Friable Integers: An Overview

Cécile Dartyge (Université de Lorraine, Vandœuvre-lès-Nancy, France)

1 Introduction

A friable (or smooth) integer is an integer without large prime factors. More precisely, let $P^+(n)$, $P^-(n)$ denote respectively the largest and the smallest prime factor of an integer n.[1] For $y \geqslant 2$, an integer n is said to be y-friable if $P^+(n) \leqslant y$ and y-sifted if $P^-(n) > y$. The term 'friable' reflects the possibility of splitting these integers into small divisors – the small prime factors. When the friable parameter y is very small, these integers can be represented as products of divisors with good size control. This flexibility in the factoring of friable integers is at the origin of several recent breakthroughs in analytic number theory. Another fundamental observation is that every integer n has a unique decomposition of the form $n = ab$ where a is y-friable and b is y-sifted. The structure of the sifted part b resembles that of a prime number while the friable part a is supposed to behave more like a standard integer. This simple idea is often an efficient starting point in the study of sums appearing in analytic number theory problems, as described, for example, in Section 5.1.

Research on friable integers really started less than a hundred years ago. Since the 1980s, the theory has been actively developing not only for its own interest, but also for the multiple applications available.

Let $S(x, y)$ denote the set of the y-friable integers not exceeding x and let $\Psi(x, y)$ be its cardinality. In the first part of this survey, we provide an account of the various methods developed for estimating $\Psi(x, y)$ according to the relative sizes of x and y and we briefly describe some results related to multiplicative properties of friable integers. Next we depict the role of friable integers in certain factoring algorithms. In the last part, we present various number theoretic-problems for which friable integers led to significant advances.

This presentation is far from exhaustive. Our aim is mainly to provide an overview of the role of friable integers in different areas from number theory. One will find more complete presentations in the impressive survey papers of Hildebrand and Tenenbaum [13], Granville [11], Pomerance [17]. We furthermore recommend the book by Crandall and Pomerance [5] and those by Tenenbaum [18, 19] which are unavoidable references respectively in algorithmic and analytic number theory.

Throughout this text the letter p with or without subscript denotes a prime number and (a, b) is the g.c.d. of the integers a and b. To simplify notation, we write $\ln_k(t)$ the kth iterated Napierian logarithm of the positive real number t, so that, in particular, for $k = 2$, $\ln_2 t = \ln\ln t$.

1 With the conventions $P^+(1) = 1$ and $P^-(1) = \infty$.

2 How to count friable integers?

2.1 The Dickman function

According to the sizes of y and x, several methods allow approximating $\Psi(x, y)$. Most of the estimates produced depend on the ratio

$$u = \frac{\ln x}{\ln y}\cdot$$

Dickman proved in 1930 that for all fixed $u > 0$, a positive proportion of integers less than x are $x^{1/u}$-friable:

$$\lim_{x\to+\infty} \frac{\Psi(x, x^{1/u})}{x} = \varrho(u)$$

where ϱ, the Dickman function, is the only continuous function on $\mathbb{R}^+$, differentiable on $]1, +\infty[$, and satisfying the delay differential equation

$$u\varrho'(u) = \varrho(u-1)$$

with initial condition $\varrho(u) = 1$ for $0 \leqslant u \leqslant 1$. This function $\varrho(u)$ corresponds to the probability of an integer less than x being $x^{1/u}$-friable. The Dickman function also appears in another context with no connection to friable integers: if $\{U_n\}_{n=1}^{\infty}$ is a sequence of independent random variables uniformly distributed in $[0, 1]$, then the series $Y = U_1 + U_1U_2 + U_1U_2U_3 + \cdots$ converges almost surely to a random variable distributed according to an absolutely continuous law with density $\mathrm{e}^{-\gamma}\varrho$, where γ is the Euler constant.[2]

Such emergence of delay differential equations is not specific to the study of friable integers. It is genuinely linked to sieve methods. Let us, traditionally, note $\Phi(x, y)$ the number of y-sifted integers not exceeding x. Then a similar process leads to proving that the probability for an integer to be y-sifted is $\omega(u)$ (see for example [19], Chapter III.6), where $\omega(u)$ is the Buchstab function. This function is defined by $\omega(u) = 1/u$ for $1 \leqslant u \leqslant 2$ and $(u\omega(u))' = \omega(u-1)$, with continuity at $u = 2$. Similar functions show up, for example, in the Rosser–Iwaniec sieve or the Jurkat–Richert sieve.

The reader will find in [19, Chapter III.5] a very precise study of the Dickman function. It is rapidly decreasing as u tends to $+\infty$, as shown by the Hildebrand–Tenenbaum estimate ([13, Corollary 2.3]), valid for $u \geqslant 1$:

$$\varrho(u) = \exp\left\{-u\left(\ln u + \ln_2(u+2) - 1 + O\Big(\frac{\ln_2(u+2)}{\ln(u+2)}\Big)\right)\right\}.$$

The function ϱ has been implemented in several pieces of mathematical software, for example Sage. The following values are given in [11]:

$$\varrho(2) \approx 3.07\times10^{-2}, \varrho(5) \approx 3.55\times10^{-4},$$
$$\varrho(10) \approx 2.77\times10^{-11}, \varrho(20) \approx 2.46\times10^{-29},$$
$$\varrho(50) \approx 6.72\times10^{-97}, \text{ etc.}$$

2 So $\gamma = \lim_{N\to\infty} \sum_{n=1}^{N} 1/n - \ln N$.

2.2 $\Psi(x,y)$ for large y, through functional equations

This initial result by Dickman has been improved by de Bruijn. The starting idea is as follows: for $z \geqslant y$, an integer n counted in $\Psi(x,z)$ is either y-friable or of type mp with $P^+(m) \leqslant p$, and the prime p lies in $]y,z]$. We deduce the *Buchstab identity*:

$$\Psi(x,y) = \Psi(x,z) - \sum_{y<p\leqslant z} \Psi(\frac{x}{p},p) \qquad (1 < y \leqslant z \leqslant x). \quad (1)$$

This leads to an iteration with (obvious) initial condition

$$\Psi(x,y) = \lfloor x \rfloor \quad (y \geqslant x).$$

Since $x/p \leqslant p$ when $\sqrt{x} \leqslant p \leqslant x$, we immediately get $\Psi(x/p,p) = \lfloor x/p \rfloor$ under this assumption. Applying this to (1) yields an asymptotic formula for $y > x^{1/2}$. Inserting this new formula back into (1), we get an estimate for $\Psi(x,y)$ when $y > x^{1/3}$ and so on. This functional approach leads to the uniform estimate for $x \geqslant y \geqslant 2$ (see for example Theorem III.5.8 of [18]):

$$\Psi(x,y) = x\varrho(u) + O\Big(\frac{x}{\ln y}\Big).$$

This formula loses accuracy for "large" values of u. For example, when $u \geqslant \ln_2 y$, the main term $x\varrho(u)$ is dominated by the error term $O(x/\ln y)$. Given the present state of zero-free regions of the Riemann zeta function, the limit of de Bruijn's method is actually the range

$$y > \exp\left((\ln x)^{5/8+\varepsilon}\right), \quad (2)$$

for any fixed $\varepsilon > 0$. The range of de Bruijn's approximation to $\Psi(x,y)$ has been improved by Hildebrand through another functional equation:

$$\Psi(x,y)\ln x = \int_1^x \Psi(t,y)\frac{\mathrm{d}\,t}{t} + \sum_{\substack{p^k\leqslant x\\ p\leqslant y}} \Psi\Big(\frac{x}{p^k},y\Big)\ln p. \quad (3)$$

This is derived by evaluating in two different ways the sum

$$S = \sum_{n\in S(x,y)} \ln n.$$

First, Abel summation provides

$$S = \Psi(x,y)\ln x - \int_1^x \Psi(t,y)\frac{\mathrm{d}\,t}{t},$$

then the additivity of the logarithm $\ln n = \sum_{p^k|n} \ln p$ furnishes the second term of the right-hand side of (3). An advantage of this formula is keeping the friable parameter y constant, so that $\Psi(x,y)$ appears as a mean value of itself in only one variable. This makes the regularization arising from the iterations more efficient. Hildebrand proved that, for all fixed $\varepsilon > 0$, the formula

$$\Psi(x,y) = x\varrho(u)\left\{1 + O\Big(\frac{\ln(u+1)}{\ln y}\Big)\right\} \quad (4)$$

holds uniformly in a range larger than (2)

$$\exp((\ln_2 x)^{5/3+\varepsilon}) \leqslant y \leqslant x. \quad (5)$$

This region (5) is closely related to the error term of the prime number theorem. Any progress on this error term implies a corresponding improvement on (5). Actually, Hildebrand proved that the Riemann hypothesis[3] is satisfied if and only if (4) holds in the region $y \geqslant (\ln x)^{2+\varepsilon}$. In the same range (5), Saias obtained an estimate for $\Psi(x,y)$ which is more precise than (4), the main term $x\varrho(u)$ being replaced by a more involved expression $\Lambda(x,y)$, already present in de Bruijn's article. Nevertheless, it is possible to obtain asymptotic formulas for $\ln(\Psi(x,y)/x)$ in wider domains. Very precise formulas may be found for example in [4], [13], or [19] . A consequence of these estimates is that for all fixed $0 < \varepsilon < 1$, and uniformly for $u \leqslant y^{1-\varepsilon}$, we have

$$\Psi(x,y) = xu^{-(1+o(1))u}, \quad (6)$$

as y and u tend to $+\infty$, . This formula gives an idea of the order of magnitude of $\Psi(x,y)$. We deduce for example that for fixed $\alpha > 1$, $\Psi(x,(\ln x)^\alpha) = x^{1-1/\alpha+o(1)}$.

2.3 Geometric method for small values of y

When y gets smaller than a power of $\ln x$, one must proceed in a different way to obtain an asymptotic formula. We observe that $\Psi(x,y)$ is the number of solutions $(m_p)_{p\leqslant y}$ in non-negative integers of the inequality $\prod_{p\leqslant y} p^{m_p} \leqslant x$. Taking logarithms, this condition becomes $\sum_{p\leqslant y} m_p \ln p \leqslant \ln x$. We are thus counting integer points inside a polytope of $\mathbb{R}^{\pi(y)}$, where $\pi(y)$ is the number of prime numbers not exceeding y. This approach is efficient for very small values of y. Ennola proved in this way that for $2 \leqslant y \leqslant \sqrt{\ln x}$,

$$\Psi(x,y) = \frac{1}{\pi(y)!}\prod_{p\leqslant y}\frac{\ln x}{\ln p}\left\{1 + O\Big(\frac{y^2}{(\ln x)(\ln y)}\Big)\right\}. \quad (7)$$

La Bretèche and Tenenbaum [4] improved on this result (range for y, quality of the error term) by employing a mixed approach resting on the residue theorem and the saddle-point method, as described in the next paragraph.

2.4 The saddle-point method

Formula (4) provides an approximation to $\Psi(x,y)$ by a regular function but this is not the case for (7) which depends on $\pi(y)$!. What happens in the domain not covered by these two estimations, that is, $\sqrt{\ln x} \leqslant y \leqslant \exp((\ln\ln x)^{5/3+\varepsilon})$?

This question has been solved by Hildebrand and Tenenbaum. They gave an estimate for $\Psi(x,y)$ by using a third approach: the saddle-point method. The indicator function of y-friable integers is a multiplicative function. Let $\zeta(s,y)$ be the associated Dirichlet series:

$$\zeta(s,y) := \sum_{P^+(n)\leqslant y}\frac{1}{n^s} = \prod_{p\leqslant y}\Big(1-\frac{1}{p^s}\Big)^{-1}.$$

By Perron's formula, ([19, Chapter II.2]), we can represent $\Psi(x,y)$ in the following way for all $\alpha > 0$ and $x \notin \mathbb{N}$:

$$\Psi(x,y) = \frac{1}{2i\pi}\int_{\alpha-i\infty}^{\alpha+i\infty}\zeta(s,y)x^s\frac{\mathrm{d}\,s}{s}.$$

The optimal choice for α corresponds to the saddle point of the integrand, that is, the unique solution of the equation

$$-\frac{\zeta'(\alpha,y)}{\zeta(\alpha,y)} = \ln x, \text{ with } -\frac{\zeta'(\alpha,y)}{\zeta(\alpha,y)} = \sum_{p\leqslant y}\frac{\ln p}{p^\alpha - 1}. \quad (8)$$

Hildebrand and Tenenbaum obtained the following formula, valid uniformly for $x \geqslant y \geqslant 2$:

$$\Psi(x,y) = \frac{x^\alpha\zeta(\alpha,y)}{\alpha\sqrt{2\pi\varphi_2(\alpha,y)}}\left\{1 + O\Big(\frac{1}{u} + \frac{\ln y}{y}\Big)\right\},$$

3 The Riemann ζ function is defined by $\zeta(s) = \sum_{n\geqslant 1} 1/n^s$ for $s \in \mathbb{C}$ with real part strictly larger than 1. It has an analytic continuation, also denoted ζ, on $\mathbb{C} \setminus \{1\}$ into a function with a simple pole at $s = 1$. The Riemann hypothesis asserts that the real parts of all non-trivial zeros of ζ are equal to 1/2.

where $\varphi_2(s,y)$ is the second derivative in s of $\ln\zeta(s,y)$.

A priori this formula provides only little directly exploitable information since the parameter α defined by (8) looks rather mysterious. However, by the prime number theorem Hildebrand and Tenenbaum obtained precise asymptotic estimates for $\alpha = \alpha(x,y)$. This enabled them to recover, and actually extend, the results of Hildebrand mentioned in the previous section. Moreover, they determined very precise estimates for the local behaviour for $\Psi(x,y)$. For example, provided y tends to $+\infty$, they proved that $\alpha(x,y) = o(1)$ if and only if $y \leqslant (\ln x)^{1+o(1)}$ where now the "$o(1)$" refers to x. A consequence is that the asymptotics $\Psi(2x,y) \sim \Psi(x,y)$ holds if, and only if, $y \leqslant (\ln x)^{1+o(1)}$.

Very recently, La Bretèche and Tenenbaum [4] obtained new estimates for $\Psi(x,y)$ in the critical range $1 \leqslant y \leqslant (\ln x)^{1+o(1)}$ elucidating completely the behaviour of this function in this region. Their results highlight some important discontinuities of $\Psi(x,y)$ when y is a prime number of size $o((\ln x)^{2/3}(\ln_2 x)^{1/3})$. The extremely friable integers are not smooth.

3 Some properties of friable integers

Before setting out the applications in cryptography and in other problems of analytic number theory, we address the following question: In what respect do friable integers behave like ordinary integers?

We will employ several criteria frequently used in analytic number theory: distribution in short intervals, in arithmetic progressions, and restricted mean values of multiplicative functions. The last paragraph in this section is devoted to the Turán–Kubilius inequality, an important tool in probabilistic number theory.

3.1 Distribution in short intervals

We expect that friable integers are evenly distributed in short intervals, that is,

$$\frac{\Psi(x+z,y)-\Psi(x,y)}{z} \sim \frac{\Psi(x,y)}{x} \tag{9}$$

in a large range for x, y, z. Hildebrand obtained an estimate of type (9) in the domain (5) and provided the interval length z satisfies $xy^{-5/12} \leqslant z \leqslant x$. Hildebrand and Tenenbaum established asymptotic estimates in larger domains in y but for longer intervals $]x, x+z[$.

For shorter intervals, Friedlander and Lagarias proved that there exists a constant $c > 0$ such that for all fixed $\alpha > 0$ and $\beta > 1-\alpha-c\alpha(1-\alpha)$, the interval $[x, x+x^\beta]$ contains a positive proportion of x^α-friable integers. Other results on the distribution of friable integers in short intervals appear in [13] and [11]. Very recently, Matomäki and Radziwiłł [15] achieved spectacular progress: they showed that for all $\varepsilon > 0$, there exists $C(\varepsilon) > 0$ such that, for large enough x, the interval $[x, x + C(\varepsilon)\sqrt{x}]$ contains at least $\sqrt{x}/(\ln x)^4$ x^ε-friable integers. We will see in Section 5 that exhibiting friable integers in such small intervals is an important step in several factoring algorithms. For example, this is the case for the quadratic sieve.

3.2 Friable integers in arithmetic progressions

We denote by $\Psi(x,y;a,q)$ the number of y-friable integers not exceeding x and congruent to a modulo q and we let $\Psi_q(x,y)$ stand for the cardinality of y-friable integers $\leqslant x$ and coprime to q. When $(a,q) \neq 1$ and q is y-friable, this cardinality is equal to $\Psi(x/d,y;a/d,q/d)$ where we have put $d = (a,q)$. It is thus sufficient to restrict to the case $(a,q) = 1$. Assuming good distribution in the invertible classes modulo q, we expect that, for $(a,q) = 1$,

$$\Psi(x,y;a,q) \sim \frac{\Psi_q(x,y)}{\varphi(q)}, \tag{10}$$

where φ is the Euler totient function: $\varphi(n) = \#(\mathbb{Z}/n\mathbb{Z})^*$. Here again there are two goals: to find estimates in domains as large as possible in terms of both q and y. In fact, determining the main term $\Psi_q(x,y)$ is already a difficult problem.

In this direction, one can find very nice results in the literature such as the works of Fouvry and Tenenbaum [9], Granville [10], and recently the articles by Harper [12] and Drappeau [8]. Harper proved that (10) holds provided the ratio $\ln x/\ln q$ tends to ∞ and $q \leqslant y^{4\sqrt{\mathrm{e}}-\varepsilon}$, $y \geqslant y_0(\varepsilon)$.

However the condition $\ln x/\ln q \to \infty$ is very restricting: it doesn't cover the case $q \approx x^\alpha$ even for very small $\alpha > 0$. This obstacle can be circumvented by considering the distribution on average over q. Indeed, Harper proved that, for suitable $c > 0$, relation (10) is satisfied for all $(\ln x)^c \leqslant y \leqslant x$ and almost all $q \leqslant \sqrt{\Psi(x,y)}(\ln x)^{-7/2}$. Furthermore, Drappeau obtained a similar result on average for $q < x^{3/5-\varepsilon}$ but with a weaker control of the uniformity in the classes a modulo q and for a friability range of type $(\ln x)^c \leqslant y \leqslant x^{c'}$ where $c' > 0$ is a very small constant.

3.3 Multiplicative functions and friable integers

In analytic number theory, one often faces the problem of estimating sums of the type

$$\Psi_f(x,y) := \sum_{n\in S(x,y)} f(n)$$

where f is a multiplicative function such that $f(mn) = f(m)f(n)$ whenever $(m,n) = 1$. Some typical examples are the τ function which furnishes the number of divisors, Dirichlet characters, the Möbius function which will be defined in Section 5.1, $r(n)$ the number of representations of n as a sum of two squares and $\varrho_P(n)$ the number of roots of P modulo n for a given polynomial $P \in \mathbb{Z}[X]$. When f is an oscillating function such that $\sum_{n\leqslant x} f(n) = o(x)$, as for example the Möbius function or Dirichlet characters, one aims to obtain the largest possible region in y for which $\Psi_f(x,y) = o(\Psi(x,y))$. When the function f is non-negative, one hopes for asymptotic formulas. In applications, the values $f(p)$ are often close on average to a real number κ: $\kappa = 2$ for $f = \tau$, $\kappa = 1$ for $f = \varrho_P$ if P is irreducible, etc.

Under such hypotheses Tenenbaum and Wu [20] proved formulas of the type

$$\Psi_f(x,y) = C_\kappa(f)x\varrho_\kappa(u)\ln(y)^{\kappa-1}(1+E(x,y)),$$

where ϱ_κ is the fractional convolution power of order κ of the Dickman function,[4] $C_\kappa(f)$ is a convergent Eulerian product

4 The function ϱ_κ is continuous on $]0,\infty[$, differentiable on $[1,\infty[$, and satisfies the delay equation $u\varrho_k'(u) + (1-\kappa)\varrho_\kappa(u) + \kappa\varrho_k(u-1) = 0$ for $u > 1$, with initial condition $\varrho_\kappa(u) = u^{\kappa-1}/\Gamma(\kappa)$ for $0 < u \leqslant 1$.

depending on f and κ, and $E(x,y)$ is an error term, which for brevity we do not define here. Under very general conditions, we have $E(x,y) = o(1)$ in ranges analogous to the Hildebrand region (5) for $\Psi(x,y)$.

3.4 The Turán–Kubilius inequality

In analytic number theory, one frequently needs to evaluate the *normal* behaviour of a given arithmetic function, i.e., the behaviour for almost all integers n, or again on a set of integers having natural density 1.[5] In particular, one would like to know whether there exists a function g with any prescribed regularity such that $|f(n) - g(n)|$ is very small for almost all integers n. In this case we say that g is a normal order of f. A famous example is the Hardy and Ramanujan theorem stating that $g(n) = \ln_2 n$ is a normal order for either of the functions $\omega(n)$ and $\Omega(n)$, giving the total number of prime factors of n, counted with or without multiplicity.

In many instances, a good candidate for g is the mean value of f, that is,

$$g(N) = E_N(f) = \frac{1}{N}\sum_{n\leqslant N} f(n).$$

We enter the domain of probabilistic number theory. An often efficient method consists in evaluating the variance

$$V_N(f) = E_N(|f(n) - E_N(f)|^2)$$

and applying the Bienaymé–Chebyshev inequality. We need however to be able to evaluate this variance and approximate the mean.

The Turán–Kubilius inequality provides a bound for the variance of additive functions. An arithmetic function h is said to be additive if $h(mn) = h(m)+h(n)$ whenever $(m,n) = 1$. The functions $\ln n$, $\omega(n)$, $\Omega(n)$, are prototypes of additive functions. Such functions are determined by their values on prime powers. The reader will find in [19] a detailed construction of the probabilistic model that can be attached to an additive function. Approximating the probability that an integer is divisible by p^k and not by p^{k+1} by $1/p^k - 1/p^{k+1} = (1-1/p)p^{-k}$, it is reasonable to expect that, for additive f, the mean value $E_N(f)$ is close to $E_N^*(f) := \sum_{p^\nu\leqslant N}(1-1/p)f(p^\nu)/p^\nu$. A corresponding approximation of the variance becomes

$$V_N^*(f) = \frac{1}{N}\sum_{1\leqslant n\leqslant N}\left|f(n) - E_N^*(f)\right|^2.$$

The Turán–Kubilius inequality states that, uniformly for all additive, complex-valued f, we have

$$V_N^*(f) \leqslant \left\{4 + O\left(\sqrt{\frac{\ln\ln N}{\ln N}}\right)\right\}B_N(f)^2,$$

where $B_N(f)^2$ is the corresponding approximation of the second moment:

$$B_N(f)^2 = \sum_{p^\nu\leqslant N}\frac{|f(p^\nu)|^2}{p^\nu}\Big(1-\frac{1}{p}\Big).$$

As an immediate application we get a quantitative version of the Hardy–Ramanujan theorem quoted above. La Bretèche and Tenenbaum [3] extended this inequality to friable integers. For $p \leqslant y$, the probability that a friable integer is exactly divisible by p^k is close to $(1-1/p^\alpha)p^{-k\alpha}$ where α is the saddle point defined by (8). The expectation, variance, and second order moment associated to this probabilistic model for friable integers are respectively

$$E_N^*(f,y) = \sum_{p^\nu\in S(N,y)}\frac{f(p^\nu)}{p^{\alpha\nu}}\Big(1-\frac{1}{p^\alpha}\Big),$$

$$V_N^*(f,y) = \frac{1}{\Psi(N,y)}\sum_{n\in S(N,y)}|f(n) - E_N^*(f,y)|^2,$$

$$B_N(f,y)^2 = \sum_{p^\nu\in S(N,y)}\frac{|f(p^\nu)|^2}{p^{\alpha\nu}}\Big(1-\frac{1}{p^\alpha}\Big).$$

La Bretèche and Tenenbaum proved that there exists an absolute constant $C > 0$ such that for all $2 \leqslant y \leqslant N$, we have

$$V_N^*(f,y) \leqslant CB_N(f,y)^2.$$

Furthermore, they established in the same optimal range the more precise inequality

$$V_N^*(f,y) \ll V(Z_f),$$

where $V(Z_f)$ is the variance of the probabilistic model Z_f associated to the additive function f on the set of friable integers. This theorem has very nice consequences for properties of friable integers. In this survey, we will present only one but the reader will find other interesting applications in [3]. Let us denote by $\{p_j(n)\}_{1\leqslant j\leq\omega(n)}$ the increasing sequence of the prime factors of an integer n. A very surprising fact is that, for almost all integers n, the order of magnitude of $p_j(n)$ depends only on j: for almost all $n \leqslant x$, we have $\ln_2(p_j(n)) \sim j$ for $J_x \leqslant j \leqslant \omega(n)$, where J_x is any function tending to infinity with x.

The friable Turán–Kubilius inequality provides a normal order for these quantities $p_j(n)$ when n runs through friable integers. La Bretèche and Tenenbaum proved that "small" prime factors of friable integers behave similarly to those of ordinary integers but that, after a certain critical index, one observes an increasing compression phenomenon. In particular, for almost all $n \in S(x,y)$ and $y \leqslant (\ln x)^{1+o(1)}$, J_x as above, we have $p_j(n) = p_j^{1+o(1)}$ for $J_x \leqslant j \leqslant \omega(n)$, where p_j is the jth prime number. Since, by the prime number theorem, $p_j \sim j\ln j$, the situation is thus very different from that of generic, normal integers.

4 Applications to algorithmic number theory and cryptography

Security of a variety of public-key cryptographic systems relies on the difficulty of factoring integers whose prime factors are large. For example, the public key in the RSA system is an integer N that is the product of two "large" prime factors, i.e., $N = pq$. Decoding is equivalent to determining the prime factors p and q.

Friable integers play a prominent role in a number of factoring algorithms, in particular in the process of finding the above p and q. Friable integers are also needed in the discrete logarithm problem and in some primality testing. In this section we provide a general idea of their use in some of these algorithms. We have selected situations that can be described

5 A subset $A \subset \mathbb{N}$ is of natural density 1 if

$$\lim_{N\to\infty}\frac{1}{N}\#\{n\leqslant N : n\in A\} = 1.$$

with little mathematical background only. In particular we will not evoke the factoring algorithms and the primality testing based on elliptic curves which are nevertheless most frequently used.

The role of the friable integers in these algorithms is analogous to the one that will be depicted in this section even if the context is different. We refer for example to Pomerance's survey in the International Mathematical Congress in Zurich in 1994 [17] and to the book by Crandall and Pomerance [5], which are important references in this context.

4.1 The quadratic sieve

The quadratic sieve was devised by Pomerance at the beginning of the eighties. We aim to find the prime factors of an integer n that we know to be composite. The starting idea is that if a and b are two integers such that $a \not\equiv \pm b \pmod n$ and $a^2 \equiv b^2 \pmod n$ then $(a-b, n)$ will be a non-trivial divisor of n. The challenge is thus to determine such integers a and b.

The first step of the quadratic sieve consists in considering the y-friable values of the polynomial $Q(t) = t^2 - n$ when t is close to $\sqrt{n}$, that is, for $|t - \sqrt{n}| < n^\varepsilon$ with some small $\varepsilon > 0$. For such t, $Q(t)$ will be small: $|Q(t)| \leqslant (2 + n^{-1/2+\varepsilon})n^{1/2+\varepsilon} < 3n^{1/2+\varepsilon}$. It is natural to expect that these values have multiplicative properties resembling those of the integers in the interval $]-3n^{1/2+\varepsilon}, 3n^{1/2+\varepsilon}[$. Why do we need so many friable values $Q(t)$? We write as $(t_1, Q(t_1)), \dots, (t_N, Q(t_N))$ the pairs for which $Q(t_i)$ is friable. The second idea of Pomerance is that if we have at least $N \geqslant \pi(y)+1$ such y-friable integers then we can build a square with these $Q(t_i)$. This can be seen with a linear algebraic argument. The factoring of each $Q(t_i)$ is of the shape $Q(t_i) = \prod_{p \leqslant y} p^{v_p(Q(t_i))}$ where $v_p(a)$ denotes the p-adic valuation of a. We can associate to each $Q(t_i)$ a vector in $\mathbb{F}_2^{\pi(y)}$ whose coordinates are the $v_p(Q(t_i)) \mod 2$, $p \leqslant y$. Since $N > \pi(y)$, the number of vectors is strictly bigger than the dimension. Thus these vectors are not independent, and there exists $J \subset \{1, \dots, N\}$ such that $\sum_{j\in J} v_p(Q(t_j)) \equiv 0 \mod 2$ for all $p \leqslant y$, in other words such that $\prod_{j\in J} \prod_{p\leqslant y} p^{v_p(Q(t_j))}$ is a square denoted by v^2. We will get

$$\Big(\prod_{j\in J} t_j\Big)^2 \equiv \prod_{j\in J}(t_j^2 - n) \equiv v^2 \pmod n.$$

If $\prod_{j\in J} t_j \not\equiv \pm v \pmod n$, $(v - \prod_{j\in J} t_j, n)$ is a non-trivial divisor of n detected with friable integers. Using (6) and under the assumption that the values of the previous polynomial behave like generic integers in the interval $]-3n^{1/2+\varepsilon}, 3n^{1/2+\varepsilon}[$, Pomerance proved that the complexity of the quadratic sieve is $L(n)^{1+o(1)}$ with $L(n) = \exp(\sqrt{\ln n \ln_2 n})$; $L(n)^{\sqrt{2}/2}$ is in fact the optimal friable limit for this algorithm.

4.2 The number field sieve

The quadratic sieve is usually used for factoring integers with less than one hundred digits. For larger integers the number field sieve, to which this section is devoted, is available.

Let n be a number to factorise. We first determine a polynomial f of degree $d \geqslant 2$ such that $f(m) \equiv 0 \pmod n$ for some integer m close to $n^{1/d}$. For example, for $m = \lfloor n^{1/d} \rfloor$, we can form the polynomial f with the expansion of n in base m: $n = m^d + c_{d-1}m^{d-1} + \cdots + c_1 m + c_0$, where the digits c_j lie between 0 and $m-1$. Next, we consider the polynomial $f(X)$. If it is reducible, we immediately obtain a factor of n. We can thus assume this does not hold. Let $\vartheta \in \mathbb{C}$ be a root of $f(X)$. We try to find a set S of pairs of coprime integers (a, b) such that

$$\prod_{(a,b)\in S} (a - b\vartheta) = \gamma^2, \quad \prod_{(a,b)\in S} (a - bm) = v^2 \tag{11}$$

for some $\gamma \in \mathbb{Z}[\vartheta]$, $v \in \mathbb{Z}$. The process to derive such squares is similar to the one described in the case of the quadratic sieve. The first step consists in finding pairs (a, b) such that $(a - bm)$ and $N(a - b\vartheta)$ are friable where N is the norm on $\mathbb{Q}(\vartheta)$.

Assuming again that the values of the considered polynomials behave like random integers, Buhler, H. Lenstra, and Pomerance proved that the complexity of the number field sieve is $\ll \exp(c(\ln n)^{1/3}(\ln_2 n)^{2/3})$, this bound being achieved for polynomials of degree $d \sim \left(\frac{3\ln n}{\ln_2 n}\right)^{1/3}$. These results are based in particular on conjectures on the distribution of friable integers in short intervals and in polynomial sequences. Currently these conjectures are out of reach, especially in the case of the number field sieve where the degree of the optimal polynomial is very high.

These last years, much research has been devoted to this subject. We saw above that binary forms of type

$$F(a, b) = (a - bm)N(a - b\vartheta)$$

play an important role in the number field sieve. For general binary forms $F \in \mathbb{Z}[X_1, X_2]$, Balog, Blomer, Tenenbaum, and the author established some lower bounds for

$$\Psi_F(x, y) = \big|\{1 \leqslant a, b \leqslant x : P^+(F(a, b)) \leqslant y\}\big|$$

when $y \geqslant x^{\alpha_F+\varepsilon}$ where α_F depends on the structure of F. In the case of irreducible binary forms, the exponent

$$\alpha_F = \deg F - 2$$

is admissible. Lachand [14] obtained asymptotic formulas valid in domains with $y = x^{o(1)}$ when f is a cubic or a product of linear terms (with an explicit expression of the previous y-exponent "o(1)" in the cubic case).

4.3 The discrete logarithm problem

The discrete logarithm is used in many cryptographic protocols. Let p be a large prime number, g a generator of $\mathbb{F}_p^*$, and $t \in \mathbb{F}_p^*$. The discrete logarithm problem[6] consists in determining ℓ such that $g^\ell = t$. We then write $\ell = \log_g t$. We start out by selecting those powers g^m having a y-friable representative. If we can find sufficiently many such powers, a linear algebra argument will enable us to determine the discrete logarithms of the primes $q \leqslant y$. After this stage, we consider the products $g^m t$ where m is a random integer. If one of the $g^m t$ is y-friable, thus of type $g^m t = \prod_{i=1}^r q_i^{a_i}$, with all the $q_i \leqslant y$, we will deduce that $\log_g t = -m + \sum_{i=1}^r a_i \log_g(q_i)$.

5 Applications of friable integers to analysis and number theory

On many occasions, friable integers opened new perspectives in problems that had remained out of reach for decades. In this section we briefly expose their use in various contexts.

6 We can work in a more general context by replacing $\mathbb{F}_p^*$ with a cyclic group.

5.1 The prime number theorem, Daboussi's theorem for multiplicative functions

In the first half of the last century, many mathematicians were convinced that an elementary proof of the prime number theorem was not possible; the qualification 'elementary' means here using only the usual tools of real analysis, and among other things avoiding complex analysis.

It was a huge surprise when Erdős and Selberg provided in 1949 an elementary but rather difficult proof of the prime number theorem.

In 1984, Daboussi [6] gave a very elegant proof by using friable integers. Let μ denote the Möbius function. This function is defined in the following way: $\mu(n) = 0$ if n is divisible by the square of a prime number, otherwise $\mu(n) = (-1)^{\omega(n)}$, where $\omega(n)$ is the number of distinct prime factors of n. A classical result in number theory asserts that the prime number theorem is equivalent to the formula

$$M(x) := \sum_{n\leqslant x} \mu(n) = o(x). \tag{12}$$

One of the ideas of Daboussi is to represent $M(x)$ in terms of sums of the Möbius function over friable integers, viz.

$$M(x,y) := \sum_{n\in S(x,y)} \mu(n).$$

Writing $n = ab$ where a is y-friable and b is y-sifted, we arrive at the formula

$$M(x) = \sum_{P^-(b)>y} \mu(b) M(x/b, y).$$

The other steps follow a more natural progression than the initial proof of Erdős and Selberg. The main ingredients are very simple estimates on sifted and on friable integers, the crucial point being the upper bound of some kind of mean value of the $M(x,y)$.

This limpid process of sifted–friable factorization combined with convolution[7] methods may be used to produce a new proof of the following theorem due to Daboussi: if f is a multiplicative function (i.e., $f(mn) = f(m)f(n)$ when m and n are coprime) with modulus at most 1 then for all real irrational α, we have

$$\lim_{x\to\infty} \frac{1}{x} \sum_{n\leqslant x} f(n) \exp(2i\pi n\alpha) = 0.$$

5.2 Friable integers and Waring's problem

Waring's problem consists in determining, given an integer $k \geqslant 2$, the smallest integer s such that all natural numbers can be represented as a sum of s kth powers. In a slightly weaker version one only asks that all sufficiently large enough integers are representable. In this latter version the number s is traditionally denoted by $G(k)$. For example, Linnik showed that $G(3) \leqslant 7$, and Davenport proved that $G(4) = 16$.

The circle method is a direct approach in which one expresses the number $R(n)$ of representations of any integer n as sum of s kth powers by a Cauchy or a Fourier integral:

$$R(n) = \int_0^1 F(\alpha)^s \exp(-2i\pi n\alpha)\, \mathrm{d}\alpha,$$

$$\text{with } F(\alpha) = \sum_{m\leqslant n^{1/k}} \exp(2i\pi m^k \alpha).$$

The main contribution to this integral arises from the so-called *major arcs* corresponding to those α close to a rational number with small denominator. Traditionally, the set of these α is denoted by $\mathcal{M}$, and its complement $\mathfrak{m} =]0,1[\setminus\mathcal{M}$ is called the set of *minor arcs*.

An important feature of this method consists in showing that the contribution from the minor arcs is negligible. One can use bounds of type

$$\int_{\mathfrak{m}} |F(\alpha)|^s \,\mathrm{d}\alpha \leqslant \max_{\alpha\in\mathfrak{m}} |F(\alpha)|^{s-2\ell} \int_0^1 |F(\alpha)|^{2\ell}\,\mathrm{d}\alpha,$$

with $2\ell \leqslant s$. The integral on the right-hand side then corresponds to the number of solutions of the Diophantine equation

$$x_1^k + \cdots + x_\ell^k = y_1^k + \cdots + y_\ell^k \quad (1 \leqslant x_i, y_i \leqslant n^{1/k}).$$

By restricting some variables to be friable, Wooley [21] settled functional inequalities between the number of solutions of a Diophantine equation in ℓ friable variables x_i, y_i and the number of solutions of an equation in $\ell - 1$ variables. The process is very complicated and not suitable for a short account, however Wooley obtained in this way many important new results, in particular that $G(k) \leqslant k\ln k + k\ln_2 k + O(k)$ as $k\to\infty$. The best previously know bound was $\leqslant (2 + o(1))k\ln k$.

5.3 Friable summation and Davenport identities

Duffin and then Fouvry and Tenenbaum [9] introduced a new summation method, called friable summation or P-summation. It consists of ordering the indexes of a series according to their largest prime factor and then studying convergence as the friable parameter approaches infinity. This leads to the following definition: a series $\sum \alpha_n$ is said to have friable sum α (or to be P-convergent to α) if

$$\sum_{P^+(n)\leqslant y} \alpha_n = \alpha + o(1) \quad (y\to+\infty).$$

A series already summable for the friable method is called *regular* if it is convergent in the usual sense and its friable sum α is equal to its ordinary sum:

$$\lim_{y\to+\infty} \sum_{P^+(n)\leqslant y} \alpha_n = \alpha = \lim_{N\to+\infty} \sum_{n=1}^{N} \alpha_n.$$

It may happen that the friable sum exists but does not coincide with the usual sum. Theorem 11 of [9] provides infinitely many examples. While, for $\vartheta \in \mathbb{R}\setminus\mathbb{Z}$, the series $\sum_{n\geqslant 1} \exp(2i\pi n\vartheta)/n$ converges (in the usual sense) to

$$\log\Big(\frac{1}{1-\exp(2i\pi\vartheta)}\Big),$$

where, here and in the sequel to this paragraph, the complex logarithm is understood as the principal determination, Fouvry and Tenenbaum proved that, for all rationals $\vartheta = a/q$ with $(a,q) = 1$, $1 \leqslant a < q$ the friable sum exists and has value

$$\lim_{y\to\infty} \sum_{P(n)\leqslant y} \frac{\exp(2i\pi na/q)}{n} = \log\Big(\frac{1}{1-\exp(2i\pi a/q)}\Big) + \frac{\Lambda(q)}{\varphi(q)},$$

where Λ is the Von Mangoldt function, defined by $\Lambda(q) = \ln p$ if $q = p^k$, $\Lambda(q) = 0$ otherwise.

Friable summation avoids the Gibbs phenomenon and can be employed to solve arduous questions in analysis such as

7 The convolution product of two arithmetic functions is defined in Section 5.3.

the problem of Davenport's identities that we will describe now.

The Dirichlet convolution product of two arithmetic functions u, v is defined by the formula

$$u * v(n) = \sum_{d|n} u(d)v(n/d) \qquad (n \geqslant 1).$$

The identity element for this composition law in the ring of arithmetic functions is often denoted by δ, so that $\delta(n) = 1$ for $n = 1$ and $\delta(n) = 0$ for $n \geqslant 2$. We denote by $\mathbf{1}$ the constant function equal to 1 for all positive integers. We then have the fundamental formula $\delta = \mu * \mathbf{1}$.

Let $B_1(t)$ be the first Bernoulli function, defined by $B_1(t) = \{t\} - 1/2$ if $t \notin \mathbb{Z}$ and $B_1(t) = 0$ for $t \in \mathbb{Z}$, where $\{t\}$ denotes the fractional part of t. It coincides everywhere with its Fourier series:

$$B_1(\vartheta) = -\sum_{k\geqslant 1} \frac{\sin(2\pi k\vartheta)}{\pi k}.$$

With this equation, we formally obtain for two arithmetic functions f and g such that $f = g * \mathbf{1}$, the beautiful identity

$$\sum_{m\geqslant 1} \frac{f(m)}{\pi m} \sin(2\pi m\vartheta) + \sum_{n\geqslant 1} \frac{g(n)}{n} B_1(n\vartheta) = 0. \qquad (13)$$

This led Davenport to formulate the following problem: given f and g, determine those real numbers ϑ such that relation (13) holds. This is a very difficult task and we do not provide a general answer.

Davenport proved that, in the case $(f, g) = (\delta, \mu)$, identity (13) is satisfied for all $\vartheta \in \mathbb{R}$. However, his argument does not work in the emblematic cases $(f, g) = (\ln, \Lambda)$, $(\tau, \mathbf{1})$, where τ is the divisor counting function. It was only sixty years later that these cases were solved by La Bretèche and Tenenbaum [2]. One of the crucial ingredients of their work is the application of friable summation which is remarkably well adapted to this problem. They proved that for $(f, g) = (\ln, \Lambda)$, (13) holds for all real ϑ, but, in the case $(f, g) = (\tau, \mathbf{1})$, they gave a criterion in terms of the continued fraction expansion of the irrational ϑ for the validity of (13). The case of the powers of convolution of $\mathbf{1}$ has then been handled by B. Martin.

5.4 Small gaps between prime numbers

We end this survey with an account of spectacular progress obtained using friable integers: the works of Zhang [22] and Maynard [16] on small gaps between the prime numbers. Zhang caused a sensation in 2013[8] by showing that there are infinitely many prime numbers $p \neq q$ such that

$$|p - q| \leqslant 70\,000\,000.$$

After this breakthrough the upper bound was reduced several times, in particular by the Polymath project. At the end of this same year there was another spectacular result: Maynard announced that this bound could be reduced to 600. The current value is 246. The twin prime conjecture, which corresponds to infinitely many prime gaps equal to 2, seems less inaccessible than ten years ago.

A key ingredient of Zhang's proof is a result on the average distribution of the primes in arithmetic progressions with friable moduli. This friable structure enables one to consider sets of integers not exceeding x and satisfying some congruence conditions modulo integers larger than $\sqrt{x}$; this was crucial in Zhang's approach to prove bounded gaps between infinitely many prime numbers.

Very recently Régis de la Bretèche [1] wrote a fascinating article for the *Gazette des Mathématiciens* on the cooperative project Polymath around the breakthroughs by Zhang [22] and Maynard [16] on this subject. The interested reader can look at [1] (in French) for more details of this wonderful progress. Undoubtedly, very beautiful mathematics is yet to be discovered along the paths of friable integers.

Acknowledgements

I am very grateful to Pierrick Gaudry, Martine and Hervé Queffélec, Anne de Roton, Gérald Tenenbaum, and to the two anonymous referees for their meticulous rereading and for all their remarks related to the French version in the *Gazette des Mathématiciens* [7]. I would also like to express warm thanks to Irène Marcovici and to Gérald Tenenbaum for all their help for this English version.

Bibliography

[1] R. de la Bretèche. Petits écarts entre les nombres premiers et polymath: une nouvelle manière de faire de la recherche en mathématiques? *SMF Gaz. des Math.*, 140:19–31, 2014.

[2] R. de la Bretèche and G. Tenenbaum. Séries trigonométriques à coefficients arithmétiques. *J. Anal. Math. 92*, 92:1–79, 2004.

[3] R. de la Bretèche and G. Tenenbaum. Entiers friables : inégalité de Turán–Kubilius et applications. *Invent. Math.*, 159:531–588, 2005.

[4] R. de la Bretèche and G. Tenenbaum. Une nouvelle approche dans la théorie des entiers friables. *Compos. Math.*, 153:453–473, 2017.

[5] R. Crandall and C. Pomerance. *Prime numbers, a computational perspective*, volume 4ème édition. Springer, 546 pp., 2001.

[6] H. Daboussi. Sur le théorème des nombres premiers. *C. R. Acad. Sc. Paris, Série I*, 298(8):161–164, 1984.

[7] C. Dartyge. Entiers friables : un tour d'horizon. *SMF Gaz. des Math.*, 156:29–39, 2018.

[8] S. Drappeau. Théorèmes de type Fouvry–Iwaniec pour les entiers friables. *Compos. Math.*, 151:828–862, 2015.

[9] E. Fouvry and G. Tenenbaum. Entiers sans grand facteur premier en progressions arithmétiques. *Proc. London Math. Soc. (3)*, 63:449–494, 1991.

[10] A. Granville. Integers, without large prime factors, in arithmetic progressions. I. *Acta Math.*, 170:255–273, 1993.

[11] A. Granville. Smooth numbers: computational number theory and beyond. *Algorithmic number theory, MSRI Proceedings*, 44:267–323, 2008.

[12] A. J. Harper. On a paper of K. Soundararajan on smooth numbers in arithmetic progressions. *J. Number Theory*, 132(1):182–199, 2012.

[13] A. Hildebrand and G. Tenenbaum. Integers without large prime factors. *Journal de Théorie des Nombres de Bordeaux*, 5(2):411–484, 1993.

[14] A. Lachand. *Entiers friables et formes binaires*. Thèse, Université de Lorraine, 2014.

[15] K. Matomäki and M. Radziwiłł. Multiplicative functions in short intervals. *Ann. of Math. (2)*, 183(3):1015–1056, 2016.

[16] J. Maynard. Small gaps between primes. *Ann. of Math. (2)*, 183(1):383–413, 2015.

[17] C. Pomerance. The role of smooth sumbers in number theoretic algorithms. *Proceedings of the international comgress*

8 The corresponding article appeared in 2014.

of mathematiciens, Zurich Switzerland 1994, 5(2):411–422, 1995.

[18] G. Tenenbaum. *Introduction à la théorie analytique et probabiliste des nombres*, volume 4ème édition. coll. Échelles, Belin, 592 pp., 2015.

[19] G. Tenenbaum. *Introduction to Analytic and Probabilistic Number Theory*, volume Third edition. Graduate Studies in Mathematics 163, Amer. Math. Soc., 2015.

[20] G. Tenenbaum and J. Wu. Moyennes de certaines fonctions multiplicatives sur les entiers friables. *J. Reine Angew. Math.*, 564:119–166, 2003.

[21] T. D. Wooley. Large improvements in Waring's problem. *Ann. of Maths.*, 135:131–164, 1992.

[22] Y. Zhang. Bounded gaps between primes. *Ann. of Math. (2)*, 179(3):1121–1174, 2014.

Cécile Dartyge [cecile.dartyge@univ-lorraine.fr] is maîtresse de conférences in the Institut Elie Cartan, Université de Lorraine, France. Her actual research interests are in analytic number theory.