Math. Proc. Camb. Phil. Soc. (2022), **173**, 715–726
doi:10.1017/S030500412200010X
First published online 28 February 2022

Sums of random multiplicative functions over function fields with few irreducible factors

BY DAKSH AGGARWAL
Department of Mathematics, Grinnell College,
1115 8*th Ave #* 3011, *Grinnell, IA, USA,* 50112
e-mail: aggarwal2@grinnell.edu

UNIQUE SUBEDI
Department of Statistics, University of Michigan,
1085 *University Ave,* 323 *West Hall, Ann Arbor, MI, USA,* 48109
e-mail: subedi@umich.edu

WILLIAM VERREAULT
Département de Mathématiques et de Statistique,
Université Laval, Québec, QC, G1V 0A6, Canada
e-mail: william.verreault.2@ulaval.ca

ASIF ZAMAN
Department of Mathematics, University of Toronto,
40 *St. George Street, Room* 6290, *Toronto, ON, Canada,* M5S 2E4
e-mail: zaman@math.toronto.edu

CHENGHUI ZHENG
Department of Statistics, University of Toronto,
100 *St.George Street, Toronto, ON, Canada,* M5S 3G3
e-mail: chenghui.zheng@mail.utoronto.ca

(*Received* 18 *August* 2021*; revised* 31 *January* 2022*; accepted* 20 *January* 2022)

Abstract

We establish a normal approximation for the limiting distribution of partial sums of random Rademacher multiplicative functions over function fields, provided the number of irreducible factors of the polynomials is small enough. This parallels work of Harper for random Rademacher multiplicative functions over the integers.

2020 Mathematics Subject Classification: 11K65 (Primary) 60F05, 60G50 (Secondary)

1. *Introduction*

Let $\mathcal{M}$ be the set of monic polynomials belonging to the polynomial ring $\mathbb{F}_q[t]$ with coefficients in the finite field $\mathbb{F}_q$ with q elements, where $q \geq 2$ is a prime power. A *random Rademacher multiplicative function* $f : \mathcal{M} \to \{-1, 0, 1\}$ over $\mathbb{F}_q[t]$ is obtained by picking

independent random variables $f(P)$ uniformly distributed on $\{\pm 1\}$ (that is, taking the value ± 1 with probability $1/2$ each) for monic irreducible polynomials P, extending f multiplicatively to all squarefree monic polynomials, and setting f to be zero for all non-squarefree monic polynomials. For example, if $F = P_1 \cdots P_k$ for distinct irreducible monic polynomials $P_1, \ldots, P_k$, then $f(F) = f(P_1) \cdots f(P_k)$. For any positive integers k and n, set

$$\mathcal{P}_k(n) = \{F \in \mathcal{M} : F \text{ squarefree}, \omega(F) = k, \text{ and } \deg(F) = n\},$$

where $\omega(F)$ is the number of distinct irreducible factors of the polynomial F, and $\deg(F)$ is the degree of F. The purpose of this paper is to establish the following theorem.

THEOREM 1. *Let f be a random Rademacher multiplicative function over $\mathbb{F}_q[t]$, where $q \geq 2$ is a fixed prime power. If $k \geq 1$ satisfies $k = o(\log n)$ as $n \to \infty$, then*

$$\frac{1}{\sqrt{|\mathcal{P}_k(n)|}} \sum_{F \in \mathcal{P}_k(n)} f(F) \tag{1·1}$$

converges in distribution to the standard normal distribution $N(0,1)$ as $n \to \infty$.

This result is motivated by the study of random multiplicative functions over the integers, which were introduced by Wintner [12] to heuristically model the Möbius function. A random Rademacher multiplicative function $f : \mathbb{N} \to \{-1, 0, 1\}$ over the integers is similarly obtained by picking independent random variables $f(p)$ for each prime p, extending it multiplicatively to all squarefree integers, and setting it to be zero for all non-squarefree integers. If $\pi_k(x)$ is the number of squarefree integers $\leq x$ with k distinct prime factors, then the sum

$$\frac{1}{\sqrt{\pi_k(x)}} \sum_{\substack{m \leq x \\ \omega(m) = k}} f(m) \tag{1·2}$$

parallels the quantity in (1·1). Indeed, integers of size x are known to heuristically correspond to polynomials in $\mathbb{F}_q[t]$ of degree $n \approx \log x$. Improving upon a result of Hough [5], Harper [4] established the following theorem which motivates our Theorem 1.

THEOREM 2. (Harper) *Let f be a random Rademacher multiplicative function over the integers. If $k \geq 1$ satisfies $k = o(\log \log x)$ as $x \to \infty$, then (1·2) converges in distribution to the standard normal distribution $N(0,1)$ as $x \to \infty$.*

Notice the range $k = o(\log \log x)$ in Theorem 2 over the integers corresponds precisely to the range $k = o(\log n)$ in Theorem 1 over the polynomial ring $\mathbb{F}_q[t]$. Theorem 1 can therefore be viewed as an extension of Theorem 2 to the function field setting. For an introduction to multiplicative functions over function fields, we refer the reader to work of Granville, Harper and Soundararajan [3], whose conventions we follow here.

The proof strategy for Theorem 1 adapts Harper's key ideas with the verification of three conditions in a martingale central limit theorem (Theorem 3). In Section 2, we prepare this strategy and define our martingale difference sequence. The analysis of this martingale allows us to efficiently reduce the theorem to a natural counting problem (Lemma 4), just as Harper did in [4, section 4·2]. However, this counting problem for function fields introduces cases which did not appear for the integers. The source of these new cases is simple: two

https://doi.org/10.1017/S030500412200010X Published online by Cambridge University Press

distinct irreducible polynomials can have the same degree, but two distinct rational primes cannot have the same size. Since our martingale is filtered based on the degree of the largest irreducible factor (similar to the size of the prime for integers), this distinction creates new terms in our sums that we must carefully treat; see the remark following Lemma 4 for details.

In Section 3, we proceed to analyse these sums and complete the proof of Theorem 1 with some technical estimates. Although these combinatorial sums are somewhat more intricate, the estimation of these sums is simpler due to the familiar analytic benefits of function fields over integers. The key technical lemma for this analysis (Lemma 5) is proved in Section 4. We use recent results on the size of $\mathcal{P}_k(n)$ by Gómez-Colunga et al. [2] and Afshar and Porritt [1], which respectively parallel classical estimates for $\pi_k(x)$ by Hardy and Ramanujan, and Sathe and Selberg.

We conclude the introduction with a few remarks on the sharpness of Theorem 1 and possible extensions. Harper showed that the range $k = o(\log\log x)$ in Theorem 2 is optimal [4, corollary 1]. He further established a normal approximation for sums like (1·2) with the looser restriction $\omega(m) \leq k$ and also for a larger class of random multiplicative functions [4, theorem 3]. It would be of interest to determine whether the range $k = o(\log n)$ in Theorem 1 is optimal and whether similar extensions hold in our setting. It seems plausible that such results carry over by similar arguments, but we did not pursue those investigations. If the range $k = o(\log n)$ is optimal as Harper's work would suggest, then this indicates that the proof of Theorem 1 is quite delicate and sensitive to even minor losses.

Notation

Let $q \geq 2$ be a prime power. Let $\mathbb{F}_q[t]$ be the polynomial ring with coefficients in the finite field $\mathbb{F}_q$ with q elements. Let $\mathcal{M}$ be the set of monic polynomials belonging to $\mathbb{F}_q[t]$. We shall use capital letters to denote a polynomial F in $\mathcal{M}$, writing $\deg(F)$ for the degree of the polynomial F, $\omega(F)$ for the number of distinct irreducible factors of F, and $\mathrm{P}^+(F)$ for the maximum degree of an irreducible dividing F. The letters P and Q will be reserved for monic irreducible polynomials. For integers $k, n \geq 1$, let $\mathcal{P}_k(n)$ be the set of squarefree polynomials F in $\mathcal{M}$ with $\omega(F) = k$ and $\deg(F) = n$. The letter f denotes a random Rademacher multiplicative function over $\mathbb{F}_q[t]$. The relation $u \ll v$ means that there exists an absolute positive constant C such that $|u| \leq Cv$. If the constant C depends on a parameter, say ε, then we shall write $u \ll_\varepsilon v$.

2. *Plan for the proof of Theorem 1*

For integers $k, n \geq 1$ and a random Rademacher multiplicative function f over $\mathbb{F}_q[t]$, define

$$S^{(k)}(n) = \sum_{F \in \mathcal{P}_k(n)} f(F).$$

Notice $\mathbb{E}[f(F)] = 0$ for any non-trivial squarefree F because f is multiplicative and $(f(P))_P$ is a sequence of independent random variables with mean zero. Hence, $S^{(k)}(n)$ has mean zero. Also, since $\mathbb{E}[f(F)f(G)] = 1$ if $F = G$ and 0 otherwise, it follows that

$$\mathbb{E}\big[S^{(k)}(n)^2\big] = \sum_{F,G \in \mathcal{P}_k(n)} \sum \mathbb{E}[f(F)f(G)] = |\mathcal{P}_k(n)|. \tag{2·1}$$

Thus the mean of (1·1) is zero and its variance is indeed one. Our goal is to prove that $S^{(k)}(n)$, normalised by its standard deviation $\sqrt{|\mathcal{P}_k(n)|}$, converges in distribution to the

https://doi.org/10.1017/S030500412200010X Published online by Cambridge University Press

standard normal as $n \to \infty$, provided $k = o(\log n)$. The strategy follows that of Harper [4] with appropriate modifications and simplifications as mentioned earlier in the introduction.

First, notice $\mathcal{P}_1(n)$ is the set of irreducible monic polynomials of degree n, so $S^{(1)}(n)$ is a sum of $|\mathcal{P}_1(n)|$ independent random variables uniform on $\{\pm 1\}$. Thus, the classical central limit theorem implies that $S^{(1)}(n)$ converges in distribution to $N(0,1)$ as $n \to \infty$. We may therefore assume throughout that $k \geq 2$.

2·1 Central limit theorem for martingale difference sequences

To prove convergence in distribution to the standard normal, we want to use a central limit theorem that gives information on the convergence of the partial sums of a martingale difference sequence. The result we use was obtained by McLeish [6], but we state it as it appeared in [4].

THEOREM 3. (McLeish) *For $n \in \mathbb{N}$, suppose that $k_n \in \mathbb{N}$, and that $X_{i,n}$, $1 \leq i \leq k_n$, is a martingale difference sequence on $(\Omega, \mathscr{F}, (\mathscr{F}_{i,n})_i, \mathbb{P})$. Write $S_n := \sum_{i \leq k_n} X_{i,n}$ and suppose that the following conditions hold:*

(i) $\sum_{i \leq k_n} \mathbb{E}[X_{i,n}^2] \to 1$ *as* $n \to \infty$;

(ii) *for each* $\varepsilon > 0$, *we have* $\sum_{i \leq k_n} \mathbb{E}\left[X_{i,n}^2 \mathbf{1}_{|X_{i,n}| > \varepsilon}\right] \to 0$ *as* $n \to \infty$;

(iii) $\limsup_{n \to \infty} \sum_{i \leq k_n} \sum_{j \leq k_n, j \neq i} \mathbb{E}\left[X_{i,n}^2 X_{j,n}^2\right] \leq 1$.

Then, S_n converges in distribution to $N(0,1)$ as $n \to \infty$.

Let us describe the martingale difference sequence in our problem. Let $n \geq k \geq 2$. Write $\mathrm{P}^+(F)$ for the maximum degree of the irreducible factors of F. For $d \geq 1$, define

$$\mathcal{P}_{k,d}(n) := \{F \in \mathcal{P}_k(n) : \mathrm{P}^+(F) = d\},$$

and set

$$S_d^{(k)}(n) := \sum_{F \in \mathcal{P}_{k,d}(n)} f(F).$$

Notice the set $\mathcal{P}_{k,n}(n)$ is empty as $k \geq 2$, so $\mathcal{P}_k(n)$ is the union of $\mathcal{P}_{k,d}(n)$ over $1 \leq d \leq n-1$ and therefore $S^{(k)}(n) = \sum_{d=1}^{n-1} S_d^{(k)}(n)$.

Writing $F \in \mathcal{P}_{k,d}(n)$ as $F = QF'$, where Q is a degree d factor of F (among possibly many), it follows by multiplicativity and independence that $\mathbb{E}[f(F)] = \mathbb{E}[f(Q)]\mathbb{E}[f(F')]$. Since $\mathbb{E}[f(Q) \mid \{f(P) : \deg P < d\}] = \mathbb{E}[f(Q)] = 0$, we get that

$$\mathbb{E}\left[f(F) \mid \{f(P) : \deg P < d\}\right] = 0,$$

and so by the linearity of expectation, it follows that

$$\mathbb{E}\left[S_d^{(k)}(n) \mid \{f(P) : \deg P < d\}\right] = 0.$$

https://doi.org/10.1017/S030500412200010X Published online by Cambridge University Press

Hence, if $\mathscr{F}_d$ denotes the sigma algebra generated by $\{f(P):\ \deg P < d\}$, then $(S_d^{(k)}(n))_{d\le n-1}$ is a martingale difference sequence with respect to $(\mathscr{F}_d)_{d\le n-1}$.

We will therefore apply Theorem 3 to the random variables $S_d^{(k)}(n)/\sqrt{|\mathcal{P}_k(n)|}$, which still form a martingale difference sequence and whose sum over $d \le n-1$ equals $S^{(k)}(n)/\sqrt{|\mathcal{P}_k(n)|}$, the quantity considered in Theorem 1. For convenience, we also use the notation

$$\mathcal{P}_{k,\le d}(n) := \bigcup_{j\le d} \mathcal{P}_{k,j}(n).$$

2·2 Reduction to some counting problems

By a computation similar to (2·1), it follows that $\mathbb{E}[S_d^{(k)}(n)^2] = \left|\mathcal{P}_{k,d}(n)\right|$, so that condition (i) of Theorem 3 holds for all n, not just in the limit. Proving Theorem 1 then boils down to verifying conditions (ii) and (iii) of Theorem 3. The second condition stated in terms of our normalised random variables asks that for all $\varepsilon > 0$,

$$\sum_{d=1}^{n-1} \mathbb{E}\left[\left(S_d^{(k)}(n)/\sqrt{|\mathcal{P}_k(n)|}\right)^2 \mathbb{1}_{\left|S_d^{(k)}(n)\right|/\sqrt{|\mathcal{P}_k(n)|}>\varepsilon}\right] \longrightarrow 0$$

as $n \to \infty$. This quantity is at most

$$\varepsilon^{-2} \sum_{d=1}^{n-1} \mathbb{E}\left[S_d^{(k)}(n)^4 / \left|\mathcal{P}_k(n)\right|^2\right].$$

Thus, it suffices to prove that

$$\sum_{d=1}^{n-1} \mathbb{E}[S_d^{(k)}(n)^4] = o(\left|\mathcal{P}_k(n)\right|^2) \tag{2·2}$$

as $n \to \infty$. The third condition becomes

$$\limsup_{n\to\infty} \sum_{d=1}^{n-1} \sum_{\substack{e=1\\ e\ne d}}^{n-1} \mathbb{E}\left[\frac{S_d^{(k)}(n)^2 S_e^{(k)}(n)^2}{\left|\mathcal{P}_k(n)\right|^2}\right] \le 1.$$

Equivalently, we will show

$$\sum_{d=1}^{n-1} \sum_{\substack{e=1\\ e\ne d}}^{n-1} \mathbb{E}\left[S_d^{(k)}(n)^2 S_e^{(k)}(n)^2\right] \le (1+o(1))\left|\mathcal{P}_k(n)\right|^2 \tag{2·3}$$

as $n \to \infty$. For any $1 \le d, e \le n-1$, it will therefore be convenient to express $\mathbb{E}\left[S_d^{(k)}(n)^2 S_e^{(k)}(n)^2\right]$ in terms of an explicit counting problem.

LEMMA 4. *With the same notation as above,*

$$\mathbb{E}\left[S_d^{(k)}(n)^2 S_e^{(k)}(n)^2\right] \le |\mathcal{P}_{k,d}(n)||\mathcal{P}_{k,e}(n)| + I_{k,d,e}(n) + J_{k,d,e}(n),$$

https://doi.org/10.1017/S030500412200010X Published online by Cambridge University Press

where

$$I_{k,d,e}(n) = \sum_{t=1}^{k-1} \sum_{\ell=1}^{n-1} \sum_{M \in \mathcal{P}_{2t,\le \min\{d,e\}}(2\ell)} \sum_{\substack{A \in \mathcal{P}_{t,\le d}(\ell) \\ A \mid M}} \sum_{\substack{U \in \mathcal{P}_{k-t,\le d}(n-\ell) \\ \mathrm{P}^+(UA)=d}} \sum_{\substack{B \in \mathcal{P}_{t,\le e}(\ell) \\ B \mid M}} \sum_{\substack{V \in \mathcal{P}_{k-t,\le e}(n-\ell) \\ \mathrm{P}^+(VB)=e}} 1, \quad (2{\cdot}4)$$

and $J_{k,d,e}(n) = 0$ *if* $d \neq e$, *otherwise*

$$J_{k,d,d}(n) = 4 \sum_{P,Q \in \mathcal{P}_1(d)} \sum_{M' \in \mathcal{P}_{2k-2}(2n-2d)} \sum_{\substack{A', B' \in \mathcal{P}_{k-1}(n-d) \\ A' \mid M', B' \mid M'}} \sum 1. \quad (2{\cdot}5)$$

Proof. Expanding out the sums and applying linearity of expectation, we get that

$$\mathbb{E}\left[S_d^{(k)}(n)^2 S_e^{(k)}(n)^2\right] = \sum_{W,X \in \mathcal{P}_{k,d}(n)} \sum_{Y,Z \in \mathcal{P}_{k,e}(n)} \mathbb{E}\left[f(W)f(X)f(Y)f(Z)\right]. \quad (2{\cdot}6)$$

Notice the expectation on the right-hand side is nonzero only when $WXYZ$ is a square, in which case it equals 1. This is the counting problem which we proceed to reformulate. We may write WX as the product of a square part U^2 and a square-free part M so $W = UA$ and $X = U(M/A)$ for some A that divides M. Note that U, A and M/A are all relatively prime and the maximum degree of their irreducible factors is $\le d$. A similar reasoning for YZ gives some other square part, say V^2, and forces their squarefree part to be M as well so $Y = VB$ and $Z = V(M/B)$ for some B that divides M. Again, V, B, and M/B are all relatively prime and the maximum degree of their irreducible factors is $\le e$.

With this notation in mind, we proceed to count the corresponding contributions according to cases. First, if $M = 1$ then $A = B = 1$, so this case contributes at most

$$\sum_{U \in \mathcal{P}_{k,d}(n)} \sum_{V \in \mathcal{P}_{k,e}(n)} 1 = |\mathcal{P}_{k,d}(n)||\mathcal{P}_{k,e}(n)|.$$

Next, we count the terms in (2·6) where $M \neq 1$ so $\deg M$ and $\omega(M)$ are both non-zero. As M is non-trivial, we have that $A \in P_{t,\le d}(\ell)$ for some $t \in \{1, \dots, k\}$ and $\ell \in \{1, \dots, n\}$. Comparing the degrees and number of irreducible factors of $W = UA$ and $X = U(M/A)$, we deduce that $\mathrm{P}^+(UA) = d$ and

$$\deg U + \deg A = n, \qquad \deg M = 2 \deg A, \qquad \omega(U) + \omega(A) = k, \qquad \omega(M) = 2\omega(A).$$

As $A \in \mathcal{P}_{t,\le d}(\ell)$, this implies that $U \in \mathcal{P}_{k-t,\le d}(n-\ell)$ and $M \in \mathcal{P}_{2t,\le d}(2\ell)$. A similar analysis holds when comparing $Y = VB$ and $Z = V(M/B)$ but, since the polynomial M is common to both arguments, it follows that A and B necessarily have the same degree and same number of prime factors and so do U and V. Hence, $B \in \mathcal{P}_{t,\le e}(\ell)$, $V \in \mathcal{P}_{k-t,\le e}(n-\ell)$, and $M \in \mathcal{P}_{2t,\le \min\{d,e\}}(2\ell)$. The terms in (2·6) with $M \neq 1, t \in \{1, \dots, k-1\}$, and $\ell \in \{1, \dots, n-1\}$ therefore contribute at most $I_{k,d,e}(n)$.

Continuing with this notation, the last case to consider is when $M \neq 1$ and $t = k$ (or equivalently $\ell = n$) in which case $U = V = 1$. Notice $U = V = 1$ implies that $WXYZ = U^2V^2M^2 = M^2$ has a prime factor of degree $\max\{d, e\}$ yet $\mathrm{P}^+(M) \le \min\{d, e\}$. If $d \neq e$, this leads to a contradiction, so this last case occurs if and only if $d = e$. Thus, M has at least two distinct

https://doi.org/10.1017/S030500412200010X Published online by Cambridge University Press

degree d factors in this case and $\mathrm{P}^+(A) = \mathrm{P}^+(B) = d$ where A and B divide M. Thus, there exists a pair of distinct irreducibles $P, Q \in \mathcal{P}_1(d)$ such that $M = PQM'$, where M' belongs to $\mathcal{P}_{2k-2}(2n-2d)$ and at least one of the following holds:

$$P \mid A \text{ and } Q \mid B, \qquad Q \mid A \text{ and } P \mid B, \qquad P \mid A \text{ and } P \mid B, \qquad Q \mid A \text{ and } Q \mid B.$$

If, say, the first situation holds, then $A = PA'$ and $B = QB'$ for $A', B' \in \mathcal{P}_{k-1}(n-d)$ dividing M'. A similar statement holds for the other cases. Combining all of these observations, we see that the terms in (2·6) with $M \neq 1$ and $t = k$ contribute at most $J_{k,d,e}(n)$, as required.

Remark. This lemma and its proof possess the key differences between the function field setting and the integers. Crucially, the product $WXYZ$ can form a square in a new way and contribute to (2·6). Namely, if $d \leq e$ then W and X do not need to share the same irreducible factor of degree d; these factors of degree d can instead pair with factors from Y and Z. This manifests in (2·4) by allowing M to have these large irreducible factors of degree $d = \min\{d, e\}$ and also by creating the additional terms (2·5) which do not appear in [4, section 4·2]. If the irreducible factors of degree d from W and X (resp. of degree e from Y and Z) are paired in a one-to-one manner, then only U (resp. V) in (2·4) would have these large factors of degree d (resp. degree e) and moreover (2·5) would not exist. This is precisely what happens for Harper in the integer setting. Namely, if integers w and x share the same largest prime factor p, then p^2 always divides wx since the size of the prime corresponds uniquely to the prime itself.

Now, using Lemma 4 with $d = e$, we see that (2·2) holds provided that

$$\sum_{d=1}^{n-1} \left(|\mathcal{P}_{k,d}(n)|^2 + I_{k,d,d}(n) + J_{k,d,d}(n) \right) = o\left(|\mathcal{P}_k(n)|^2 \right).$$

Similarly, (2·3) holds provided that

$$\sum_{d=1}^{n-1} \sum_{e=1}^{n-1} \left(|\mathcal{P}_{k,d}(n)||\mathcal{P}_{k,e}(n)| + I_{k,d,e}(n) \right) \leq (1 + o(1))|\mathcal{P}_k(n)|^2.$$

Since

$$\sum_{d=1}^{n-1} \sum_{e=1}^{n-1} |\mathcal{P}_{k,d}(n)||\mathcal{P}_{k,e}(n)| = |\mathcal{P}_k(n)|^2,$$

both (2·2) and (2·3) will therefore be satisfied provided

$$\sum_{d=1}^{n-1} |\mathcal{P}_{k,d}(n)|^2 + \sum_{d=1}^{n-1} \sum_{e=1}^{n-1} I_{k,d,e}(n) + \sum_{d=1}^{n-1} J_{k,d,d}(n) = o(|\mathcal{P}_k(n)|^2) \tag{2·7}$$

as $n \to \infty$. This establishes Theorem 1 assuming (2·7) holds.

3. *Completing the proof of Theorem 1*

It remains to prove (2·7), which rests on the following key technical lemma whose proof is postponed to Section 4.

https://doi.org/10.1017/S030500412200010X Published online by Cambridge University Press

LEMMA 5. *Fix an integer $r \geq 1$. If k and n are integers such that $r \leq k \leq (\log n)/3$, then*

$$\sum_{\substack{k_1,n_1,\dots,k_r,n_r\geq 1\\ k_1+\cdots+k_r=k\\ n_1+\cdots+n_r=n}} \sum\cdots\sum \sum\sum |\mathcal{P}_{k_1}(n_1)|^2\cdots|\mathcal{P}_{k_r}(n_r)|^2 \ll_r \frac{q^{2n}(\log n+2-\log 2)^{2k-2r}}{n^2(k-r)!^2}. \tag{3·1}$$

In particular, if $r \geq 2$ is fixed and $k = o(\log n)$ as $n \to \infty$, then the above is $o(|\mathcal{P}_k(n)|^2)$.

Assuming Lemma 5, it suffices to show that each of the three sums in (2·7) are $o(|\mathcal{P}_k(n)|^2)$ provided $k = o(\log n)$ as $n \to \infty$. We deal with each estimate in separate subsections.

3·1 Estimate for $\sum_{d=1}^{n-1} \left|\mathcal{P}_{k,d}(n)\right|^2$

For $F \in \mathcal{P}_{k,d}(n)$, one has $F = PF'$ for some $P \in \mathcal{P}_1(d)$ and $F' \in \mathcal{P}_{k-1}(n-d)$. This implies that $|\mathcal{P}_{k,d}(n)| \leq |\mathcal{P}_1(d)||\mathcal{P}_{k-1}(n-d)|$ and so

$$\sum_{d=1}^{n-1} \left|\mathcal{P}_{k,d}(n)\right|^2 \leq \sum_{d=1}^{n-1} |\mathcal{P}_1(d)|^2\,|\mathcal{P}_{k-1}(n-d)|^2\,.$$

This is a subsum of Lemma 5 with $r=2$ so it is $o(|\mathcal{P}_k(n)|^2)$ as $n \to \infty$, as required.

3·2 Estimate for $\sum_{d=1}^{n-1}\sum_{e=1}^{n-1} I_{k,d,e}(n)$

Consider the definition of $I_{k,d,e}(n)$ in (2·4). The condition $\mathrm{P}^+(UA) = d$ implies that at least one of the following holds: $\mathrm{P}^+(U) = d$ or $\mathrm{P}^+(A) = d$. Summing over d and, in some cases, dropping the requirement that the maximum degree of the irreducible factors of our polynomials is $\leq d$ or $\leq e$, this implies that $\sum_{d=1}^{n-1} I_{k,d,e}(n)$ is at most

$$\sum_{t=1}^{k-1}\sum_{\ell=1}^{n-1}\sum_{M\in\mathcal{P}_{2t}(2\ell)}\sum_{d=1}^{n-1}\left(\sum_{\substack{A\in\mathcal{P}_{t,d}(\ell)\\ A|M}}\sum_{U\in\mathcal{P}_{k-t}(n-\ell)} + \sum_{\substack{A\in\mathcal{P}_t(\ell)\\ A|M}}\sum_{U\in\mathcal{P}_{k-t,d}(n-\ell)}\right)\sum_{\substack{B\in\mathcal{P}_{t,\leq e}(\ell)\\ B|M}}\sum_{\substack{V\in\mathcal{P}_{k-t,\leq e}(n-\ell)\\ \mathrm{P}^+(VB)=e}} 1$$
$$= 2\sum_{t=1}^{k-1}\sum_{\ell=1}^{n-1}\sum_{M\in\mathcal{P}_{2t}(2\ell)}\sum_{\substack{A\in\mathcal{P}_t(\ell)\\ A|M}}\sum_{U\in\mathcal{P}_{k-t}(n-\ell)}\sum_{\substack{B\in\mathcal{P}_{t,\leq e}(\ell)\\ B|M}}\sum_{\substack{V\in\mathcal{P}_{k-t,\leq e}(n-\ell)\\ \mathrm{P}^+(VB)=e}} 1.$$

Applying the same argument to the condition $\mathrm{P}^+(VB) = e$ and summing over e, it follows that $\sum_{d=1}^{n-1}\sum_{e=1}^{n-1} I_{k,d,e}(n)$ is at most

$$4\sum_{t=1}^{k-1}\sum_{\ell=1}^{n-1}\left(\sum_{M\in\mathcal{P}_{2t}(2\ell)}\sum_{\substack{A\in\mathcal{P}_t(\ell)\\ A|M}}\sum_{\substack{B\in\mathcal{P}_t(\ell)\\ B|M}} 1\right)\left(\sum_{U\in\mathcal{P}_{k-t}(n-\ell)}\sum_{V\in\mathcal{P}_{k-t}(n-\ell)} 1\right). \tag{3·2}$$

Fix $t \in \{1,\dots,k-1\}$ and $\ell \in \{1,\dots,n-1\}$. Notice that the double sum with U and V is equal to $|\mathcal{P}_{k-t}(n-\ell)|^2$. Next, consider the triple sum with M, A, and B. Writing $G = \gcd(A,B)$, we have that $A = GA'$, $B = GB'$, and $M = GA'B'M'$ for some M' coprime to A', B', and G. Since A and B have the same degree and same number of prime factors (and hence so do A' and B'), it follows that G and M' must have the same degree and same number of

https://doi.org/10.1017/S030500412200010X Published online by Cambridge University Press

prime factors. Namely, if $G \in \mathcal{P}_j(g)$ for some integer $0 \le j \le t$ and some integer $0 \le g \le \ell$, then $M' \in \mathcal{P}_j(g)$ and $A', B' \in \mathcal{P}_{t-j}(\ell - g)$. Note the case $j = 0$ (and hence $g = 0$) occurs when $G = M' = 1$ so $M = AB$, and the case $j = t$ (and hence $g = \ell$) occurs when $A' = B' = 1$ so $M = GM'$. Combining these observations implies that

$$\begin{aligned}\sum_{M\in\mathcal{P}_{2t}(2\ell)}\sum_{\substack{A\in\mathcal{P}_t(\ell)\\ A|M}}\sum_{\substack{B\in\mathcal{P}_t(\ell)\\ B|M}} 1 &\le 2|\mathcal{P}_t(\ell)|^2 + \sum_{j=1}^{t-1}\sum_{g=1}^{\ell-1}\sum\sum_{G,M'\in\mathcal{P}_j(g)}\sum\sum_{A',B'\in\mathcal{P}_{t-j}(\ell-g)} 1\\ &= 2|\mathcal{P}_t(\ell)|^2 + \sum_{j=1}^{t-1}\sum_{g=1}^{\ell-1}|\mathcal{P}_j(g)|^2|\mathcal{P}_{t-j}(\ell-g)|^2.\end{aligned} \tag{3·3}$$

Inserting these estimates in (3·2), we conclude that $\sum_{d=1}^{n-1}\sum_{e=1}^{n-1} I_{k,d,e}(n)$ is at most

$$8\sum_{t=1}^{k-1}\sum_{\ell=1}^{n-1}|\mathcal{P}_t(\ell)|^2|\mathcal{P}_{k-t}(n-\ell)|^2 + 4\sum_{t=1}^{k-1}\sum_{\ell=1}^{n-1}\sum_{j=1}^{t-1}\sum_{g=1}^{\ell-1}|\mathcal{P}_{k-t}(n-\ell)|^2|\mathcal{P}_j(g)|^2|\mathcal{P}_{t-j}(\ell-g)|^2.$$

Since $k = o(\log n)$ as $n \to \infty$, both of these sums are $o(|\mathcal{P}_k(n)|^2)$ by Lemma 5, as required.

3·3 Estimate for $\sum_{d=1}^{n-1} J_{k,d,d}(n)$

From (2·5), we have that

$$\sum_{d=1}^{n-1} J_{k,d,d}(n) = 4\sum_{d=1}^{n-1}|\mathcal{P}_1(d)|^2 \sum_{M'\in\mathcal{P}_{2k-2}(2n-2d)}\sum\sum_{\substack{A',B'\in\mathcal{P}_{k-1}(n-d)\\ A'|M',\,B'|M'}} 1.$$

Notice the inner triple sum is the same as (3·3) with $\ell = n - d$ and $t = k - 1$. Thus, $\sum_{d=1}^{n-1} J_{k,d,d}(n)$ is at most

$$8\sum_{d=1}^{n-1}|\mathcal{P}_1(d)|^2|\mathcal{P}_{k-1}(n-d)|^2 + 4\sum_{d=1}^{n-1}\sum_{j=1}^{k-2}\sum_{g=1}^{n-d-1}|\mathcal{P}_1(d)|^2|\mathcal{P}_j(g)|^2|\mathcal{P}_{k-j-1}(n-d-g)|^2.$$

Since $k = o(\log n)$ as $n \to \infty$, all of these sums are $o(|\mathcal{P}_k(n)|^2)$ by Lemma 5. This completes the proof of (2·7) and the proof of Theorem 1.

4. *Proof of Lemma 5*

All that remains is to prove Lemma 5. To do so, we shall first require an estimate for the size of $\mathcal{P}_k(n)$ that is uniform for all integers k and n. Gómez-Colunga et al. [2] have recently established such a result.

PROPOSITION 6. (Gómez-Colunga–Kavaler–McNew–Zhu) *Uniformly for all* $k, n \ge 1$,

$$|\mathcal{P}_k(n)| \le \frac{q^n}{n}\frac{(\log n + 2 - \log 2)^{k-1}}{(k-1)!}.$$

https://doi.org/10.1017/S030500412200010X Published online by Cambridge University Press

This corresponds to a classical result of Hardy and Ramanujan [7] for the integers: there exists a constant $B > 0$ such that, for all $k \geq 1$ and $x \geq 2$, we have

$$\pi_k(x) \ll \frac{x}{\log x} \frac{(\log\log x + B)^{k-1}}{(k-1)!},$$

where $\pi_k(x)$ is the number of squarefree integers up to x with k prime factors.

Sathe [8, 9] and Selberg [10] famously derived an asymptotic estimate for $\pi_k(x)$ when $k = o(\log\log x)$. We shall also need an asymptotic estimate for $|\mathcal{P}_k(n)|$ that is valid when $k = o(\log n)$. Although the estimate $|\mathcal{P}_k(n)| \sim q^n(\log n)^{k-1}/n(k-1)!$ (see, e.g., [11]) suffices for our purposes, we state here the strongest and most recent result on an estimate for $|\mathcal{P}_k(n)|$, which is a so-called Sathe–Selberg formula for function fields established by Afshar and Porritt [1].

PROPOSITION 7. (Afshar–Porritt). *Let $A > 1$. Uniformly for all $n \geq 2$ and $1 \leq k \leq A \log n$,*

$$|\mathcal{P}_k(n)| = \frac{q^n(\log n)^{k-1}}{n(k-1)!}\left(G\left(\frac{k-1}{\log n}\right) + O_A\left(\frac{k}{(\log n)^2}\right)\right),$$

where

$$G(z) = \frac{1}{\Gamma(1+z)} \prod_{\substack{P \in \mathcal{M} \\ P \text{ irreducible}}} \left(1 + \frac{z}{q^{\deg P}}\right)\left(1 - \frac{1}{q^{\deg P}}\right)^z,$$

and $\Gamma(\cdot)$ is the Gamma function defined as $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx$.

Propositions 6 and 7 imply our key technical lemma.

Proof of Lemma 5. The second estimate follows from (3·1) since

$$\frac{q^{2n}(\log n + 2 - \log 2)^{2k-2r}}{n^2(k-r)!^2} \leq \frac{q^{2n}(\log n)^{2k-2}}{n^2(k-1)!^2} \cdot \frac{k^{2r-2}}{(\log n)^{2r-2}}\left(1 + \frac{2-\log 2}{\log n}\right)^{2k-2r},$$

and Proposition 7 implies that if $k = o(\log n)$ as $n \to \infty$, then $|\mathcal{P}_k(n)| \sim q^n(\log n)^{k-1}/n(k-1)!$.

To prove (3·1), we proceed by induction on r. For $r = 1$, the claim follows immediately from Proposition 6. For $r \geq 2$, if $n_1 + \cdots + n_r = n$, then at least one of $n_1, \ldots, n_r$ is at most $\lfloor n/r \rfloor$. By symmetry, we may assume it is n_r so the left-hand side of (3·1) is at most

$$\ll_r \sum_{k_r=1}^{k-r+1} \sum_{n_r=1}^{\lfloor n/r \rfloor} |\mathcal{P}_{k_r}(n_r)|^2 \left(\sum\sum \cdots \sum\sum_{\substack{k_1, n_1, \ldots, k_{r-1}, n_{r-1} \geq 1 \\ k_1 + \cdots + k_{r-1} = k - k_r \\ n_1 + \cdots + n_{r-1} = n - n_r}} |\mathcal{P}_{k_1}(n_1)|^2 \cdots |\mathcal{P}_{k_{r-1}}(n_{r-1})|^2 \right).$$

Notice that $n - n_r \geq n/2$ as $r \geq 2$. Since $k \leq (\log n)/3$ by assumption, this implies that $k - k_r \leq k - 1 \leq \log(n/2)/3 \leq \log(n - n_r)/3$. Thus, by the inductive hypothesis, the above is

$$\ll_r \sum_{k_r=1}^{k-r+1} \sum_{n_r=1}^{\lfloor n/r \rfloor} |\mathcal{P}_{k_r}(n_r)|^2 \frac{q^{2(n-n_r)}(\log(n-n_r) + c)^{2k-2k_r-2r+2}}{(n-n_r)^2(k-k_r-r+1)!^2},$$

https://doi.org/10.1017/S030500412200010X Published online by Cambridge University Press

where for brevity we have set $c = 2 - \log 2$. Applying Proposition 6, we see that this is at most

$$\sum_{k_r=1}^{k-r+1} \frac{q^{2n}}{(k_r-1)!^2(k-k_r-r+1)!^2} \sum_{n_r=1}^{\lfloor n/r \rfloor} \frac{(\log n_r + c)^{2k_r-2}(\log (n-n_r)+c)^{2k-2k_r-2r+2}}{n_r^2(n-n_r)^2}$$
$$\ll_r \frac{q^{2n}}{n^2} \sum_{k_r=1}^{k-r+1} \frac{(\log n + c)^{2k-2k_r-2r+2}}{(k_r-1)!^2(k-k_r-r+1)!^2} \sum_{n_r=1}^{\lfloor n/r \rfloor} \frac{(\log n_r + c)^{2k_r-2}}{n_r^2}. \quad (4{\cdot}1)$$

Note that for any integer $m \geq 0$,

$$\sum_{j=1}^{\infty} \frac{(\log j + c)^m}{j^2} \ll \int_1^{\infty} \frac{(\log t + c)^m}{t^2}\, dt = \int_c^{\infty} t^m e^{c-t}\, dt \ll \int_0^{\infty} t^m e^{-t}\, dt = m!.$$

Using this estimate on the inner sum over n_r, it follows that (4·1) is

$$\ll_r \frac{q^{2n}}{n^2} \sum_{k_r=1}^{k-r+1} \frac{(2k_r-2)!(\log n + c)^{2k-2k_r-2r+2}}{(k_r-1)!^2(k-k_r-r+1)!^2}.$$

For the final sum over k_r, notice that the ratio of consecutive summands is equal to

$$\frac{2k_r(2k_r-1)}{k_r^2} \frac{(k-k_r-r+1)^2}{(\log n + c)^2} \leq \frac{4k^2}{(\log n)^2} \leq \frac{4}{9},$$

since $k \leq (\log n)/3$ by assumption. Hence, the final sum over k_r is dominated by its value at the endpoint $k_r = 1$, yielding the desired estimate. This establishes Lemma 5.

Acknowledgments. This research was conducted as part of the 2020 Fields Undergraduate Summer Research Program. The authors are grateful to the Fields Institute for their financial support and facilitating their online collaboration.

REFERENCES

[1] A. AFSHAR and S. PORRITT. The function field Sathe–Selberg formula in arithmetic progressions and 'short intervals'. *Acta Arih.* **187** (2019), 101–124.
[2] A. GÓMEZ-COLUNGA, C. KAVALER, N. MCNEW and M. ZHU. On the size of primitive sets in function fields. *Finite Fields Appl.* **64** (2020), 101658.
[3] A. GRANVILLE, A. J.HARPER and K.SOUNDARARAJAN. Mean values of multiplicative functions over function fields. *Research in Number Theory*, **1** (2015), 25.
[4] A. J.HARPER. On the limit distributions of some sums of a random multiplicative function. *J. Reine Angew. Math.* **678** (2013).
[5] B.HOUGH. Summation of a random multiplicative function on numbers having few prime factors. *Math. Proc. Camb. Phil. Soc.*, **150**(2) (2011), 193–214.
[6] D. L.MCLEISH. Dependent central limit theorems and invariance principles. *Ann. Probab.*, **2**(4) (1974), 620–628.
[7] S.RAMANUJAN and G.HARDY. The normal number of prime factors of a number *n*. *Quarterly J. Math.* **48** (1917), 76–92.
[8] L. G.SATHE. On a problem of Hardy on the distribution of integers having a given number of prime factors, I, II. *J. Indian Math. Soc.(NS)* **17** (1953), 63–141.
[9] L. G.SATHE. On a problem of Hardy on the distribution of integers having a given number of prime factors, III, IV. *J. Indian Math. Soc.(NS)* **18** (1954), 27–81.

https://doi.org/10.1017/S030500412200010X Published online by Cambridge University Press

[10] A. SELBERG. Note on a paper by L. G. Sathe. *J. Indian Math. Soc.*, **1** (1954), 83–87.
[11] R. WARLIMONT. Arithmetical semigroups. IV. Selberg's analysis. *Arch. Math. (Basel)* **60**(1) (1993), 58–72.
[12] A. WINTNER. Random factorisations and Riemann's hypothesis. *Duke Math. J.*, **11**(2) (1944), 267–275.