A joint central limit theorem for the sum-of-digits function, and asymptotic divisibility of Catalan-like sequences

Michael Drmota and Christian Krattenthaler

Technische Universität Wien; Universität Wien

2-divisibility of central binomial coefficients

We all know that

$$\binom{2n}{n} \equiv 0 \pmod{2}$$

2-divisibility of central binomial coefficients

We all know that

$$\binom{2n}{n} \equiv 0 \pmod 2 \quad \text{for } n \geq 1.$$

2-divisibility of central binomial coefficients

We all know that

$$\binom{2n}{n} \equiv 0 \pmod 2 \quad \text{for } n \geq 1.$$

We all know that

$$\binom{2n}{n} \equiv 0 \pmod 4 \quad \text{for } n \geq 2.$$

2-divisibility of central binomial coefficients

We all know that

$$\binom{2n}{n} \equiv 0 \pmod 2 \quad \text{for } n \geq 1.$$

We all know that

$$\binom{2n}{n} \equiv 0 \pmod 4 \quad \text{for } n \geq 3.$$

2-divisibility of central binomial coefficients

We all know that

$$\binom{2n}{n} \equiv 0 \pmod 2 \quad \text{for } n \geq 1.$$

We all know that not always

$$\binom{2n}{n} \equiv 0 \pmod 4.$$

More precisely, the above holds if and only if n is not a power of 2.

2-divisibility of central binomial coefficients

We all know that

$$\binom{2n}{n} \equiv 0 \pmod 2 \quad \text{for } n \geq 1.$$

We all know that not always

$$\binom{2n}{n} \equiv 0 \pmod 4.$$

More precisely, the above holds if and only if n is not a power of 2.
In particular, this implies that

$$\lim_{N\to\infty} \frac{1}{N} \# \left\{ n < N : \binom{2n}{n} \equiv 0 \ (\text{mod } 4) \right\} = 1.$$

How about

$$\lim_{N\to\infty} \frac{1}{N} \# \left\{ n < N : \binom{2n}{n} \equiv 0 \pmod 8 \right\} = ?$$

2-divisibility of central binomial coefficients

How about

$$\lim_{N\to\infty} \frac{1}{N} \# \left\{ n < N : \binom{2n}{n} \equiv 0 \pmod 8 \right\} = ?$$

We have

$$\frac{1}{10} \# \left\{ n < 10 : \binom{2n}{n} \equiv 0 \pmod 8 \right\} = 0.1$$

2-divisibility of central binomial coefficients

How about

$$\lim_{N\to\infty} \frac{1}{N} \# \left\{ n < N : \binom{2n}{n} \equiv 0 \pmod 8 \right\} = ?$$

We have

$$\frac{1}{50} \# \left\{ n < 50 : \binom{2n}{n} \equiv 0 \pmod 8 \right\} = 0.56$$

2-divisibility of central binomial coefficients

How about

$$\lim_{N\to\infty} \frac{1}{N} \# \left\{ n < N : \binom{2n}{n} \equiv 0 \pmod 8 \right\} =?$$

We have

$$\frac{1}{100} \# \left\{ n < 100 : \binom{2n}{n} \equiv 0 \pmod 8 \right\} = 0.71$$

2-divisibility of central binomial coefficients

How about

$$\lim_{N\to\infty} \frac{1}{N} \# \left\{ n < N : \binom{2n}{n} \equiv 0 \pmod 8 \right\} = ?$$

We have

$$\frac{1}{1000} \# \left\{ n < 1000 : \binom{2n}{n} \equiv 0 \pmod 8 \right\} = 0.944$$

2-divisibility of central binomial coefficients

How about

$$\lim_{N\to\infty} \frac{1}{N} \# \left\{ n < N : \binom{2n}{n} \equiv 0 \pmod 8 \right\} = ?$$

We have

$$\frac{1}{10000} \# \left\{ n < 10000 : \binom{2n}{n} \equiv 0 \pmod 8 \right\} = 0.9896$$

2-divisibility of central binomial coefficients

How about

$$\lim_{N\to\infty} \frac{1}{N} \# \left\{ n < N : \binom{2n}{n} \equiv 0 \pmod 8 \right\} = ?$$

We have

$$\frac{1}{10000} \# \left\{ n < 10000 : \binom{2n}{n} \equiv 0 \pmod 8 \right\} = 0.9896$$

Apparently, again

$$\lim_{N\to\infty} \frac{1}{N} \# \left\{ n < N : \binom{2n}{n} \equiv 0 \pmod 8 \right\} = 1.$$

2-divisibility of central binomial coefficients

How about

$$\lim_{N\to\infty} \frac{1}{N}\#\left\{n < N : \binom{2n}{n} \equiv 0 \pmod 8\right\} =?$$

We have

$$\frac{1}{10000}\#\left\{n < 10000 : \binom{2n}{n} \equiv 0 \pmod 8\right\} = 0.9896$$

Apparently, again

$$\lim_{N\to\infty} \frac{1}{N}\#\left\{n < N : \binom{2n}{n} \equiv 0 \pmod 8\right\} = 1.$$

The same observation works modulo 16, modulo 32, etc.

2-divisibility of Catalan numbers

We all (?) know that

$$C_n = \frac{1}{n+1}\binom{2n}{n} \equiv 0 \pmod 2$$

if and only if $n \neq 2^e - 1$, $e = 0, 1, 2, \dots$.

2-divisibility of Catalan numbers

We all (?) know that

$$C_n = \frac{1}{n+1}\binom{2n}{n} \equiv 0 \pmod 2$$

if and only if $n \neq 2^e - 1$, $e = 0, 1, 2, \dots$.
In particular, this implies that

$$\lim_{N\to\infty} \frac{1}{N} \# \left\{ n < N : \frac{1}{n+1}\binom{2n}{n} \equiv 0 \ (\text{mod } 2) \right\} = 1.$$

2-divisibility of Catalan numbers

How about

$$\lim_{N\to\infty} \frac{1}{N} \# \left\{ n < N : \frac{1}{n+1}\binom{2n}{n} \equiv 0 \pmod 4 \right\} = ?$$

How about

$$\lim_{N\to\infty} \frac{1}{N} \# \left\{ n < N : \frac{1}{n+1}\binom{2n}{n} \equiv 0 \pmod 4 \right\} = ?$$

We have

$$\frac{1}{10} \# \left\{ n < 10 : \frac{1}{n+1}\binom{2n}{n} \equiv 0 \pmod 4 \right\} = 0.1$$

2-divisibility of Catalan numbers

How about

$$\lim_{N\to\infty} \frac{1}{N} \# \left\{ n < N : \frac{1}{n+1}\binom{2n}{n} \equiv 0 \pmod 4 \right\} = ?$$

We have

$$\frac{1}{50} \# \left\{ n < 50 : \frac{1}{n+1}\binom{2n}{n} \equiv 0 \pmod 4 \right\} = 0.58$$

2-divisibility of Catalan numbers

How about

$$\lim_{N\to\infty} \frac{1}{N} \# \left\{ n < N : \frac{1}{n+1}\binom{2n}{n} \equiv 0 \pmod 4 \right\} = ?$$

We have

$$\frac{1}{100} \# \left\{ n < 100 : \frac{1}{n+1}\binom{2n}{n} \equiv 0 \pmod 4 \right\} = 0.72$$

How about

$$\lim_{N\to\infty} \frac{1}{N} \# \left\{ n < N : \frac{1}{n+1}\binom{2n}{n} \equiv 0 \pmod 4 \right\} =?$$

We have

$$\frac{1}{1000} \# \left\{ n < 1000 : \frac{1}{n+1}\binom{2n}{n} \equiv 0 \pmod 4 \right\} = 0.945$$

2-divisibility of Catalan numbers

How about

$$\lim_{N\to\infty} \frac{1}{N} \# \left\{ n < N : \frac{1}{n+1}\binom{2n}{n} \equiv 0 \pmod 4 \right\} =?$$

We have

$$\frac{1}{10000} \# \left\{ n < 10000 : \frac{1}{n+1}\binom{2n}{n} \equiv 0 \pmod 4 \right\} = 0.9897$$

How about

$$\lim_{N\to\infty} \frac{1}{N} \# \left\{ n < N : \frac{1}{n+1} \binom{2n}{n} \equiv 0 \pmod 4 \right\} = ?$$

We have

$$\frac{1}{10000} \# \left\{ n < 10000 : \frac{1}{n+1} \binom{2n}{n} \equiv 0 \pmod 4 \right\} = 0.9897$$

Apparently, again

$$\lim_{N\to\infty} \frac{1}{N} \# \left\{ n < N : \frac{1}{n+1} \binom{2n}{n} \equiv 0 \pmod 4 \right\} = 1,$$

and the same observation holds modulo 8, modulo 16, etc.

5-divisibility of Catalan numbers

However, there is nothing special about the modulus 2:

5-divisibility of Catalan numbers

However, there is nothing special about the modulus 2:

Here are the first few Catalan numbers:

$$1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862, 16796, 58786, 208012, 742900,$$
$$2674440, 9694845, 35357670, 129644790, 477638700, 1767263190,$$
$$6564120420, 24466267020, 91482563640, 343059613650, 1289904147324,$$
$$4861946401452, 18367353072152, 69533550916004, 263747951750360,$$
$$1002242216651368, 3814986502092304, 14544636039226909,$$
$$55534064877048198, 212336130412243110, 812944042149730764,$$
$$3116285494907301262, 11959798385860453492, 45950804324621742364,$$
$$176733862787006701400, 680425371729975800390, 2622127042276492108820,$$
$$10113918591637898134020, 39044429911904443959240,$$
$$150853479205085351660700, 583300119592996693088040,$$
$$2257117854077248073253720, 8740328711533173390046320,$$
$$33868773757191046886429490, 131327898242169365477991900, \dots$$

5-divisibility of Catalan numbers

However, there is nothing special about the modulus 2:

We have

$$\frac{1}{10000}\#\left\{n < 10000 : \binom{2n}{n} \equiv 0 \pmod{25}\right\} = 0.702$$

5-divisibility of Catalan numbers

However, there is nothing special about the modulus 2:

We have

$$\frac{1}{100000}\#\left\{n < 100000 : \binom{2n}{n} \equiv 0 \pmod{25}\right\} = 0.82612$$

5-divisibility of Catalan numbers

However, there is nothing special about the modulus 2:

We have

$$\frac{1}{100000}\#\left\{n < 100000 : \binom{2n}{n} \equiv 0 \pmod{25}\right\} = 0.82612$$

More calculations indicate that

$$\lim_{N\to\infty} \frac{1}{N}\#\left\{n < N : \frac{1}{n+1}\binom{2n}{n} \equiv 0 \pmod{5^\alpha}\right\} = 1,$$

for any α.

p-divisibility of Catalan numbers

In a series of preprints on the $\mathtt{ar}\chi\mathtt{iv}$, Rob Burns investigated divisibility properties of combinatorial numbers. In particular, using an automata method of Eric Rowland and Reem Yassawi, he proved that

$$\lim_{N\to\infty} \frac{1}{N} \# \left\{ n < N : \frac{1}{n+1} \binom{2n}{n} \equiv 0 \pmod{p} \right\} = 1,$$

for any prime number p.

p-divisibility of Catalan numbers

In a series of preprints on the $\mathtt{ar}\chi\mathtt{iv}$, Rob Burns investigated divisibility properties of combinatorial numbers. In particular, using an automata method of Eric Rowland and Reem Yassawi, he proved that

$$\lim_{N\to\infty} \frac{1}{N} \# \left\{ n < N : \frac{1}{n+1}\binom{2n}{n} \equiv 0 \pmod{p} \right\} = 1,$$

for any prime number p.

Together with Michael Drmota, I decided to "do this properly".

p-divisibility of Catalan numbers

In a series of preprints on the $\mathtt{ar}\chi\mathtt{iv}$, Rob Burns investigated divisibility properties of combinatorial numbers. In particular, using an automata method of Eric Rowland and Reem Yassawi, he proved that

$$\lim_{N\to\infty} \frac{1}{N} \# \left\{ n < N : \frac{1}{n+1} \binom{2n}{n} \equiv 0 \pmod{p} \right\} = 1,$$

for any prime number p.

Together with Michael Drmota, I decided to "do this properly".

– Prove the same result for any prime power.

p-divisibility of Catalan numbers

In a series of preprints on the arχiv, Rob Burns investigated divisibility properties of combinatorial numbers. In particular, using an automata method of Eric Rowland and Reem Yassawi, he proved that

$$\lim_{N\to\infty} \frac{1}{N} \# \left\{ n < N : \frac{1}{n+1}\binom{2n}{n} \equiv 0 \pmod{p} \right\} = 1,$$

for any prime number p.

Together with Michael Drmota, I decided to "do this properly".

- Prove the same result for any prime power.
- Prove this kind of result for a large(r) class of sequences.

How to “do this properly”

How to “do this properly”

Let $v_p(N)$ denote the *p-adic valuation* of the integer N, which by definition is the maximal exponent α such that p^α divides N.

Legendre’s formula for the p-adic valuation of factorials implies

$$v_p(n!) = \frac{1}{p-1}\big(n - s_p(n)\big),$$

where $s_p(N)$ denotes the *p-ary sum-of-digits function*

$$s_p(N) = \sum_{j \ge 0} \varepsilon_j(N),$$

with $\varepsilon_j(N)$ denoting the j-th digit in the p-adic representation of N.

How to "do this properly"

Let $v_p(N)$ denote the *p-adic valuation* of the integer N, which by definition is the maximal exponent α such that p^α divides N.

Legendre's formula for the p-adic valuation of factorials implies

$$v_p(n!) = \frac{1}{p-1}\big(n - s_p(n)\big),$$

where $s_p(N)$ denotes the *p-ary sum-of-digits function*

$$s_p(N) = \sum_{j\ge 0} \varepsilon_j(N),$$

with $\varepsilon_j(N)$ denoting the j-th digit in the p-adic representation of N.

Hence, we have

$$v_p\left(\frac{1}{n+1}\binom{2n}{n}\right) = \frac{1}{p-1}\big(2s_p(n) - s_p(2n)\big) - v_p(n+1).$$

How to “do this properly”

Hence, we have

$$v_p\left(\frac{1}{n+1}\binom{2n}{n}\right)=\frac{1}{p-1}\big(2s_p(n)-s_p(2n)\big)-v_p(n+1).$$

Hence, we have

$$v_p\left(\frac{1}{n+1}\binom{2n}{n}\right)=\frac{1}{p-1}\big(2s_p(n)-s_p(2n)\big)-v_p(n+1).$$

We see that, in order to prove that $v_p\left(\frac{1}{n+1}\binom{2n}{n}\right)$ “becomes large” for most n (and the same for similar — “Catalan-like” — sequences), we need sufficiently precise results on the distribution of linear combinations of the form

$$c_1 s_q(A_1 n)+c_2 s_q(A_2 n)+\cdots+c_d s_q(A_d n),\qquad n<N,$$

with real numbers c_j and integers $A_j\ge 1$, $1\le j\le d$.

How to "do this properly"

Hence, we have

$$v_p\left(\frac{1}{n+1}\binom{2n}{n}\right) = \frac{1}{p-1}\left(2s_p(n) - s_p(2n)\right) - v_p(n+1).$$

We see that, in order to prove that $v_p\left(\frac{1}{n+1}\binom{2n}{n}\right)$ "becomes large" for most n (and the same for similar — "Catalan-like" — sequences), we need sufficiently precise results on the distribution of linear combinations of the form

$$c_1 s_q(A_1 n) + c_2 s_q(A_2 n) + \cdots + c_d s_q(A_d n), \qquad n < N,$$

with real numbers c_j and integers $A_j \geq 1$, $1 \leq j \leq d$.

Equivalently, we need sufficiently precise results on the distribution of the vector

$$(s_q(A_1 n), s_q(A_2 n), \ldots, s_q(A_d n)), \qquad n < N.$$

The general divisibility result

Theorem

Let p be a given prime number, α a positive integer, $P(n)$ a polynomial in n with integer coefficients, and $(C_i)_{1\le i\le r}$, $(D_i)_{1\le i\le s}$, $(E_i)_{1\le i\le t}$, $(F_i)_{1\le i\le t}$ given integer sequences with $C_i, D_i > 0$ and $p \nmid \gcd(E_i, F_i)$ for all i, $\sum_{i=1}^r C_i = \sum_{i=1}^s D_i$, and $\{C_i : 1 \le i \le r\} \ne \{D_i : 1 \le i \le s\}$. If all elements of the sequence $\big(S(n)\big)_{n\ge 0}$, defined by

$$S(n) := \frac{P(n)}{\prod_{i=1}^t (E_i n + F_i)} \frac{\prod_{i=1}^r (C_i n)!}{\prod_{i=1}^s (D_i n)!},$$

are integers, then

$$\lim_{N\to\infty} \frac{1}{N} \# \{n < N : S(n) \equiv 0 \pmod{p^\alpha}\} = 1.$$

The general divisibility result

Corollary

Let m be a positive integer, $P(n)$ a polynomial in n with integer coefficients, and $(C_i)_{1\le i\le r}$, $(D_i)_{1\le i\le s}$, $(E_i)_{1\le i\le t}$, $(F_i)_{1\le i\le t}$ given integer sequences with $C_i, D_i > 0$ and $p \nmid \gcd(E_i, F_i)$ for all i and primes p dividing m, $\sum_{i=1}^{r} C_i = \sum_{i=1}^{s} D_i$, and $\{C_i : 1 \le i \le r\} \ne \{D_i : 1 \le i \le s\}$. If all elements of the sequence $\big(S(n)\big)_{n\ge 0}$, defined by

$$S(n) := \frac{P(n)}{\prod_{i=1}^{t}(E_i n + F_i)} \frac{\prod_{i=1}^{r}(C_i n)!}{\prod_{i=1}^{s}(D_i n)!},$$

are integers, then

$$\lim_{N\to\infty} \frac{1}{N} \# \{n < N : S(n) \equiv 0 \pmod{m}\} = 1.$$

The general divisibility result

This theorem covers:

The general divisibility result

This theorem covers:

(1) Binomial coefficients such as the central binomial coefficients $\binom{2n}{n}$, or more generally $\binom{(a+b)n}{an}$ for positive integers a and b, including variations such as $\binom{2n}{n-1}$, etc.

The general divisibility result

This theorem covers:

(1) Binomial coefficients such as the central binomial coefficients $\binom{2n}{n}$, or more generally $\binom{(a+b)n}{an}$ for positive integers a and b, including variations such as $\binom{2n}{n-1}$, etc.

(2) Multinomial coefficients such as $\frac{((a_1+a_2+\cdots+a_s)n)!}{(a_1n)!\,(a_2n)!\cdots(a_sn)!}$, etc.

The general divisibility result

This theorem covers:

(1) Binomial coefficients such as the central binomial coefficients $\binom{2n}{n}$, or more generally $\binom{(a+b)n}{an}$ for positive integers a and b, including variations such as $\binom{2n}{n-1}$, etc.

(2) Multinomial coefficients such as $\frac{((a_1+a_2+\cdots+a_s)n)!}{(a_1 n)!\,(a_2 n)!\cdots(a_s n)!}$, etc.

(3) Fuß–Catalan numbers. These are defined by $\frac{1}{n}\binom{(m+1)n}{n-1}$, where m is a given positive integer.

The general divisibility result

This theorem covers:

(1) Binomial coefficients such as the central binomial coefficients $\binom{2n}{n}$, or more generally $\binom{(a+b)n}{an}$ for positive integers a and b, including variations such as $\binom{2n}{n-1}$, etc.

(2) Multinomial coefficients such as $\frac{((a_1+a_2+\cdots+a_s)n)!}{(a_1 n)!\,(a_2 n)!\cdots(a_s n)!}$, etc.

(3) Fuß–Catalan numbers. These are defined by $\frac{1}{n}\binom{(m+1)n}{n-1}$, where m is a given positive integer.

(4) Gessel's super ballot numbers (often also called super-Catalan numbers) $\frac{(2n)!\,(2m)!}{n!\,m!\,(m+n)!}$ for non-negative integers m, or for $m = an$ with a a positive integer.

The general divisibility result

This theorem covers:

(1) Binomial coefficients such as the central binomial coefficients $\binom{2n}{n}$, or more generally $\binom{(a+b)n}{an}$ for positive integers a and b, including variations such as $\binom{2n}{n-1}$, etc.

(2) Multinomial coefficients such as $\frac{((a_1+a_2+\dots+a_s)n)!}{(a_1 n)!\,(a_2 n)!\cdots(a_s n)!}$, etc.

(3) Fuß–Catalan numbers. These are defined by $\frac{1}{n}\binom{(m+1)n}{n-1}$, where m is a given positive integer.

(4) Gessel's super ballot numbers (often also called super-Catalan numbers) $\frac{(2n)!\,(2m)!}{n!\,m!\,(m+n)!}$ for non-negative integers m, or for $m = an$ with a a positive integer.

(5) Many counting sequences in tree and map enumeration such as $\frac{m+1}{n((m-1)n+2)}\binom{mn}{n-1}$, $\frac{2\cdot 3^n}{(n+2)(n+1)}\binom{2n}{n}$, $\frac{2}{(3n-1)(3n-2)}\binom{3n-1}{n}$, $\frac{2}{(3n+1)(n+1)}\binom{4n+1}{n}$, $\frac{1}{2(n+2)(n+1)}\binom{2n}{n}\binom{2n+2}{n+1}$.

The actual main result

Theorem (Central limit theorem)

Let $q \geq 2$ be an integer, and let $A_1, A_2, \ldots, A_d$ be positive integers. Then the vector

$$(s_q(A_1 n), s_q(A_2 n), \ldots, s_q(A_d n)), \qquad 0 \leq n < N,$$

satisfies a d-dimensional central limit theorem with asymptotic mean vector $((q-1)/2, \ldots, (q-1)/2) \cdot \log_q N$ and asymptotic covariance matrix $\Sigma \cdot \log_q N$, where Σ is positive semi-definite. If we further assume that q is prime and that the integers $A_1, A_2, \ldots, A_d$ are not divisible by q, then Σ is explicitly given by

$$\Sigma = \left(\frac{(q^2-1)}{12} \frac{\gcd(A_i, A_j)^2}{A_i A_j} \right)_{1 \leq i,j \leq d}.$$

The actual main result

Theorem (Central limit theorem)

Let $q \geq 2$ be an integer, and let $A_1, A_2, \ldots, A_d$ be positive integers. Then the vector

$$(s_q(A_1 n), s_q(A_2 n), \ldots, s_q(A_d n)), \qquad 0 \leq n < N,$$

satisfies a d-dimensional central limit theorem with asymptotic mean vector $((q-1)/2, \ldots, (q-1)/2) \cdot \log_q N$ and asymptotic covariance matrix $\Sigma \cdot \log_q N$, where Σ is positive semi-definite. If we further assume that q is prime and that the integers $A_1, A_2, \ldots, A_d$ are not divisible by q, then Σ is explicitly given by

$$\Sigma = \left(\frac{(q^2-1)}{12} \frac{\gcd(A_i, A_j)^2}{A_i A_j} \right)_{1 \leq i,j \leq d}.$$

The actual main result

Theorem (Central limit theorem)

Let $q \geq 2$ be an integer, and let $A_1, A_2, \dots, A_d$ be positive integers. Then the vector

$$(s_q(A_1 n), s_q(A_2 n), \dots, s_q(A_d n)), \qquad 0 \leq n < N,$$

satisfies a d-dimensional central limit theorem with asymptotic mean vector $((q-1)/2, \dots, (q-1)/2) \cdot \log_q N$ and asymptotic covariance matrix $\Sigma \cdot \log_q N$, where Σ is positive semi-definite. If we further assume that q is prime and that the integers $A_1, A_2, \dots, A_d$ are not divisible by q, then Σ is explicitly given by

$$\Sigma = \left(\frac{(q^2-1)}{12} \frac{\gcd(A_i, A_j)^2}{A_i A_j} \right)_{1 \leq i,j \leq d}.$$

For $q = 2$, this had been proved earlier by (Johannes) Schmid and (Wolfgang) Schmidt, independently.

The actual main result

What goes into the proof?

The actual main result

What goes into the proof?

– One shows that $f(n) = s_q(An)$, with A a positive integer, is a *q-quasi-additive function*, meaning that there exists $r \geq 0$ such that

$$f(q^{k+r}a + b) = f(a) + f(b) \qquad \text{for all } b < q^k.$$

The actual main result

What goes into the proof?

– One shows that $f(n) = s_q(An)$, with A a positive integer, is a *q-quasi-additive function*, meaning that there exists $r \geq 0$ such that

$$f(q^{k+r}a + b) = f(a) + f(b) \qquad \text{for all } b < q^k.$$

– Kropf and Wagner had shown that a q-quasi-additive function $f(n)$ of at most logarithmic growth satisfies a central limit theorem of the form

$$\frac{1}{N}\#\left\{n < N : f(n) \leq \mu \log_q N + t\sqrt{\sigma^2 \log_q N}\right\} = \Phi(t) + o(1),$$

where $\Phi(t)$ denotes the distribution function of the standard Gaußian distribution, for appropriate constants μ and σ^2. This implies the claim about the limit law and its expectation.

The actual main result

What goes into the proof?

– One shows that $f(n) = s_q(An)$, with A a positive integer, is a *q-quasi-additive function*, meaning that there exists $r \geq 0$ such that

$$f(q^{k+r}a + b) = f(a) + f(b) \qquad \text{for all } b < q^k.$$

– Kropf and Wagner had shown that a q-quasi-additive function $f(n)$ of at most logarithmic growth satisfies a central limit theorem of the form

$$\frac{1}{N}\#\left\{n < N : f(n) \leq \mu \log_q N + t\sqrt{\sigma^2 \log_q N}\right\} = \Phi(t) + o(1),$$

where $\Phi(t)$ denotes the distribution function of the standard Gaußian distribution, for appropriate constants μ and σ^2. This implies the claim about the limit law and its expectation.

– For the variance, one has to do a nasty calculation involving exponential sums.

Main ingredients of the proof of the divisibility result

Main ingredients of the proof of the divisibility result

Theorem

Let p be a given prime number, α a positive integer, $P(n)$ a polynomial in n with integer coefficients, and $(C_i)_{1\le i\le r}$, $(D_i)_{1\le i\le s}$, $(E_i)_{1\le i\le t}$, $(F_i)_{1\le i\le t}$ given integer sequences with $C_i, D_i > 0$ and $p \nmid \gcd(E_i, F_i)$ for all i, $\sum_{i=1}^r C_i = \sum_{i=1}^s D_i$, and $\{C_i : 1 \le i \le r\} \ne \{D_i : 1 \le i \le s\}$. If all elements of the sequence $\big(S(n)\big)_{n\ge 0}$, defined by

$$S(n) := \frac{P(n)}{\prod_{i=1}^t (E_i n + F_i)} \frac{\prod_{i=1}^r (C_i n)!}{\prod_{i=1}^s (D_i n)!},$$

are integers, then

$$\lim_{N\to\infty} \frac{1}{N} \# \{n < N : S(n) \equiv 0 \pmod{p^\alpha}\} = 1.$$

Main ingredients of the proof of the divisibility result

Here is our sequence:

$$S(n) := \frac{P(n)}{\prod_{i=1}^{t}(E_i n + F_i)} \frac{\prod_{i=1}^{r}(C_i n)!}{\prod_{i=1}^{s}(D_i n)!}.$$

Main ingredients of the proof of the divisibility result

Here is our sequence:

$$S(n) := \frac{P(n)}{\prod_{i=1}^{t}(E_i n + F_i)} \frac{\prod_{i=1}^{r}(C_i n)!}{\prod_{i=1}^{s}(D_i n)!}.$$

We have to consider

$$\begin{aligned} v_p\big(S(n)\big) &= v_p\big(P(n)\big) - \sum_{i=1}^{t} v_p(E_i n + F_i) + \sum_{i=1}^{r} v_p\big((C_i n)!\big) \\ &\qquad - \sum_{i=1}^{s} v_p\big((D_i n)!\big) \\ &\ge -\sum_{i=1}^{t} v_p(E_i n + F_i) - \frac{1}{p-1}\sum_{i=1}^{r} s_p(C_i n) + \frac{1}{p-1}\sum_{i=1}^{s} s_p(D_i n). \end{aligned}$$

Main ingredients of the proof of the divisibility result

$$v_p\big(S(n)\big) \ge -\sum_{i=1}^{t} v_p(E_i n + F_i) - \frac{1}{p-1}\sum_{i=1}^{r} s_p(C_i n) + \frac{1}{p-1}\sum_{i=1}^{s} s_p(D_i n).$$

Main ingredients of the proof of the divisibility result

$$v_p\big(S(n)\big) \ge -\sum_{i=1}^{t} v_p(E_i n + F_i) \\ - \frac{1}{p-1}\sum_{i=1}^{r} s_p(C_i n) + \frac{1}{p-1}\sum_{i=1}^{s} s_p(D_i n).$$

– It follows from an analysis of Bober (using Landau's criterion) that, if $S(n)$ is integral for all n, then $r < s$.

Main ingredients of the proof of the divisibility result

$$v_p\big(S(n)\big) \ge -\sum_{i=1}^{t} v_p(E_i n + F_i) - \frac{1}{p-1}\sum_{i=1}^{r} s_p(C_i n) + \frac{1}{p-1}\sum_{i=1}^{s} s_p(D_i n).$$

– It follows from an analysis of Bober (using Landau's criterion) that, if $S(n)$ is integral for all n, then $r < s$.

– One shows furthermore that, if $v_p(En + B)$ is considered as a random variable for n in the integer interval $[0, N-1]$, then

$$\mathbf{E}_N\big(v_p(En+F)\big) = \begin{cases} 0, & \text{if } p \mid E, \\ \frac{1}{p-1} + o(1), & \text{if } p \nmid E, \end{cases} \qquad \text{as } N \to \infty,$$

and

$$\mathbf{Var}_N\big(v_p(En+F)\big) = \begin{cases} 0, & \text{if } p \mid E, \\ \frac{p}{(p-1)^2} + o(1), & \text{if } p \nmid E, \end{cases} \qquad \text{as } N \to \infty.$$

Main ingredients of the proof of the divisibility result

$$v_p\big(S(n)\big) \ge -\sum_{i=1}^{t} v_p(E_i n + F_i) - \frac{1}{p-1}\sum_{i=1}^{r} s_p(C_i n) + \frac{1}{p-1}\sum_{i=1}^{s} s_p(D_i n).$$

Main ingredients of the proof of the divisibility result

$$v_p\big(S(n)\big) \ge -\sum_{i=1}^{t} v_p(E_i n + F_i) - \frac{1}{p-1}\sum_{i=1}^{r} s_p(C_i n) + \frac{1}{p-1}\sum_{i=1}^{s} s_p(D_i n).$$

Let $T(n)$ denote the right-hand side of the inequality. From the previous considerations it follows that

$$\mathbf{E}_N\big(T(n)\big) = \Omega\big(\log_p(N)\big), \qquad \text{as } N \to \infty$$

and

$$\mathbf{Var}_N\big(T(n)\big) = O\big(\log_p(N)\big), \qquad \text{as } N \to \infty$$

Main ingredients of the proof of the divisibility result

$$v_p\big(S(n)\big) \geq -\sum_{i=1}^{t} v_p(E_i n + F_i)$$
$$-\frac{1}{p-1}\sum_{i=1}^{r} s_p(C_i n) + \frac{1}{p-1}\sum_{i=1}^{s} s_p(D_i n).$$

Let $T(n)$ denote the right-hand side of the inequality. From the previous considerations it follows that

$$\mathbf{E}_N\big(T(n)\big) = \Omega\big(\log_p(N)\big), \qquad \text{as } N \to \infty$$

and

$$\mathbf{Var}_N\big(T(n)\big) = O\big(\log_p(N)\big), \qquad \text{as } N \to \infty$$

Chebyshev's inequality

$$\mathbf{P}\big(|X - \mathbf{E}(X)| < \varepsilon\big) > 1 - \frac{1}{\varepsilon^2}\mathbf{Var}(X).$$

with $\varepsilon = \big(\log_p(n)\big)^{3/4}$ and $X = T(n)$ then finishes the argument.

The general divisibility result

Theorem

Let p be a given prime number, α a positive integer, $P(n)$ a polynomial in n with integer coefficients, and $(C_i)_{1\le i\le r}$, $(D_i)_{1\le i\le s}$, $(E_i)_{1\le i\le t}$, $(F_i)_{1\le i\le t}$ given integer sequences with $C_i, D_i > 0$ and $p \nmid \gcd(E_i, F_i)$ for all i, $\sum_{i=1}^r C_i = \sum_{i=1}^s D_i$, and $\{C_i : 1 \le i \le r\} \ne \{D_i : 1 \le i \le s\}$. If all elements of the sequence $\big(S(n)\big)_{n\ge 0}$, defined by

$$S(n) := \frac{P(n)}{\prod_{i=1}^t (E_i n + F_i)} \frac{\prod_{i=1}^r (C_i n)!}{\prod_{i=1}^s (D_i n)!},$$

are integers, then

$$\lim_{N\to\infty} \frac{1}{N} \# \{n < N : S(n) \equiv 0 \pmod{p^\alpha}\} = 1.$$