An Erdős–Kac law for local solubility in families of varieties

D. Loughran[1] · E. Sofos[2]

Abstract

We study probability distributions arising from local obstructions to the existence of p-adic points in families of varieties. In certain cases we show that an Erdős–Kac type normal distribution law holds.

Mathematics Subject Classification 14G05 · 60F05 · 14D10 · 11N36

Contents

1 Introduction

1.1 A central limit theorem for fibrations

Let V be a smooth projective variety over $\mathbb{Q}$ equipped with a dominant morphism $\pi : V \to \mathbb{P}^n$ with geometrically integral generic fibre. We view π as defining a family of varieties given by the fibres of π.

A natural problem is to study the distribution of the varieties in the family with a rational point. In the case of families of conics, this problem was studied by Serre

✉ E. Sofos
efthymios.sofos@glasgow.ac.uk

D. Loughran
dtl32@bath.ac.uk

1   Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, UK

2   The Mathematics and Statistics Building, University of Glasgow, G12 8QQ Glasgow, UK

Ⓝ Birkhäuser

[20]. He obtained precise upper bounds for the counting function

$$N(\pi, B) := \sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : x \in \pi(V(\mathbb{Q})), H(x) \leq B\}$$

of the number of varieties in the family with a rational point. (Here H is the usual naive height on $\mathbb{P}^n(\mathbb{Q})$). For example, his results show that for families of conics if π admits no section over $\mathbb{Q}$ then $N(\pi, B) = o(B^{n+1})$, i.e. 100% of the fibres of π have no rational point. He did this by showing that 100% of the fibres fail to be p-adically soluble for all primes p.

This subject has been studied in various settings by different authors; the reader is referred to [13,14] for a history of the subject. The main result from [14] generalised Serre's result from families of conics to arbitrary families of varieties $\pi : V \to \mathbb{P}^n$ for any smooth projective variety V. Again the authors considered the closely related problem of counting the number of varieties in the family which are everywhere locally soluble. They proved an upper bound of the shape

$$\sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : x \in \pi(V(\mathbf{A}_{\mathbb{Q}})), H(x) \leq B\} \ll \frac{B^{n+1}}{(\log B)^{\Delta(\pi)}}, \tag{1.1}$$

for an explicit non-negative $\Delta(\pi) \in \mathbb{Q}$. (Here $\mathbf{A}_{\mathbb{Q}}$ denotes the adeles of $\mathbb{Q}$.) Moreover, they conjectured in [14, Conj. 1.6] that the upper bound (1.1) is sharp, under the necessary assumptions that the set being counted is non-empty and that the fibre over every codimension 1 point of $\mathbb{P}^n$ contains an irreducible component of multiplicity 1. As it will occur frequently in our results, we recall the definition of $\Delta(\pi)$ here.

Definition 1.1 Let $\pi : V \to X$ be a dominant proper morphism of smooth irreducible varieties over a field k of characteristic 0. For each (scheme-theoretic) point $x \in X$ with residue field $\kappa(x)$, the absolute Galois group $\mathrm{Gal}(\overline{\kappa(x)}/\kappa(x))$ of the residue field acts on the irreducible components of $\pi^{-1}(x)_{\overline{\kappa(x)}} := \pi^{-1}(x) \times_{\kappa(x)} \overline{\kappa(x)}$ of multiplicity 1. We choose some finite group Γ_x through which this action factors. Then we define

$$\delta_x(\pi) = \frac{\sharp\left\{\gamma \in \Gamma_x : \begin{array}{l} \gamma \text{ fixes an irreducible component} \\ \text{of } \pi^{-1}(x)_{\overline{\kappa(x)}} \text{ of multiplicity 1} \end{array}\right\}}{\sharp\Gamma_x}$$

and $\Delta(\pi) = \sum_{D \in X^{(1)}} (1 - \delta_D(\pi)),$

where $X^{(1)}$ denotes the set of codimension 1 points of X.

These invariants are defined by group theoretic data which can often be calculated in practice. In this paper we consider the following problem which is closely related to Serre's:

Given a family of varieties $\pi : V \to \mathbb{P}^n$ and $j \in \mathbb{Z}_{\geq 0}$, what is the distribution of varieties in the family which fail to have a p-adic point for exactly j primes p?

To study this problem, for $x \in \mathbb{P}^n(\mathbb{Q})$ we consider the function

$$\omega_\pi(x) := \sharp\{\text{primes } p : \pi^{-1}(x)(\mathbb{Q}_p) = \emptyset\}. \tag{1.2}$$

Note that $\omega_\pi(x)$ need not be finite in general; however it is finite if $\pi^{-1}(x)$ is geometrically integral, as follows from the Lang–Weil estimates [16] and Hensel's lemma. If the generic fibre of π is geometrically integral, then $\omega_\pi(x)$ is finite for all x outside of some proper Zariski closed set. (In practice, we restrict to the smooth fibres of π.) One can also consider variants of the function ω_π from (1.2), by considering real solubility or by dropping conditions at finitely many primes; we discuss this possibility in Sect. 4.6.

As is clear from (1.1), if $\Delta(\pi) > 0$ then the function $\omega_\pi(x)$ is almost always positive. Our first result gives more specific information about the distribution of $\omega_\pi(x)$, and is an analogue of the Erdős–Kac theorem [9] in our setting. Recall that this states that the function

$$\omega(m) := \sharp\{\text{primes } p : p \mid m\} \tag{1.3}$$

behaves likes a normal distribution with mean and variance $\log\log n$; more formally, for every interval $\mathcal{J} \subset \mathbb{R}$ one has

$$\lim_{x\to\infty} \frac{1}{x}\sharp\left\{1 \le m \le x : \frac{\omega(m) - \log\log m}{\sqrt{\log\log m}} \in \mathcal{J}\right\} = \frac{1}{\sqrt{2\pi}}\int_{\mathcal{J}} e^{-\frac{t^2}{2}}\,dt.$$

This theorem is one of the foundational results in probabilistic number theory.

For our analogue, we need some notation. For each $B \in \mathbb{R}_{\ge 1}$ and $A \subseteq \mathbb{P}^n(\mathbb{Q})$ we define

$$\nu_B(A) := \frac{\sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \le B, x \in A\}}{\sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \le B\}}. \tag{1.4}$$

If $\lim_{B\to\infty} \nu_B(A)$ exists then its value is to be conceived as the "density" of A. Our result is the following. (Here, and in what follows, we also commit the minor abuse of implicitly excluding the finitely many rational points x with $\log H(x) \le 1$.)

Theorem 1.2 *Let V be a smooth projective variety over $\mathbb{Q}$ equipped with a dominant morphism $\pi : V \to \mathbb{P}^n$ with geometrically integral generic fibre and $\Delta(\pi) \ne 0$. Let H be the usual naive height on $\mathbb{P}^n$. Then for any interval $\mathcal{J} \subset \mathbb{R}$ we have*

$$\lim_{B\to\infty} \nu_B\left(\left\{x \in \mathbb{P}^n(\mathbb{Q}) : \frac{\omega_\pi(x) - \Delta(\pi)\log\log H(x)}{\sqrt{\Delta(\pi)\log\log H(x)}} \in \mathcal{J}\right\}\right) = \frac{1}{\sqrt{2\pi}}\int_{\mathcal{J}} e^{-\frac{t^2}{2}}\,dt.$$

Note that the probability distribution obtained only depends on the invariant $\Delta(\pi)$ from Definition 1.1; the geometric properties of the smooth members of the family are irrelevant. A measure-theoretic interpretation of Theorem 1.2 is as follows: It says that

$$\mathcal{J} \mapsto \lim_{B\to\infty} \nu_B\left(x; \frac{\omega_\pi(x) - \Delta(\pi)\log\log H(x)}{\sqrt{\Delta(\pi)\log\log H(x)}} \in \mathcal{J}\right)$$

defines a probability measure on $\mathbb{R}$ which equals the standard Gaussian measure. Informally, it says $\omega_\pi(x)$ is normally distributed with mean and variance $\Delta(\pi) \log \log H(x)$.

Theorem 1.2 is proved by studying the moments

$$\mathcal{M}_r(\pi, B) := \sum_{\substack{x \in \mathbb{P}^n(\mathbb{Q}), H(x) \le B \\ \pi^{-1}(x) \text{ smooth}}} \left(\frac{\omega_\pi(x) - \Delta(\pi) \log \log B}{\sqrt{\Delta(\pi) \log \log B}} \right)^r, \quad (r \in \mathbb{Z}_{\ge 0}). \quad (1.5)$$

Theorem 1.3 *Keep the assumptions of Theorem 1.2. Then for each $r \in \mathbb{Z}_{\ge 0}$ we have*

$$\frac{\mathcal{M}_r(\pi, B)}{\sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \le B\}} = \mu_r + O_r\left(\frac{\log \log \log \log B}{(\log \log B)^{1/2}}\right), \quad \textit{where}$$

$$\mu_r := \begin{cases} \frac{r!}{2^{r/2}(r/2)!}, & r \text{ even,} \\ 0, & r \text{ odd.} \end{cases}$$

Here μ_r is the r-th moment of the standard normal distribution. Our main tool in the proof of Theorem 1.3 is the result of Granville and Soundararajan [11]. Theorem 1.2 is proved from Theorem 1.3 via a standard argument, which rests on the fact that the normal distribution is determined by its moments

There are general conditions under which one can prove an Erdős–Kac law for certain additive functions defined on $\mathbb{Z}^{n+1}$, see [10, §12] for example. In principle, these results could be extended to cover additive arithmetic functions restricted to values of a general polynomial, i.e. $\sum_{p|f(x)} h(p)$ for integer polynomials f and functions h of certain growth over the primes; see the work of Xiong [25]. However, ω_π does not admit any such interpretation, as the following example shows.

Example 1.4 Consider the following family of conics

$$ax^2 + by^2 = cz^2 \quad \subset \mathbb{P}^2 \times \mathbb{P}^2,$$

equipped with the projection π to $(a : b : c)$. Take $(a, b, c) \in \mathbb{Z}^3$ pairwise coprime, square-free and all congruent to 1 mod 4. A Hilbert symbol calculation shows that

$$\omega_\pi(a : b : c) = \left(\frac{1}{2} \sum_{p|a} \left(1 - \left(\frac{bc}{p}\right)\right)\right) + \left(\frac{1}{2} \sum_{p|b} \left(1 - \left(\frac{ac}{p}\right)\right)\right)$$

$$+ \left(\frac{1}{2} \sum_{p|c} \left(1 - \left(\frac{-ab}{p}\right)\right)\right), \quad (1.6)$$

where $\left(\frac{\cdot}{p}\right)$ is the Legendre symbol (cf. [12, p. 13]). One cannot directly apply the aforementioned general results here, since the function in (1.6) is not the restriction of an additive function to the values of a polynomial. Nevertheless Theorem 1.2 implies that the function ω_π has normal order $\frac{3}{2} \log \log H(a : b : c)$ in this case.

We also give an application of our results to a family of curves of genus 1.

Example 1.5 Let $c, d \in \mathbb{Z}$ be such that $cd(c - d) \neq 0$ and let $f(t) \in \mathbb{Z}[t]$ be a square-free polynomial of even degree. Consider the variety

$$W : \quad x^2 - cw^2 = f(t)y^2, \quad x^2 - dw^2 = f(t)z^2 \quad \subset \mathbb{P}^2 \times \mathbb{A}^1.$$

Let $\pi : V \to \mathbb{P}^1$ be a non-singular compactification of the natural projection $W \to \mathbb{P}^1$ to the t-coordinate. The generic fibre of π is a smooth intersection of two quadrics in $\mathbb{P}^3$, hence is a genus 1 curve. The singular fibres occur over the closed points corresponding to the irreducible polynomials dividing f. Moreover, by [8, Prop. 4.1], the fibre over every such closed point P is a double fibre, hence $\delta_P(\pi) = 0$. Theorem 1.2 therefore implies that ω_π has normal order $r(f) \log \log H(1 : t)$ in this case, where $r(f)$ is the number of irreducible polynomials dividing f.

This last example is particularly interesting, as the upper bound (1.1) is conjecturally sharp only if the fibre over every codimension 1 point contains an irreducible component of multiplicity 1. No such assumptions are required in the statements of our theorems.

The next example illustrates how to (essentially) recover the usual ω (1.3) as a special case of our ω_π.

Example 1.6 Let V be a smooth projective variety over $\mathbb{Q}$ equipped with a dominant morphism $\pi : V \to \mathbb{P}^1$ such that:

(1) The fibre over $(0 : 1)$ has multiplicity $m > 1$, i.e. we have $\pi^*((0 : 1)) = mD$ for some divisor D on V.
(2) All other fibres are geometrically integral.

Examples of such varieties are "unnodal Halphen surfaces of index m" [6, §2].

Let now (x_0, x_1) be a primitive integer vector and $P = (x_0 : x_1) \in \mathbb{P}^1(\mathbb{Q})$. Then our methods will yield the existence of some $A > 0$ such that for all primes $p > A$ we have

$$v_p(x_0) = 0 \implies \pi^{-1}(P)(\mathbb{Q}_p) \neq \emptyset, \quad v_p(x_0) = 1 \implies \pi^{-1}(P)(\mathbb{Q}_p) = \emptyset,$$

where v_p denotes the p-adic valuation. Thus if x_0 is square-free and $p \nmid x_0$ for all $p \leq A$, then $\omega_\pi(P) = \omega(x_0)$. (We shall see that small primes and primes of higher multiplicity do not effect the overall probabilistic behaviour, so our results essentially recover the usual Erdős–Kac theorem.)

1.2 The pseudo-split case

Our results from Sect. 1.1 only apply when $\Delta(\pi) \neq 0$. It turns out that a normal distribution does not hold when $\Delta(\pi) = 0$. We refer to the case $\Delta(\pi) = 0$ as the "pseudo-split case". This is because the condition $\Delta(\pi) = 0$ is equivalent to the condition that the fibre over every codimension 1 point of $\mathbb{P}^n$ is *pseudo-split*, in the sense of [15, Def. 1.3]. The pseudo-split case is interesting from an arithmetic

perspective, as these are exactly the families of varieties for which a positive proportion of the fibres can be everywhere locally soluble (see [14, Thm. 1.3]).

In the pseudo-split case there is a discrete probability distribution, in a sense that is made precise in the following theorem. For $j \in \mathbb{Z}_{\geq 0}$ and $B \geq 1$ we define

$$\tau_\pi(j, B) := \frac{\sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \leq B, \pi^{-1}(x) \text{ smooth}, \omega_\pi(x) = j\}}{\sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \leq B\}}. \quad (1.7)$$

Theorem 1.7 *Let V be a smooth projective variety over $\mathbb{Q}$ equipped with a dominant morphism $\pi : V \to \mathbb{P}^n$ with geometrically integral generic fibre and $\Delta(\pi) = 0$. Then*

$$\tau_\pi : \mathbb{Z} \to \mathbb{R}, \quad j \mapsto \tau_\pi(j) := \lim_{B \to \infty} \tau_\pi(j, B) \quad (1.8)$$

is well-defined and defines a probability measure on $\mathbb{Z}$. Moreover, for every $j \in \mathbb{Z}_{\geq 0}$ we have the following upper bound

$$\tau_\pi(j) \ll_\pi \frac{1}{(1+j)^j (\log(2+j))^{j/2}}, \quad (1.9)$$

where the implied constant depends at most on π.

One way to interpret Theorem 1.7 is that $\omega_\pi(x)$ has a *limit law*. A limit law is originally defined for functions defined in the integers (see [24, Def. 2.2, p. 427]), however, the definition easily extends to functions defined in $\mathbb{P}^n(\mathbb{Q})$: We say that a function $f : \mathbb{P}^n(\mathbb{Q}) \to \mathbb{R}$ has a limit law with distribution function F if

$$\lim_{B \to +\infty} \nu_B(x \in \mathbb{P}^n(\mathbb{Q}) : f(x) \leq z) = F(z)$$

holds for a function $F : \mathbb{R} \to [0, 1]$ which is non-decreasing, right-continuous and satisfies $F(-\infty) = 0$, $F(+\infty) = 1$, for all $z \in \mathbb{R}$ for which F is continuous at z. The function ω_π takes values in $\mathbb{Z}_{\geq 0}$ and for such functions $f : \mathbb{P}^n(\mathbb{Q}) \to \mathbb{Z}_{\geq 0}$ the definition of the limit law is equivalent to the existence of the limit

$$\lim_{B \to +\infty} \nu_B(x \in \mathbb{P}^n(\mathbb{Q}) : f(x) = j)$$

for every fixed $j \in \mathbb{Z}_{\geq 0}$ and the property

$$\sum_{j=0}^{\infty} \lim_{B \to +\infty} \nu_B(x \in \mathbb{P}^n(\mathbb{Q}) : f(x) = j) = 1.$$

These are the two properties that are verified in Theorem 1.7, in addition to a bound in terms of j for the limits.

We illustrate Theorem 1.7 with some examples.

Example 1.8 Let $d, n > 1$. Let $\pi : V \to \mathbb{P}_{\mathbb{Q}}^{N-1}$ be the family of all hypersurfaces of degree d in $\mathbb{P}_{\mathbb{Q}}^n$, where $N = \binom{n+d}{d}$. (Note that V is regular.) If $(d, n) = (2, 2)$, i.e. the family of all plane conics, then $\Delta(\pi) = 1/2$ [20, Ex. 4] and Theorem 1.2 applies. If however $(d, n) \neq (2, 2)$, then the fibre over every codimension 1 point is geometrically integral, thus $\Delta(\pi) = 0$ and Theorem 1.7 applies. (See the proof of [19, Thm. 3.6] for this fact.) We deduce that when $(d, n) \neq (2, 2)$, the probability that a smooth hypersurface has no p-adic point for exactly j many primes p is well-defined and exists.

A particularly interesting case is when $n \geq d^2$. Here the Ax–Kochen theorem [1] implies that the map $V(\mathbb{Q}_p) \to \mathbb{P}^{N-1}(\mathbb{Q}_p)$ is surjective for all but finitely many primes p. In particular, we have $\tau_\pi(j) = 0$ for all but finitely many $j \in \mathbb{Z}$.

An example where the measure τ_π has infinite support is the following.

Example 1.9 Let

$$V : \quad \sum_{i=0}^{3} y_i x_i^3 = 0 \qquad \subset \mathbb{P}^3 \times \mathbb{P}^3$$

and let $\pi : V \to \mathbb{P}^3$ be the projection onto the y-coordinate; here π is the family of all diagonal cubic surfaces. In Sect. 4.2 we will show that there exists an absolute constant $c > 0$ such that $\tau_\pi(j) > c(1 + j)^{-3j}$ for all $j \in \mathbb{Z}_{\geq 0}$. This shows that τ_π has infinite support in this case and that (1.9) cannot be significantly improved.

It turns out that one has the following characterisation for when the measure τ_π has finite support; it happens if and only if an Ax–Kochen-type property holds.

Theorem 1.10 *Keep the assumptions of Theorem 1.7. Then the measure τ_π has finite support if and only if $V(\mathbb{Q}_p) \to \mathbb{P}^n(\mathbb{Q}_p)$ is surjective for all but finitely many primes p.*

Families for which $V(\mathbb{Q}_p) \to \mathbb{P}^n(\mathbb{Q}_p)$ is surjective for all but finitely many p were studied in [15]. A geometric criterion for when this holds can be found in [15, Thm. 1.4].

Our methods also allow us to prove the following local-global principle for existence of varieties in the family which are non-locally soluble at *exactly* a given finite set of places.

Theorem 1.11 *Keep the assumptions of Theorem 1.7. Let S be a finite set of places of $\mathbb{Q}$. Assume that $\pi(V(\mathbb{Q}_v)) \neq \mathbb{P}^n(\mathbb{Q}_v)$ for all $v \in S$ and that $\pi(V(\mathbb{Q}_v)) \neq \emptyset$ for all $v \notin S$. Then there exists $x \in \mathbb{P}^n(\mathbb{Q})$ such that $\pi^{-1}(x)$ is smooth and*

$$\pi^{-1}(x)(\mathbb{Q}_v) = \emptyset \quad \Longleftrightarrow \quad v \in S.$$

Note that for conics Hilbert's version of quadratic reciprocity implies that a conic over $\mathbb{Q}$ fails to have a $\mathbb{Q}_v$-point at exactly a set of places S of even cardinality, despite there being a conic C_v over every $\mathbb{Q}_v$ with $C_v(\mathbb{Q}_v) = \emptyset$. Theorem 1.11 shows that for

families of varieties with $\Delta(\pi) = 0$ there is no such reciprocity law. (This phenomenon was first observed in the case of curves of genus at least 1 by Poonen and Stoll [18].)

One of the major differences between the case $\Delta(\pi) > 0$ and $\Delta(\pi) = 0$ is that the function $\omega_\pi(x)$ becomes arbitrarily large on average only when $\Delta(\pi) > 0$. To make this precise we study the moments of ω_π. Define for $r \in \mathbb{Z}_{\geq 0}$ the function

$$\mathcal{N}_r(\pi, B) := \sum_{\substack{x \in \mathbb{P}^n(\mathbb{Q}), H(x) \leq B \\ \pi^{-1}(x) \text{ smooth}}} \omega_\pi(x)^r. \tag{1.10}$$

Note that an obvious consequence of Theorem 1.3 is that if $\Delta(\pi) > 0$ then

$$\mathcal{N}_r(\pi, B) B^{-n-1} \gg_{\pi,r} (\log \log B)^r.$$

In contrast, if $\Delta(\pi) = 0$ then for all $r \geq 0$ the function $\mathcal{N}_r(\pi, B) B^{-n-1}$ remains bounded as $B \to \infty$; specifically we have the following counterpart of Theorem 1.3.

Theorem 1.12 *Keep the assumptions of Theorem* 1.7. *Then for every* $r \in \mathbb{Z}_{\geq 1}$ *we have*

$$\lim_{B \to \infty} \frac{\mathcal{N}_r(\pi, B)}{\sharp\{x \in \mathbb{P}^n(\mathbb{Q}), H(x) \leq B\}} = \sum_{j=0}^{\infty} j^r \tau_\pi(j). \tag{1.11}$$

Note that, apart from very special cases, existence of moments does not automatically imply existence of a limit law or vice versa.

1.3 Layout of the paper and proof ingredients

We begin in Sect. 2 with an elementary result on counting rational points in $\mathbb{P}^n(\mathbb{Q})$ which lie in a given residue class.

We prove Theorem 1.3 in Sect. 3. For this we show in Proposition 3.9 that the moments of a 'truncated' version of ω_π are approximated by the moments of the standard normal distribution. The proof is based on equidistribution properties of the fibers of π that are verified during the earlier stages in Sect. 3 and subsequently fed into work of Granville and Soundararajan [11]. We finish the proof of Theorem 1.3 in Sect. 3.4 by showing that the moments of ω_π and the moments of the truncated version of ω_π have the same asymptotic behaviour. Theorem 1.2 is then deduced from Theorem 1.3 in Sect. 3.5.

In Sect. 4 we prove the results from Sect. 1.2. The most difficult part of the proof of Theorem 1.7 is establishing the existence of the limit (1.8), which we achieve via Bhargava's effective version of the Ekedahl sieve [3]. Theorems 1.10 and 1.12 are proved using similar methods and the results from Sect. 2. We finish Sect. 4 by briefly explaining how our results generalise in a straightforward manner to minor variants given by considering real solubility or by dropping conditions at finitely many primes.

Notation For an integral homogeneous polynomial f, a point $x \in \mathbb{P}^n(\mathbb{Q})$ and $Q \in \mathbb{N}$, we say that "$f(x) \equiv 0 \bmod Q$" if $f(\mathbf{x}) \equiv 0 \bmod Q$ for some primitive representative $\mathbf{x} \in \mathbb{Z}^{n+1}$ of x. We use the notation "$Q \mid f(x)$" analogously.

The quantities $\delta_x(\pi)$ and $\Delta(\pi)$ are introduced in Definition 1.1, the function $\omega_\pi(x)$ is defined in (1.2) and the indicator function $\theta_p(x)$ is introduced in (3.2). The arithmetic functions ω, μ, φ respectively denote the number of prime divisors, the Möbius function and the Euler totient function, respectively.

2 Explicit equidistribution on projective space

2.1 Counting with congruences

We will be required to count rational points in projective space which satisfy imposed congruence conditions. To state our result, we let

$$c_n = \lim_{B \to \infty} \frac{\sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \le B\}}{B^{n+1}} = \frac{2^n}{\zeta(n+1)}, \tag{2.1}$$

where $\zeta(s)$ denotes the Riemann zeta function.

Proposition 2.1 *Let $B > 1$, $Q \in \mathbb{N}$ and $\Upsilon \subseteq \mathbb{P}^n(\mathbb{Z}/Q\mathbb{Z})$. Then*

$$\sharp\left\{x \in \mathbb{P}^n(\mathbb{Q}) : \begin{array}{l} H(x) \le B, \\ x \bmod Q \in \Upsilon \end{array}\right\}$$
$$= \frac{c_n \sharp \Upsilon}{\sharp \mathbb{P}^n(\mathbb{Z}/Q\mathbb{Z})} B^{n+1} + O\left(Q \sharp \Upsilon \left(B + \frac{B^n}{Q^n} (\log B)^{[1/n]}\right)\right),$$

where $[\cdot]$ denotes the integer part.

Proof Let $\widehat{\Upsilon} = \{\mathbf{x} \in (\mathbb{Z}/Q\mathbb{Z})^{n+1} : \mathbf{x} \not\equiv \mathbf{0} \bmod p \; \forall p \mid Q, (x_0 : \cdots : x_n) \in \Upsilon\}$ be the affine cone of Υ. Applying Möbius inversion we see that the cardinality in question is

$$\frac{1}{2}\sharp\{\mathbf{x} \in \mathbb{Z}^{n+1} : \max\{|x_0|, \ldots, |x_n|\} \le B, \gcd(x_0, \ldots, x_n) = 1, \mathbf{x} \bmod Q \in \widehat{\Upsilon}\} + O(1)$$

$$= \frac{1}{2} \sum_{\substack{k \le B \\ \gcd(k,Q)=1}} \mu(k)\sharp\{\mathbf{x} \in \mathbb{Z}^{n+1} : \max\{|x_0|, \ldots, |x_n|\} \le B/k, k\mathbf{x} \bmod Q \in \widehat{\Upsilon}\} + O(1)$$

$$= \frac{1}{2} \sum_{\mathbf{y} \in \widehat{\Upsilon}} \sum_{\substack{k \le B \\ \gcd(k,Q)=1}} \mu(k)\sharp\{\mathbf{x} \in \mathbb{Z}^{n+1} : \max\{|x_0|, \ldots, |x_n|\} \le B/k, \mathbf{x} \equiv k^{-1}\mathbf{y} \bmod Q\} + O(1)$$

$$= \frac{1}{2} \sum_{\mathbf{y} \in \widehat{\Upsilon}} \sum_{\substack{k \le B \\ \gcd(k,Q)=1}} \mu(k)\left(\left(\frac{2B}{kQ}\right) + O(1)\right)^{n+1} + O(1).$$

Continuing, we use the estimate $\sum_{k>B} k^{-n-1} \ll B^{-n}$ to find that the main term is

$$\frac{2^n \sharp \widehat{\Upsilon} B^{n+1}}{Q^{n+1}} \sum_{\substack{k \in \mathbb{N} \\ \gcd(k,Q)=1}} \frac{\mu(k)}{k^{n+1}} + O\left(\frac{\sharp \widehat{\Upsilon} B}{Q^{n+1}}\right)$$

$$= \frac{c_n \sharp \widehat{\Upsilon}}{Q^{n+1} \prod_{p|Q}(1 - 1/p^{n+1})} \cdot B^{n+1} + O\left(\frac{\sharp \widehat{\Upsilon} B}{Q^{n+1}}\right).$$

Recall that

$$\sharp \mathbb{P}^n(\mathbb{Z}/p^k\mathbb{Z}) = p^{n(k-1)} \cdot \frac{p^{n+1} - 1}{p - 1}. \tag{2.2}$$

From this it is easy to establish

$$Q^{n+1} \prod_{p|Q}(1 - 1/p^{n+1}) = \varphi(Q) \sharp \mathbb{P}^n(\mathbb{Z}/Q\mathbb{Z}),$$

where φ is Euler's totient function. We obtain the main term as stated in our proposition via $\sharp \widehat{\Upsilon} = \varphi(Q) \sharp \Upsilon$. To deal with the error terms observe that

$$\left(\left(\frac{2B}{kQ}\right) + O(1)\right)^{n+1} - \left(\frac{2B}{kQ}\right)^{n+1} \ll \sum_{0 \le \ell \le n} \left(\frac{B}{kQ}\right)^\ell \ll 1 + \left(\frac{B}{kQ}\right)^n,$$

where the last inequality stems from $\sum_{0 \le \ell \le n} t^\ell \le (n+1)(1+t^n)$, valid for every $t \ge 0$ and $n \in \mathbb{N}$. We obtain the error term

$$\sum_{\mathbf{y} \in \widehat{\Upsilon}} \sum_{\substack{k \le B \\ \gcd(k,Q)=1}} |\mu(k)| \left\{ \left(\left(\frac{2B}{kQ}\right) + O(1)\right)^{n+1} - \left(\frac{2B}{kQ}\right)^{n+1} \right\}$$

$$\ll \sharp \widehat{\Upsilon} \left\{ B + (B/Q)^n (\log B)^{[1/n]} \right\}.$$

Using $\sharp \widehat{\Upsilon} \ll Q \sharp \Upsilon$ completes the proof. $\square$

Before continuing we record an elementary lemma here.

Lemma 2.2 *For $n, Q \in \mathbb{N}$ we have*

$$Q^n \le \sharp \mathbb{P}^n(\mathbb{Z}/Q\mathbb{Z}) \le 2^{\omega(Q)} Q^n.$$

Proof It suffices to prove the result when $Q = p^k$ for some prime p. By (2.2) we have

$$p^{kn} \le \sharp \mathbb{P}^n(\mathbb{Z}/p^k\mathbb{Z}) = p^{n(k-1)} \cdot \frac{p^{n+1} - 1}{p - 1} \le p^{n(k-1)} \cdot (2p^n) = 2p^{kn}.$$

$\square$

2.2 Some probability measures

2.2.1 Measures on $\mathbb{P}^n(\mathbb{Q}_p)$

Let p be a prime. The finite sets $\mathbb{P}^n(\mathbb{Z}/p^k\mathbb{Z})$ come with a natural uniform probability measure. Taking the limit $\mathbb{P}^n(\mathbb{Z}_p) = \lim_{k\to\infty} \mathbb{P}^n(\mathbb{Z}/p^k\mathbb{Z})$ we obtain a well-defined probability measure ϑ_p on $\mathbb{P}^n(\mathbb{Z}_p)$ (this measure differs from Peyre's local Tamagawa measure [17, §2.2] by a constant). These measures admit the following explicit description. Let $\Upsilon \subset \mathbb{P}^n(\mathbb{Z}/p^k\mathbb{Z})$. Then

$$\vartheta_p\left(\{x \in \mathbb{P}^n(\mathbb{Q}_p) : x \bmod p^k \in \Upsilon\}\right) = \frac{\sharp\Upsilon}{\sharp\mathbb{P}^n(\mathbb{Z}/p^k\mathbb{Z})}. \tag{2.3}$$

These "residue disks" generate the σ-algebra on $\mathbb{P}^n(\mathbb{Z}_p)$, hence the measure ϑ_p is uniquely determined by (2.3). Proposition 2.1 may be viewed as an effective version of equidistribution of rational points on $\mathbb{P}^n$ with respect to the measures ϑ_p.

One relates the measure ϑ_p to the usual Haar measure on $\mathbb{Z}_p$ via the following.

Lemma 2.3 *Let μ_p denote the Haar probability measure on $\mathbb{Z}_p^{n+1}$. Let $\Upsilon \subset \mathbb{P}^n(\mathbb{Z}_p)$ and let $\widehat{\Upsilon} \subset \mathbb{Q}_p^{n+1}$ be its affine cone. Then $\vartheta_p(\Upsilon) = \mu_p(\widehat{\Upsilon} \cap \mathbb{Z}_p^{n+1})$.*

Proof It suffices to prove the result for the residue disks

$$\Upsilon = \{x \in \mathbb{P}^n(\mathbb{Z}_p) : x \equiv (a_0 : a_1 : \cdots : a_n) \bmod p^k\}$$

for some $a_i \in \mathbb{Z}_p$; we have $\vartheta_p(\Upsilon) = 1/\sharp\mathbb{P}^n(\mathbb{Z}/p^k\mathbb{Z})$. Note that one of the a_i may be taken to be a unit; for simplicity we assume this is a_0. A moment's thought reveals that

$$\widehat{\Upsilon} \cap \mathbb{Z}_p^{n+1} = \left\{\mathbf{x} \in \mathbb{Z}_p^{n+1} : \left|\frac{x_i}{p^{v_p(x_0)}} - \frac{a_i x_0}{a_0 p^{v_p(x_0)}}\right|_p \leq p^{-k}, i = 1, \ldots, n\right\}.$$

Thus, by (2.2) we find that

$$\mu_p(\widehat{\Upsilon} \cap \mathbb{Z}_p^{n+1})$$
$$= \sum_{m=0}^{\infty} \{\mathbf{x} \in \mathbb{Z}_p^{n+1} : v_p(x_0) = m, |x_i - a_i x_0/a_0|_p \leq p^{-k-m}, i = 1, \ldots, n\}$$
$$= \sum_{m=0}^{\infty} \frac{1}{p^m}\left(1 - \frac{1}{p}\right)\left(\frac{1}{p^{k+m}}\right)^n = \frac{1}{p^{n(k-1)}}\left(\frac{p-1}{p^{n+1}-1}\right) = \frac{1}{\sharp\mathbb{P}^n(\mathbb{Z}/p^k\mathbb{Z})}.$$

$\square$

2.2.2 Measure on $\mathbb{P}^n(\mathbb{R})$

We let ϑ_∞ be the pushforward of the usual probability measure on the n-sphere S^n via the quotient map $S^n \to \mathbb{P}^n(\mathbb{R})$.

3 An Erdős–Kac theorem for fibrations

3.1 Set-up

We begin the proof of the results from Sect. 1.1. Let V be a smooth projective variety over $\mathbb{Q}$ with a dominant morphism $\pi : V \to \mathbb{P}^n_{\mathbb{Q}}$ with geometrically integral generic fibre. (We assume $\Delta(\pi) > 0$ from Sect. 3.3.) We choose a *model* for π, i.e. a proper scheme $\mathcal{V}$ over $\mathbb{Z}$ together with a proper morphism $\pi : \mathcal{V} \to \mathbb{P}^n_{\mathbb{Z}}$ (also denoted π by abuse of notation), which extends $V \to \mathbb{P}^n_{\mathbb{Q}}$. In what follows, all implied constants are allowed to depend on π, the choice of model, and the A and f occurring in Lemma 3.2.

We begin by studying the basic properties of $\omega_\pi(x)$. We first show that it enjoys analogous bounds to the usual ω.

Lemma 3.1 *There exists $D > 0$ with the following property. Let $x \in \mathbb{P}^n(\mathbb{Q})$ be such that $\pi^{-1}(x)$ is smooth. Then*

$$\omega_\pi(x) \ll \frac{\log H(x)}{\log \log H(x)}, \quad \max\{p : \pi^{-1}(x)(\mathbb{Q}_p) = \emptyset\} \ll H(x)^D.$$

Proof Let $S \subset \mathbb{P}^n_{\mathbb{Q}}$ denote the non-smooth locus of π; this is a proper closed subset of $\mathbb{P}^n_{\mathbb{Q}}$. Let $\mathcal{S}$ be the closure of S in $\mathbb{P}^n_{\mathbb{Z}}$ and choose a finite collection of homogeneous polynomials $f_1, \ldots, f_s$ which generate the ideal of $\mathcal{S}$. Let $D = \max_i\{\deg f_i\}$.

For all sufficiently large primes p, the fibre $\pi^{-1}(x \bmod p)$ is smooth if and only if $x \bmod p \notin \mathcal{S}$, which happens if and only if $p \nmid f_i(x)$ for some i. Moreover, by the Lang–Weil estimates [16] and Hensel's lemma, for all sufficiently large primes p (independently of x) if $\pi^{-1}(x \bmod p)$ is smooth then $\pi^{-1}(x)(\mathbb{Q}_p) \neq \emptyset$. It follows that

$$\sharp\{p : \pi^{-1}(x)(\mathbb{Q}_p) = \emptyset\} \leq \sharp\{p : p \mid f_i(x) \,\forall\, i \in \{1, \ldots, s\}\} + O(1).$$

Letting **x** be a primitive representative of x and using the bound $\omega(n) \ll (\log n)(\log \log n)^{-1}$, we obtain

$$\omega_\pi(x) \le \sum_{i=1}^{r} \omega(f_i(\mathbf{x})) + O(1) \ll \frac{\log(H(x)^D)}{\log\log(H(x)^D)},$$

$$\max\{p : \pi^{-1}(x)(\mathbb{Q}_p) = \emptyset\} \ll H(x)^D.$$

□

To simplify notation it will be easier to work with some choice of polynomial which vanishes on the singular locus, rather than the whole singular locus. The proof of the following is a minor adaptation of the proof of Lemma 3.1 (just choose $f = f_1$).

Lemma 3.2 *Let $f \in \mathbb{Z}[x_0, \ldots, x_n]$ be a homogeneous square-free polynomial such that π is smooth away from the divisor $f(x) = 0 \subset \mathbb{P}_{\mathbb{Q}}^n$. Then there exists $A > 0$ such that for all primes $p > A$ the following hold.*

1. *The restriction of π to $\mathbb{P}_{\mathbb{F}_p}^n$ is smooth away from the divisor $f(x) = 0 \subset \mathbb{P}_{\mathbb{F}_p}^n$.*
2. *Let $x \in \mathbb{P}^n(\mathbb{Q})$. If $\pi^{-1}(x)(\mathbb{Q}_p) = \emptyset$ then $p \mid f(x)$.*

In Sect. 3.2, we allow ourselves to increase A as necessary to take care of bad behaviour at small primes.

3.2 Equidistribution properties in the fibres

The next step is to translate the condition $\pi^{-1}(x)(\mathbb{Q}_p) = \emptyset$ into something amenable to tools from analytic number theory. We do this by using the tools developed in [14]. The key result is [14, Thm. 2.8], which is a valuative criterion for non-existence of a p-adic point in a fibre, for sufficiently large primes p. In the special case of the conic bundle over $\mathbb{Q}$

$$x^2 + y^2 = tz^2 \qquad \subset \mathbb{P}^2 \times \mathbb{A}^1,$$

the criterion [14, Thm. 2.8] says that if $p \equiv 3 \bmod 4$ and the p-adic valuation of t is 1, then the fibre over t has no $\mathbb{Q}_p$-point, as is familiar from the theory of Hilbert symbols.

We introduce the quantity which will arise in this analysis. For any prime p let

$$\sigma_p := \frac{\sharp\{x \in \mathbb{P}^n(\mathbb{F}_p) : \pi^{-1}(x) \text{ is non-split}\}}{\sharp\mathbb{P}^n(\mathbb{F}_p)}. \tag{3.1}$$

(Here we use the term "non-split" in the sense of Skorobogatov [23, Def. 0.1].)

Lemma 3.3 *Let A and f be as in Lemma 3.2 and $p > A$. Then*

$$\sigma_p \le \frac{d}{p}, \qquad where \ d = \deg f.$$

Proof A non-split fibre is necessarily singular. Hence by Lemma 3.2, for $p > A$ we have

$$\sharp\{x \in \mathbb{P}^n(\mathbb{F}_p) : \pi^{-1}(x) \text{ is non-split}\} \leq \sharp\{x \in \mathbb{P}^n(\mathbb{F}_p) : f(x) = 0\}.$$

Projecting to a suitable hyperplane, this is at most $d\sharp\mathbb{P}^{n-1}(\mathbb{F}_p)$. The result follows. □

We now use results from [14] to deduce an equidistribution result for θ_p. To simplify notation, we denote the characteristic function of the p-adically insoluble fibres by

$$\theta_p(x) := \begin{cases} 1, & \text{if } \pi^{-1}(x)(\mathbb{Q}_p) = \emptyset, \\ 0, & \text{otherwise.} \end{cases} \tag{3.2}$$

Our result is the following asymptotic upper and lower bounds. (Here c_n is as in (2.1).)

Proposition 3.4 *Let $d = \deg f$. Enlarging A if necessary, there exists $\alpha \geq 0$ with the following property. Let $Q \in \mathbb{N}$ be square-free with $p \nmid Q$ for all $p \leq A$. Then*

$$\pm \sum_{\substack{x \in \mathbb{P}^n(\mathbb{Q}) \\ H(x) \leq B \\ \pi^{-1}(x) \text{ smooth}}} \prod_{p \mid Q} \theta_p(x)$$

$$\leq \pm c_n \Big(\prod_{p \mid Q} (\sigma_p \pm \alpha/p^2) \Big) B^{n+1} + O\left((4d)^{\omega(Q)} (Q^{2n+1}B + QB^n(\log B)^{[1/n]}) \right),$$

where the implied constant is independent of B and Q.

Proof Let $p > A$ be a prime. Enlarging A if necessary, the Lang–Weil estimates and Hensel's lemma show that if $\pi^{-1}(x \bmod p)$ is split then $\pi^{-1}(x)$ has a $\mathbb{Q}_p$-point. Thus the sum in the proposition is

$$\leq \sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \leq B, \pi^{-1}(x \bmod p) \text{ is non-split } \forall p \mid Q\}.$$

Applying Proposition 2.1 with $\Upsilon = \{x \in \mathbb{P}^n(\mathbb{Z}/Q\mathbb{Z}): \pi^{-1}(x \bmod p) \text{ is non-split} \forall p \mid Q\}$, we infer that the above cardinality equals

$$c_n B^{n+1} \prod_{p \mid Q} \frac{\sharp\{x \in \mathbb{P}^n(\mathbb{F}_p) : \pi^{-1}(x) \text{ is non-split}\}}{\sharp\mathbb{P}^n(\mathbb{F}_p)}$$

$$+ O\left(Q\sharp\Upsilon \left(B + \frac{B^n}{Q^n}(\log B)^{[1/n]} \right) \right).$$

One has $\sharp\Upsilon = \sharp\mathbb{P}^n(\mathbb{Z}/Q\mathbb{Z}) \prod_{p \mid Q} \sigma_p$, thus Lemmas 2.2 and 3.3 imply that $\sharp\Upsilon \ll (2d)^{\omega(Q)} Q^{n-1}$, which is satisfactory for the upper bound in the proposition.

For the lower bound, we apply the sparsity result of [14, Thm. 2.8]. This gives a square-free homogeneous polynomial $g \in \mathbb{Z}[x_0, \ldots, x_n]$ which is coprime with f

and contains the singular locus of f such that, enlarging A if necessary, the sum in the proposition is

$$\geq \sharp \left\{ x \in \mathbb{P}^n(\mathbb{Q}) : \begin{array}{l} H(x) \leq B, \; f(x) \equiv 0 \bmod p, \; f(x) \not\equiv 0 \bmod p^2, \\ g(x) \not\equiv 0 \bmod p, \; \pi^{-1}(x \bmod p) \text{ is non-split}, \; \forall p \mid Q \end{array} \right\}.$$

We now apply Proposition 2.1 with

$$\Upsilon' = \left\{ x \in \mathbb{P}^n(\mathbb{Z}/Q^2\mathbb{Z}) : \begin{array}{l} f(x) \equiv 0 \bmod p, \; f(x) \not\equiv 0 \bmod p^2, \\ g(x) \not\equiv 0 \bmod p, \; \pi^{-1}(x \bmod p) \text{ is non-split}, \; \forall p \mid Q \end{array} \right\}$$

to see that the sum in the proposition is

$$\geq \frac{c_n \sharp \Upsilon'}{\sharp \mathbb{P}^n(\mathbb{Z}/Q^2\mathbb{Z})} B^{n+1} + O\left(Q^2 \sharp \Upsilon' \left(B + \frac{B^n}{Q^{2n}} (\log B)^{[1/n]} \right) \right). \tag{3.3}$$

As $g(x) = 0$ contains the singular locus of f, we may apply [5, Prop. 2.3] to find that

$$\sharp \left\{ x \in \mathbb{P}^n(\mathbb{Z}/p^2\mathbb{Z}) : \begin{array}{l} f(x) \equiv 0 \bmod p, \; f(x) \not\equiv 0 \bmod p^2, \\ g(x) \not\equiv 0 \bmod p, \; \pi^{-1}(x \bmod p) \text{ is non-split} \end{array} \right\}$$
$$= \sharp \left\{ x \in \mathbb{P}^n(\mathbb{F}_p) : f(x) = 0, g(x) \neq 0, \pi^{-1}(x) \text{ is non-split} \right\} (p^n + O(p^{n-1}))$$
$$= \sharp \{ x \in \mathbb{P}^n(\mathbb{F}_p) : \pi^{-1}(x) \text{ is non-split} \} p^n + O(p^{2n-2}).$$

Here the last line follows from the fact that if $\pi^{-1}(x)$ is non-split then necessarily $f(x) = 0$ by Lemma 3.2, together with the fact that $\sharp\{x \in \mathbb{P}^n(\mathbb{F}_p) : f(x) = g(x) = 0\} \ll p^{n-2}$ as f and g share no common factor. Recalling that $\sharp\mathbb{P}^n(\mathbb{Z}/p^2\mathbb{Z}) = p^n \sharp \mathbb{P}^n(\mathbb{F}_p)$ (2.2), shows that

$$\prod_{p \mid Q} (\sigma_p - \alpha/p^2) \leq \frac{\sharp \Upsilon'}{\sharp \mathbb{P}^n(\mathbb{Z}/Q^2\mathbb{Z})} \leq \prod_{p \mid Q} (\sigma_p + \alpha/p^2) \tag{3.4}$$

for some $\alpha > 0$. This yields the correct main term for the lower bound. For the error term, enlarging A if necessary we have $\alpha/p^2 < d/p$ for all $p > A$. Thus (3.4) and Lemmas 2.2 and 3.3 give

$$\sharp \Upsilon' \leq \sharp \mathbb{P}^n(\mathbb{Z}/Q^2\mathbb{Z}) \prod_{p \mid Q} (\sigma_p + \alpha/p^2) \leq (2^{\omega(Q)} Q^{2n}) \cdot (2d)^{\omega(Q)}/Q,$$

which yields the required error term in (3.3). □

We now fix the choice of A. Note that as $\min\{p : p \mid Q\} \to \infty$, the lower bound in Proposition 3.4 converges to the upper bound.

Lemma 3.5 *We have*

$$\sum_{p \le B} \sharp \{x \in \mathbb{P}^n(\mathbb{F}_p) : \pi^{-1}(x) \text{ is non-split}\} = \Delta(\pi)\frac{B^n}{\log(B^n)} + O\left(\frac{B^n}{(\log B)^2}\right).$$

Proof This follows from an easy modification of the proof of [14, Prop. 3.10]. (*Loc. cit.* states an asymptotic formula without an error term; one obtains an error term via the version of the Chebotarev density theorem given in [22, Thm. 9.11].) □

We will also require the following.

Proposition 3.6 *There exists a constant β_π such that*

$$\sum_{p \le B} \sigma_p = \Delta(\pi)(\log \log B) + \beta_\pi + O((\log B)^{-1})$$

Proof The proof is a simple application of Lemma 3.5 and partial summation. First, let $a_p := \sharp \{x \in \mathbb{P}^n(\mathbb{F}_p) : \pi^{-1}(x) \text{ is non-split}\}$, define $S(B) := \sum_{2 < p \le x} a_p$ and let

$$R(B) := S(B) - \Delta(\pi)\frac{B^n}{\log(B^n)}.$$

Lemma 3.5 is equivalent to the estimate $R(B) \ll \frac{B^n}{(\log B)^2}$. By partial summation we obtain

$$\sum_{2 < p \le B} \frac{a_p}{p^n} = \frac{S(B)}{B^n} + n \int_2^t \frac{\left(\Delta(\pi)\frac{u^n}{\log(u^n)} + R(u)\right)}{u^{n+1}} \mathrm{d}u.$$

Lemma 3.5 directly gives $S(B)/B^n \ll 1/\log B$. We furthermore have $\int_2^\infty \frac{|R(u)|}{u^{n+1}} \mathrm{d}u < \infty$ due to $|R(B)| \ll \frac{B^n}{(\log B)^2}$. Hence, we may write

$$\int_2^t \frac{R(u)}{u^{n+1}} \mathrm{d}u = \int_2^\infty \frac{R(u)}{u^{n+1}} \mathrm{d}u + O\left(\int_t^\infty \frac{|R(u)|}{u^{n+1}} \mathrm{d}u\right)$$

$$= \int_2^\infty \frac{R(u)}{u^{n+1}} \mathrm{d}u + O\left(\int_t^\infty \frac{1}{u(\log u)^2} \mathrm{d}u\right)$$

$$= \int_2^\infty \frac{R(u)}{u^{n+1}} \mathrm{d}u + O\left(\frac{1}{\log t}\right).$$

Recalling that $\int_2^t (u \log u)^{-1} \mathrm{d}t = \log \log t - \log \log 2$ and letting

$$\gamma_\pi := -\Delta(\pi)(\log \log 2) + \frac{a_2}{2^n} + \int_2^\infty \frac{R(u)}{u^{n+1}} \mathrm{d}u$$

we have proved

$$\sum_{p \leq B} \frac{a_p}{p^n} = \Delta(\pi) \log \log B + \gamma_p + O\left(\frac{1}{\log B}\right). \tag{3.5}$$

Using this in the simple form $\sum_{p \leq B} \frac{a_p}{p^n} \ll \log \log B$ then by partial summation we obtain

$$\sum_{p \leq B} \frac{a_p}{p^{n+1}} \ll \frac{\log \log B}{B} + \int_1^B \frac{\log \log u}{u^2} du \ll 1.$$

We thus obtain that $\sum_p \frac{a_p}{p^{n+1}}$ converges and that the tail is at most

$$\sum_{p > B} \frac{a_p}{p^{n+1}} \ll \frac{\log \log B}{B} + \int_B^\infty \frac{\log \log u}{u^2} du \ll \frac{\log \log B}{B}. \tag{3.6}$$

Let us now define the function ϵ_p for primes p via the equation

$$\frac{1}{\sharp \mathbb{P}^n(\mathbb{F}_p)} = \frac{1}{p^n} + \frac{\epsilon_p}{p^n}.$$

Recalling (3.1) and making use of (3.5), we see that this gives

$$\sum_{p \leq B} \sigma_p = \sum_{p \leq B} \frac{a_p}{p^n} + \sum_{p \leq B} \frac{a_p \epsilon_p}{p^n} = \Delta(\pi) \log \log B + \gamma_p + O\left(\frac{1}{\log B}\right) + \sum_{p \leq B} \frac{a_p \epsilon_p}{p^n}.$$

At this point we use (2.2) to obtain

$$\frac{1}{\sharp \mathbb{P}^n(\mathbb{F}_p)} = \frac{p-1}{p^{n+1}-1} = \frac{1}{p^n} \frac{1}{(1 + \frac{1}{p} + \cdots + \frac{1}{p^n})} = \frac{1}{p^n}\left(1 + O\left(\frac{1}{p}\right)\right),$$

from which we get $\epsilon_p \ll 1/p$. By (3.6) we see that $\sum_p a_p \epsilon_p p^{-n}$ converges and that

$$\sum_{p \leq B} \frac{a_p \epsilon_p}{p^n} = \sum_p \frac{a_p \epsilon_p}{p^n} + O\left(\frac{\log \log B}{B}\right).$$

Taking $\beta_\pi := \gamma_\pi + \sum_p a_p \epsilon_p p^{-n}$ concludes the proof. $\qquad \square$

3.3 Moments of a truncated version of ω_π

We assume from now on that $\Delta(\pi) > 0$. In what follows $t_0, t_1 : \mathbb{R} \to \mathbb{R}$ are two functions that satisfy

$$1 < t_0(B) < t_1(B) < B, \quad \lim_{B\to\infty} t_0(B) = \lim_{B\to\infty} t_1(B) = \infty.$$

Both functions $t_0(B), t_1(B)$ will be chosen optimally at a later stage. Define the function

$$\omega_\pi^\flat(x, B) := \sum_{t_0(B)<p\leq t_1(B)} (\theta_p(x) - \sigma_p), \tag{3.7}$$

where σ_p is as in (3.1). We need to estimate asymptotically the moments of $\omega_\pi(x)$; it turns out that it is easier to work with the 'truncated' version $\omega_\pi^\flat(x, B)$ of $\omega_\pi(x)$. Introducing $t_1(B)$ deals with the dependence on Q in the error term of Proposition 3.4, whilst $t_0(B)$ is used to control the error $\pm\alpha/p^2$ in the leading constant in Proposition 3.4.

To study the degree to which $\omega_\pi(x)$ is affected by the primes in the interval $(t_0(B), t_1(B)]$ we begin by observing that Proposition 3.6 provides us with

$$\sum_{p\leq t_0(B)} \sigma_p \ll \log\log t_0(B), \qquad \sum_{t_1(B)<p\leq B} \sigma_p \ll \left(\log\frac{\log B}{\log t_1(B)}\right) + \frac{1}{\log t_1(B)}. \tag{3.8}$$

We define

$$\mathcal{M}_r^\flat(\pi, B) := \sum_{\substack{x\in\mathbb{P}^n(\mathbb{Q}),H(x)\leq B \\ \pi^{-1}(x)\text{ smooth}}} \left(\frac{\omega_\pi^\flat(x, B)}{\sqrt{\Delta(\pi)\log\log B}}\right)^r, \quad (r\in\mathbb{Z}_{\geq 0}). \tag{3.9}$$

Note that $\mathcal{M}_r^\flat(\pi, B)$ depends on $t_0(B)$ and $t_1(B)$ due to the presence of $\omega_\pi^\flat(x, B)$. The estimates in Proposition 3.6 and (3.8) yield

$$\sum_{t_0(B)<p\leq t_1(B)} \sigma_p = \Delta(\pi)\log\log B + O\left(\left(\log\frac{\log B}{\log t_1(B)}\right) + \log\log t_0(B)\right). \tag{3.10}$$

Furthermore, we have $\sigma_p - \sigma_p^2 = \sigma_p + O(p^{-2})$ due to Lemma 3.3. This shows that

$$\sum_{t_0(B)<p\leq t_1(B)} (\sigma_p - \sigma_p^2) = \Delta(\pi)\log\log B + O\left(\left(\log\frac{\log B}{\log t_1(B)}\right) + \log\log t_0(B)\right). \tag{3.11}$$

Lemma 3.7 *Let $\psi \in \{-1, 1\}$. In the situation of Proposition* 3.4 *we have*

$$\prod_{p|Q}(\sigma_p + \psi\alpha/p^2) = \prod_{p|Q}\sigma_p + O\left(\frac{(2\alpha d)^{\omega(Q)}}{Q\min\{p : p \mid Q\}}\right).$$

Proof We prove the inequality for one choice of sign, namely $\psi = 1$, the other choice being similar. Denoting $Q = p_1 \cdots p_k$ with $p_i < p_{i+1}$, we have by Lemma 3.3 that

$$\left(\prod_{p|Q}(\sigma_p + \alpha/p^2)\right) - \left(\prod_{p|Q}\sigma_p\right) = \sum_{S\subsetneq\{1,\ldots,k\}}\left(\prod_{i\in S}\sigma_{p_i}\right)\left(\prod_{i\notin S}\frac{\alpha}{p_i^2}\right)$$

$$\leq \sum_{S\subsetneq\{1,\ldots,k\}}\left(\prod_{i\in S}\frac{d}{p_i}\right)\left(\prod_{i\notin S}\frac{\alpha}{p_i^2}\right) \leq (\alpha d)^k \sum_{S\subsetneq\{1,\ldots,k\}}\left(\prod_{i\in S}\frac{1}{p_i}\right)\left(\prod_{i\notin S}\frac{1}{p_i^2}\right)$$

$$= \frac{(\alpha d)^k}{Q}\sum_{S\subsetneq\{1,\ldots,k\}}\left(\prod_{i\notin S}\frac{1}{p_i}\right) \leq \frac{(2\alpha d)^k}{Qp_1}.$$

$\square$

Before proceeding we recall [11, Prop. 3].

Lemma 3.8 (Granville–Soundararajan) *Let $\mathcal{P}$ be a finite set of primes and let $\mathcal{A} :=$ $\{a_1, \ldots, a_y\}$ be a multiset of y natural numbers. For $Q \in \mathbb{N}$ define*

$$\mathcal{A}_Q := \sharp\{m \leq y : Q \mid a_m\}.$$

Let h be a real-valued, non-negative multiplicative function such that for square-free Q we have $0 \leq h(Q) \leq Q$. For any $r \in \mathbb{N}$ we let

$$C_r = \Gamma(r + 1)/(2^{r/2}\Gamma(1 + r/2)) \text{ and } \mathcal{E}_{\mathcal{P}}(\mathcal{A}, h, r) := \sum_{\substack{Q\in\mathbb{N},\mu(Q)^2=1 \\ p|Q\Rightarrow p\in\mathcal{P} \\ \omega(Q)\leq r}}\left|\mathcal{A}_Q - \frac{h(Q)}{Q}y\right|.$$

Defining

$$\mu_{\mathcal{P}}(h) := \sum_{p\in\mathcal{P}}\frac{h(p)}{p} \text{ and } \sigma_{\mathcal{P}}(h) := \left(\sum_{p\in\mathcal{P}}\frac{h(p)}{p}\left(1 - \frac{h(p)}{p}\right)\right)^{1/2},$$

we have uniformly for all $r \leq \sigma_{\mathcal{P}}(h)^{2/3}$ that

$$\sum_{a\in\mathcal{A}}\left(\sharp\{p \in \mathcal{P} : p \mid a\} - \mu_{\mathcal{P}}(h)\right)^r$$

$$= C_r y\sigma_{\mathcal{P}}(h)^r + O(C_r y\sigma_{\mathcal{P}}(h)^{r-2}r^3 + \mu_{\mathcal{P}}(h)^r\mathcal{E}_{\mathcal{P}}(\mathcal{A}, h, r)) \qquad (3.12)$$

if r is even, and

$$\sum_{a \in \mathcal{A}} \left(\sharp\{p \in \mathcal{P} : p \mid a\} - \mu_{\mathcal{P}}(h) \right)^r \ll C_r y \sigma_{\mathcal{P}}(h)^{r-1} r^{3/2} + \mu_{\mathcal{P}}(h)^r \mathcal{E}_{\mathcal{P}}(\mathcal{A}, h, r)$$

(3.13)

if r is odd.

We apply this result to study the moments of $\omega_\pi^{\flat}(x, B)$.

Proposition 3.9 *Fix a positive integer r and let $t_0(B)$ and $t_1(B)$ be given by*

$$t_0(B) := (\log \log B)^{2r} \quad and \quad t_1(B) = B^{\frac{1}{5r(n+1)}}.$$

(3.14)

Then we have

$$\frac{\mathcal{M}_r^{\flat}(\pi, B)}{c_n B^{n+1}} = \begin{cases} \mu_r + O_r\left(\dfrac{\log \log \log \log B}{\log \log B}\right), & \textit{if r is even,} \quad (3.15) \\[4mm] O_r\left(\dfrac{1}{(\log \log B)^{\frac{1}{2}}}\right), & \textit{if r is odd.} \quad (3.16) \end{cases}$$

Proof We apply Lemma 3.8 with

$$\mathcal{A} := \left\{ a_x := \prod_{\substack{p \text{ prime} \\ \pi^{-1}(x)(\mathbb{Q}_p) = \emptyset}} p : x \in \mathbb{P}^n(\mathbb{Q}), H(x) \le B, \pi^{-1}(x) \text{ smooth} \right\}.$$

Lemma 3.1 ensures that a_x is well-defined. The key property of a_x is that for any square-free Q we have

$$Q \mid a_x \iff (p \mid Q \Rightarrow \pi^{-1}(x)(\mathbb{Q}_p) = \emptyset).$$

Therefore, if we let

$$\mathcal{P} := \{p \in (t_0, t_1] : p \text{ prime}\}, h(Q) := Q \prod_{p \mid Q} \sigma_p,$$

$$y := \sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \le B, \pi^{-1}(x) \text{ smooth}\},$$

then

$$\mathcal{A}_Q - \frac{h(Q)}{Q} y = \left(\sum_{\substack{x \in \mathbb{P}^n(\mathbb{Q}) \\ H(x) \le B \\ \pi^{-1}(x) \text{ smooth}}} \prod_{p \mid Q} \theta_p(x) \right) - \left(\prod_{p \mid Q} \sigma_p \right) y.$$

Note that $y = c_n B^{n+1} + O(B^n (\log B)^{[1/n]})$; indeed

$$\sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \le B, \pi^{-1}(x) \text{ singular}\}$$
$$\le \sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \le B, f(x) = 0\} \ll B^n$$

by Lemma 3.2 and [21, Thm. 13.4]. To study $\mathcal{E}_\mathcal{P}(\mathcal{A}, h, r)$ we use this and Lemma 3.3 to show that if Q is square-free and is divided only by primes $p > A$, then

$$\mathcal{A}_Q - \frac{h(Q)}{Q} y = \left(\sum_{\substack{x \in \mathbb{P}^n(\mathbb{Q}) \\ H(x) \le B \\ \pi^{-1}(x) \text{ smooth}}} \prod_{p | Q} \theta_p(x) \right) - \left(\prod_{p|Q} \sigma_p \right) c_n B^{n+1}$$
$$+ O\left(d^{\omega(Q)} \frac{B^n (\log B)^{[1/n]}}{Q} \right).$$

We can now employ Proposition 3.4 and Lemma 3.7 to see that

$$\sum_{\substack{x \in \mathbb{P}^n(\mathbb{Q}) \\ H(x) \le B \\ \pi^{-1}(x) \text{ smooth}}} \prod_{p | Q} \theta_p(x) - \left(\prod_{p|Q} \sigma_p \right) c_n B^{n+1}$$
$$\ll \frac{B^{n+1} (2\alpha d)^{\omega(Q)}}{Q \min\{p : p \mid Q\}} + (4d)^{\omega(Q)} (Q^{2n+1} B + Q B^n (\log B)^{[1/n]}).$$

Noting that $A^{\omega(Q)}/Q \ll Q^{-0.9} \ll Q^{0.9} \ll (4d)^{\omega(Q)} Q$ we deduce that

$$\left| \mathcal{A}_Q - \frac{h(Q)}{Q} y \right| \ll \frac{B^{n+1} (2\alpha d)^{\omega(Q)}}{Q \min\{p : p \mid Q\}} + (4d)^{\omega(Q)} (Q^{2n+1} B + Q B^n (\log B)^{[1/n]}).$$

For any square-free Q that is divisible by at most r primes, all lying in $(t_0, t_1]$, we have $Q \le t_1^r$. Therefore, in the notation of Lemma 3.8 we have

$$\mathcal{E}_\mathcal{P}(\mathcal{A}, h, r) \ll_r B^{n+1} \left(\sum_{\substack{Q \in \mathbb{N}, \mu(Q)^2 = 1 \\ p|Q \Rightarrow p \in \mathcal{P} \\ \omega(Q) \le r}} \frac{1}{Q \min\{p : p \mid Q\}} \right)$$
$$+ (t_1^{r(2n+1)} B + t_1^r B^n (\log B)^{[1/n]}) t_1^r,$$

where we used the estimate

$$\sum_{\substack{Q \in \mathbb{N}, \mu(Q)^2 = 1 \\ \omega(Q) \le r, p|Q \Rightarrow p \in \mathcal{P}}} 1 \le \sharp\{Q \in \mathbb{N} : Q \le t_1^r\} = t_1^r.$$

Writing $Q = p_1 \cdots p_r$ with $p_i < p_{i+1}$ we have

$$\sum_{\substack{Q \in \mathbb{N}, \mu(Q)^2 = 1 \\ \omega(Q) \leq r, p|Q \Rightarrow p \in \mathcal{P}}} \frac{1}{Q \min\{p : p \mid Q\}}$$

$$= \sum_{t_0 < p_r \leq t_1} \frac{1}{p_r} \sum_{t_0 < p_{r-1} < p_r} \frac{1}{p_{r-1}} \cdots \sum_{t_0 < p_2 < p_3} \frac{1}{p_2} \sum_{t_0 < p_1 < p_2} \frac{1}{p_1^2},$$

which can be seen to be $\ll \frac{(\log\log t_1)^{r-1}}{t_0}$ due to $\sum_{p \leq t_1} p^{-1} \ll \log\log t_1$ and $\sum_{p > t_0} p^{-2} \ll t_0^{-1}$. Using assumption (3.14) we obtain

$$\mathcal{E}_{\mathcal{P}}(\mathcal{A}, h, r) \ll_r \frac{B^{n+1}(\log\log B)^{r-1}}{t_0} + (t_1^{r(2n+1)} B + t_1^r B^n (\log B)^{[1/n]}) t_1^r$$

$$\ll_r B^{n+1}(\log\log B)^{-r-1} + B^{n+1/2}$$

$$\ll_r B^{n+1}(\log\log B)^{-r-1}.$$

Define

$$\hat{\mu}(B) := \sum_{t_0 < p \leq t_1} \sigma_p \quad \text{and} \quad \hat{\sigma}(B) := \left(\sum_{t_0 < p \leq t_1} (\sigma_p - \sigma_p^2) \right)^{1/2}.$$

Note that by (3.10)–(3.11) and (3.14) we have

$$\hat{\mu}(B) = \Delta(\pi) \log\log B + O(\log\log\log\log B). \tag{3.17}$$

Furthermore, Lemma 3.3 shows that $\sum_p \sigma_p^2 = O(1)$, hence

$$\hat{\sigma}(B)^2 = \hat{\mu}(B) + O_r(1).$$

By (3.17) we get $\hat{\sigma}(B) = (\Delta(\pi) \log\log B)^{1/2}(1 + O_r((\log\log\log\log B)/\log\log B))^{1/2}$, hence using the estimate $(1 + \epsilon)^{1/2} = 1 + O(\epsilon)$ that is valid for all $0 < \epsilon < 1$, we obtain

$$\hat{\sigma}(B) = (\Delta(\pi) \log\log B)^{1/2} + O_r\left(\frac{\log\log\log\log B}{(\log\log B)^{1/2}} \right). \tag{3.18}$$

We therefore see that the error term in (3.12) is

$$\ll_r B^{n+1}(\log\log B)^{r/2-1} + B^{n+1}(\log\log B)^{-1} \ll_r B^{n+1}(\log\log B)^{r/2-1}.$$

Noting that

$$\sum_{a \in \mathcal{A}} \left(\sharp\{p \in \mathcal{P} : p \mid a\} - \mu_{\mathcal{P}}(h) \right)^r = \sum_{\substack{x \in \mathbb{P}^n(\mathbb{Q}), H(x) \le B \\ \pi^{-1}(x) \text{ smooth}}} \omega_\pi^\flat(x, B)^r$$

establishes

$$\frac{\mathcal{M}_r^\flat(\pi, B)}{c_n B^{n+1}} = \mu_r \frac{\hat{\sigma}(B)^r}{(\Delta(\pi) \log \log B)^{r/2}} + O_r\left(\frac{1}{\log \log B}\right).$$

The proof of (3.15) can now be concluded by using (3.18) to verify

$$\frac{\hat{\sigma}(B)^r}{(\Delta(\pi) \log \log B)^{r/2}} = \left(1 + O_r\left(\frac{\log \log \log \log B}{\log \log B}\right)\right)^{r/2}$$
$$= 1 + O_r\left(\frac{\log \log \log \log B}{\log \log B}\right).$$

The proof of (3.16) can be performed in an entirely analogous manner by using (3.13). $\qquad\square$

3.4 Proof of Theorem 1.3

We first require the following preparatory lemma.

Lemma 3.10 *Let* $y(B), z(B)$ *be two functions satisfying*

$$z(B) > 1, \lim_{B \to \infty} y(B) = \infty \quad and \quad \lim_{B \to \infty} \frac{\log y(B)}{\log B} = 0.$$

Let $m \in \mathbb{Z}_{\ge 0}$ *and let* $F \in \mathbb{Z}[x_0, \dots, x_n]$ *be a primitive homogeneous polynomial. Then*

$$\sum_{\substack{x \in \mathbb{P}^n(\mathbb{Q}) \\ H(x) \le B \\ F(x) \ne 0}} \left(z(B) + \sum_{\substack{p \mid F(x) \\ p \le y(B)}} 1 \right)^m \ll_{F,m} B^{n+1} (z(B) + \log \log y(B))^m.$$

Proof It suffices to show that for every $r \in \mathbb{Z} \cap [0, m]$ we have

$$\sum_{\substack{x\in\mathbb{P}^n(\mathbb{Q}),H(x)\leq B \\ F(x)\neq 0}}\left(\sum_{\substack{p|F(x) \\ p\leq y(B)}}1\right)^r \ll_{F,r} B^{n+1}(\log\log y(B))^r, \qquad (3.19)$$

as the result will then easily follow from the binomial theorem. We have

$$\sum_{\substack{x\in\mathbb{P}^n(\mathbb{Q}),H(x)\leq B \\ F(x)\neq 0}}\left(\sum_{\substack{p|F(x) \\ p\leq y(B)}}1\right)^r \leq \sum_{p_1,\ldots,p_r\leq y(B)}\sum_{\substack{x\in\mathbb{P}^n(\mathbb{Q}),H(x)\leq B \\ 1\leq i\leq r\Rightarrow p_i|F(x)}}1.$$

Letting Q be the least common multiple of the primes $p_1,\ldots,p_r$ we see that $\omega(Q)\leq r$ and $\mu(Q)^2=1$. Furthermore, for every $Q\in\mathbb{N}$ having these two properties there are at most r^r vectors $(p_1,\ldots,p_r)$ with every prime p_i satisfying $p_i\leq y(B)$ and with Q being the least common multiple of the p_i. This is because for each $1\leq i\leq r$ the prime p_i must divide Q, so the number of available p_i is at most $\omega(Q)\leq r$. This shows that

$$\sum_{p_1,\ldots,p_r\leq y(B)}\sum_{\substack{x\in\mathbb{P}^n(\mathbb{Q}),H(x)\leq B \\ 1\leq i\leq r\Rightarrow p_i|F(x)}}1 \ll_r \sum_{\substack{Q\in\mathbb{N} \\ \omega(Q)\leq r}}\mu(Q)^2\sum_{\substack{x\in\mathbb{P}^n(\mathbb{Q}),H(x)\leq B \\ Q|F(x) \\ p|Q\Rightarrow p\leq y(B)}}1.$$

Letting

$$\Upsilon = \{x\in\mathbb{P}^n(\mathbb{Z}/Q\mathbb{Z}) : F(x)\equiv 0 \bmod Q\}$$

we may obtain the following via Lemma 2.2 and following similar steps as in the proof of Lemma 3.3,

$$\sharp\Upsilon = \prod_{p|Q}\sharp\{x\in\mathbb{P}^n(\mathbb{F}_p) : F(x)=0\} \leq \prod_{p|Q}(\deg F\cdot\sharp\mathbb{P}^{n-1}(\mathbb{F}_p))$$

$$\leq Q^{n-1}(2\deg F)^{\omega(Q)} \ll_r Q^{n-1}.$$

Noting that the assumption $\log y(B)=o(\log B)$ shows that $y(B)\ll_\epsilon B^\epsilon$ for every $\epsilon>0$. Hence, we have $Q\leq y(B)^r\ll_\epsilon B^\epsilon$ and invoking Proposition 2.1 with (2.2) we obtain

$$\sum_{\substack{x\in\mathbb{P}^n(\mathbb{Q}),H(x)\leq B \\ Q|F(x)}}1 \ll_{\epsilon,r} \frac{Q^{n-1}}{\sharp\mathbb{P}^n(\mathbb{Z}/Q\mathbb{Z})}B^{n+1} + Q^n\left(B+\frac{B^n}{Q^n}(\log B)^{[1/n]}\right)$$

$$\ll_{\epsilon,r} \frac{1}{Q}B^{n+1} + B^{n+\frac{1}{10}},$$

thus

$$\sum_{\substack{Q \in \mathbb{N}, \omega(Q) \leq r \\ p | Q \Rightarrow p \leq y(B)}} \mu(Q)^2 \sum_{\substack{x \in \mathbb{P}^n(\mathbb{Q}), H(x) \leq B \\ Q | F(x)}} 1$$

$$\ll_{\epsilon, r} B^{n+1} \left(\sum_{\substack{Q \in \mathbb{N}, \omega(Q) \leq r \\ p | Q \Rightarrow p \leq y(B)}} \frac{\mu(Q)^2}{Q} \right) + B^{n + \frac{1}{10}} \left(\sum_{\substack{Q \in \mathbb{N}, \omega(Q) \leq r \\ p | Q \Rightarrow p \leq y(B)}} \mu(Q)^2 \right).$$

The last sum over Q is at most $y(B)^r \ll_\epsilon B^\epsilon$, while the previous satisfies

$$\sum_{\substack{Q \in \mathbb{N}, \omega(Q) \leq r \\ p | Q \Rightarrow p \leq y(B)}} \frac{\mu(Q)^2}{Q} \leq \left(\sum_{p \leq y(B)} \frac{1}{p} \right)^r \ll_r (\log \log y(B))^r.$$

This verifies (3.19) and thus concludes the proof.   □

We begin the proof of Theorem 1.3 by noting that

$$\mathcal{M}_r(\pi, B) = \sum_{\substack{x \in \mathbb{P}^n(\mathbb{Q}), H(x) \leq B \\ f(x) \neq 0}} \left(\frac{\omega_\pi(x) - \Delta(\pi) \log \log B}{\sqrt{\Delta(\pi) \log \log B}} \right)^r + O_r(B^n (\log B)^r). \tag{3.20}$$

This is because by Lemma 3.1 we have

$$\sum_{\substack{x \in \mathbb{P}^n(\mathbb{Q}), H(x) \leq B \\ \pi^{-1}(x) \text{ smooth} \\ f(x)=0}} \left(\frac{\omega_\pi(x) - \Delta(\pi) \log \log B}{\sqrt{\Delta(\pi) \log \log B}} \right)^r \ll_r \sum_{\substack{x \in \mathbb{P}^n(\mathbb{Q}) \\ H(x) \leq B \\ f(x)=0}} (\log B)^r \ll B^n (\log B)^r.$$

We continue the proof of Theorem 1.3 by applying Proposition 3.9. For every x in the sum on the right side of (3.20), Lemma 3.1 shows that

$$\omega_\pi(x) - \Delta(\pi) \log \log B$$
$$= \omega_\pi^\flat(x, B) + \sum_{p \leq t_0(B)} \theta_p(x) + \sum_{t_1(B) < p \ll B^D} \theta_p(x)$$
$$+ \left(\left\{ \sum_{t_0(B) < p \leq t_1(B)} \sigma_p \right\} - \Delta(\pi) \log \log B \right).$$

Owing to (3.17) the last term is $\ll_{\mathcal{C},\epsilon_1}$ log log log log B. Using Lemma 3.2 and the trivial bound $0 \le \theta_p(x) \le 1$ we see that

$$\sum_{p \le t_0(B)} \theta_p(x) = \sum_{p \le A} \theta_p(x) + \sum_{\substack{p | f(x) \\ A < p \le t_0(B)}} \theta_p(x) \ll 1 + \sum_{\substack{p | f(x) \\ p \le t_0(B)}} 1.$$

Observe that

$$m \in \mathbb{Z}\setminus\{0\}, z \in \mathbb{R}_{>1} \Rightarrow \sharp\{p > z : p \mid m\} \le \frac{\log|m|}{\log z} \tag{3.21}$$

because $z^{\sharp\{p>z:p|m\}} \le \prod_{p|m} p \le m$. Hence, whenever $x \in \mathbb{P}^n(\mathbb{Q})$ is such that $H(x) \le B$ and $f(x) \ne 0$ we deduce by Lemma 3.2 that

$$\sum_{p > t_1(B)} \theta_p(x) = \sum_{\substack{p | f(x) \\ p > t_1(B)}} \theta_p(x) \le \sum_{\substack{p | f(x) \\ p > t_1(B)}} 1 \le \frac{\log|f(x)|}{\log t_1(B)} \ll \frac{\log B}{\log t_1(B)} \ll_{\epsilon_1} 1,$$

where we used the fact that $\log f(x) \ll \log H(x)$. We are thus led to the conclusion that for any x on the right side of (3.20) we have

$$\omega_\pi(x) - \Delta(\pi) \log\log B = \omega_\pi^\flat(x, B) + E_\pi(x, B),$$

for some function $E_\pi(x, B)$ satisfying

$$|E_\pi(x, B)| \ll_{\mathcal{C},\epsilon_1} \log\log\log\log B + \sum_{p|f(x), p \le t_0(B)} 1. \tag{3.22}$$

Therefore, we obtain for $r \ne 0$ that

$$(\omega_\pi(x) - \Delta(\pi) \log\log B)^r = \omega_\pi^\flat(x, B)^r + \sum_{k=0}^{r-1} \binom{r}{k} \omega_\pi^\flat(x, B)^k E_\pi(x, B)^{r-k}.$$

This allows the comparison with the "truncated moment" (3.9), to find via (3.20) that

$$\left| \mathcal{M}_r(\pi, B) - \mathcal{M}_r^\flat(\pi, B) \right|$$

$$\ll_r \frac{B^n}{(\log B)^{-r}} + \sum_{k=0}^{r-1} \sum_{\substack{x \in \mathbb{P}^n(\mathbb{Q}), H(x) \le B \\ f(x) \ne 0}} \frac{\omega_\pi^\flat(x, B)^k E_\pi(x, B)^{r-k}}{(\log\log B)^{r/2}}.$$

By Cauchy's inequality we see that the last sum is

$$\ll_r \frac{1}{(\log\log B)^{r/2}} \sum_{k=0}^{r-1} \left(\sum_{\substack{x\in\mathbb{P}^n(\mathbb{Q}), H(x)\leq B \\ f(x)\neq 0}} \omega_\pi^\flat(x,B)^{2k} \right)^{1/2}$$

$$\times \left(\sum_{\substack{x\in\mathbb{P}^n(\mathbb{Q}), H(x)\leq B \\ f(x)\neq 0}} E_\pi(x,B)^{2(r-k)} \right)^{1/2}.$$

We apply Proposition 3.9 with $r = 2k$ to obtain

$$\left| \mathcal{M}_r(\pi,B) - \mathcal{M}_r^\flat(\pi,B) \right|$$
$$\ll_r \frac{B^n}{(\log B)^{-r}} + \sum_{k=0}^{r-1} \frac{B^{(n+1)/2}}{(\log\log B)^{(r-k)/2}} \left(\sum_{\substack{x\in\mathbb{P}^n(\mathbb{Q}) \\ H(x)\leq B \\ f(x)\neq 0}} E_\pi(x,B)^{2(r-k)} \right)^{1/2}.$$

Recalling (3.22) and applying Lemma 3.10 with

$$m = 2(r-k), \ z(B) := \log\log\log\log B \ \text{ and } \ y(B) := t_0(B),$$

we see that, in light of $z(B) + \log\log y(B) \ll_r \log\log\log\log B$, one has

$$\frac{\mathcal{M}_r(\pi,B) - \mathcal{M}_r^\flat(\pi,B)}{B^{n+1}} \ll_r \frac{(\log B)^r}{B} + \frac{\log\log\log\log B}{(\log\log B)^{1/2}}.$$

We conclude that

$$\mathcal{M}_r(\pi,B) = \mathcal{M}_r^\flat(\pi,B) + O_r\left(B^{n+1} \frac{\log\log\log\log B}{(\log\log B)^{1/2}} \right).$$

An application of Proposition 3.9 completes the proof of Theorem 1.3. □

3.5 Proof of Theorem 1.2

Lemma 3.11 *There exists a set $\mathcal{S} \subset \{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \leq B\}$ with*

$$\sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \leq B, x \notin \mathcal{S}\} = O\left(\frac{B^{n+1}}{(\log\log\log B)^2} \right) \tag{3.23}$$

and such that for all $x \in \mathcal{S}$ we have

$$\frac{\omega_\pi(x) - \Delta(\pi) \log \log B}{\sqrt{\Delta(\pi) \log \log B}} = \frac{\omega_\pi(x) - \Delta(\pi) \log \log H(x)}{\sqrt{\Delta(\pi) \log \log H(x)}} + O\left(\frac{1}{\sqrt{\log B}}\right). \quad (3.24)$$

Proof Denote $\Xi(B) = \log \log \log B$ and define

$$\mathcal{S} := \left\{ x \in \mathbb{P}^n(\mathbb{Q}) : \begin{array}{l} \frac{B}{\log B} < H(x) \leq B, \pi^{-1}(x) \text{ smooth,} \\ \left| \frac{\omega_\pi(x) - \Delta(\pi) \log \log B}{\sqrt{\Delta(\pi) \log \log B}} \right| \leq \Xi(B) \end{array} \right\}.$$

The cardinality of those $x \in \mathbb{P}^n(\mathbb{Q})$ with $H(x) \leq B$, $\pi^{-1}(x)$ smooth and

$$\left| \frac{\omega_\pi(x) - \Delta(\pi) \log \log B}{\sqrt{\Delta(\pi) \log \log B}} \right| > \Xi(B)$$

is at most

$$\sum_{\substack{x \in \mathbb{P}^n(\mathbb{Q}), H(x) \leq B \\ \pi^{-1}(x) \text{ smooth}}} \left(\frac{\omega_\pi(x) - \Delta(\pi) \log \log B}{\Xi(B) \sqrt{\Delta(\pi) \log \log B}} \right)^2 = \frac{c_n B^{n+1}}{\Xi(B)^2} (1 + o(1)) \ll \frac{B^{n+1}}{\Xi(B)^2},$$

where the case $r = 2$ of Theorem 1.3 has been used. This provides us with

$$\sharp\mathcal{S} = c_n B^{n+1} + O\left(\frac{B^{n+1}}{\Xi(B)^2}\right).$$

Now note that for all $x \in \mathbb{P}^n(\mathbb{Q})$ with $B/\log B < H(x) \leq B$ we have

$$\log \log H(x) = \log \log B + O\left(\frac{\log \log B}{\log B}\right),$$

therefore $(\log \log H(x) / \log \log B)^{1/2} = 1 + O(1/\log B)$. Thus, for each such x we get

$$\frac{\omega_\pi(x) - \Delta(\pi) \log \log B}{\sqrt{\Delta(\pi) \log \log B}} = \frac{\omega_\pi(x) - \Delta(\pi) \log \log B}{\sqrt{\Delta(\pi) \log \log H(x)}}$$

$$+ O\left(\frac{|\omega_\pi(x) - \Delta(\pi) \log \log B|}{\log B \sqrt{\log \log B}}\right).$$

We deduce that if $x \in \mathcal{S}$ then this is

$$\frac{\omega_\pi(x) - \Delta(\pi)\log\log B}{\sqrt{\Delta(\pi)\log\log H(x)}} + O\left(\frac{\Xi(B)}{\log B}\right) = \frac{\omega_\pi(x) - \Delta(\pi)\log\log H(x)}{\sqrt{\Delta(\pi)\log\log H(x)}}$$
$$+ O\left(\frac{\sqrt{\log\log B}}{\log B} + \frac{\Xi(B)}{\log B}\right),$$

which is sufficient for our lemma. □

We are now in place to prove Theorem 1.2. For $z \in \mathbb{R}$ we denote the distribution function of the standard normal distribution by

$$\Phi(z) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}}\, dt.$$

Recall the definition of the probability measure ν_B in (1.4) and note that the set $\mathbb{P}^n(\mathbb{Q})$ becomes a probability space once equipped with the measure ν_B. (The measure ν_B is supported on the rational points of height at most B.) For any $B \in \mathbb{R}_{\geq 3}$ we consider the random variable X_B defined on $\mathbb{P}^n(\mathbb{Q})$ as follows,

$$\mathrm{X}_B(x) := \begin{cases} \frac{\omega_\pi(x) - \Delta(\pi)\log\log B}{\sqrt{\Delta(\pi)\log\log B}}, & \pi^{-1}(x) \text{ smooth}, \\ 0, & \text{otherwise}. \end{cases}$$

For $r \in \mathbb{Z}_{\geq 0}$ the r-th moment of X_B is by definition equal to

$$\int_{-\infty}^{+\infty} \mathrm{X}_B^r d\nu_B = \sum_{\substack{x \in \mathbb{P}^n(\mathbb{Q}) \\ \pi^{-1}(x) \text{ smooth}}} \left(\frac{\omega_\pi(x) - \Delta(\pi)\log\log B}{\sqrt{\Delta(\pi)\log\log B}}\right)^r \frac{\nu_B(\{x\})}{\sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \leq B\}}$$

and recalling (1.5) we see that this coincides with $\mathcal{M}_r(\pi, B)/\sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \leq B\}$. Theorem 1.3 shows that

$$\lim_{B \to +\infty} \int_{-\infty}^{+\infty} \mathrm{X}_B^r d\nu_B$$

exists and is equal to the r-th moment of the standard normal distribution. By [4, Th. 30.2] we get that X_B converges in law to the standard normal distribution, i.e. for every $y \in \mathbb{R}$ we have

$$\lim_{B \to +\infty} \nu_B\left(\{x \in \mathbb{P}^n(\mathbb{Q}) : \mathrm{X}_B \leq y\}\right) = \Phi(y). \tag{3.25}$$

Next, for every fixed $\epsilon > 0$, $z \in \mathbb{R}$ and all sufficiently large B we see that the error term in (3.24) has modulus at most ϵ, therefore (3.24) gives

$$\nu_B \left(\{ x \in \mathcal{S} : \mathrm{X}_B(x) \leq z - \epsilon \} \right) \leq \nu_B \left(\left\{ x \in \mathcal{S} : \frac{\omega_\pi(x) - \Delta(\pi) \log \log H(x)}{\sqrt{\Delta(\pi) \log \log H(x)}} \leq z \right\} \right)$$

(3.26)

and

$$\nu_B \left(\{ x \in \mathcal{S} : \mathrm{X}_B(x) \leq z + \epsilon \} \right) \geq \nu_B \left(\left\{ x \in \mathcal{S} : \frac{\omega_\pi(x) - \Delta(\pi) \log \log H(x)}{\sqrt{\Delta(\pi) \log \log H(x)}} \leq z \right\} \right)$$

(3.27)

for all sufficiently large B. Observe that the set $\mathcal{S}$ in Lemma 3.11 satisfies the following as $B \to +\infty$ due to (3.23),

$$\nu_B(\mathcal{S}) = 1 + O \left(\frac{1}{(\log \log \log B)^2} \right) = 1 + o(1). \tag{3.28}$$

In light of (3.25) this means that $\lim_{B \to +\infty} \nu_B (\{ x \in \mathcal{S} : \mathrm{X}_B \leq y \}) = \Phi(y)$, which, when applied to $y = z - \epsilon$ and $y = z + \epsilon$ and combined with (3.26) and (3.27), yields

$$\liminf_{B \to +\infty} \nu_B \left(\left\{ x \in \mathcal{S} : \frac{\omega_\pi(x) - \Delta(\pi) \log \log H(x)}{\sqrt{\Delta(\pi) \log \log H(x)}} \leq z \right\} \right) \geq \Phi(z - \epsilon)$$

and

$$\limsup_{B \to +\infty} \nu_B \left(\left\{ x \in \mathcal{S} : \frac{\omega_\pi(x) - \Delta(\pi) \log \log H(x)}{\sqrt{\Delta(\pi) \log \log H(x)}} \leq z \right\} \right) \leq \Phi(z + \epsilon).$$

Letting $\epsilon \to 0$ and using the fact that Φ is continuous we obtain

$$\lim_{B \to +\infty} \nu_B \left(\left\{ x \in \mathcal{S} : \frac{\omega_\pi(x) - \Delta(\pi) \log \log H(x)}{\sqrt{\Delta(\pi) \log \log H(x)}} \leq z \right\} \right) = \Phi(z),$$

which, by (3.28) implies that

$$\lim_{B \to +\infty} \nu_B \left(\left\{ x \in \mathbb{P}^n(\mathbb{Q}) : \frac{\omega_\pi(x) - \Delta(\pi) \log \log H(x)}{\sqrt{\Delta(\pi) \log \log H(x)}} \leq z \right\} \right) = \Phi(z).$$

Since this holds for every fixed $z \in \mathbb{R}$ it gives Theorem 1.2 as an immediate consequence. $\qquad \square$

4 The pseudo-split case

4.1 Proof of Theorem 1.7

We let $\pi : V \to \mathbb{P}^n$ be as in Theorem 1.7 and choose a model for π as in Sect. 3.1.

4.1.1 Existence of the limit

We first prove the existence of (1.8) using the versions of the sieve of Ekedahl given in [14, §4.1], [2, §3] and [3, §3]. We begin with a strengthening of Lemma 3.2. (Here it is crucial that $\Delta(\pi) = 0$.)

Lemma 4.1 *There exists a closed subset $Z \subset \mathbb{P}^n_{\mathbb{Z}}$ of codimension at least 2 and a constant $A > 0$ with the following property. Let $p > A$ be a prime and $x \in \mathbb{P}^n(\mathbb{Z}_p)$ such that $\pi^{-1}(x)(\mathbb{Z}_p) = \emptyset$. Then $x \bmod p \in Z(\mathbb{F}_p)$.*

Proof This is a special case of [14, Prop. 4.1]. ☐

Lemma 4.2 *Let A be as in Lemma 4.1 Then for every $B, M > 1$ we have*

$$\sharp\left\{x \in \mathbb{P}^n(\mathbb{Q}) : \begin{array}{l} H(x) \leq B, \pi^{-1}(x) \text{ smooth,} \\ \exists p > M \text{ s.t. } \pi^{-1}(x)(\mathbb{Q}_p) = \emptyset \end{array}\right\} \ll \frac{B^{n+1}}{M \log M} + B^n,$$

where the implied constant depends at most on A and π.

Proof This follows immediately from Lemma 4.1 and Bhargava's effective version of the Ekedahl sieve [3, Thm. 3.3]. ☐

We now prove a strengthening of Proposition 3.4 in the case $\Delta(\pi) = 0$. The crucial point about the next proposition is that it gives an asymptotic formula for a counting problem which has local conditions imposed at *every* place v, whereas Proposition 3.4 only imposes conditions at *finitely* many primes. In what follows we use the measures ϑ_v from Sect. 2.2.

Proposition 4.3 *Let S be a finite set of places of $\mathbb{Q}$. Then*

$$\lim_{B \to \infty} \frac{\sharp\left\{x \in \mathbb{P}^n(\mathbb{Q}) : \begin{array}{l} H(x) \leq B, \pi^{-1}(x) \text{ smooth,} \\ v \in S \Leftrightarrow \pi^{-1}(x)(\mathbb{Q}_v) = \emptyset \end{array}\right\}}{\sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \leq B\}}$$
$$= \prod_{v \in S} \vartheta_v(\mathbb{P}^n(\mathbb{Q}_v) \backslash \pi(V(\mathbb{Q}_v)) \prod_{v \notin S} \vartheta_v(\pi(V(\mathbb{Q}_v))),$$

where the right hand side is a convergent Euler product.
Moreover, let $Q \in \mathbb{N}$ be square-free. Then there exists $K_0 > 0$ such that

$$\sharp\left\{x \in \mathbb{P}^n(\mathbb{Q}) : \begin{array}{l} H(x) \leq B, \pi^{-1}(x) \text{ smooth,} \\ \pi^{-1}(x)(\mathbb{Q}_p) = \emptyset \,\forall\, p \mid Q \end{array}\right\}$$
$$\ll K_0^{\omega(Q)}\left(\frac{B^{n+1}}{Q^2} + BQ^{n-1} + \frac{B^n \log B}{Q}\right). \tag{4.1}$$

Proof The asymptotic formula is proved using an adaptation of [2, Thm. 3.8], via the version of the sieve of Ekedahl given in [2, Prop. 3.4]. That the condition (3.5) from [2, Prop 3.4] is satisfied follows from Lemma 4.1 and [2, Lem. 3.5]. If $\pi(V(\mathbb{Q}_v)) \neq \emptyset$ then [2, Lem. 3.9] implies that $\pi(V(\mathbb{Q}_v))$ is measurable, has positive measure and has boundary of measure 0. Moreover $\pi(V(\mathbb{Q}_v)) \subset \mathbb{P}^n(\mathbb{Q}_v)$ is closed as π is proper. It follows that $\mathbb{P}^n(\mathbb{Q}_v)\backslash\pi(V(\mathbb{Q}_v))$ is measurable, has boundary of measure 0, and has positive measure if non-empty. Therefore the measurability hypotheses in [2, Prop. 3.4] are all satisfied. Applying [2, Prop. 3.4] gives the asymptotic formula. (Note that [2, Lem. 4.8] works with the Haar measure on $\mathbb{Q}_v^{n+1}$, whereas in the statement we use the measure ϑ_v. One easily obtains our statement using Lemma 2.3.)

Next, by the Lang–Weil estimates there exists $K_0 > 0$ such that $\sharp Z(\mathbb{F}_p) \leq K_0 \, p^{n-2}$ for all p. Therefore we have $\sharp Z(\mathbb{Z}/Q\mathbb{Z}) \leq K_0^{\omega(Q)} Q^{n-2}$. Lemma 4.1 and Proposition 2.1 now show that the left side of (4.1) is

$$\ll \sharp \left\{ x \in \mathbb{P}^n(\mathbb{Q}) : \begin{array}{l} H(x) \leq B, \pi^{-1}(x) \text{ smooth}, \\ p \mid Q \Rightarrow x \bmod p \in Z(\mathbb{F}_p) \end{array} \right\}$$

$$\ll \frac{\sharp Z(\mathbb{Z}/Q\mathbb{Z})}{\sharp \mathbb{P}^n(\mathbb{Z}/Q\mathbb{Z})} B^{n+1} + Q \sharp Z(\mathbb{Z}/Q\mathbb{Z}) \left(B + \frac{B^n}{Q^n} \log B \right)$$

$$\ll K_0^{\omega(Q)} Q^{-2} B^{n+1} + K_0^{\omega(Q)} Q^{n-1} B + K_0^{\omega(Q)} Q^{-1} B^n \log B.$$

$\square$

We now show the existence of the limit (1.8). In fact, we prove the following explicit formula. (Recall the definition of $\tau_\pi(B, j)$ from (1.7).)

Proposition 4.4 *We have*

$$\tau_\pi(j) := \lim_{B \to \infty} \tau_\pi(B, j)$$

$$= \sum_{p_1 < p_2 < \cdots < p_j} \prod_{p \mid p_1 \cdots p_j} \vartheta_p(\mathbb{P}^n(\mathbb{Q}_p)\backslash\pi(V(\mathbb{Q}_p))) \prod_{p \nmid p_1 \cdots p_j} \vartheta_p(\pi(V(\mathbb{Q}_p))),$$

where the sum and products are convergent.

Proof If $Q \in [1, B^{1/3})$ is a square-free integer then one can immediately see from (4.1) that

$$\sharp \left\{ x \in \mathbb{P}^n(\mathbb{Q}) : \begin{array}{l} H(x) \leq B, \pi^{-1}(x) \text{ smooth}, \\ p \mid Q \Rightarrow \pi^{-1}(x)(\mathbb{Q}_p) = \emptyset \end{array} \right\} \ll \frac{K_0^{\omega(Q)}}{Q^2} B^{n+1}. \qquad (4.2)$$

Combining the upper bound and the asymptotic provided by Proposition 4.3 one sees that for every square-free $Q \neq 0$ we have

$$\prod_{p \mid Q} \vartheta_p(\mathbb{P}^n(\mathbb{Q}_p) \setminus \pi(V(\mathbb{Q}_p))) \prod_{p \nmid Q} \vartheta_p(\pi(V(\mathbb{Q}_p))) \ll \frac{K_0^{\omega(Q)}}{Q^2}, \qquad (4.3)$$

with an implied constant independent of Q. Fix any $M > 1$. By Lemma 4.2 we see that

$$\tau_\pi(B, j) = \frac{\sharp\left\{x \in \mathbb{P}^n(\mathbb{Q}) : \begin{array}{l} H(x) \le B, \pi^{-1}(x) \text{ smooth}, \omega_\pi(x) = j, \\ \pi^{-1}(x)(\mathbb{Q}_p) = \emptyset \Rightarrow p \le M \end{array}\right\}}{\sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \le B\}}$$
$$+ O\left(\frac{1}{M \log M} + \frac{1}{B}\right),$$

with an implied constant that is independent of j, M and B. We infer that $\tau_\pi(B, j)$ equals

$$\sum_{p_1 < p_2 < \cdots < p_j \le M} \frac{\sharp\left\{x \in \mathbb{P}^n(\mathbb{Q}) : \begin{array}{l} H(x) \le B, \pi^{-1}(x) \text{ smooth}, \\ \pi^{-1}(x)(\mathbb{Q}_p) = \emptyset \Leftrightarrow p \mid p_1 \cdots p_j \end{array}\right\}}{\sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \le B\}}$$
$$+ O\left(\frac{1}{M \log M} + \frac{1}{B}\right).$$

Fixing the value of M and taking the limit as $B \to \infty$, we see from Proposition 4.3 that

$$\limsup_{B \to \infty} \left| \tau_\pi(B, j) - \sum_{p_1 < \cdots < p_j \le M} \prod_{p \mid p_1 \cdots p_j} \vartheta_p(\mathbb{P}^n(\mathbb{Q}_p) \backslash \pi(V(\mathbb{Q}_p))) \prod_{p \nmid p_1 \cdots p_j} \vartheta_p(\pi(V(\mathbb{Q}_p))) \right| \ll \frac{1}{M}.$$

(4.4)

Note that the infinite series

$$\sum_{p_1 < p_2 < \cdots < p_j} \prod_{p \mid p_1 \cdots p_j} \vartheta_p(\mathbb{P}^n(\mathbb{Q}_p) \backslash \pi(V(\mathbb{Q}_p))) \prod_{p \nmid p_1 \cdots p_j} \vartheta_p(\pi(V(\mathbb{Q}_p)))$$

converges owing to the bound

$$\prod_{p \mid p_1 \cdots p_j} \vartheta_p(\mathbb{P}^n(\mathbb{Q}_p) \backslash \pi(V(\mathbb{Q}_p))) \prod_{p \nmid p_1 \cdots p_j} \vartheta_p(\pi(V(\mathbb{Q}_p)))$$
$$\ll \frac{K_0^j}{(p_1 \cdots p_j)^2},$$

(4.5)

that follows from (4.3). Taking M to be arbitrarily large in (4.4) proves the result. □

4.1.2 Probability measure

We now show that (1.8) indeed defines a probability measure on $\mathbb{Z}$. To do so, it suffices to show that

$$\sum_{j \in \mathbb{Z}} \tau_\pi(j) = 1. \tag{4.6}$$

Partitioning all possible values for $\omega_\pi(x)$ we have

$$\sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \le B, \pi^{-1}(x) \text{ smooth}\}$$
$$= \sum_{j=0}^{\infty} \sharp\left\{x \in \mathbb{P}^n(\mathbb{Q}) : \begin{array}{l} H(x) \le B, \omega_\pi(x) = j, \\ \pi^{-1}(x) \text{ smooth} \end{array}\right\}.$$

Fix $j_0 \in \mathbb{N}$ and note that if $\omega_\pi(x) > j_0$ then the largest prime p such that $\pi^{-1}(x)(\mathbb{Q}_p) = \emptyset$ exceeds the j_0-th largest prime, therefore it is at least j_0. This shows that

$$\sum_{j > j_0} \sharp\left\{x \in \mathbb{P}^n(\mathbb{Q}) : \begin{array}{l} H(x) \le B, \omega_\pi(x) = j, \\ \pi^{-1}(x) \text{ smooth} \end{array}\right\}$$
$$\le \sharp\left\{x \in \mathbb{P}^n(\mathbb{Q}) : \begin{array}{l} H(x) \le B, \pi^{-1}(x) \text{ smooth}, \\ \exists p > j_0 \text{ s.t. } \pi^{-1}(x)(\mathbb{Q}_p) = \emptyset \end{array}\right\}$$

which is $O(B^{n+1} j_0^{-1} + B^n)$ by Lemma 4.2. Dividing by $\sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \le B\}$ gives

$$\frac{\sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \le B, \pi^{-1}(x) \text{ smooth}\}}{\sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \le B\}} = \sum_{0 \le j \le j_0} \tau_\pi(B, j) + O\left(\frac{1}{j_0} + \frac{1}{B}\right),$$

with an implied constant that is independent of j_0. Letting $B \to \infty$ we obtain

$$\sum_{0 \le j \le j_0} \tau_\pi(j) = 1 + O\left(\frac{1}{j_0}\right).$$

Letting $j_0 \to \infty$ we infer that the sum over j converges to 1, thus verifying (4.6).

4.1.3 Upper bounds

We now prove (1.9). Combining (4.5) and Proposition 4.4 shows that

$$\tau_\pi(j) = \lim_{B \to \infty} \tau_\pi(j, B) \ll_\pi K_0^j \sum_{p_1 < \ldots < p_j} \frac{1}{p_1^2 \cdots p_j^2}.$$

Let us denote the primes in ascending order as $q_1 = 2$, $q_2 = 3$, etc. Writing

$$K_0^j \sum_{p_1 < ... < p_j} \frac{1}{p_1^2 \cdots p_j^2} = \sum_{p_1 \geq 2} \frac{K_0}{p_1^2} \sum_{p_2 > p_1} \frac{K_0}{p_2^2} \cdots \sum_{p_j > p_{j-1}} \frac{K_0}{p_j^2},$$

we observe that the sum over p_2 contains all primes $p \geq q_2$, the sum over p_3 contains all primes $p \geq q_3$ and so on. Therefore, one has

$$\tau_\pi(j) \ll \prod_{i=1}^j \left(\sum_{p \geq q_i} \frac{K_0}{p^2} \right).$$

By the prime number theorem and partial summation we see that $\sum_{p \geq z} p^{-2} \leq c_0/(z \log z)$ for some absolute $c_0 > 1$, thus $\tau_\pi(j) \prod_{i=1}^j (q_i \log q_i) \ll (c_0 K_0)^j$. Using $q_i \sim i \log i$ and the prime number theorem with partial summation we obtain

$$
\begin{aligned}
\log \left(\prod_{i=1}^j q_i \log q_i \right) &= \sum_{p \leq q_j} \log p + \sum_{p \leq q_j} \log \log p \\
&= q_j + O\left(\frac{q_j}{\log q_j} \right) + \frac{q_j \log \log q_j}{\log q_j} + O\left(\frac{q_j \log \log q_j}{(\log q_j)^2} \right) \\
&= j \log j + j \log \log j + o(j \log \log j) \\
&\geq j \log j + \frac{3}{4} j \log \log j,
\end{aligned}
$$

for all sufficiently large j. We deduce that for all large j one has

$$\tau_\pi(j) \leq (c_0 K_0)^j \prod_{i=1}^j \frac{1}{q_i \log q_i} \leq \frac{(c_0 K_0)^j}{j^j (\log j)^{\frac{3j}{4}}} \leq \frac{1}{j^j (\log j)^{j/2}},$$

from which (1.9) follows. This completes the proof of Theorem 1.7.   □

4.2 The family of diagonal cubic surfaces

We now return to Example 1.9, and prove the claim that there exists an absolute constant $c > 0$ such that $\tau_\pi(j) > c(1 + j)^{-3j}$.

Let $y = (y_0 : y_1 : y_2 : y_3) \in \mathbb{P}^3(\mathbb{Z}_p)$ where $(y_0, \ldots, y_3)$ is primitive. By the criterion in [7, p.28], if a prime $p \equiv 1 \bmod 3$ satisfies $p \nmid y_0 y_1$, $p \| y_2$, $p \| y_3$ and neither $-y_1/y_0$ nor $-y_3/y_2$ are cubes, then $\pi^{-1}(y)$ has no p-adic point. It is easy to see that there exists an absolute constant $K_1 > 0$ such that the measure of this with respect to ϑ_p is at least $K_1 p^{-2}$. Hence, denoting by q_i the i-th largest prime being

1 mod 3, Proposition 4.4 gives

$$\tau_\pi(j) \geq K_2 \prod_{i=1}^{j} \frac{K_1}{q_i^2},$$

for some constant $K_2 > 0$ (the product in Proposition 4.4 being convergent). By the prime number theorem for arithmetic progressions we have $q_i \sim 2i \log i$, hence

$$\log \tau_\pi(j)^{-1} \leq -\log K_2 - j(\log K_1) + 2 \sum_{\substack{p \leq (2j \log j)(1+o(1)) \\ p \equiv 1 \bmod 3}} \log p = (2j \log j)(1 + o(1)) \leq 3j \log j$$

for all sufficiently large j. This proves the claim.   □

4.3 Proof of Theorem 1.10

The implication $\Leftarrow$ is clear. For the other, assume that $V(\mathbb{Q}_p) \to \mathbb{P}^n(\mathbb{Q}_p)$ is not surjective for infinitely many primes p. Let S be a finite set of such primes and let $x_p \in \mathbb{P}^n(\mathbb{Q}_p) \backslash \pi(V(\mathbb{Q}_p))$ for $p \in S$. By Proposition 2.1, a positive proportion of $x \in \mathbb{P}^n(\mathbb{Q}_p)$ are arbitrarily close to the x_p for all $p \in S$. Moreover, as π is proper the set $\pi(V(\mathbb{Q}_p))$ is closed with respect to the p-adic topology. It follows that provided the x are sufficiently close to the x_p we have $\pi^{-1}(x)(\mathbb{Q}_p) = \emptyset$ for all $p \in S$. Hence for such x we have $\omega_\pi(x) \geq \sharp S$. As S can be chosen sufficiently large, the result follows.   □

4.4 Proof of Theorem 1.11

As shown in the proof of Proposition 4.3, the sets $\pi(V(\mathbb{Q}_v))$ and $\mathbb{P}^n(\mathbb{Q}_v) \backslash \pi(V(\mathbb{Q}_v))$ are measurable with respect to ϑ_v, and have positive measure if non-empty. The result now follows as the Euler product in Proposition 4.3 is convergent.   □

4.5 Proof of Theorem 1.12

Let $\mathcal{N}_r(\pi, B)$ be as in (1.10). We begin with the following.

Lemma 4.5 *For every $r \in \mathbb{Z}_{\geq 0}$ we have $\mathcal{N}_r(\pi, B) \ll_r B^{n+1}$.*

Proof Recall f and A from Lemma 3.2. By Lemma 3.1 we have

$$\mathcal{N}_r(\pi, B) = \sum_{\substack{x \in \mathbb{P}^n(\mathbb{Q}), H(x) \leq B \\ f(x) \neq 0}} \omega_\pi(x)^r + O\left(B^n (\log B)^r\right),$$

where we have used the evident bound $\sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \leq B, f(x) = 0\} \ll B^n$. For any $\epsilon > 0$ and any x with $f(x) \neq 0$ we have via (3.21) that

$$\sharp\{p > B^\epsilon : \pi^{-1}(x)(\mathbb{Q}_p) = \emptyset\} \leq \sharp\{p > B^\epsilon : p \text{ divides } f(x)\} \leq \frac{\log |f(x)|}{\log(B^\epsilon)}.$$

As $H(x) \leq B$ implies $|f(x)| \ll B^{\deg(f)}$, we thus find that

$$\omega_\pi(x) = O_\epsilon(1) + \sum_{A < p \leq B^\epsilon} \theta_p(x).$$

Let us now define $\epsilon(r) := (3r)^{-1}$. Then

$$\mathcal{N}_r(\pi, B) = \sum_{\substack{x \in \mathbb{P}^n(\mathbb{Q}), H(x) \leq B \\ f(x) \neq 0}} \left(O_r(1) + \sum_{A < p \leq B^{\epsilon(r)}} \theta_p(x) \right)^r$$

$$\ll_r \sum_{m=0}^{r} \sum_{\substack{x \in \mathbb{P}^n(\mathbb{Q}), H(x) \leq B \\ f(x) \neq 0}} \left(\sum_{A < p \leq B^{\epsilon(r)}} \theta_p(x) \right)^m.$$

To prove the lemma it therefore suffices to show that

$$\sum_{\substack{x \in \mathbb{P}^n(\mathbb{Q}), H(x) \leq B \\ f(x) \neq 0}} \left(\sum_{A < p \leq B^{\epsilon(r)}} \theta_p(x) \right)^m \ll_m B^{n+1}.$$

Using the multinomial theorem the sum over x above equals

$$\sum_{\substack{(m_p)_{p \in (A, B^{\epsilon(r)}]} \\ \sum_{p \in (A, B^{\epsilon(r)}]} m_p = m}} \frac{m!}{\prod_{p \in (A, B^{\epsilon(r)}]} m_p!} \sum_{\substack{x \in \mathbb{P}^n(\mathbb{Q}), H(x) \leq B \\ f(x) \neq 0}} \prod_{\substack{p \in (A, B^{\epsilon(r)}] \\ m_p \neq 0}} \theta_p(x).$$

Letting k be the number of $p \in (A, B^{\epsilon(r)}]$ with $m_p \neq 0$, shows that the last quantity is

$$\ll_m \sum_{k=1}^{m} \sum_{A < p_1 < p_2 < \cdots < p_k \leq B^{\epsilon(r)}} \sum_{\substack{x \in \mathbb{P}^n(\mathbb{Q}), H(x) \leq B \\ f(x) \neq 0}} \prod_{i=1}^{k} \theta_{p_i}(x).$$

By (4.2) and the fact that $k \leq m \leq r$ and $\epsilon(r) \leq (3m)^{-1}$ we see that the inner sum over x is $\ll B^{n+1} \prod_{i=1}^{k}(K_0 p_i^{-2})$. We obtain that

$$\sum_{\substack{A<p_1<p_2<\cdots<p_k\leq B^{\epsilon(r)}}} \sum_{\substack{x\in\mathbb{P}^n(\mathbb{Q}),H(x)\leq B \\ f(x)\neq 0}} \prod_{i=1}^{k}\theta_{p_i}(x)$$

$$\ll B^{n+1} \sum_{A<p_1<p_2<\cdots<p_k\leq B^{\epsilon(r)}} \prod_{i=1}^{k}(K_0 p_i^{-2}) \ll B^{n+1},$$

thus concluding our proof. $\square$

Now observe that

$$\frac{\mathcal{N}_r(\pi, B)}{\sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \leq B\}} = \sum_{j=0}^{\infty} j^r \tau_\pi(B, j). \qquad (4.7)$$

However, for any $M > 1$ we may use the inequality $\mathbf{1}_{\{\omega_\pi(x)>M\}}(x) \leq \omega_\pi(x)/M$ to find that

$$\sum_{j>M} j^r \tau_\pi(B, j) = \frac{\sum_{\substack{x\in\mathbb{P}^n(\mathbb{Q}),H(x)\leq B \\ \pi^{-1}(x) \text{ smooth}}} \omega_\pi(x)^r \mathbf{1}_{\{\omega_\pi(x)>M\}}(x)}{\sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \leq B\}} \ll \frac{\mathcal{N}_{r+1}(\pi, B)}{MB^{n+1}}.$$

$$(4.8)$$

We now infer from (4.7), (4.8) and Lemma 4.5 that

$$\left| \frac{\mathcal{N}_r(\pi, B)}{\sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \leq B\}} - \sum_{0\leq j\leq M} j^r \tau_\pi(B, j) \right| \ll_r \frac{1}{M}, \qquad (4.9)$$

where the implied constant is independent of B and M. Fixing M and taking $B \to \infty$ we are led to the conclusion that

$$\limsup_{B\to\infty} \left| \frac{\mathcal{N}_r(\pi, B)}{\sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \leq B\}} - \sum_{0\leq j\leq M} j^r \tau_\pi(j) \right| \ll \frac{1}{M}.$$

By (1.9) the sum over j is convergent as $M \to \infty$, which completes the proof. $\square$

4.6 Generalisations

One can consider variants of the function ω_π from (1.2), by considering real solubility or by dropping conditions at finitely many primes. Namely, let S be a finite set of

places of $\mathbb{Q}$. Then we define

$$\omega_{\pi,S}(x) := \sharp\{\text{places } v \notin S : \pi^{-1}(x)(\mathbb{Q}_v) = \emptyset\}.$$

We have considered the case $\omega_\pi = \omega_{\pi,\infty}(x)$ for simplicity of exposition, but a minor variant of our arguments yields the following generalisation of Theorem 1.7 (the important point being that the asymptotic in Proposition 4.3 applies to arbitrary S).

Theorem 4.6 *Let V be a smooth projective variety over $\mathbb{Q}$ equipped with a dominant morphism $\pi : V \to \mathbb{P}^n$ with geometrically integral generic fibre and $\Delta(\pi) = 0$. Let S be a finite set of places of $\mathbb{Q}$ and $j \in \mathbb{Z}$. Then the limit*

$$\lim_{B \to \infty} \frac{\sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \leq B, \pi^{-1}(x) \text{ smooth}, \omega_{\pi,S}(x) = j\}}{\sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \leq B\}}$$

exists and defines a probability measure on $\mathbb{Z}$.

The analogues of the other results from Sect. 1.2 also hold for the modified $\omega_{\pi,S}$. Also of course the analogue of Theorem 1.2 and the other results in Sect. 1.1 trivially hold with ω_π replaced by $\omega_{\pi,S}$, since $\omega_{\pi,S} = \omega_\pi + O(1)$.

References

1. Ax, J., Kochen, S.: Diophantine problems over local fields. I. Am. J. Math. **87**, 605–630 (1965)
2. Bright, M.J., Browning, T.D., Loughran, D.: Failures of weak approximation in families. Compos. Math. **152**(7), 1435–1475 (2016)
3. Bhargava, M.: The geometric sieve and the density of squarefree values of invariant polynomials, arXiv:1402.0031 (2014)
4. Billingsley, P.: The probability theory of additive arithmetic functions. Ann. Probab. **2**, 749–791 (1974)
5. Transactions of the American Mathematical Society, 371(8), 5757–5785 (2019)
6. Cantat, S., Dolgachev, I.: Rational surfaces with a large group of automorphisms. J. Am. Math. Soc. **25**(3), 863–905 (2012)
7. Colliot-Thélène, J.-L., Kanevsky, D., Sansuc, J.-J.: Arithmétique des surfaces cubiques diagonales, Diophantine approximation and transcendence theory (Bonn, : Lecture Notes in Math., vol. 1290. Springer, Berlin, 1987, 1–108 (1985)
8. Colliot-Thélène, J.-L., Skorobogatov, A.N., Swinnerton-Dyer, P.: Double fibres and double covers: paucity of rational points. Acta Arith. **79**(2), 113–135 (1997)
9. Erdős, P., Kac, M.: The Gaussian law of errors in the theory of additive number theoretic functions. Am. J. Math. **62**, 738–742 (1940)

10. Elliott, P.: Probabilistic Number Theory. II, vol. 240. Springer, Berlin (1980)
11. Granville, A., Soundararajan, K.: Sieving and the Erdős–Kac theorem, Equidistribution in number theory, an introduction, NATO Sci. Ser. II Math. Phys. Chem., vol. 237, Springer, Dordrecht, pp. 15–27 (2007)
12. Hooley, C.: On ternary quadratic forms that represent zero. Glasgow Math. J. **35**(1), 13–23 (1993)
13. Loughran, D.: The number of varieties in a family which contain a rational point. J. Eur. Math. Soc. (to appear) (2013)
14. Loughran, D., Smeets, A.: Fibrations with few rational points. Geom. Funct. Anal. **26**(5), 1449–1482 (2016)
15. Loughran, D., Skorobogatov, A., Smeets, A.: Pseudo-split fibres and arithmetic surjectivity, arXiv:1705.10740 (2017)
16. Lang, S., Weil, A.: Number of points of varieties in finite fields. Am. J. Math. **76**, 819–827 (1954)
17. Peyre, E.: Hauteurs et mesures de Tamagawa sur les variétés de Fano. Duke Math. J. **79**(1), 101–218 (1995)
18. Poonen, B., and Stoll, M.: A local-global principle for densities, Topics in number theory (University Park, PA, 1997) Math. Appl., vol. 467. Kluwer Acadamic Publication, Dordrecht, pp. 241–244 (1999)
19. Poonen, B., Voloch, F., Diophantine, Random, equations, Arithmetic of higher-dimensional algebraic varieties (Palo Alto, CA, 2002) Progr. Math., vol. 226. Birkhäuser Boston, Boston, MA, pp. 175–184 (2004)
20. Serre, J.-P.: Spécialisation des éléments de $Br_2(\mathbf{Q}(T_1, \ldots, T_n))$. C. R. Acad. Sci. Paris Sér. I Math. **311**(7), 397–402 (1990)
21. Serre, J.-P.: Lectures on the Mordell–Weil Theorem, Aspects of Mathematics, 3rd edn. Friedr. Vieweg, Braunschweig (1997)
22. Serre, J.-P.: Lectures on $N_X(p)$, Chapman & Hall/CRC Research Notes in Mathematics, vol. 11. CRC Press, Boca Raton, FL (2012)
23. Skorobogatov, A.N.: Descent on fibrations over the projective line. Am. J. Math. **118**(5), 905–923 (1996)
24. Tenenbaum, G.: Introduction to Analytic and Probabilistic Number Theory, Graduate Studies in Mathematics, vol. 163, 3rd edn. American Mathematical Society, Providence, RI (2015)
25. Xiong, M.: The Erdős–Kac theorem for polynomials of several variables. Proc. Am. Math. Soc. **137**(8), 2601–2608 (2009)