Subgroup growth and zeta functions

B. Sury

Let G be a group, and let $a_n(G) = |\{H \leq G : [G : H] = n\}|$.
We want to study the arithmetic function $n \mapsto a_n(G)$, or the function $n \mapsto s_n(G) = \Sigma_{m \leq n} a_m(G)$. This is called the growth function of G. Notice that $a_n < \infty$ when G is finitely generated, but this is not necessary.

The concept was inspired by the 'word growth' of Gromov et al. In word growth, one studies the following function. If G is generated by a finite set $\Sigma, b_n = |\{g : l_\Sigma(g) \leq n\}|$. We will say more about this a little later.

For any function $f : \mathbb{N} \to \mathbb{N}$, the group $G = \Pi_p(\mathbb{Z}/p)^{1+\log_p f(p)}$ satisfies $f(n) \leq s_n(G) < \infty$.

Hence, it is better to restrict our study to finitely generated groups to get interesting results. However, it should be kept in mind that there are examples of groups which are not finitely generated but the growth of a_n is like a polynomial and a major open problem is to characterise these.

Subgroup growth is a subject that took off in a major way about 10 years ago. It has unexpected connections with subjects like model theory as well as with the traditional ones like number theory, algebraic and arithmetic groups, classification of finite simple groups, combinatorics and probabilistic number theory. Recently, deep analytic number-theoretic methods have come into play.

The first result of the subject is that of Marshall Hall in 1949. He obtained a recursive formula for $a_n(G)$ when G is a free group of rank at least 2. From that formula, one concludes that $a_n \sim n(n!)^{r-1}$ where r is the rank. This growth is sometimes called superexponential. Since each subgroup of index n corresponds to a transitive, r-generated subgroup of S_n, it follows that 'most' r-generated subgroups of S_n are transitive.

Note that, for any group $G, a_n(G) = a_n(G/R(G))$ where $R(G) = \cap\{H \leq G : [G : H] < \infty\}$. Thus, for our study we may assume without loss of generality

that the group G is *residually finite* i.e., $R(G) = \{1\}$. It is well-known (and easy to prove) that all finitely generated, linear groups are residually finite.

Three questions that obviously arise:
Q1. What are the possible growth types?
Q2. Given G, what is its growth type?
Q3. Given a type of growth, which groups have it?

(We say that the growth type is a function $f(n)$ if, $a_n \leq f(n)^a$ for all n, and $a_n > f(n)^b$ for infinitely many n, for some constants $a, b > 0$.)
Since any d-generated group G is an image of the free group F_d, its growth is at most as fast as that of F_d. So, the fastest possible growth type is the superexponential type n^n.

On the other hand, for finite groups, s_n is eventually constant. For $\mathbb{Z}$, $s_n = n$. One might ask whether this is the slowest possible growth type for infinite groups. The answer is 'yes' but is surprisingly hard to prove and uses what is thought of as one of the major results in the subject so far:
We say that G has polynomial subgroup growth (**PSG**) if there is a constant $c > 0$ so that $s_n(G) \leq n^c$ for all n.

Theorem (Lubotzky-Mann-Segal)
A finitely generated, residually finite group G has PSG if, and only if, it is virtually solvable, and of finite rank. Equivalently, these things happen if, and only if, G is virtually solvable and is linear over $\mathbf{Q}$.
Finite rank means (even for infinitely generated groups) that there is a constant $d > 0$ such that all finitely generated subgroups are d-generated. A property holds virtually if a subgroup of finite index has that property.

As we saw, the growth type n^n is the fastest possible one for finitely generated groups. The theorem shows, in particular, that the slowest possible type for infinite groups is the type n. For, if G is a counterexample to the assertion, it must have PSG. Therefore, it must be virtually solvable and will have a subgroup H of finite index which has a quotient isomorphic to $\mathbf{Z}$. If $[G : H] = d$, then $s_n(G) \geq s_{[n/d]}(H) \geq [n/d]$ for every n.

The Lubotzky-Mann-Segal theorem was first proved by Lubotzky & Mann (Inventiones Math. 104, 1991) for the main case of linear groups over characteristic 0 fields. The main point of the proof is that for finitely generated,

linear groups which are not virtually solvable, $a_n \geq n^{c \log n / \log \log n}$ for some constant $c > 0$. In particular, this implies that there is a gap in the possible growths of finitely generated, linear groups. We shall prove it shortly.

Zeta functions
For a group G such that $a_n(G) < \infty$, let us define its zeta function as $\zeta_G(s) = \sum_{n=1}^{\infty} \frac{a_n(G)}{n^s}$. This was introduced for nilpotent groups in 1988 by Grunewald, Segal and Smith. For instance, $\zeta_{\mathbb{Z}}$ is the Riemann Zeta function.

For nilpotent G, the zeta function has some nice features:
(i) G has PSG - equivalently, $\zeta_G(s)$ actually converges in some right half plane,
(ii) $\zeta_G(s) = \Pi_p \zeta_{G,p}(s)$, where $\zeta_{G,p}(s) = \sum_{i \geq 0} a_{p^i}(G) p^{-is}$. Thus, ζ_G has an Euler product decomposition - this restates the fact that a finite nilpotent group is the product of its Sylow subgroups.

Thus, for nilpotent G, the zeta function shares some of the nice properties of the Riemann (or the Dedekind) zeta function.

Remarks
(i) We could also study the zeta function which counts the subgroups of a given index isomorphic to the whole group or the zeta function that counts normal subgroups of a given index. It is still unclear which might be the best analogue.
(ii) For general finitely generated G, it turns out that the analogy of ζ_G with Dedekind zeta functions is too simplistic and very recently, it has been revealed that a better analogy is with the Weil zeta function of an algebraic variety over $\mathbf{Z}$. We shall discuss this in detail later.
(iii) There are examples of non-nilpotent groups whose zeta function has an Euler product expansion.

Theorem

$$\zeta_{\mathbb{Z}^r}(s) = \zeta(s)\zeta(s-1)\cdots\zeta(s-r+1)$$

It would be difficult to show without using the zeta function that $s_n(\mathbb{Z}^2) \sim \pi^2 n/12$.
Several proofs of the above theorem are known. Here is one.

Proof. If $H \leq \mathbb{Z}^r, [\mathbb{Z}^r : H] < \infty$, then $H = g\mathbb{Z}^r$, for some $g \in M_r(\mathbb{Z}) \cap$

$GL_r(\mathbb{Q})$. Moreover, $g_1\mathbb{Z}^r = g_2\mathbb{Z}^r \Leftrightarrow g_1^{-1}g_2 \in GL_r(\mathbb{Z})$. Therefore, $\zeta_{\mathbb{Z}^r}(s) = \sum_{g\in C} |\det g|^{-s}$ where C is the set $M_r(\mathbb{Z}) \cap GL_r(Q)/GL_r(\mathbb{Z})$. One can take for C, the set of lower triangular matrices where the entries a_{ij} are nonnegative integers satisfying $a_{ij} < a_{ii}$ for all $i > j$, and with $a_{ii} \geq 1$. A simple counting gives the expression in the theorem.

(Curious) **Corollary**

$$np(n) = \sum_{i=1}^{n-1} \sigma(i)p(n-i) + \sigma(n)$$

Proof. First, we observe for an arbitrary group G that $a_n = t_n/(n-1)!$, where t_n is the number of transitive actions of G on $\{1, 2, \cdots, n\}$. Further, if $h_n = |\text{Hom }(G, S_n)|$, then one has the relation

$$h_n = \sum_{k=1}^{n-1} \binom{n-1}{k-1} t_k h_{n-k} + t_n$$

(for each $1 \leq k \leq n$, there are $\binom{n-1}{k-1}$ ways to choose the orbit of 1, t_k transitive actions on it, and h_{n-k} actions on its complement).
Rewriting the relation in terms of the a_n, one has

$$a_n = h_n/(n-1)! - \sum_{k=1}^{n-1} \frac{h_{n-k}}{(n-k)!} a_k \cdots (\spadesuit)$$

Note that for $G = F_r$, this give Hall's formula

$$a_n(F_r) = n(n!)^{r-1} - \sum_{k=1}^{n-1} (n-k)!^{r-1} a_k(F_r)$$

Let us apply this to our case which is $\mathbb{Z}^2$; one has $h_n(\mathbb{Z}^2) = n!p(n)$ (x can be arbitrarily chosen in S_n, and y chosen in its centraliser $C_{S_n}(x)$, so that $h_n = \sum_x |C_{S_n}(x)| = |S_n| \sum 1/|[x]| = |S_n||[x]|1/|[x]| = |S_n|p(n)$.)
Also, $\zeta(s)\zeta(s-1) = \sum \frac{\sigma(n)}{n^s}$ so that $a_n = \sigma(n)$. Now, the equation $\spadesuit$ is precisely the assetion of the corollary.

The Zeta function for a nonabelian group
One can calculate ζ_G for the Heisenberg group G consisting of the matrices

$\begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix}$ with $a, b, c \in \mathbb{Z}$. It turns out that

$$\zeta_G(s) = \frac{\zeta(s)\zeta(s-1)\zeta(2s-2)\zeta(2s-3)}{\zeta(3s-3)}$$

It is curious to recall Ramanujan's formula

$$\sum \frac{\sigma_a(n)\sigma_b(n)}{n^s} = \frac{\zeta(s)\zeta(s-a)\zeta(s-b)\zeta(s-a-b)}{\zeta(2s-a-b)}$$

(I don't know whether there is a connection.)

Because of the double pole at $s = 2$, we get for this Heisenberg group that $s_n \sim \frac{\zeta(2)^2 n^2 logn}{2\zeta(3)}$.

Very recently, Grunewald & duSautoy have proved the following remarkable theorem using certain tauberian theorems:

Theorem. *Let G be a finitely generated nilpotent infinite group. Then,*
(i) the abscissa of convergence $\alpha(G)$ of $\zeta_G(s)$ is a rational number and $\zeta_G(s)$ can be meromorphically continued to $Re(s) > \alpha(G) - \delta$ for some $\delta > 0$. The continued function is holomorphic on the line $Re(s) = \alpha(G)$ except for a pole at $s = \alpha(G)$;
(ii) there exist a nonnegative integer $b(G)$ and real numbers c, c' such that

$$s_n \sim cn^{\alpha(G)}(logn)^{b(G)}$$
$$s_n^{\alpha(G)} \sim c'(logn)^{b(G)+1}$$

as $n \to \infty$.

This also uses the following earlier result of duSautoy:

Theorem. *Let G be finitely generated, nilpotent, torsion-free. Then, for any prime $p, \zeta_{G,p}(s) = \Sigma_{n \geq 0} a_{p^n}(G) p^{-ns}$ is a ration function of p^{-s}.*

This implies, in particular, the sequence $\{a_{p^n}\}$ satisfies a linear recurrence. The proof uses model theory in an essential manner.

The proof of the LMS-theorem
Now, we talk in some detail about the proof of the theorem that finitely generated, residually finite PSG groups are exactly the virtually solvable groups

of finite rank. We notice that the class of finitely generated, residually finite groups is a rich one, including in it all finitely generated, linear groups. Moreover, many results which fail for abstract, finitely generated groups hold good for the residually finite ones among them. A glaring example is the Burnside problem which was solved in the affirmative by Efim Zelmanov: *A finitely generated, residually finite group of finite exponent is finite.* Residual finiteness is equivalent to the profinite topology being Hausdorff. One completes G with respect to this topology to get what is called the profinite completion $\hat{G}$ of G. Similarly, one can have other kinds of completions like the pro-p completion $G^{(p)}$ for a prime p. The latter topology is defined by taking the subgroups of p-power index to be a fundamental system of neighbourhoods of the identity. One has $G \leq \hat{G}$ if G is residually finite and $G \leq G^{(p)}$ if G is residually-p.

For a profinite group, by finite generation one means topological finite generation, and one measures the growth of open subgroups. PSG has the obvious meaning. It is evident that $a_n(G) = a_n(\hat{G})$. Hence if G has PSG, then so does $\hat{G}$ as a profinite group. The proof of the LMS-theorem uses the following characterisation of p-adic Lie groups which is of independent interest.

Theorem. *A finitely generated, pro-p group G is a p-adic Lie group if, and only if, G has PSG.*

The LMS-proof proceeds by first reducing to the case of residually-p groups. On using the above theorem, it follows that $G \leq G^{(p)} \leq GL_n(\mathbb{Q}_p) \leq GL_n(\mathbb{C})$. By standard algebraic group-theoretic arguments, one further reduces to the case of a finitely generated subgroup Γ of $GL_n(\mathbb{Z}_s)$ for some finite set S of primes. The main part of the theorem is to prove the theorem for such groups Γ. *The idea is to show that if Γ is not virtually solvable, even the subclass of 'congruence subgroups' of a given index grows faster than polynomially.* More precisely, if Γ is not virtually solvable, its Zariski closure G is a non-solvable algebraic group. A deep theorem of Nori and, independently, Mathews, Vaserstein & Weisfeiler implies that Γ has strong approximation i.e., the closure of Γ in the S-congruence topology of $G(\mathbb{Z}_S)$ is open in $\Pi_{p \notin S} G(\mathbb{Z}_p)$. Equivalently, the closure of Γ in $G(\mathbb{Z}_S)$ is a subgroup Γ_0 of finite index in $G(\mathbb{Z}_S)$. From this, one gets $s_n(\Gamma) = \Sigma_{m \leq n} a_m(\Gamma)$ = subgroups of index at most n in $\Gamma \geq c_n(\Gamma_0)$ = congruence subgroups of index at most n in Γ_0. It should be noted that the classification of finite simple groups is used in the

proof of the above strong approximation result.

Main claim. $c_n(G(\mathbb{Z}_s)) \geq n^{c \log n/(\log\log n)^2}$ for some $c > 0$.

In the proof of the claim, we will use the prime number theorem. As a matter of fact, one has the following stronger result using stronger versions of the prime number theorem in arithmetic progressions. The proof by Lubotzky had a gap which has been filled recently by him and Goldfeld.

Let Γ be a finitely generated linear group. Then, either Γ has PSG or there exists $c > 0$ such that $s_n(\Gamma) \geq n^{clogn/loglogn}$.

Proof of the main claim.
We call $R = \mathbb{Z}_S$ for simplicity. For any positive integer m, denote by $G(m)$ the congruence subgroup mod m i.e., the kernel of the map $G(R) \to G(R/mR)$. Strong approximation means that $G(R)$ surjects onto $G(R/pR)$ for almost all primes p. Also, for almost all primes p, the finite group $G(R/pR)$ has even order (by Lang's theorem, the group has a split torus - this contributes a subgroup of order $p-1$). We augment S by adjoining these finitely many exceptional primes to S and call it S again. Let N be a large positive integer. By the prime number theorem, the number of primes outside S which are $\leq N$ is like $l = N/\log N$. Moreover, the prime number theorem says that $\Sigma_{p \notin S, p \leq N} \log p \sim N$. Hence $M = \prod_{p \notin S, p \leq N} p \sim e^N$. By the Chinese remainder theorem, $[G(R)/G(M)] = \Pi_{p/M}[G(R)/G(p)]$ contains an elementary abelian subgroup of order 2^l. But, an elementary abelian 2-group of rank l has at least $2^{l^2/4}$ subgroups of rank $[l/2]$. Thus, $G(R)$ has atleast $2^{l^2/4}$ congruence subgroups containing $G(M)$. Now, $[G : G(M)] \leq M^d$ where $d = \dim G$. Hence, $c_{M^d}(G(R)) \geq 2^{l^2/4}$. Rewriting this as

$$\frac{\log c_{M^d}}{\log(M^d)} = \frac{\log c_{M^d}}{dN} \geq \frac{l^2 \log 2}{4dN} \sim c\frac{N}{(\log N)^2}$$

this claim is proved.

Lubotzky proved that the growth of congruence subgroups in characteristic 0 satisfies the inequalities

$$n^{c_1 logn/loglogn} \leq c_n \leq n^{c_2 logn/loglogn}$$

This gives a very interesting group-theoretic characterisation of the arithmetic property called the congruence subgroup property (CSP) in characteristic 0. Indeed, if CSP does not hold good, then $s_n \geq n^{clogn}$ for some c and infinitely many n. In other words, the growth rate of all subgroups of finite index is strictly greater than the growth rate of the congruence of subgroups among them. This makes it possible to formulate an analogue of the CSP for non-arithmetic lattices. In particular, it is natural to conjecture that $s_n \geq n^{c \log n}$ for all lattices in $SO(n,1)$.

It is appropriate to point out here a rather important open problem (called 'quite a challenging problem' by Lubotzky, Pyber and Shalev):
• *Does there exist a gap in the growth of abstract, finitely generated groups? In particular, is it true that is* $a_n \geq n^{c \log n/(\log \log n)^2}$ *for infinitely many* n*, for such groups?*
We draw attention to the fact that Lubotzky, Pyber and Shalev have proved that there exist nonlinear finitely generated groups whose growth type is $n^{logn/(loglogn)^2}$. *Their example involves the profinite group* $\prod_{n \geq 5} A_n$ *which is topologically finitely generated.*

Remark. *For an arithmetic group* Γ *in positive characteristic, one only knows the weak inequality* $n^{c \log n} \leq s_n \leq n^{c'(\log n)^2}$.

It is not known (even for $SL_2(\mathbb{Z})$*) whether* $\lim_{n \to \infty} \frac{\log c_n}{(\log n)^2/\log \log n}$ *exists.*

Probabilistic methods.
Two major problems have been to characterise:
(i) profinite PSG groups,
(ii) abstract PSG groups which are not finitely generated. It is interesting to note that, in contrast to (ii) above, profinite PSG groups are automatically finitely generated. The proof of this involves two interesting new notions which turn out to be equivalent!

A profinite group G *is positively finitely generated (PFG) if the probability* $P(G,k) > 0$ *for some positive integer* k*. Here* $P(G,k)$ *is the measure of the set of* k*-tuples of* G *which generate* G*. The measure is relative to the* k*-fold product of the Haar measure of* G *normalised to be a probability measure. Analogous to PSG one can define PMSG - polynomial maximal subgroup growth. Obviously PSG* $\Rightarrow$ *PMSG.*

Lemma. $PMSG \Rightarrow PFG$.
Proof. *Let G be a profinite group with PMSG and let k be a positive integer. Since G has PMSG, there is $c > 0$ such that the number m_n of open maximal subgroups of index n is bounded by n^c. Now, a k-tuple of elements of G generates a proper subgroup if, and only if, these elements lie in a maximal subgroup. The probability of this is at most $\Sigma_{n\geq 2}\frac{m_n}{n^k} \leq \Sigma_{n\geq 2} n^{c-k}$. Evidently, for $k \geq c+2$, the sum is less than 1; so $P(G,k) > 0$.*

It turns out that the reverse implication is true although the proof is much deeper and uses the classification of finite simple groups. In fact, it is based on their result that a finite simple group has at most $n^{1.875+o(1)}$ maximal subgroups of index n.

Theorem *(Mann-Shalev). $PMSG \Leftrightarrow PFG$.*
Thus, $PSG \Rightarrow PMSG \Leftrightarrow PFG \Rightarrow f.g$ for a profinite group.

It can be shown that for an arithmetic group satisfying the CSP, in characteristic 0, the profinite completion has PFG. It is unknown whether this characterises the CSP.

Degree of growth.
For an abstract or profinite PSG group G, one can define $\deg(G) = \limsup \frac{\log a_n(G)}{\log n}$. *In other words,* $\deg(G)$ *is the 'smallest' positive real number c such that $a_n(G) = O(n^{c+\epsilon}) \forall\ \epsilon > 0$.*

The following theorems of Shalev rely heavily on some deep and delicate analytic number-theoretic results on the distribution of primes such as the fundamental lemma from sieve theory and Bombieri-type short intervals theorem.

Theorem *(Shalev).*
$\forall c \geq 1$, *there is a 2-generated profinite group G with PSG with* $\deg(G) = c$.
In fact, G can be taken to be a product of $PSL_2(\mathbb{F}_p)$ over some set of primes. Moreover, the degree cannot lie in (0,1).

Contrastingly,

Theorem *(Shalev).*
For a finitely generated, abstract group G with PSG,
(i) the degree cannot lie in the intervals (0,1), (1,3/2), (3/2,5/3).
(ii) If $H \leq G$ has finite index, then $\deg(H) \leq \deg(G) \leq \deg(H) + 1$.

This is unlike Gromov's word growth where commensurable groups will have the same degree of word growth. Moreover, unlike word growth, where the degree is always an integer, our degree could be $\notin \mathbb{Z}$. An example is the one we computed earlier viz, for the Heisenberg group whose degree is 3/2. It is not known whether the degree could be irrational and, whether the set of degrees forms a countable set. No formula is known for computing the degree for nilpotent groups. For word growth, a formula was given by Bass for the degree of nilpotent groups.

In the above-mentioned paper, Shalev also proves the following two results of independent interest:

Theorem.
Let G be a finitely generated residually finite group. Then,
$a_n(G) = O(n)$ if, and only if, G is virtually cyclic.
Moreover, if $a_n(G) = o(nloglogn)$, then G is virtually cyclic and the subgroup growth is at most linear. Therefore, there is a gap from growth cn to growth $cnloglogn$.

Theorem. *Let G be a finitely generated residually finite group. Then $deg(G) = 1$ if, and only if, G is equivalent to one of the following groups:*
(i) the dihedral group D_∞,
(ii) a plane crystallographic group which does not contain a 180 degree rotation,
(iii) $\mathbb{Z}[h, 1/h]. < h >$, where $h \in \mathbf{Q} \setminus \{0, 1, -1\}$,
(iv) $\mathbb{Z}[h, 1/h]. < h, -1 >$, where $h \in \mathbf{Q} \setminus \{0, 1, -1\}$.

An interesting computation on which the first theorem is based is:

Proposition. *Let K be a number field and $R \subset K$ a finitely generated subring with quotient field K. Let $U \leq R^*$ be a torsion-free subgroup of the group of units. Let $G = R.U$. Set $d = [K : \mathbf{Q}]$ and $r = d(U)$, the minimal number of generators of U. Then, $deg(G) = d + r - 1$.*

As mentioned earlier, using model-theoretic techniques, duSautoy proved the following deep result:

Theorem.
The zeta function of a compact p-adic Lie group is a rational function of p^{-s}. Moreover, the degree is always rational.

A 'challenging open question' is to:
• *Characterise non-f.g., PSG abstract groups.*
In this regard, Shalev has proved:
The following are equivalent for a residually finite (possibly infinitely generated) group G:
(i) $a_n(G) = o(n)$,
(ii) G has a central subgroup C of finite index whose finite quotients are all cyclic,
(iii) the sequence $\{a_n\}$ is bounded.

Immediately from this, we see that:
The degree of a PSG group cannot lie in $(0, 1)$.
Proof. *$deg(G) = limsup_{n\to\infty} \frac{loga_n}{logn}$. Thus, $\alpha \geq deg(G)$ if, and only if, $a_n(G) = O(n^{\alpha+\epsilon})$ for any $\epsilon > 0$. Clearly, if $deg(G) < 1$, then $a_n(G) = o(n)$ and so, by the implication (i) $\Rightarrow$ (iii), we have $deg(G) = 0$.*

In this regard, one has the following:
Question. *If G is a profinite group in which $a_n(G) < n$ for all sufficiently large n, does it follow that G has a procyclic central open subgroup?*

Finally, another open problem which is considered to be important is:
• *Characterise PSG profinite groups.*

We end with a rather curious (and not superficial) connection of subgroup growth (specificaly of the zeta function) with Mersenne primes viz.:

Theorem.
Let $\Delta = \Pi_p PSL_n(\mathbb{F}_p)$. (that this is a finitely generated profinite group is easy to see). Then,
(i) If $n \geq 3$ or if $n = 2, p \geq 3$, the zeta function $\zeta_{\Delta,p}(s)$ is a rational function of p^{-s}.
(ii) If $n = 2$, then $\zeta_{\Delta,2}(s)$ is a rational function of 2^{-s} if, and only if, there

are only finitely many Mersenne primes.

The reason that Mersenne primes enter the picture is that $PSL_2(\mathbf{F}_q)$ has a subgroup of index a power of 2 if, and only if, q is a Mersenne prime.

We summarise the discussion with some open questions in addition to the ones we have already mentioned:

• What are the possible degrees of PSG abstract groups? Do they form a countable set? Can a group have irrational degree?
• Is there a formula for the degree of a $f \cdot g.$ nilpotent group? For instance, is the degree the supremum over the degrees of the pro-p completions? Are the latter numbers (which are rational by duSautoy's theorem) constant along p in arithmetic progressions?