite field  $\mathbb{F}_p$ , where  $a \in \mathbb{F}_p$  is fixed and  $b_n$  are independent and identically distributed random variables in  $\mathbb{F}_p$ . Conditionally on the Riemann hypothesis for all Dedekind zeta functions, we show that the chain exhibits a cut-off phenomenon for most primes  $p$  and most values of  $a \in \mathbb{F}_p$ . We also obtain weaker, but unconditional, upper bounds for the mixing time.

## 1. INTRODUCTION

Let  $p$  be a prime number and  $\mathbb{F}_p$  the field with  $p$  elements. Consider the Markov chain on  $\mathbb{F}_p$

$$x_{n+1} = ax_n + b_n, \quad (1.1)$$

where the multiplier  $a$  is non-zero and the  $b_n$ 's are independent random variables taking values in  $\mathbb{F}_p$  with a common law  $\mu$ . We will assume that the support of  $\mu$  has at least two elements, a condition which is easily seen to be equivalent to the aperiodicity of the Markov chain. The chain will then mix and eventually equidistribute towards a stationary measure, which must clearly be the uniform distribution  $u$  on  $\mathbb{F}_p$ . For  $\delta \in (0, 1)$ , it is customary to define the mixing time  $T(\delta) \in \mathbb{N}$  as

$$T(\delta) := \inf\{n \in \mathbb{N} \mid \|\mu_a^{(n)} - u\|_{TV} \leq \delta\}.$$

Here  $\mu_a^{(n)}$  denotes the law of  $x_n$  on  $\mathbb{F}_p$  (starting from  $x_0 = 0$  say) and the norm is the total variation distance:

$$\|\pi_1 - \pi_2\|_{TV} = \sup_{A \subset \mathbb{F}_p} |\pi_1(A) - \pi_2(A)| = \frac{1}{2} \|\pi_1 - \pi_2\|_1 \in [0, 1]$$

for any two probability measures  $\pi_1, \pi_2$  on  $\mathbb{F}_p$ .

In this paper, we will be interested in giving good bounds on  $T(\delta)$  in terms of  $p$ . When  $a = 1$  or when  $a$  has small multiplicative order in  $\mathbb{F}_p^\times$ , then the Markov chain exhibits a diffusive behaviour and  $T(\delta)$  will typically be of order  $p^2$ . However, if  $a$  is more typical in  $\mathbb{F}_p^\times$ , we

---

EB has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 617129); PV has received funding from the Royal Society and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 803711).

expect the mixing time to be of order  $\log p$ , which is of course optimal up to a multiplicative constant when the size of the support  $\text{Supp}(\mu)$  of  $\mu$  is bounded independently of  $p$ , because only at most  $|\text{Supp}(\mu)|^n$  sites can be visited by the Markov chain after  $n$  steps (see (1.4) below).

This problem, inspired by the pseudo-random generators used in the first electronic computers and based on the idea, due to D.H. Lehmer [17], of iterating the map  $x \mapsto ax + 1$  on  $\mathbb{F}_p$ , has been studied in a beautiful paper of Chung-Diaconis-Graham [4] in the special case when  $a = 2 \pmod p$  for all  $p$  and  $b_n$  is uniform in  $\{-1, 0, 1\}$ . Their work was extended by Hildebrand [11, 14] to the case when  $a$  is allowed to be random, i.e.  $a = \lambda_n \pmod p$  and  $b_n = \beta_n \pmod p$ , where the  $\lambda_n$ 's and  $\beta_n$ 's are two independent sequences of i.i.d. integer valued random variables, each chosen according to some fixed probability measure supported on  $\mathbb{Z}$ . In this case, under mild assumptions of the measures, mixing occurs in  $O(\log p \log \log p)$  steps, and this bound is sharp in some cases (e.g.  $a = 2$ ,  $\mu$  uniform on  $\{-1, 0, 1\}$  and  $p$  is a Mersenne prime, i.e. of the form  $p = 2^n - 1$ ). See [4, Theorem 1] as well as [12]. Still, the sharpness seems to be occurring only for a thin, density zero set of primes (recall that Mersenne primes have density zero and form a conjecturally infinite set of primes). This hints at the possibility that an upper bound of the form  $O(\log p)$  might hold for almost all primes and indeed Chung-Diaconis-Graham prove in [4] the analogous statement in  $\mathbb{Z}/p\mathbb{Z}$ , where  $p$  is an arbitrary integer not required to be prime.

We do not answer this specific question, but rather study what happens when  $a$  is an arbitrary element of  $\mathbb{F}_p$  (as opposed to being the mod  $p$  quotient of a fixed value  $a \in \mathbb{Z}$  independent of  $p$  as in the above mentioned works). It was observed by Bukh, Harper, Helfgott and Lindenstrauss (see [10, Exercises 3.13 and 3.14]), that a general polylogarithmic upper bound on the mixing time, valid for all primes  $p$  and for all  $a$  of sufficiently large multiplicative order, can be deduced easily from some estimates of Konyagin [15]. This gives

$$T(\delta) = O((\log p)^{2+o(1)}) \tag{1.2}$$

provided the multiplicative order of  $a$  is at least  $O(\log p \log \log p)$  in  $\mathbb{F}_p$ . See Section 2 below, where a more precise statement is given. Improving on this bound would touch upon some delicate number theoretical issues: for example proving a  $O(\log p)$  upper bound for most primes  $p$  and for all values of  $a$  of sufficiently large multiplicative order would already imply the celebrated Lehmer conjecture (see Section 7 and [3]).

In this paper, we prove three results valid for most values of the multiplier  $a$ . First we prove that the  $O(\log p \log \log p)$  upper bound holds for all primes (Theorem 1). Then we will show a stronger logarithmic  $O(\log p)$  upper bound, which is only valid for most (in a certain weak sense) primes  $p$  (Theorem 2). Finally, conditionally on the Riemann

hypothesis for Dedekind zeta functions of number fields, we will establish that the Markov chain exhibits a cut-off phenomenon for most (this time density 1) primes  $p$  (Theorem 3). A sequence of Markov chains is said to exhibit the cut-off phenomenon if there is a sharp phase transition to equidistribution, that is we have  $T(\delta_1)/T(\delta_2) \rightarrow 1$  as we go through the sequence of chains for any  $\delta_1, \delta_2 \in (0, 1)$ . We refer the reader to [6, 18] and references therein for an introduction to cut-off phenomena for Markov chains.

**Theorem 1.** *Let  $p$  be a prime number and  $\mu$  a probability measure on  $\mathbb{F}_p$  supported on at least two elements. Let  $\varepsilon \in (0, 1)$ . Then for all but an  $\varepsilon$ -fraction of all  $a \in \mathbb{F}_p^\times$ , the mixing times of the Markov chain (1.1) satisfy for any  $\delta \in (0, \frac{1}{2})$ :*

$$T(\delta) \leq T_2(\delta) \leq C_{\varepsilon, \mu} \log p (\log \log p + \log(C_{\varepsilon, \mu} \delta^{-1})),$$

where  $C_{\varepsilon, \mu} = O(\varepsilon^{-1} H_2(\mu)^{-1})$  for some absolute implied constant.

Here  $T_2(\delta) := \inf\{n \in \mathbb{N} | p \|\mu_a^{(n)} - u\|_2^2 \leq \delta^2\}$  is the  $\ell^2$ -mixing time, and  $H_2(\mu) := -\log \|\mu\|_2^2$  is the Rényi entropy of  $\mu$  of order 2, where  $\|\mu\|_2^2 = \sum_{x \in \mathbb{F}_p} \mu(x)^2$ .

We recall that for all  $\delta > 0$  we have  $T(\delta) \leq T_2(2\delta) \leq T_2(\delta)$ , which is a direct consequence of the Cauchy-Schwarz inequality:

$$\|\mu_a^{(n)} - u\|_1^2 \leq p \|\mu_a^{(n)} - u\|_2^2.$$

In the next results we will improve the above bound for a typical prime. For this reason we need to restrict somewhat the class of chains of the form (1.1) that we consider by assuming that the  $b_i$ 's are the mod  $p$  values of a fixed sequence of integer valued i.i.d. random variables whose law  $\mu$  is supported on  $\mathbb{Z}$  and independent of  $p$ .

**Theorem 2.** *Fix a probability measure  $\mu$  on  $\mathbb{Z}$  supported on a finite set of at least two elements, and consider the Markov chain (1.1). There is a possibly empty exceptional set of primes  $B$  such that for every  $X \geq 1$ ,*

$$\sum_{p \in B \cap [e^{X/2}, e^X]} \frac{1}{p} \leq \frac{(\log X)^6}{\sqrt{X}} \quad (1.3)$$

with the property that given  $\varepsilon, \delta \in (0, 1)$ , if  $p$  is a prime larger than a constant depending only on  $\mu$  and these parameters, and  $p \notin B$ , then for all but an  $\varepsilon$ -fraction of all  $a \in \mathbb{F}_p$  we have:

$$T(\delta) \leq T_2(\delta) \leq \frac{5}{H_2(\mu)} \log p,$$

where  $H_2(\mu) := -\log \sum_{z \in \mathbb{Z}} \mu(z)^2$ .

Note that condition (1.3) implies that  $\sum_{p \in B} p^{-1}$  converges, while it is well-known that  $\sum_p p^{-1}$  diverges and that indeed  $\sum_{p \in [e^{X/2}, e^X]} p^{-1}$

lies between two positive constants independent of  $X > 0$ . This is the sense in which the complement of  $B$  contains most primes.

Finally we state our main result establishing a cut-off under GRH. We define the entropy of  $\mu$  to be

$$H(\mu) = \sum_{z \in \mathbb{Z}} \mu(z) \log \frac{1}{\mu(z)}.$$

We also set  $C_\mu = \max\{1, \log(\sup \mu / \inf \mu)\}$ , where  $\sup$  and  $\inf$  are taken over the support of  $\mu$ .

**Theorem 3.** *Assume that the Riemann hypothesis holds for the Dedekind zeta function of an arbitrary number field. Fix a probability measure  $\mu$  on  $\mathbb{Z}$  supported on a finite set of at least two elements, and consider the Markov chain (1.1). Fix  $\varepsilon, \kappa, \theta > 0$ . For  $X > 0$ , let  $n$  be the integer part of  $(X + \theta\sqrt{X})/H(\mu)$ . Then for all but an  $\varepsilon$ -fraction of all primes  $p$  in  $[e^X, (1 + \kappa)e^X]$  and all but an  $\varepsilon$ -fraction of all  $a \in \mathbb{F}_p$  we have*

$$\|\mu_a^{(n)} - u\|_{TV} \leq 5\varepsilon^{-1} e^{-\frac{H(\mu)\theta^2}{16C_\mu^2}},$$

provided  $X \geq 2 \log \varepsilon^{-1} + C(\theta^2 \log \theta)^2 + C$ , where  $C$  is a constant depending only on  $\kappa$  and  $\mu$ .

On the other hand, there is a universal lower bound on  $\|\mu_a^{(n)} - u\|_{TV}$  in terms of the entropy of the walk. Mixing cannot occur before the entropy of the walk  $H(\mu_a^{(n)}) \leq nH(\mu)$  approaches the maximal entropy, namely  $H(u) = \log p$ . In fact, we will show the following in Section 5. Let  $p$  be prime,  $a \in \mathbb{F}_p$  and  $\theta \in (0, \sqrt{\log p}/2)$ . If  $n \geq 0$  is the integer part of  $(\log p - \theta\sqrt{\log p})/H(\mu)$ , we have

$$\|\mu_a^{(n)} - u\|_{TV} \geq 1 - e^{-\theta^2} - 2e^{-\frac{H(\mu)\theta^2}{C_\mu^2}}. \quad (1.4)$$

Combined with Theorem 3, this yields the following.

**Corollary 4.** *Under the assumptions of Theorem 3 the Markov chain (1.1) exhibits a cut-off phenomenon in total variation at  $\frac{1}{H(\mu)} \log p$  for most primes  $p$  and most multipliers  $a \in \mathbb{F}_p$ . More precisely, for any given  $\varepsilon \in (0, 1)$  there is an exceptional set of primes  $E$  of density at most  $\varepsilon$ , such that for all  $p \notin E$  and for all but an  $\varepsilon$ -fraction of all  $a \in \mathbb{F}_p$  we have:*

$$|T(\delta) - \frac{1}{H(\mu)} \log p| \leq C_{\mu, \varepsilon} \sqrt{\log p},$$

for all  $\delta \in (\varepsilon, 1 - \varepsilon)$ , where  $C_{\mu, \varepsilon} = O_\mu(\sqrt{\log(1/\varepsilon)})$ .

We can also let  $\varepsilon$  tend to zero and get a set of primes with density 1 for which a cut-off takes place, at the cost of loosening a bit the control on the cut-off window. See Corollary 10.

We will also prove an  $\ell^2$  version of these results. The  $\ell^2$  cut-off window is even narrower (of bounded size). In fact Theorem 3 will be

proven after we handle the analogous  $\ell^2$  version by a small variation of that argument, see Section 4.

**1.1. Motivation.** Our motivation for studying the Markov chain (1.1) is manifold. First of all this chain is the natural “mod  $p$  analogue” of the classical Bernoulli convolutions we studied in [1], where a close relationship between the entropy of the random walk and the Mahler measure of the multiplier was established. This enabled us to give a “mod  $p$ ” reformulation of the Lehmer conjecture in [3]. On the other hand the chain (1.1) is intimately related to random polynomials of large degree and is a crucial tool in our recent paper on irreducibility of random polynomials [2], where Konyagin’s estimate (1.2) is used to establish irreducibility under GRH<sup>1</sup>. The latter paper and the current one can be seen as companion papers, since the main irreducibility result from [2] will be the key ingredient in the proof of the cut-off in Theorem 3. Finally the chain (1.1), which can also be described in terms of random walks on the affine group of the line, can be seen as a toy model for many more sophisticated random walks on groups or homogeneous spaces of algebraic origin.

**1.2. Strategy.** We now briefly describe the main strategy behind the proof of the above results. It starts with the observation that the chain (1.1) at time  $n$  is given by the value at  $a \in \mathbb{F}_p^\times$  of a random polynomial

$$P(x) = b_0x^{n-1} + \dots + b_{n-2}x + b_{n-1} \quad (1.5)$$

where the  $b_i \in \mathbb{Z}$ ’s are as before i.i.d. random variables with law  $\mu$ . In particular the law of  $P(a)$  is exactly  $\mu_a^{(n)}$ . Furthermore, we can write

$$\|\mu_a^{(n)}\|_2^2 = \sum_{y \in \mathbb{F}_p} \mu_a^{(n)}(y)^2 = \sum_{y \in \mathbb{F}_p} \mathbb{P}(P_1(a) = P_2(a) = y) = \mathbb{P}(P_1(a) = P_2(a))$$

where  $P_1$  and  $P_2$  are two independent random polynomials as in (1.5)

Summing over  $a \in \mathbb{F}_p$ , we get

$$\sum_{a \in \mathbb{F}_p} \|\mu_a^{(n)}\|_2^2 = \mathbb{E}(N_p(P)), \quad (1.6)$$

where  $P \in \mathbb{Z}[x]$  is a random polynomial with the same law as  $P_1 - P_2$  and  $N_p(P)$  is the number of distinct roots of  $P$  in  $\mathbb{F}_p$ . If  $P = 0$ , then  $N_p(P) = p$  of course, while if  $P \neq 0$ , we can use the crude upper bound  $N_p(P) \leq n$ . This yields

$$\sum_{a \in \mathbb{F}_p} \|\mu_a^{(n)}\|_2^2 \leq n + \mathbb{P}(P = 0)p.$$

---

<sup>1</sup>Here and everywhere in this paper, when we say that we assume GRH, we mean that we assume that the Dedekind zeta function of an arbitrary number field satisfies the Riemann hypothesis, namely has its non trivial zeroes on the critical line.

Together with a further trick exploiting the self-similarity of the law

$$\mu_a^{(n+m)} = \mu_a^{(n)} * a^n \cdot \mu_a^{(m)}$$

this will yield Theorem 1.

In order to prove the cut-off in Theorem 3, we need to perform a further averaging over primes and use the following key arithmetical ingredient, which is a special case of Chebotarev's theorem or just the prime ideal theorem: if  $P \in \mathbb{Z}[x]$  is irreducible over  $\mathbb{Q}$ , then, on average over the primes  $p$ ,  $N_p(P)$  is equal to 1. In other words as  $X \rightarrow +\infty$

$$\mathbb{E}_p(N_p(P)) = 1 + \text{error}$$

where  $\mathbb{E}_p$  denotes the average over primes  $p \leq X$  and the error tends to 0 as  $X \rightarrow +\infty$ . The quality of the error depends on what is known about the zeroes of the Dedekind zeta function  $\zeta_K(z)$  of the number field  $K := \mathbb{Q}(x)/(P)$  near  $z = 1$ . Assuming GRH, we get a very good error term. It is important to control the error term uniformly as  $P$  varies, so explicit bounds in terms of the degree and discriminant of  $P$  are necessary.

We proved in our previous paper [2] that  $P$  is irreducible with high probability, and this implies that the main contribution to the right hand side of (1.6) comes when  $P$  is either irreducible or equal to 0. In the latter case we have  $N_p(P) = p$ , as we remarked above. So we obtain

$$\mathbb{E}_p \left( \sum_{a \in \mathbb{F}_p} \|\mu_a^{(n)}\|_2^2 \right) \simeq 1 + \mathbb{P}(P = 0) \mathbb{E}_p(p)$$

from (1.6). Note that  $\mathbb{P}(P = 0) = \|\mu\|_2^{2n}$  since the  $b_i$ 's are i.i.d. and that  $p \|\mu_a^{(n)}\|_2^2 - 1 = p \|\mu_a^{(n)} - u\|_2^2 \geq 0$ . Now by the Markov inequality the previous average estimate will hold for most values of the prime  $p$  and the residue  $a \in \mathbb{F}_p$ . This means that

$$p \|\mu_a^{(n)} - u\|_2^2 \ll p \|\mu\|_2^{2n}$$

for most  $p$  and most  $a$ 's. In other words (since  $\|\mu_a^{(n)}\|_2 \geq \|\mu\|_2^n$ )

$$\|\mu\|_2^{2n} p - 1 \leq p \|\mu_a^{(n)} - u\|_2^2 \ll \|\mu\|_2^{2n} p.$$

What this means is precisely that we will see an  $\ell^2$  cut-off exactly when  $\|\mu\|_2^{2n} p$  becomes small, i.e. when  $n \simeq \frac{1}{H_2(\mu)} \log p$ .

The cut-off in total variation of Theorem 3 is deduced from the  $\ell^2$  cut-off estimate we have just described by restricting to a large part of the sample space where  $P$  is roughly uniform. Finally Theorem 2 is the result of a more precise analysis of what happens if we do not assume GRH and how we can make do with a weaker error term and no irreducibility of  $P$  using instead estimates on the number of zeroes of Dedekind zeta functions and Dobrowolski's Mahler measure lower bound to get a bound on the number of irreducible factors of  $P$ .

**1.3. Organization.** The paper is organized as follows. In Section 2, we discuss the general upper bound (1.2), which can be deduced from Konyagin’s estimates. We include this for comparison with our new results. In Section 3, we prove Theorem 1 using a multiplicity trick exploiting the “self-similarity” feature of the chain (1.1). In Section 4, we prove Theorem 7, which establishes the cut-off in  $\ell^2$  under GRH with a bounded cut-off window. Section 5 is devoted to the proof of Theorem 3 and Corollary 4. Theorem 2 will be established in Section 6. Finally, in Section 7, we mention mixing in  $\ell^q$ -norm, diameter bounds and connections with the Lehmer problem, and we indicate some directions of future research.

## 2. THE GENERAL UPPER BOUND

In this section, we discuss a general polylogarithmic upper bound on the mixing time. It was observed by Bukh, Harper, Helfgott and Lindenstrauss (see [10, Exercises 3.13 and 3.14]) that such a bound can be deduced easily from Konyagin’s main lemma in [15]. More precisely we aim at the following.

**Theorem 5** (Konyagin, Bukh, Harper, Helfgott, Lindenstrauss). *There is an absolute constant  $C > 0$  such that the following holds. Let  $p$  be a prime and  $\mu$  a probability measure on  $\mathbb{F}_p$  supported on at least two elements. Consider the Markov chain (1.1) for a residue  $a \in \mathbb{F}_p^\times$  with multiplicative order at least  $C \log p(\log \log \log p)$ . Then for every  $\delta \in (0, \frac{1}{2})$ , we have*

$$T(\delta) \leq T_2(\delta) \leq \frac{C}{1 - \|\mu\|_2^2} \log(\delta^{-1}p) \log p(\log \log p)^5.$$

We include this result, because it provides a good comparison standpoint for the theorems of our paper that deal with mixing bounds for most values of the multiplier  $a$  rather than all values as in Theorem 5. We will not use this result directly in this paper; however, a more general version of it (proved in Section 5.2 of [2]) is used crucially in our companion paper [2], which is needed for Theorem 3.

We will use the following notation: for  $u \in \mathbb{R}$  we set  $\tilde{u}$  to be the unique representative of  $u$  modulo  $\mathbb{Z}$  in the interval  $(-\frac{1}{2}, \frac{1}{2}]$ . It clearly depends only on the class of  $u$  modulo  $\mathbb{Z}$ . For ease of notation, we identify the additive group of  $\mathbb{F}_p$  with a subgroup of  $\mathbb{R}/\mathbb{Z}$  by viewing  $x \bmod p\mathbb{Z}$  as  $x/p \in \mathbb{R}/\mathbb{Z}$ . We write  $\log^{(n)}$  for the  $n$ -th iteration of  $\log$ .

**Lemma 6** (Konyagin). *There is an absolute constant  $C \geq 1$  such that if  $p$  is a prime and  $a \in \mathbb{F}_p^\times$ , for every  $x_0 \in \mathbb{F}_p^\times$  and every  $m \geq C \log p(\log^{(2)} p)^4$ , either  $a$  has multiplicative order at most  $C \log p \log^{(3)} p$ , or we have*

$$\sum_{i=0}^m \left( \frac{\widetilde{x_0 a^i}}{p} \right)^2 \geq \frac{1}{C \log^{(2)} p}. \quad (2.1)$$

This lemma is a very slight improvement of [15, Lemma 6] and a special case of [2, Proposition 24]. The proof in [2] closely follows Konyagin's original argument in [15] and in the setting of Lemma 6 it does not introduce any new ideas.

For the convenience of the reader, we explain how to deduce Lemma 6 from [2, Proposition 24]. We take  $M = m_1 = D = 1$ ,  $Q = p$ ,  $\beta = x_0$  and  $\alpha = a$  in [2, Proposition 24], see [2, Section 5.1] for the notation. Taking  $L = m$  in [2, Proposition 24], we get that either (2.1) holds, or there is a polynomial  $P \in \mathbb{Z}[x]$  of degree at most  $3 \log p$  with Mahler measure at most

$$M(P) \leq (\log p)^{30 \log p/L} \leq \exp(30C^{-1}(\log \log p)^{-3}) \quad (2.2)$$

such that  $P(a) = 0$ . Recall that if  $Q \in \mathbb{Z}[x]$  is a polynomial  $Q = a_0 + a_1x + \dots + a_nx^n$ , then  $M(Q)$  is defined as

$$M(Q) = |a_n| \prod_1^n \max\{1, |\alpha_i|\} \quad (2.3)$$

where the  $\alpha_i$ 's are the complex roots of  $Q$ . By the classical Dobrowolski bound [7] on Mahler measure, either

$$M(P) \geq \exp\left(c \left(\frac{\log \log \deg P}{\log \deg P}\right)^3\right) \geq \exp\left(c \frac{1}{(\log \log p)^3}\right)$$

for some absolute constant  $c$ , or  $P$  is a product of cyclotomic polynomials. If we choose  $C$  sufficiently large in (2.2), then we must have the second case, which makes the multiplicative order of  $a$  less than  $C \log p \log^{(3)} p$  (recall that the degree of the  $n$ -cyclotomic polynomial is Euler's totient function, which satisfies the estimate  $\phi(n) \gg n/\log^{(2)} n$ .)

*Proof of Theorem 5.* We aim at proving a good upper bound on the size of the Fourier coefficients of  $\mu_a^{(n)}$ . For  $a \in \mathbb{F}_p$ , set  $\pi_a : \mathbb{F}_p[x] \rightarrow \mathbb{F}_p$  be the evaluation map  $P \mapsto P(a)$ . Recall that  $\mu_a^{(n)} = \pi_a(\mu^{(n)})$  is the law of  $P(a)$ , where  $P(x) = b_0x^{n-1} + \dots + b_{n-2}x + b_{n-1}$  is a random polynomial with law  $\mu^{(n)}$ , that is, the coefficients  $b_i$ 's are i.i.d. with law  $\mu$  supported on  $\mathbb{F}_p$ .

The Fourier coefficient

$$\widehat{\mu_a^{(n)}}(\xi) := \mathbb{E}\left(e\left(\frac{\xi P(a)}{p}\right)\right),$$

where  $e(u) := \exp(2i\pi u)$ , can be written, by independence of the  $b_i$ 's as

$$\widehat{\mu_a^{(n)}}(\xi) := \prod_{i=0}^{n-1} \phi(\xi a^i), \quad (2.4)$$

where

$$\phi(t) = \mathbb{E}(e(\frac{tb_0}{p})) = \sum_{x \in \mathbb{F}_p} \mu(x) e(\frac{tx}{p})$$

for  $t \in \mathbb{F}_p$ . We claim that

$$|\phi(t)| \leq \exp(-4 \sum_{x \neq 0} \nu(x) \left(\frac{\widetilde{tx}}{p}\right)^2), \quad (2.5)$$

where  $\nu$  is the law of  $b_0 - b_1$ . Indeed, by easy calculus  $\cos(2\pi u) \leq 1 - 8\widetilde{u}^2$  for all  $u \in \mathbb{R}$ . Then

$$1 - |\phi(t)|^2 = 1 - \mathbb{E}(e(\frac{t(b_0 - b_1)}{p})) = \sum_{x \neq 0} \nu(x) (1 - \cos(2\pi \frac{tx}{p})) \geq 8 \sum_{x \neq 0} \nu(x) \left(\frac{\widetilde{tx}}{p}\right)^2$$

and (since  $1 - 2x \leq e^{-x}$  for  $x \geq 0$ ) (2.5) follows.

Now (2.4) yields

$$|\widehat{\mu_a^{(n)}}(\xi)| \leq \exp(-4 \sum_{x \neq 0} \nu(x) \sum_{i=0}^{n-1} \left(\frac{\xi x a^i}{p}\right)^2).$$

Note that  $\nu(0) = \|\mu\|_2^2$ . Now splitting the sum of length  $n$  into intervals of length  $C \log p (\log^{(2)} p)^4$  to which we apply Lemma 6, we obtain

$$|\widehat{\mu_a^{(n)}}(\xi)|^2 \leq \exp(-\frac{4n(1 - \|\mu\|_2^2)}{C^2 \log p (\log^{(2)} p)^5}),$$

provided  $\xi \neq 0$  and  $n \geq C \log p (\log^{(2)} p)^4$ . This bound becomes  $\leq \frac{\delta^2}{p}$ , provided

$$n \geq \frac{C^2}{(1 - \|\mu\|_2^2)} \log(\delta^{-1} p) \log p (\log^{(2)} p)^5.$$

If so, then by Parseval's identity

$$\|\mu_a^{(n)} - u\|_1^2 \leq p \|\mu_a^{(n)} - u\|_2^2 = \sum_{\xi \neq 0} |\widehat{\mu_a^{(n)}}(\xi)|^2 \leq \delta^2.$$

□

### 3. MULTIPLICITY OF FOURIER COEFFICIENTS

In this section, we prove Theorem 1. The proof relies on the following observations about the Fourier coefficients of  $\mu_a^{(n)}$  that come from the self-similar nature of the measure (a key feature of Bernoulli convolutions). For every  $n, m \geq 0$ ,

$$\mu_a^{(n+m)} = \mu_a^{(n)} * a^n \cdot \mu_a^{(m)} \quad (3.1)$$

where  $*$  is the convolution product and  $a^n \cdot \mu_a^{(m)}$  is the image of  $\mu_a^{(m)}$  under the multiplication map  $x \mapsto a^n x$ . This also plays a role in the paper [4]. From (3.1), it follows that for every  $\xi \in \mathbb{F}_p$ , if  $0 \leq r \leq s$ ,

$$|\widehat{\mu_a^{(s)}}(\xi)| \leq |\widehat{\mu_a^{(r)}}(\xi)|.$$

Moreover, for  $\xi \neq 0$ , we have

$$|\widehat{\mu_a^{(n+m)}}(\xi)| \leq |\widehat{\mu_a^{(n)}}(\xi)| \sup_{x \neq 0} |\widehat{\mu_a^{(m)}}(x)|, \quad (3.2)$$

and iterating for any  $k \geq 1$ ,

$$|\widehat{\mu_a^{(km)}}(\xi)| \leq [\sup_{x \neq 0} |\widehat{\mu_a^{(m)}}(x)|]^k. \quad (3.3)$$

Similarly, using  $\mu_a^{(n+m)} = \mu_a^{(i)} * a^i \cdot \mu_a^{(n)} * a^{n+i} \cdot \mu_a^{(m-i)}$  in place of (3.1),

$$|\widehat{\mu_a^{(n+m)}}(\xi)| \leq \inf_{i=0, \dots, m} |\widehat{\mu_a^{(n)}}(a^i \xi)|. \quad (3.4)$$

We can interpret (3.4) as a form of high multiplicity of the Fourier coefficients of  $\mu_a^{(n)}$ . Indeed it implies that if  $\mu_a^{(n+m)}$  has one large Fourier coefficient, then  $\mu_a^{(n)}$  must have at least  $m$  large Fourier coefficients provided  $a$  has order at least  $m$ . From (3.2) we have on the one hand

$$p \|\mu_a^{(2n+m)} - u\|_2^2 = \sum_{\xi \neq 0} |\widehat{\mu_a^{(2n+m)}}(\xi)|^2 \leq \sum_{\xi \neq 0} |\widehat{\mu_a^{(n)}}(\xi)|^2 \sup_{\xi \neq 0} |\widehat{\mu_a^{(n+m)}}(\xi)|^2$$

and if  $a$  has order  $> m$  in  $\mathbb{F}_p^\times$ , by (3.4)

$$(m+1) \sup_{\xi \neq 0} |\widehat{\mu_a^{(n+m)}}(\xi)|^2 \leq \sup_{\xi \neq 0} \sum_{0 \leq i \leq m} |\widehat{\mu_a^{(n)}}(a^i \xi)|^2 \leq \sum_{\xi \neq 0} |\widehat{\mu_a^{(n)}}(\xi)|^2. \quad (3.5)$$

In particular, combining (3.5) with (3.3) we get for all  $k, m, n \geq 0$ :

$$p \|\mu_a^{(n+k(m+n))} - u\|_2^2 \leq \sum_{\xi \neq 0} |\widehat{\mu_a^{(n)}}(\xi)|^2 \left( \frac{\sum_{\xi \neq 0} |\widehat{\mu_a^{(n)}}(\xi)|^2}{m+1} \right)^k \leq \frac{(p \|\mu_a^{(n)}\|_2^2)^{k+1}}{(m+1)^k} \quad (3.6)$$

provided  $a$  has order at least  $m+1$  in  $\mathbb{F}_p^\times$ .

In conclusion we have shown that, when  $a$  has sufficiently large multiplicative order, the self-similarity (3.1) of the law implies an automatic decay of Fourier coefficients. We can exploit this to prove Theorem 1.

*Proof of Theorem 1.* Recall our previous notation. For  $a \in \mathbb{F}_p$ ,  $\pi_a : \mathbb{F}_p[x] \rightarrow \mathbb{F}_p$  is the evaluation map  $P \mapsto P(a)$  and  $\mu_a^{(n)} = \pi_a(\mu^{(n)})$  is the law of  $P_1(a)$ , where  $P_1(x) = b_0 x^{n-1} + \dots + b_{n-2} x + b_{n-1}$  is a random

polynomial whose coefficients  $b_i$ 's are i.i.d. with law  $\mu$  supported on  $\mathbb{F}_p$ . Then we make the crucial observation

$$\sum_{a \in \mathbb{F}_p} \|\mu_a^{(n)}\|_2^2 = \sum_{a \in \mathbb{F}_p} \mathbb{P}(P_1(a) = P_2(a)) = \mathbb{E}(N_p(P)),$$

where  $N_p(P)$  is the number of distinct roots of  $P$  in  $\mathbb{F}_p$  and  $P = P_1 - P_2$  for two independent random polynomials  $P_1, P_2$  distributed as above. If  $P = 0$ ,  $N_p(P) = p$ , while if  $P \neq 0$ ,  $N_p(P) \leq n$ . Hence

$$\sum_{a \in \mathbb{F}_p} \|\mu_a^{(n)}\|_2^2 \leq p \|\mu\|_2^{2n} + n \leq 1 + n,$$

if  $n \geq \frac{1}{H_2(\mu)} \log p$ . We thus choose to set  $n = \lceil \frac{1}{H_2(\mu)} \log p \rceil$ . By the Markov inequality, it follows that for all but an  $\varepsilon/2$ -fraction of all residues  $a \in \mathbb{F}_p$  we have

$$p \|\mu_a^{(n)}\|_2^2 \leq 2\varepsilon^{-1}(1+n).$$

Pick  $m$  so that  $10\varepsilon^{-1}(n+1) \geq m+1 \geq 8\varepsilon^{-1}(n+1)$ . Then  $p \|\mu_a^{(n)}\|_2^2 / (m+1) \leq e^{-1}$  and choosing the least  $k$  such that  $e^{-k} \leq \delta^2 \varepsilon / (n+1)$ , we may apply (3.6) to conclude

$$\|\mu_a^{(M)} - u\|_1^2 \leq p \|\mu_a^{(M)} - u\|_2^2 \leq \delta^2$$

with

$$M = n + k(m+n) = O(\varepsilon^{-1} n \log(\varepsilon^{-1} \delta^{-1} n)).$$

To apply (3.6), we need that the multiplicative order of  $a$  is at least  $m$ , which holds for all but at most  $m^2$  choices of  $a$ . Excluding these, the claim still holds for all but an  $\varepsilon$ -fraction of the residues  $a \in \mathbb{F}_p$ , provided  $p$  is large enough depending on  $\varepsilon$ , which we may assume, otherwise the theorem is vacuous. The result follows.  $\square$

#### 4. CUT-OFF IN $\ell^2$

We start by establishing an  $\ell^2$  version of Theorem 3. The method of proof will be adapted in the next section to establish Theorem 3. Recall that  $H_2(\mu) = -\log \|\mu\|_2^2$ , where  $\|\mu\|_2^2 = \sum_{z \in \mathbb{Z}} \mu(z)^2$ .

**Theorem 7.** *Assume GRH. Fix a probability measure  $\mu$  on  $\mathbb{Z}$  supported on a finite set of at least two elements, and consider the Markov chain (1.1). Fix  $\varepsilon, \kappa > 0$  and  $\theta > 0$ . For  $X > 0$ , let  $n$  be the integer part of  $\frac{1}{H_2(\mu)}(X + \theta)$ . Set  $D_{\mu, \kappa, \varepsilon} = 8(1 + \kappa)\varepsilon^{-2} \|\mu\|_2^{-2}$ . Then for all but an  $\varepsilon$ -fraction of all primes  $p$  in  $[e^X, (1 + \kappa)e^X]$  and all but an  $\varepsilon$ -fraction of all  $a \in \mathbb{F}_p$ , we have*

$$p \|\mu_a^{(n)} - u\|_2^2 \leq D_{\mu, \kappa, \varepsilon} e^{-\theta},$$

provided  $X \geq 2 \log \varepsilon^{-1} + C(\theta \log \theta)^2 + C$ , where  $C$  is a constant depending only on  $\kappa$  and  $\mu$ .

Note on the other hand that we have a universal lower bound

$$p\|\mu_a^{(n)} - u\|_2^2 = p\|\mu_a^{(n)}\|_2^2 - 1 \geq p\|\mu\|_2^{2n} - 1 \geq e^{-\theta} - 1$$

which is valid when  $n = \lfloor \frac{1}{H_2(\mu)}(X + \theta) \rfloor$  for all  $X > 0$ ,  $\theta \in \mathbb{R}$ , and all  $p \in [e^X, (1 + \kappa)e^X]$ . In particular we see that the Markov chain is far from  $\ell^2$ -equidistribution if  $\theta < 0$ , but becomes very close to equidistribution if  $\theta > 0$  is a fixed large number. Hence, for those primes  $p$  and residues  $a$  as in Theorem 7 the  $\ell^2$ -mixing time satisfies  $T_2(\delta) = \frac{1}{H_2(\mu)}(1 + o_\delta(1)) \log p$  as  $p \rightarrow +\infty$ , and hence that an  $\ell^2$ -cut-off takes place at  $\frac{1}{H_2(\mu)} \log p$ . The phase transition is very rapid and occurs within a small (bounded) window.

We will deduce Theorem 7 from Proposition 8 below, which is the analogous result for the expectation of  $p\|\mu_a^{(n)} - u\|_2^2$  (when averaging over  $p$  and  $a$ ) by a simple use of the Markov inequality. Given  $n$  and  $p$ , we will say that a residue  $a \in \mathbb{F}_p$  is *admissible* if it is non-zero and not a root of a cyclotomic polynomial of degree at most  $3n$ .

**Proposition 8.** *Assume GRH. Fix a probability measure  $\mu$  on  $\mathbb{Z}$  supported on a finite set of at least two elements, and consider the Markov chain (1.1). Fix  $\kappa > 0$ . For  $n, X > 0$ ,*

$$\mathbb{E}_{a,p}(p\|\mu_a^{(n)}\|_2^2 - p\|\mu\|_2^{2n}) \leq 1 + O(n^4 X^2 e^{-X/2}) + O(e^{-O(\sqrt{n}/\log n)}),$$

where  $\mathbb{E}_{a,p}$  denotes the (weighted) average over all primes  $p \in [e^X, (1 + \kappa)e^X]$  and all admissible residues  $a \in \mathbb{F}_p$ , and the implied constants depend only on  $\kappa$  and  $\mu$ .

By *weighted average over primes*, we mean that we attribute the weight  $\log p$  to each prime  $p$  when counting.

*Proof of Theorem 7.* Let  $Y = p\|\mu_a^{(n)} - u\|_2^2 = p\|\mu_a^{(n)}\|_2^2 - 1$ . By Proposition 8, we have

$$\mathbb{E}_{a,p}(Y) \leq \mathbb{E}_{a,p}(p\|\mu\|_2^{2n}) + O_\mu(\theta^4 X^2 e^{-X/2}) + O(e^{-O_\mu(\sqrt{X}/\log X)}).$$

Note that  $\mathbb{E}_{a,p}(p\|\mu\|_2^{2n}) \leq \alpha$ , where  $\alpha := (1 + \kappa)e^{-\theta}/\|\mu\|_2^2$ . Therefore  $\mathbb{E}_{a,p}(Y) \leq 2\alpha$  provided  $X \geq C_{\kappa,\mu}(1 + (\theta \log \theta)^2)$ .

We write  $\mathbb{E}_{a,p}(Y) = \mathbb{E}_p(\mathbb{E}_a(Y|p))$  and by the Markov inequality

$$\mathbb{P}_p(\mathbb{E}_a(Y|p) \geq 4\alpha/\varepsilon) \leq \varepsilon/2.$$

Let  $B$  be the set of primes  $p \in [e^X, (1 + \kappa)e^X]$  and  $A$  the subset of those such that  $\mathbb{E}_a(Y|p) > 4\alpha/\varepsilon$ . Then

$$|A|X \leq \sum_{p \in A} \log p \leq \frac{\varepsilon}{2} \sum_{p \in B} \log p \leq \frac{\varepsilon}{2}(X + \log(1 + \kappa))|B| \leq \varepsilon X|B|,$$

provided  $X \geq \log(1 + \kappa)$ . So  $|A|/|B| \leq \varepsilon$ .

Now for  $p \in B \setminus A$ , by the Markov inequality again,  $\mathbb{P}_a(Y > 8\alpha/\varepsilon^2) \leq \varepsilon/2$ . And we deduce that for all but an  $\varepsilon$ -fraction of primes  $p \in [e^X, (1 + \kappa)e^X]$ , for all but an  $\varepsilon/2$ -fraction of admissible residues  $a \in \mathbb{F}_p$ , we

have  $Y \leq 8\alpha/\varepsilon^2$ . On the other hand, there are at most  $O(n^3)$  non-admissible residues. So again as soon as  $X$  is large enough (say  $X > 2\log(\theta\varepsilon^{-1}) + O(1)$ , so that  $n^3 \ll \varepsilon e^X$ ), this represents at most an  $\varepsilon/2$ -fraction of all residues modulo  $p$ . This yields the upper bound in Theorem 7.  $\square$

We now turn to the proof of Proposition 8. It will follow easily from the main theorem of [2], whose proof assumed the validity of the Riemann hypothesis for Dedekind zeta functions of number fields, which we recall now.

**Theorem 9** ([2, Theorem 2]). *Assume GRH. Fix a probability measure  $\nu$  on  $\mathbb{Z}$  supported on a finite set of at least two elements. Let  $\{A_i\}_{0 \leq i \leq n}$  be independent random variables with common law  $\nu$  and set  $P = A_n x^n + \dots + A_1 x + A_0 \in \mathbb{Z}[x]$ . Then, with probability at least  $1 - \exp(-O_\nu(n^{1/2}/\log n))$ ,*

$$P = \Phi \tilde{P},$$

where  $\tilde{P}$  is irreducible in  $\mathbb{Q}[x]$ ,  $\deg \Phi = O_\nu(\sqrt{n})$  and  $\Phi$  is a product of cyclotomic polynomials and a monomial  $x^m$  for some  $m \in \mathbb{Z}_{\geq 0}$ .

Below we denote by  $P = P_1 - P_2$  the difference between two independent copies  $P_1, P_2$  of a polynomial  $a_{n-1}x^{n-1} + \dots + a_1x + a_0$ , where the  $a_i$ 's are i.i.d random variables chosen according to the law  $\mu$ . We also denote by  $\tilde{P}$  the quotient of  $P$  by all its monomial and cyclotomic divisors. We will write  $Ad$  for the set of admissible residues in  $\mathbb{F}_p$ . Also  $N_p^{Ad}(P)$  will be the number of admissible roots of  $P$ . We may write

$$\begin{aligned} \sum_{a \in Ad} \|\mu_a^{(n)}\|_2^2 &= \sum_{a \in Ad} \mathbb{P}(P_1(a) = P_2(a)) & (4.1) \\ &= |Ad| \mathbb{P}(P_1 = P_2) + \mathbb{E}(1_{\tilde{P} \text{ irred.}} N_p^{Ad}(P)) + O(n) \mathbb{P}(\tilde{P} \text{ reducible}). & (4.2) \end{aligned}$$

The  $O(n)$  on the right hand side comes from the trivial bound on the number of roots of  $P$  in  $\mathbb{F}_p$ .

We denote by  $\nu$  the law of  $x - y$ , where  $x$  and  $y$  are independent random variables with law  $\mu$ . This way, the coefficients of  $P = P_1 - P_2$  are i.i.d. and distributed according to  $\nu$ . Note that admissible roots of  $P$  in  $\mathbb{F}_p$  are roots of  $\tilde{P}$  in  $\mathbb{F}_p$ . Hence

$$N_p^{Ad}(P) \leq N_p(\tilde{P}),$$

where  $N_p(\tilde{P})$  is the number of distinct roots of  $\tilde{P}$  in  $\mathbb{F}_p$ . Moreover, we observe that  $\mathbb{P}(P_1 = P_2) = \mathbb{P}(P = 0) = \nu(0)^n = \|\mu\|_2^{2n}$ . The  $k$ -th cyclotomic polynomial is of degree  $\phi(k)$ , and if  $\phi(k) \leq 3n$ , then  $k = O(n \log \log n)$ . In particular there are at most  $O(n \log \log n)^2 = O(n^3)$

non-admissible roots in  $\mathbb{F}_p$ , so  $p/|Ad| \leq 1 + O(n^3/p)$ . In view of these observations, multiplying both sides of (4.2) by  $p/|Ad|$  yields

$$\mathbb{E}_a(p\|\mu_a^{(n)}\|_2^2) \leq p\|\mu\|_2^{2n} + \mathbb{E}(1_{\tilde{P} \text{ irred.}} N_p(\tilde{P})) + O(n^4/p) + O(n^4)\mathbb{P}(\tilde{P} \text{ reducible})$$

with absolute implied constants, where  $\mathbb{E}_a$  denotes the average over admissible residues.

Now using Theorem 9, we write

$$\mathbb{E}_a(p\|\mu_a^{(n)}\|_2^2) \leq p\|\mu\|_2^{2n} + \mathbb{E}(1_{\tilde{P} \text{ irred.}} N_p(\tilde{P})) + O(n^4 e^{-X}) + O(\exp(-O(\frac{n^{1/2}}{\log n}))), \quad (4.3)$$

where the implied constants depend only on  $\mu$ .

We now perform the average over primes  $p \in [e^X, (1 + \kappa)e^X]$ . The main point is the following well-known fact: *when  $p$  varies among primes, the average number of distinct roots lying in  $\mathbb{F}_p$  of a given non-constant polynomial  $Q \in \mathbb{Z}[x]$  converges to the number of irreducible factors of  $Q$ .* This is an instance of the prime ideal theorem for number fields and effective estimates on the error term can be given in terms of the height and degree of the polynomial and the location of zeroes of the Dedekind zeta functions of the associated number fields. This observation was at the basis of the proof of Theorem 9 which we gave in our previous article [2], and it will be used here as well.

If  $Q \in \mathbb{Z}[x]$  is irreducible over  $\mathbb{Q}$  and  $p$  is a prime not dividing the discriminant of  $Q$ , then the number  $N_p(Q)$  of roots of  $Q$  in  $\mathbb{F}_p$  coincides with the number  $A_p(Q)$  of prime ideals of norm  $p$  in the number field  $\mathbb{Q}[x]/(Q)$  (see e.g. [5, Theorem 4.8.13]). Moreover, under GRH, we have the following for all  $X > 0$

$$\sum_{p \in [e^X, (1+\kappa)e^X]} A_p(Q) \log(p) = \kappa e^X + O_\kappa(e^{X/2} X^2 \log |\Delta_Q|),$$

where  $\Delta_Q$  is the discriminant of  $Q$ . This is a classical result in analytic number theory originally due to Landau. For a proof in this exact form (when  $\kappa = 1$ ), see [2, Prop. 9].

We may replace  $A_p(Q)$  by  $N_p(Q)$  in the above sum without seriously worsening the error term, because at most  $O(\log |\Delta_Q|)$  primes can ever divide  $\Delta_Q$ , and both  $A_p(Q)$  and  $N_p(Q)$  are bounded by  $d = \deg(Q)$ . This implies that

$$\sum_{p \in [e^X, (1+\kappa)e^X]} N_p(Q) \log(p) = \kappa e^X + O_\kappa(e^{X/2} X^2 d \log |\Delta_Q|). \quad (4.4)$$

We can apply this estimate to  $Q = x$  to count ordinary primes and also to  $Q = \tilde{P}$ . Note that  $\Delta_{\tilde{P}}$  divides  $\Delta_P$  and that  $|\Delta_P| \leq (C_\mu n)^{2n}$ , where  $C_\mu$  is an upper bound on the absolute value of coefficients of  $P$ , i.e.  $C_\mu = \max\{|z|, \mu(z) \neq 0\}$ . Hence if  $\tilde{P}$  is irreducible, (4.4) yields

$$\mathbb{E}_p(N_p(\tilde{P})) = 1 + O_{\kappa, \mu}(e^{-X/2} X^2 n^3), \quad (4.5)$$

where we used the weighted average over primes in  $[e^X, (1 + \kappa)e^X]$ .

Combining this with (4.3), we get

$$\mathbb{E}_{a,p}(p\|\mu_a^{(n)}\|_2^2) \leq \mathbb{E}_p(p\|\mu\|_2^{2n}) + 1 + O_{\kappa,\mu}(e^{-X/2}X^2n^3 + n^4e^{-X}) + O_\mu(e^{-O_\mu(\frac{n^{1/2}}{\log n})})$$

and this yields the desired error term in Proposition 8.

## 5. PROOF OF CUT-OFF IN TOTAL VARIATION

We now pass to the proof of Theorem 3. The main idea is to use the Cauchy-Schwarz inequality to relate the total variation norm to the  $\ell^2$  norm, and then to modify the proof of the  $\ell^2$ -cut-off given in the previous section by discarding rare events from the sample space. This is performed via the classical Chernoff-Hoeffding inequality, which says that if  $X_1, \dots, X_n$  are independent zero mean real random variables taking values in an interval  $[a, b]$ , then for all  $t \geq 0$

$$\mathbb{P}\left(\frac{1}{n}|X_1 + \dots + X_n| \geq t\right) \leq 2 \exp\left(-\frac{2nt^2}{(b-a)^2}\right).$$

We denote by  $\mu^{(n)}$  the law of a random polynomial  $P$  as in (1.5) with independent integer random coordinates  $b_i$ 's with law  $\mu$ . Then setting  $X_i = -\log \mu(b_i) - H(\mu)$  gives that for any  $n, t > 0$  the subset  $A_n$  of all  $P \in \mathbb{Z}[x]$  such that

$$\left|\frac{1}{n} \log \mu^{(n)}(P) + H(\mu)\right| \leq t \tag{5.1}$$

has  $\mu^{(n)}$ -probability at least  $1 - 2e^{-2t^2n/C_\mu^2}$ , where

$$C_\mu := \log \sup_{x \in \text{Supp}(\mu)} \mu(x) - \log \inf_{x \in \text{Supp}(\mu)} \mu(x).$$

Let  $\mu_{|A_n}^{(n)}$  be the measure (possibly of mass  $< 1$ ) obtained by restricting  $\mu^{(n)}$  to the event  $A_n$ . Let  $\pi_a : \mathbb{Z}[x] \rightarrow \mathbb{F}_p$  be the evaluation map sending  $P$  to  $P(a)$ , so that  $\mu_a^{(n)} = \pi_a(\mu^{(n)})$ . We may write

$$\begin{aligned} \|\mu_a^{(n)} - u\|_1 &\leq \|\pi_a(\mu_{|A_n}^{(n)}) - u\|_1 + \|\pi_a(\mu_{|A_n}^{(n)} - \mu^{(n)})\|_1 \\ &\leq \|\pi_a(\mu_{|A_n}^{(n)}) - u\|_1 + \|\mu_{|A_n}^{(n)} - \mu^{(n)}\|_1 \\ &= \|\pi_a(\mu_{|A_n}^{(n)}) - u\|_1 + \mu^{(n)}(A_n^c). \end{aligned} \tag{5.2}$$

On the other hand, by Cauchy-Schwarz,

$$\|\pi_a(\mu_{|A_n}^{(n)}) - u\|_1^2 \leq p \|\pi_a(\mu_{|A_n}^{(n)}) - u\|_2^2. \tag{5.3}$$

We write

$$\begin{aligned} p \|\pi_a(\mu_{|A_n}^{(n)}) - u\|_2^2 &= p \|\pi_a(\mu_{|A_n}^{(n)})\|_2^2 - 2\mu^{(n)}(A_n) + 1 \\ &= p \|\pi_a(\mu_{|A_n}^{(n)})\|_2^2 - 1 + 2\mu^{(n)}(A_n^c). \end{aligned} \tag{5.4}$$

Combining (5.2), (5.3) and (5.4), we get

$$\|\mu_a^{(n)} - u\|_1^2 \leq p \|\pi_a(\mu_{|A_n}^{(n)})\|_2^2 - 1 + 5\mu^{(n)}(A_n^c). \quad (5.5)$$

Similarly to (4.1), we write

$$\sum_{a \in Ad} \|\pi_a(\mu_{|A_n}^{(n)})\|_2^2 = \mathbb{E}_{|A_n}^{(n)}(N_p^{Ad}(P)), \quad (5.6)$$

where, as before, we write  $P = P_1 - P_2$  for the difference of two independent copies  $P_1, P_2$  of a random polynomial with law  $\mu^{(n)}$ , and  $N_p^{Ad}(P)$  is the number of admissible roots of  $P$  in  $\mathbb{F}_p$ . Also,  $\mathbb{E}_{|A_n}^{(n)}(X)$  is a shorthand for  $\mathbb{E}^{(n)}(X \cdot 1_{A_n \times A_n})$  and  $\mathbb{E}^{(n)}$  is the expectation for the probability law  $\mu^{(n)} \times \mu^{(n)}$ . Writing  $\tilde{P}$  again for the quotient of  $P$  by all its monomial and cyclotomic factors, we have

$$\mathbb{E}_{|A_n}^{(n)}(N_p^{Ad}(P)) \leq |Ad| \mathbb{P}_{|A_n}^{(n)}(P = 0) + \mathbb{E}^{(n)}(1_{\tilde{P} \text{ irred.}} N_p^{Ad}(P)) + O(n) \mathbb{P}^{(n)}(\tilde{P} \text{ reducible}).$$

We plug this into (5.6), multiply both sides by  $p/|Ad| \leq 1 + O(n^3 e^{-X})$ , and apply Theorem 9 as before,

$$\mathbb{E}_a(p \|\pi_a(\mu_{|A_n}^{(n)})\|_2^2) \leq p \|\mu_{|A_n}^{(n)}\|_2^2 + \mathbb{E}^{(n)}(1_{\tilde{P} \text{ irred.}} N_p(\tilde{P})) + O(n^4 e^{-X}) + O(\exp(-O(\frac{n^{1/2}}{\log n}))),$$

where  $\mathbb{E}_a$  denotes the average over admissible residues, and the implied constants depend only on  $\mu$ . We average over primes as before, and using (4.5), we get

$$\mathbb{E}_{a,p}(p \|\pi_a(\mu_{|A_n}^{(n)})\|_2^2) \leq \mathbb{E}_p(p \|\mu_{|A_n}^{(n)}\|_2^2) + 1 + O_{\kappa,\mu}(e^{-X/2} X^2 n^4) + O_\mu(e^{-O_\mu(\frac{n^{1/2}}{\log n})}). \quad (5.7)$$

Now recall that  $n$  is the integer part of  $\frac{1}{H(\mu)}(X + \theta\sqrt{X})$ . We combine (5.5) and (5.7) to write

$$\mathbb{E}_{a,p}(\|\mu_a^{(n)} - u\|_1^2) \leq \mathbb{E}_p(p \|\mu_{|A_n}^{(n)}\|_2^2) + 5\mu^{(n)}(A_n^c) + O_{\kappa,\mu}(\theta^4 X^2 e^{-X/2}) + O_\mu(e^{-O_\mu(\frac{\sqrt{X}}{\log X})}). \quad (5.8)$$

Using (5.1),

$$\begin{aligned} p \|\mu_{|A_n}^{(n)}\|_2^2 &= \sum_{\omega \in A_n} p \mu^{(n)}(\omega)^2 \leq p e^{-(H(\mu)-t)n} \\ &\leq (1 + \kappa) e^X e^{-(H(\mu)-t)(\frac{1}{H(\mu)}(X + \theta\sqrt{X}) - 1)} \end{aligned}$$

We set  $t = \frac{\theta H(\mu)}{2\sqrt{X}}$  and obtain

$$p \|\mu_{|A_n}^{(n)}\|_2^2 = O_{\kappa,\mu}(e^{-\frac{\theta}{2}\sqrt{X}} e^{\frac{\theta^2}{2}}),$$

while

$$\mu^{(n)}(A_n^c) \leq 2e^{-2t^2 n / C_\mu^2} \leq 2e^{-\frac{H(\mu)\theta^2}{4C_\mu^2}}.$$

Combining the last two estimates with (5.8), we obtain

$$\mathbb{E}_{a,p}(\|\mu_a^{(n)} - u\|_1^2) \leq O_{\kappa,\mu}(e^{\frac{\theta^2}{2}} e^{-\frac{\theta}{2}\sqrt{X}}) + 10e^{-\frac{H(\mu)\theta^2}{4C_\mu^2}} + O_{\kappa,\mu}(\theta^4 X^2 e^{-X/2}) + O_\mu(e^{-O_\mu(\sqrt{X}/\log X)}). \quad (5.9)$$

Writing  $\alpha = 10e^{-\frac{H(\mu)\theta^2}{4C_\mu^2}}$ , we get

$$\mathbb{E}_{a,p}(\|\mu_a^{(n)} - u\|_1^2) \leq 2\alpha,$$

provided  $X > C(\theta^2 \log \theta)^2 + C$  for a suitably large  $C$  (dependent on  $\mu, \kappa$  only). Now applying the Markov inequality twice in exactly the same way as in the proof of Theorem 7, we find that for all but an  $\varepsilon$  proportion of the primes  $p$  in  $[e^X, (1 + \kappa)e^X]$ , and for all but an  $\varepsilon$  proportion of  $a \in \mathbb{F}_p$ , we have

$$\|\mu_a^{(n)} - u\|_1^2 \leq 8\alpha/\varepsilon^2.$$

Taking square roots and dividing by 2, we get the desired estimate. This ends the proof of Theorem 3.

*Proof of Corollary 4.* We begin with the proof of (1.4). Set  $t = H(\mu)\theta/\sqrt{\log p}$  and consider (5.1). We have

$$|\pi_a(A_n)| \min_{\omega \in A_n} \mu^{(n)}(\omega) \leq \mu^{(n)}(A_n) \leq 1.$$

Hence

$$|\pi_a(A_n)| \leq e^{n(H(\mu)+t)}.$$

Since  $n = \lfloor (\log p - \theta\sqrt{\log p})/H(\mu) \rfloor$ , this means that most of the walk is confined to a small part of  $\mathbb{F}_p$ , and hence

$$\begin{aligned} \|\mu_a^{(n)} - u\|_{TV} &\geq |\mu_a^{(n)}(\pi_a(A_n)^c) - u(\pi_a(A_n)^c)| \\ &\geq \frac{|\pi_a(A_n)^c|}{p} - \mu_a^{(n)}(\pi_a(A_n)^c) \\ &\geq 1 - \frac{1}{p} e^{n(H(\mu)+t)} - \mu^{(n)}(A_n^c) \\ &\geq 1 - \frac{1}{p} e^{(\log p - \theta\sqrt{\log p})H(\mu) - 1(H(\mu) + H(\mu)\theta/\sqrt{\log p})} - 2e^{-2t^2 n/C_\mu^2} \\ &\geq 1 - \frac{1}{p} e^{\log p - \theta^2} - 2e^{-2H(\mu)^2 \theta^2 (\log p)^{-1} \log p / 2H(\mu)C_\mu^2} \\ &\geq 1 - e^{-\theta^2} - 2e^{-\frac{H(\mu)\theta^2}{C_\mu^2}}. \end{aligned}$$

This establishes (1.4).

We now move on to the proof of Corollary 4. Set  $\theta = D_\mu \sqrt{\log \varepsilon^{-1}}$  for some constant  $D_\mu > 0$  to be determined later. First, the lower bound for  $T(\delta)$  follows easily from (1.4). Indeed, if  $n \leq (\log p - \theta\sqrt{\log p})/H(\mu)$ , then (1.4) implies that

$$\|\mu_a^{(n)} - u\|_{TV} \geq 1 - e^{-O_\mu(\theta^2)} \geq 1 - \varepsilon \geq \delta$$

provided  $D_\mu$  is large enough.

Now for the upper bound, set  $\kappa = 1$  and, for every integer  $m \geq 1$ ,  $X = \log 2^m$  in Theorem 3. Let  $E_m$  be the set of exceptional primes given by Theorem 3 in the interval  $[2^m, 2^{m+1}]$ . If  $p \notin E_m$  but  $\log 2^m \geq \theta^5$ , then Theorem 3 yields:

$$\|\mu_a^{(n)} - u\|_{TV} \leq 8\varepsilon^{-1} e^{-\frac{H(\mu)\theta^2}{16C_\mu^2}} \leq \varepsilon \leq \delta$$

provided  $D_\mu$  is large enough and  $n \geq (\log p + \theta\sqrt{\log p})/H(\mu)$ . This establishes the corollary with  $E := \cup_{m \geq \theta^5} E_m$ . Note that  $E$  has density at most  $\varepsilon$  among all primes, since  $E_m$  has density  $\leq \varepsilon$  among primes in  $[2^m, 2^{m+1}]$  for all  $m$ .  $\square$

We also state another variant of Corollary 4. Its proof is entirely similar, but in the application of Theorem 3, one needs to take  $\theta^2 = \sqrt{X}/(D_\mu \log X)$  for a suitable constant  $D_\mu$  and  $\varepsilon = \exp(H(\mu)\theta^2/32C_\mu^2)$ . We leave the details to the reader.

**Corollary 10.** *Assume GRH. There is an exceptional set of primes  $F$  of density 0, such that for all primes  $p \notin F$ , for every  $\eta \in (0, 1)$  and for all but an  $e^{-O_\mu(\sqrt{\log p}/\log \log p)}$ -fraction of all  $a \in \mathbb{F}_p$  we have:*

$$|T(\delta) - \frac{1}{H(\mu)} \log p| \leq \frac{\eta}{H(\mu)} \log p,$$

for all  $\delta \in (\eta, 1 - \eta)$ , provided  $\log p > C(\mu, \eta)$ .

Moreover, the density of the exceptional primes  $F$  can be estimated as

$$|\{p \in F \cap [1, e^X]\}| = O_\mu(e^{-O_\mu(\sqrt{X}/\log X)}) \frac{e^X}{X}.$$

## 6. UNCONDITIONAL UPPER BOUND

In this section, we prove Theorem 2. Its proof, which no longer assumes GRH, elaborates, on the one hand, on the bounds obtained in Section 4, and, on the other hand, on the multiplicity trick already used in Section 3. First we show the following.

**Proposition 11.** *Fix a probability measure  $\mu$  on  $\mathbb{Z}$  supported on a finite set of at least two elements, set  $H_2(\mu) = -\log \|\mu\|_2^2$  and consider the Markov chain (1.1). For  $X$  be a sufficiently large (depending on  $\mu$ ) number and let  $n$  be the integer part of  $\frac{1}{H_2(\mu)}X$ . Then*

$$\sum_{p \in [e^{X/2}, e^X]} \mathbb{E}_a(\|\mu_a^{(n)}\|_2^2) = O_\mu((\log X)^5),$$

where the average  $\mathbb{E}_a$  is taken over all admissible residues  $a \in \mathbb{F}_p$ .

Before we start the proof of this proposition, let us go back to the setting of Section 4, where we averaged over primes in order to get an upper bound on the  $\ell^2$  norm. Without GRH at our disposal, we can no longer guarantee that a random polynomial is a product of a

single irreducible polynomial together with monomial and cyclotomic factors. However, we can easily give an upper bound on the number of irreducible factors. Indeed, if  $P \in \mathbb{Z}[x]$  and  $P = \Phi \tilde{P}$ ,  $\tilde{P} = P_1 \dots P_k$ , where the  $P_i$ 's are the non-cyclotomic irreducible factors, then

$$M(P) = \prod_1^k M(P_i),$$

where  $M(P)$  is the Mahler measure of  $P$ , see (2.3) for the definition. By the classical Dobrowolski bound [7] on Mahler measure,

$$M(P_i) \geq e^{c(\frac{\log \log d_i}{\log d_i})^3} \geq e^{\frac{c}{(\log n)^3}},$$

where  $d_i = \deg P_i$  and  $c$  is an absolute constant. In particular,

$$M(P) \geq \exp(ck/(\log n)^3).$$

On the other hand,  $M(Q) \leq (\sum |a_i|^2)^{1/2}$  if  $Q = \sum_i a_i x^i \in \mathbb{Z}[x]$ . In our setting,  $P = P_1 - P_2$ , where  $P_1, P_2$  are distributed according to  $\mu^{(n)}$ , so  $|a_i| \leq 2H$  with  $H := \max\{|z|, \mu(z) \neq 0\}$ . Hence  $M(P) \leq 2H\sqrt{n}$ , which, compared with the previous bound gives

$$k = O_\mu((\log n)^4). \quad (6.1)$$

We may thus perform the average over primes, as in Section 4, in order to give a reasonably good upper bound on  $p\|\mu_a^{(n)}\|_2^2$ . To achieve this without GRH requires a slightly different kind of prime averaging as reflected in the statement of Theorem 2.

*Proof of Proposition 11.* Proceeding as in (4.3), we write

$$\mathbb{E}_a(p\|\mu_a^{(n)}\|_2^2) = p\|\mu\|_2^{2n} + \frac{p}{|Ad|} \mathbb{E}(1_{P \neq 0} N_p^{Ad}(P)),$$

where  $Ad \subset \mathbb{F}_p$  is the set of admissible residues, and  $P = P_1 - P_2$ , where  $P_1, P_2$  are distributed according to  $\mu^{(n)}$ , and  $P = \Phi \tilde{P}$ ,  $\tilde{P} = P_1 \dots P_k$ , where the  $P_i$ 's are the non-cyclotomic irreducible factors. Since  $p - |Ad| = O(n^3)$ , we get

$$\mathbb{E}_a(p\|\mu_a^{(n)}\|_2^2) = p\|\mu\|_2^{2n} + \mathbb{E}(1_{P \neq 0} N_p^{Ad}(P)) + O(n^4/p). \quad (6.2)$$

We are now in a position to sum over primes. Assuming GRH we had (4.4) with an excellent error term. Here we no longer assume GRH, but we can nevertheless make do with a weaker estimate. We have:

**Lemma 12.** *If  $Q \in \mathbb{Z}[x]$  is irreducible, of degree  $d$  and discriminant  $\Delta_Q$ , and if, as earlier,  $N_p(Q)$  denotes the number of distinct roots of  $Q$  in  $\mathbb{F}_p$ , then for all  $X > 0$ ,*

$$\sum_{p \in [e^{X/2}, e^X]} \frac{N_p(Q)}{p} = O\left(\frac{\log(2|\Delta_Q|)}{d} \left(1 + \frac{d^4}{X^4}\right)\right),$$

where the implied constant is absolute.

This lemma is unconditional and should be compared to (4.4), which assumed GRH. Recall that for all large enough  $X$ ,  $\sum_{p \in [e^{X/2}, e^X]} \frac{1}{p} \in [c, 1/c]$  for some absolute constant  $c \in (0, 1)$ , as follows directly from the prime number theorem or from Chebychev's bounds. Thus the lemma can be interpreted as a way to express the fact that  $N_p(Q)$  remains bounded by  $O(\log(2|\Delta_Q|)/d)$  on average over the primes in large windows  $[\sqrt{x}, x]$ ,  $\log x > d$ , where the average assigns a weight  $1/p$  to each prime  $p$ .

We postpone the proof of this lemma, and, first, explain how to complete the proof of Proposition 11. From (6.2), summing over primes in  $[e^{X/2}, e^X]$ , we get:

$$\sum_{p \in [e^{X/2}, e^X]} \mathbb{E}_a(\|\mu_a^{(n)}\|_2^2) \leq e^X \|\mu\|_2^{2n} + \mathbb{E}(1_{P \neq 0} \sum_{p \in [e^{X/2}, e^X]} \frac{N_p(\tilde{P})}{p}) + O(e^{-X/2} n^4). \quad (6.3)$$

Now, if  $\tilde{P} = P_1 \dots P_k$ , by Lemma 12, bearing in mind that  $n = \lfloor \frac{1}{H_2(\mu)} X \rfloor$ ,

$$\sum_{p \in [e^{X/2}, e^X]} \frac{N_p(\tilde{P})}{p} = \sum_{i=1}^k \sum_{p \in [e^{X/2}, e^X]} \frac{N_p(P_i)}{p} = \sum_{i=1}^k O_\mu\left(\frac{\log |\Delta_{P_i}|}{\deg P_i}\right).$$

On the other hand, by Mahler's bound [19, Theorem 1],

$$(\log |\Delta_{P_i}|) / \deg(P_i) \leq \log \deg(P_i) + 2 \log M(P_i).$$

But  $M(P_1) \dots M(P_k) = M(P) \leq 2H\sqrt{n}$ , where  $H := \max\{|z|, \mu(z) \neq 0\}$ . Hence we get

$$\begin{aligned} \sum_{p \in [e^{X/2}, e^X]} \frac{N_p(\tilde{P})}{p} &= \sum_{i=1}^k \sum_{p \in [e^{X/2}, e^X]} \frac{N_p(P_i)}{p} = \sum_{i=1}^k O_\mu\left(\frac{\log |\Delta_{P_i}|}{\deg P_i}\right) \\ &\leq O_\mu(1 + (k+1) \log n) = O_\mu((\log n)^5), \end{aligned}$$

where we used (6.1) in the last line. Plugging this back into (6.3), we obtain the desired estimate.  $\square$

*Proof of Lemma 12.* The proof is a straightforward variation of the proof of Proposition 13 in [2] and for this reason we shall be brief.

The argument relies on the explicit formula in analytic number theory relating the zeroes of the Dedekind zeta function and the count of prime ideals in a number field with explicit error terms. This is a standard tool in analytic number theory, which is extensively discussed in the literature in various settings. We will refer to [2], because it contains the formulae in the precise form we need them.

Let  $K$  be a number field,  $d$  its degree, and  $\Delta_K$  its discriminant. Let  $A(n)$  be the number of prime ideals with norm  $n$ . Given a  $C^2$

compactly supported function  $g$  on  $(0, +\infty)$ , and  $s \in \mathbb{C}$  we set

$$G(s) = \int_{\mathbb{R}} \exp(su)g(u)du.$$

The explicit formula (see e.g. [2, Theorem 10]) asserts that

$$\sum_{n,m \geq 1} A(n) \log(n)g(\log(n^m)) = G(1) - \sum_{\rho} G(\rho), \quad (6.4)$$

where the last sum on the right hand side extends over all zeroes  $\rho$  of the Dedekind zeta function of  $K$  counted with multiplicity.

First, we focus on estimating the right hand side of (6.4). We make the following choice for  $g$ . We set  $g(u) = e^{-u} \frac{8}{X} \chi^{*4}(\frac{8u}{X} - 6)$ , where  $\chi = 1_{[-\frac{1}{2}, \frac{1}{2}]}$  is the indicator function of the unit length interval  $[-\frac{1}{2}, \frac{1}{2}]$  and  $\chi^{*4}$  is its four-fold convolution product. This corresponds to the choice  $g = g_{X,k}$  with  $k = 4$  in [2, Lemma 11]. In that lemma, the following elementary properties of  $G(s)$  were verified for all complex  $s$  with  $\operatorname{Re}(s) \leq 1$ ,

$$|G(s)| \leq e^{(\operatorname{Re}(s)-1)X/2}, \quad (6.5)$$

$$|G(s)| \leq \left(\frac{16}{X|s-1|}\right)^4. \quad (6.6)$$

We also have at our disposal standard explicit bounds on the number (counting multiplicity) of zeroes of the zeta function  $\zeta_K$  near 1 and elsewhere. All zeros satisfy  $\operatorname{Re} \rho < 1$ , and for every  $0 \leq r \leq 1$ , we have

$$|\{\rho : \zeta_K(\rho) = 0, |1 - \rho| < r\}| \leq \frac{3}{2} + 3r \log |\Delta_K|, \quad (6.7)$$

while there is an absolute constant  $C > 0$ , such that for every  $r > 1$ ,

$$|\{\rho : \zeta_K(\rho) = 0, |1 - \rho| < r\}| \leq C \log |\Delta_K| + Cdr \log r. \quad (6.8)$$

For those we refer to [2, Lemma 12] or [20].

Note also that  $\log |2\Delta_K| \geq cd$  for some absolute constant  $c > 0$  by Minkowski's bound. Then combining (6.7) with (6.5), we get

$$\sum_{\rho, |1-\rho| < d^{-1}} |G(\rho)| = O\left(\frac{\log 2|\Delta_K|}{d}\right). \quad (6.9)$$

We now partition the remainder of the complex plane in annuli around 1 as follows for  $j \in \mathbb{Z}_{\geq 0}$

$$R_j := \{\rho : \zeta_K(\rho) = 0, 2^j d^{-1} \leq |1 - \rho| < 2^{j+1} d^{-1}\}.$$

Using (6.7) and (6.8), this yields that for all  $j \in \mathbb{Z}_{\geq 0}$ ,

$$|R_j| = O(4^j d^{-1} \log 2|\Delta_K|).$$

Now, by (6.6),

$$\sum_{\rho \in R_j, j \geq 0} |G(\rho)| \leq \sum_{j \geq 0} |R_j| \left(\frac{16}{X d^{-1} 2^j}\right)^4 \leq \sum_{j \geq 0} O(d^{-1} \log 2|\Delta_K|) 4^{-j} \left(\frac{16}{X d^{-1}}\right)^4.$$

Now, combining this estimate with (6.9), yields

$$|G(1) - \sum_{\rho} G(\rho)| \leq 1 + \sum_{\rho} |G(\rho)| = O\left(\frac{\log 2|\Delta_K|}{d}\left(1 + \frac{d^4}{X^4}\right)\right). \quad (6.10)$$

Now, we turn to the left hand side of (6.4). Observe that  $g$  is supported in the interval  $[X/2, X]$ , and that  $0 \leq g(u) \leq e^{-u}$  for all  $u > 0$ . Therefore, the main contribution in the above sum on the left hand side comes from prime ideals of prime norm. Indeed, it is shown in the proof of [2, Proposition 13] (see the bound (2.5) therein and subsequent estimates) that

$$\left| \sum_{n,m \geq 1} A(n) \log(n) g(\log(n^m)) - \sum_{p \text{ prime}} A(p) \log(p) g(\log(p)) \right| = O(dX e^{-X/4}), \quad (6.11)$$

where the implied constant is absolute.

We apply (6.4) to the number field  $K = \mathbb{Q}[x]/(Q)$ . We note that  $\Delta_K$  divides  $\Delta_Q$ , and, as already mentioned in Section 4,  $A(p)$  may differ from  $N_p(Q)$  only for those primes  $p$  that divide the discriminant  $\Delta_Q$ . This accounts for at most  $O(\log |\Delta_Q|)$  primes, hence:

$$\left| \sum_{p \text{ prime}} N_p(Q) \log(p) g(\log(p)) - \sum_{p \text{ prime}} A(p) \log(p) g(\log(p)) \right| = O(de^{-X/2} \log |\Delta_Q|) \quad (6.12)$$

Putting (6.4), (6.10), (6.11) and (6.12) together, we obtain

$$\sum_{p \text{ prime}} N_p(Q) \log(p) g(\log(p)) = O\left(\frac{\log 2|\Delta_Q|}{d}\left(1 + \frac{d^4}{X^4}\right)\right).$$

Now, observe from our definition of  $g$  that if  $p \in [e^{3X/5}, e^{4X/5}]$ , then  $g(\log p) \geq c \frac{1}{pX}$  for some absolute constant  $c > 0$ . Hence setting  $Y = 4X/5$ , we get

$$\sum_{p \in [e^{3Y/4}, e^Y]} \frac{N_p(Q)}{p} = O\left(\frac{\log 2|\Delta_Q|}{d}\left(1 + \frac{d^4}{Y^4}\right)\right).$$

Applying this bound to  $Y = X$ ,  $Y = 3X/4$  and  $Y = 9X/16$ , we conclude that, indeed,

$$\sum_{p \in [e^{X/2}, e^X]} \frac{N_p(Q)}{p} = O\left(\frac{\log 2|\Delta_Q|}{d}\left(1 + \frac{d^4}{X^4}\right)\right),$$

as desired. □

*Proof of Theorem 2.* Let  $C_\mu$  be the implied constant in the big- $O$  appearing in Proposition 11 and let  $B$  be the set of primes  $p$  such that

$$\mathbb{E}_a(p) \|\mu_a^{(\lfloor 2 \frac{1}{H_2(\mu)} \log p \rfloor)}\|_2^2 > 4C_\mu \sqrt{\log p} (\log \log p)^{-1}.$$

By Proposition 11, using that  $n \mapsto \|\mu_a^{(n)}\|_2$  is non-increasing, we get

$$\sum_{p \in [e^{X/2}, e^X]} \frac{1_{p \in B}}{p} \leq \frac{(\log X)^6}{\sqrt{X}}.$$

Let  $p \notin B$  and set  $n = \lfloor 2 \frac{1}{H_2(\mu)} \log p \rfloor$ . Then

$$\mathbb{E}_a(p \|\mu_a^{(n)}\|_2^2) \leq 4C_\mu \sqrt{\log p} (\log \log p)^{-1}.$$

In particular, by the Markov inequality, for all but an  $\varepsilon/2$ -fraction of all admissible residues, and hence for all but an  $\varepsilon$ -fraction of all residues  $a \in \mathbb{F}_p$ , we have

$$p \|\mu_a^{(n)}\|_2^2 \leq 8\varepsilon^{-1} C_\mu \sqrt{\log p} (\log \log p)^{-1}.$$

Using (3.6), with  $k = 1$ , and  $m = \lfloor \frac{1}{H_2(\mu)} \log p \rfloor$ , we may now bound the right hand side, and obtain

$$\|\mu_a^{(\lfloor 5 \frac{1}{H_2(\mu)} \log p \rfloor)} - u\|_1^2 \leq p \|\mu_a^{(2n+m)} - u\|_2^2 = O_\mu\left(\frac{\varepsilon^{-2}}{(\log \log p)^2}\right),$$

and the claimed bounds on mixing times follow.  $\square$

## 7. CONCLUDING REMARKS

**7.1. Mixing and cut-off in  $\ell^q$ .** It is natural to ask about mixing times  $T_q(\delta)$  in  $\ell^q$ -norm for any  $q \in [1, +\infty]$ . Here

$$T_q(\delta) := \inf\{n \in \mathbb{N}, \|\mu_a^{(n)} - u\|_q \leq \delta \|u\|_q\}.$$

As is well-known, by convexity, the function  $q \mapsto \|\mu_a^{(n)} - u\|_q / \|u\|_q$  is non-decreasing. In particular,  $q \mapsto T_q(\delta)$  is non-decreasing (see e.g. [18, Section 4.7]). On the other hand, Cauchy-Schwarz implies that  $\|\mu_a^{(2n)} - u\|_\infty \leq \|\mu_a^{(n)} - u\|_2^2$  for all  $n$ . And it follows that for all  $q \in [1, +\infty]$  we have

$$T_q(\delta) \leq 2T_2(\delta^2). \quad (7.1)$$

This means that we can apply our main theorems to get a good upper bound on  $T_q(\delta)$ .

Regarding lower bounds, we have, as before, the trivial lower bound  $\|\mu_a^{(n)} - u\|_q / \|u\|_q \geq \|\mu^{(n)}\|_q / \|u\|_q - 1 = \|\mu\|_q^n / \|u\|_q - 1$ , which implies that

$$T_q(\delta) \geq \frac{1}{H_q(\mu)} (\log p - O_q(\delta)),$$

where  $H_q(\mu) = \frac{q}{q-1} \log \|\mu\|_q^{-1}$  is the Rényi entropy of order  $q$  of  $\mu$ .

If  $\mu$  is uniform (i.e.  $\mu(x) = \|\mu\|_\infty$  for all  $x \in \text{Supp}(\mu)$ ), then  $H_q(\mu) = H_\infty(\mu) = H(\mu)$  for all  $q$ . In particular, we see that mixing in  $\ell^1$  and in  $\ell^2$  occur at the same time, and hence so does mixing in  $\ell^q$  norm for any  $q \in [1, 2]$ . Consequently, in this situation, cut-off happens in  $\ell^q$ -norm in the setting of Corollary 4 for instance. However, when  $q > 2$ , it seems less clear how to establish an analogous cut-off in  $\ell^q$ -norm.

**7.2. The exceptional set of multipliers.** As mentioned in the introduction, the multiplier  $a$  must be of large enough multiplicative order in  $\mathbb{F}_p^\times$  for the chain (1.1) to equidistribute in  $O(\log p)$  steps, let alone satisfy a cut-off at  $(\log p)/H(\mu)$  as in Corollary 4. Even a multiplicative order of size more than  $\log p$  is not enough; indeed, Chung-Diaconis-Graham [4] show that when  $a = 2$  and  $\mu$  is uniform on  $\{-1, 0, 1\}$ , if  $p$  is of the form  $2^n - 1$ , then at least  $cn \log n$  steps are necessary for mixing to occur. In this case,  $a$  has order  $n \simeq \log p$ . Their example can easily be modified to obtain exceptional  $a$ 's with order at least  $(\log p)^2$ .

Now, as far as cut-off at  $(\log p)/H(\mu)$  is concerned, the exceptional set of multipliers is even larger, and contains residues of large multiplicative order. Indeed, as remarked in [13], when  $a = 2$ , the support of  $\mu_a^{(n)}$  has at most  $2^{n+1}$  elements, and hence for every fixed  $\delta > 0$ , for all  $n \leq (1 - \delta) \log_2(p)$ ,  $\mu_a^{(n)}$  is far from uniform in total variation. However, if  $\mu$  is uniform on  $\{-1, 0, 1\}$ , as above, then  $H(\mu) = \log 3 > \log 2$ . This means that the value  $a = 2$  is exceptional (i.e. belongs to the  $\varepsilon$ -fraction of bad multipliers in Theorem 3 and Corollary 4) for all primes  $p$ . This happens in general whenever the residue  $a \in \mathbb{F}_p$  has a representative in  $\mathbb{Z}$  of size  $< \exp H(\mu)$  (or more generally if  $a$  is the reduction modulo a prime ideal of norm  $p$  of a fixed algebraic number with Mahler measure  $< \exp H(\mu)$ ). On the other hand, it is well-known and easy to check that apart from a density zero family of primes, any given  $a \in \mathbb{Z} \setminus \{-1, 0, 1\}$  has multiplicative order at least  $c\sqrt{p}/\log p$ , and it is also known modulo GRH that given any function  $\varepsilon(p) > 0$  tending to 0 when  $p \rightarrow +\infty$ ,  $a$  has multiplicative order even at least  $\varepsilon(p)p$  for a density one set of primes, see [8].

**7.3. Diameter bounds.** The support of the distribution is the following subset of  $\mathbb{F}_p$

$$\text{Supp}(\mu_a^{(n)}) = \{b_0 + b_1 a + \dots + b_{n-1} a^{n-1}, b_i \in \text{Supp}(\mu)\} \subset \mathbb{F}_p.$$

It is of arithmetical interest to determine an upper bound on the diameter  $D_a(p)$ , i.e. the first  $n$ , for which the support is all of  $\mathbb{F}_p$ . Clearly  $D_a(p)$  is bounded above by the uniform mixing time  $T_\infty(\delta)$  for any  $\delta < 1$ . But as recalled in (7.1),  $T_\infty(\delta) \leq 2T_2(\delta^2)$ , which is a general fact about Markov chains. In particular, the following is true.

**Proposition 13.** *For all primes  $p$  and all  $a \in \mathbb{F}_p$  we have*

$$\text{Supp}(\mu_a^{(n)}) = \mathbb{F}_p$$

*for all  $n > 2T_2(1)$ . In other words  $D_a(p) \leq 2T_2(1) + 1$ .*

We may thus apply the upper bounds on  $T_2(\delta)$  given by Theorems 1, 2, 5 or 7 to deduce corresponding diameter bounds in each setting.

**7.4. Connections with Lehmer's problem.** Recall that the Mahler measure of polynomial  $P \in \mathbb{Z}[x]$  of degree  $n$  and dominant term  $a_n$  is the real number

$$M(P) = |a_n| \prod_1^n \max\{1, |\alpha_i|\}$$

where the  $\alpha_i$ 's are the complex roots of  $P$ . The celebrated Lehmer problem [16] asks whether there is an absolute constant  $\varepsilon_0 > 0$  such that

$$M(P) > 1 + \varepsilon_0$$

for every polynomial  $P$  that is not a product of cyclotomic or monomial factors.

In [3], we gave an equivalent formulation of the Lehmer problem, and proved that it is equivalent to the existence of a constant  $\delta_0 > 0$  such that for a density one set of primes  $p$ , for every  $a \in \mathbb{F}_p^\times$  of multiplicative order at least  $(\log p)^2$  we have

$$|\text{Supp}(\mu_a^{(\lfloor \log p \rfloor)})| \geq p^{\delta_0}. \quad (7.2)$$

In particular, we see that if Lehmer's problem has a negative answer, then there is a function  $B(p)$  with  $B(p) \rightarrow +\infty$  as  $p \rightarrow +\infty$  such that for a density one set of primes, there will be residues  $a \in \mathbb{F}_p^\times$  with multiplicative order at least  $(\log p)^2$ , (we can in fact go up to  $p^{1/2-o(1)}$  and even up to any  $o(p)$  assuming GRH), for which the diameter, and hence the mixing times, are at least  $B(p) \log p$ . This justifies the claim made in the introduction that an  $O(\log p)$  mixing time for almost all primes and all residues of large enough multiplicative order would imply the Lehmer conjecture.

We do not know about the converse. Konyagin's universal upper bound from Theorem 5 (whose proof uses Dobrowolski's bound towards the Lehmer problem) does not seem to improve significantly (at least not its quadratic nature in  $\log p$ ) assuming a positive answer to Lehmer. With such an assumption, the best we can do at the moment is the bound (7.2) or a similar lower bound on  $H(\mu_a^{(\log p)})$ , but this is not quite enough to bound from above the mixing time.

**7.5. The associated reversible chain.** The chain (1.1) is not reversible. It is natural to ask what happens to the symmetrized chain, whose transition probabilities are

$$q_a(x, y) := \frac{1}{2}(p_a(x, y) + p_a(y, x)),$$

where  $x, y \in \mathbb{F}_p$  and  $p_a(x, y) = \mu(\{b \in \mathbb{Z}, y = ax + b \pmod{p}\})$  is the transition probability of the chain (1.1). It would be interesting to perform for the symmetrized chain  $q_a$  a similar analysis as the one we did for  $p_a$  in this paper.

It is clear, however, that  $q_a$  is of a different nature compared to  $p_a$ , as it will typically take much longer to mix. To see this, observe that it is a quotient of the simple random walk on the wreath product  $\mathbb{Z} \wr \mathbb{Z}$ , and its distribution at time  $n$  is  $\nu_a^{(n)} = \pi_a(\nu^{(n)})$ , where  $\pi_a : \mathbb{Z}[x, x^{-1}] \rightarrow \mathbb{F}_p$  is the evaluation map  $P \mapsto P(a)$  and  $\nu^{(n)}$  is the associated law on the Laurent polynomials  $\mathbb{Z}[x, x^{-1}]$ . In particular,

$$p \|\nu_a^{(n)} - u\|_2^2 = p \|\nu_a^{(n)}\|_2^2 - 1 \geq p \|\nu^{(n)}\|_2^2 - 1.$$

But  $\|\nu^{(n)}\|_2^2$  can be bounded below by the return probability of a certain symmetric random walk on  $\mathbb{Z} \wr \mathbb{Z}$ , which is of order  $e^{-n^{1/3}(\log n)^{2/3}}$  ([9, Thm 2]). It follows that the  $\ell^2$  mixing time of the symmetrized chain is at least  $(\log p)^{3-o(1)}$ .

**Acknowledgements.** It is a pleasure to thank Boris Bukh, Persi Diaconis, Jonathan Hermon and Ariel Rapaport for helpful conversations. The first named author is grateful to the Newton Institute in Cambridge for its hospitality while part of this work was conducted. We would also like to thank the referee for their very careful examination of our manuscript.

## REFERENCES

- [1] E. Breuillard and P. P. Varjú. Entropy of bernoulli convolutions and uniform exponential growth for linear groups, 2018. arXiv:1510.04043v3, to appear in *J. d'Analyse Math.* [↑5](#)
- [2] ———. Irreducibility of random polynomials of large degree, 2018. arXiv:1810.13360. [↑5, 6, 7, 8, 13, 14, 20, 21, 22](#)
- [3] ———. On the Lehmer conjecture and counting in finite fields. *Discrete Analysis* 5, 2019. [↑2, 5, 25](#)
- [4] F. R. K. Chung, P. Diaconis, and R. L. Graham. Random walks arising in random number generation. *Ann. Probab.* 15 (3):1148–1165, 1987. [↑2, 10, 24](#)
- [5] H. Cohen. *A course in computational algebraic number theory*. Graduate Texts in Mathematics, vol. 138. Springer-Verlag, Berlin, 1993. [↑14](#)
- [6] P. Diaconis. The cutoff phenomenon in finite Markov chains. *Proc. Nat. Acad. Sci. U.S.A.* 93 (4):1659–1664, 1996. [↑3](#)
- [7] E. Dobrowolski. On a question of Lehmer and the number of irreducible factors of a polynomial. *Acta Arith.* 34 (4):391–401, 1979. [↑8, 19](#)
- [8] P. Erdős and M. R. Murty. On the order of  $a \pmod{p}$ . In *Number theory (Ottawa, ON, 1996)*, pages 87–97. 1999. [↑24](#)
- [9] A. Erschler. Isoperimetry for wreath products of Markov chains and multiplicity of selfintersections of random walks. *Probab. Theory Related Fields* 136 (4):560–586, 2006. [↑26](#)
- [10] H. A. Helfgott. Growth and expansion in algebraic groups over finite fields, 2019. arXiv:1902.06308v3. [↑2, 7](#)
- [11] M. Hildebrand. Random processes of the form  $X_{n+1} = a_n X_n + b_n \pmod{p}$ . *Ann. Probab.* 21 (2):710–720, 1993. [↑2](#)
- [12] ———. On the Chung-Diaconis-Graham random process. *Electron. Comm. Probab.* 11:347–356, 2006. [↑2](#)
- [13] ———. On a lower bound for the Chung-Diaconis-Graham random process. *Statist. Probab. Lett.* 152:121–125, 2019. [↑24](#)

- [14] M. V. Hildebrand. *Rates of convergence of some random processes on finite groups*. ProQuest LLC, Ann Arbor, MI, 1990. Thesis (Ph.D.)—Harvard University. [↑2](#)
- [15] S. V. Konyagin. Estimates for Gaussian sums and Waring’s problem modulo a prime. *Trudy Mat. Inst. Steklov.* 198:111–124, 1992. [↑2](#), [7](#), [8](#)
- [16] D. H. Lehmer. Factorization of certain cyclotomic functions. *Ann. of Math. (2)* 34 (3):461–479, 1933. [↑25](#)
- [17] ———. Mathematical methods in large-scale computing units. In *Proceedings of a Second Symposium on Large-Scale Digital Calculating Machinery*, 1949, pages 141–146. 1951. [↑2](#)
- [18] D. A. Levin and Y. Peres. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2017. Second edition. [↑3](#), [23](#)
- [19] K. Mahler. An inequality for the discriminant of a polynomial. *Michigan Math. J.* 11:257–262, 1964. [↑20](#)
- [20] H. M. Stark. Some effective cases of the Brauer-Siegel theorem. *Invent. Math.* 23:135–152, 1974. [↑21](#)

CENTRE FOR MATHEMATICAL SCIENCES, WILBERFORCE ROAD, CAMBRIDGE  
CB3 0WB, UK

*Email address:* `emmanuel.breuillard@maths.cam.ac.uk`

CENTRE FOR MATHEMATICAL SCIENCES, WILBERFORCE ROAD, CAMBRIDGE  
CB3 0WB, UK

*Email address:* `pv270@dpms.cam.ac.uk`