Probability Measures
and
Computational Complexity

Alexandre Borovik

Lyon, 14 April 2003

Motivation

- “Average case” complexity
- “Fast checks”: easy answers to most instances of hard problems

Applications:

Group theoretic cryptography

Quotient tests

We compute in a finitely presented group

$$G = \langle x_1, \ldots, x_m \mid r_1, \ldots, r_n \rangle = F/R.$$

Elements $x, y \in G$ are freely reduced words in the free group

$$F = \langle x_1, \ldots, x_m \rangle$$

with a routine checking whether

$$x = y \text{ in } G \Longleftrightarrow xy^{-1} \in R.$$

Quotient tests for identity in G:

work with a smaller factor group $H = F/Q$, $R < Q$, with a quicker identity routine:

$$x \notin Q \Longrightarrow x \notin R.$$

How one can measure the relative efficiency of a quotient test?

A case study

$G = B_n$ braid group given by $\binom{n}{2}$ "band generators" s_{ij}

Two tests: which is best?

$$B_n \longrightarrow \mathrm{Sym}_n \qquad s_{ij} \mapsto (ij)$$

$$B_n \longrightarrow \mathbb{Z} \qquad s_{ij} \mapsto 1$$

$$w \mapsto \begin{cases} \text{the sum of} \\ \text{exponents} \\ \text{of generators} \end{cases}$$

Given a "random" element

$$x \in F_{\frac{n(n-1)}{2}} = \langle x_1, \ldots, x_{\frac{n(n-1)}{2}} \rangle,$$

which is more likely:

$$x = 0 \ \text{ in } \ \mathbb{Z} \qquad \text{or} \qquad x = 1 \ \text{ in } \ \mathrm{Sym}_n?$$

Random element: produced by a lazy random walk of length k.

We need "laziness" to avoid the odd/even length disparity:

at every step, pause with probability $\frac{1}{n}$

Probability to hit 1:

$$r_k(\mathbb{Z}) \;\sim\; \frac{1}{\sqrt{2\pi k}} \cdot \left(1 - \frac{1}{n}\right)^{-1}$$

$$r_k(\mathrm{Sym}_n) \;\leqslant\; \frac{1}{n!} + ane^{-\frac{2k}{n}}$$

$a \approx 1$ does not depend on n

[Diaconis-Shashahani]

For $k \approx 1000$, $n \approx 100$ the test

$$B_n \longrightarrow \mathrm{Sym}_n$$

is incomparably better than

$$B_n \longrightarrow \mathbb{Z}$$

Testing conjugacy:

$p_k =$ probability for two random walks of length k to produce conjugate elements

$$p_k(\mathbb{Z}) \;\approx\; \frac{1}{2\sqrt{\pi k}}$$

$$p_k(\mathrm{Sym}_n) \;\approx\; \frac{1}{n^2} \qquad \textbf{[Babai and Pyber]}$$

Now the test

$$B_n \longrightarrow \mathrm{Sym}_n$$

is not better than

$$B_n \longrightarrow \mathbb{Z}$$

Some lessons:

- there is need for practical methods for engineering style, "back of the envelope" calculations
- even in simplest problems we are at the frontiers of existing theories
- "asymptotically best" does not mean "best"

On the free group F_m, we need

families of atomic probabilty measures of varying mean length of words.

- Should be natural in the context of computational group theory.
- Easy to study asymptotic behaviour when the mean word length tends to infinity.

- Easy, 'back of an envelope' estimates of probabilities of 'standard' subsets:

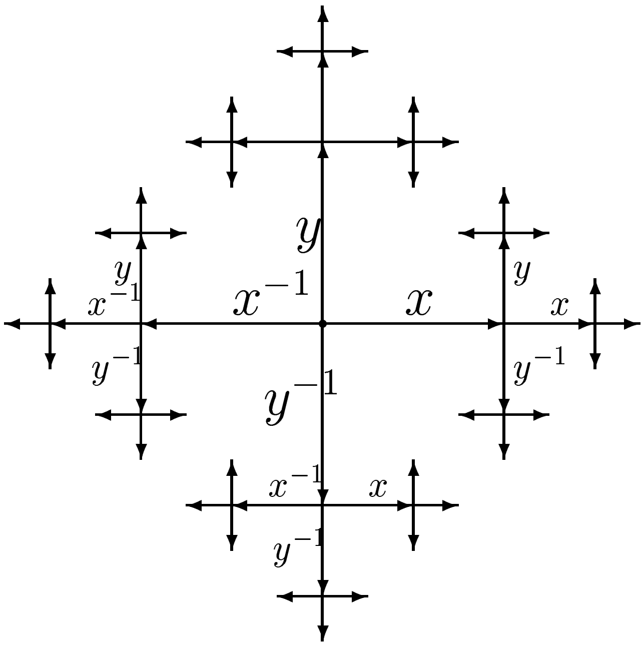
  - (normal) subgroups,
  - cosets with respect to subgroups,
  - conjugacy classes,
  - sets of words of special nature (say, squares or commutators).

- Many sets we wish to measure are context free languages.

- An important subclass: *regular* subsets (produced by deterministic finite automata):
  - finitely generated subgroups and their cosets;
  - **cones**: sets of all words which start with the given initial segment.
  - Closed under
    - Boolean operations,
    - translation by elements of F,
    - conjugation.

A natural choice:

Measures generated by random walks



A no-return random walk on the Cayley tree for the free group $F = \langle x, y \rangle$.

No return random walk on the tree as a random word generator

- Start at 1
  - either return 1 with probability s,
  - or move to one of the $2m$ adjacent vertices with probability

$$\frac{1-s}{2m}.$$

- When at a vertex $v \neq 1$,
  - either return v with probability s,
  - or move, with probability
$$\frac{1-s}{2m-1},$$
    to one of the $2m-1$ adjacent vertices lying away from 1.

In other words:

take random freely reduced words w of random lengths $|w|$ distributed by the exponential law

$$P(|w| = k) = s(1-s)^k.$$

We get the atomic probability measure

$$\mu_s(w) = \frac{s(1-s)^{|w|}}{2m \cdot (2m-1)^{|w|-1}} \quad \text{for } w \neq 1$$

$$\mu_s(1) = s$$

Mean length: L_s such that

$$s = \frac{1}{L_s + 1}$$

$$L_s \to \infty \Longleftrightarrow s \to 0$$

Renormalisation:

$$\mu_s^*(w) = \left(\frac{2m}{2m-1} \cdot \frac{1}{s}\right) \cdot \mu_s(w).$$

μ_s^* is **multiplicative**:

$$\mu_s^*(u \circ v) = \mu_s^*(u)\mu_s^*(v);$$

$u \circ v$ denotes the product such that $|uv| = |u| + |v|$ i.e. there is no cancellation.

μ_s is **almost multiplicative**:

$$\mu_s(u \circ v) = c\mu_s(u)\mu_s(v)$$

A similar approach, **Flajolet et al.:**

"Boltzmann Sampler" for random combinatorial objects:

If a thing A is made from things B and C then

$$p(A) = p(B) \cdot p(C)$$

For $R \subset F$, its **measure** $\mu(R)$ is the analytic function

$$\begin{aligned} \mu(R) : (0,1) &\to \mathbb{R} \\ s &\mapsto \mu_s(R) \end{aligned}$$

We are interested in its behaviour at

$$s \to 0^+$$

which is the same as the mean length

$$L_s \to \infty$$

Sphere $S_k = \{ w \mid |w| = k \}$

Relative frequencies of $R \subset F$**:**

$$f_k = \frac{|R \cap S_k|}{|S_k|}$$

Frequency measure:

$$\begin{aligned}
\lambda(w) &= \frac{1}{2m(2m-1)^{|w|-1}}, \quad \text{if} \quad w \neq 1, \\
\lambda(1) &= 1 \\
\lambda(R) &= \sum_{k=1}^{\infty} f_k(R)
\end{aligned}$$

Generating functions

Set $t = \mu_s^*(x_i)$ (x_i a generator)

Let $R \subset F \setminus \{1\}$, $\quad r_k = |R \cap S_k|$.

$$\mu^*(R) = \sum_{k=0}^{\infty} r_k t^k$$

the **generating function** for the number of elements of length k in R.

$$\mu(R) = s \sum_{k=0}^{\infty} f_k (1-s)^k$$

is essentially the generating function for the **relative frequencies**

$$f_k = \frac{|R \cap S_k|}{|S_k|}.$$

Linear approximation at $s = 0$:

$$\mu(R) = \mu_0 + \mu_1 s + o(s)$$

Hardy and Littlewood:

$$\mu_0(R) = \lim_{n\to\infty} \frac{1}{n} \left(f_1 + \cdots + f_n \right).$$

(the **Cesaro limit**)

If $H < F$ and $|F : H| < \infty$,

$$\mu_0(H) = \frac{1}{|F : H|}$$

For the subgroup H consisting of all elements of even length,

$$\{f_k\} = 1, 0, 1, 0, 1, 0, \ldots$$

$$\begin{aligned} \mu_0(H) &= \lim_{n\to\infty} \frac{1}{2n}\,(1 + 0 + \cdots + 1 + 0) \\ &= \frac{1}{2} \\ &= \frac{1}{|F : H|} \end{aligned}$$

Linear approximation at $s = 0$:

$$\mu(R) = \mu_0 + \mu_1 s + o(s)$$

Tauberian theorem by Hardy and Littlewood:

If $\mu_0(R) = 0$, then

$$\mu_1(R) \; = \; \sum_{k=1}^{\infty} f_k(R)$$

Linear approximation at $s = 0$:

$$\mu(R) = \mu_0 + \mu_1 s + o(s)$$

If the frequency measure

$$\lambda(R) = \sum_{k=1}^{\infty} f_k(R)$$

exists, then

$$\mu_0(R) = 0$$

and

$$\mu_1(R) = \lambda(R)$$

Classification of sets by size:

- *Thick*: $\mu_0(R)$ exists, $\mu_0(R) > 0$ and
$$\mu(R) = \mu_0(R) + o(1)$$
- *Thin*: $\mu_0(R) = 0$, $\mu_1(R) = \sum f_k$ exists and
$$\mu(R) = \mu_1(R)s + o(s)$$
- *Sparse*: $\mu_0(R) = 0$, $\mu_1(R) = \sum f_k$ exists and
$$\limsup \sqrt[k]{f_k} < 1$$
(exponential decline of f_k with a global bound)

- *Intermediate density* subsets:
  $\mu_0(R) = 0$ but $\mu_1(R)$ does not exist.

- *Singular* sets: $\mu_0(R)$ does not exist.

Degrees of polynomial growth

$$f : F \longrightarrow \mathbb{R}^{\geqslant 0}$$

The average value f on the sphere S_k:

$$E_k = \sum_{|w|=k} \frac{f(w)}{|S_k|}$$

$$E_k = \sum_{|w|=k} \frac{f(w)}{|S_k|}$$

For every fixed $s \in (0, 1)$, the mean value $E_f(s)$ of f with respect to μ_s:

$$\begin{aligned} E_f(s) &= \sum_{w \in F} f(w)\mu_s(w) \\ &= s \sum_{k=0}^{\infty} E_k (1-s)^k. \end{aligned}$$

$$
\begin{aligned}
E_f(s) &= \sum_{w \in F} f(w)\mu_s(w) \\
&= s \sum_{k=0}^{\infty} E_k (1-s)^k.
\end{aligned}
$$

f has **polynomial growth of degree d on average** if

$$\int_0^1 \frac{E_f(s)}{s^d} ds$$

converges when

$$s \to 0^+$$

The definition is natural;

$f(w) = |w|^d$ has growth of degree d on average.

Let χ be the characteristic function of $R \subset F$.

R is **polynomially negligible** if, for every d,

$$|w|^d \chi(w)$$

has growth of degree at most 0 on average.

(Can be ignored in 'cut and paste' computations of average case complexity.)

Theorem Let $R \lhd F$ be a normal subgroup.

- If $G = F/R$ is not recurrent then R is **thin**:

$$\mu(R) = \mu_1(s)s + o(s).$$

- If $G = F/R$ is not amenable then R is **sparse** and polynomially negligible.

Co-diagonal subgroup $D \lhd F = \langle x_1, \ldots, x_m \rangle$:

the kernel of the homomorphism

$$F \longrightarrow \mathbb{Z}, \qquad x_i \mapsto 1$$

$$\mu(D) = \frac{(m-1)}{m\sqrt{2 - \frac{2}{m}}} \cdot \sqrt{s} + o(\sqrt{s})$$

D is not polynomially negligible.

Quotient test revisited:

$$B_n \longrightarrow \mathbb{Z} = F/D$$

$$B_n \longrightarrow \mathrm{Sym}_n = F/S$$

$$\mu(D) \approx \frac{\sqrt{s}}{\sqrt{2}}$$

$$\mu(S) \approx \frac{1}{n!} + \mu_1 s + o(s)$$

Can be shown: $\mu_1 < 10$

Theorem If $K \lhd F$, $F/K = G$ finite then

$$\mu(K) = \mu_0 + \mu_1(s) + o(s)$$

$$\mu_0 = \frac{1}{|G|}$$

$$\mu_1 = \frac{1}{2m} + \frac{1}{|G|} \cdot \frac{2m-1}{2m} \cdot \sum_{i=2}^{|G|} \frac{1}{1-\lambda_i}$$

where

$$1 = \lambda_1 > \lambda_2 \geqslant \cdots \geqslant \lambda_{|G|}$$

are the eigenvalues of the simple random walk on G.

Of course,

$1 - \lambda_2$ is the **spectral gap** of G:

determines the mixing time of random walk on $G = F/K$

Multiplicative measure:

logarithmic rescaling of convergence

Quotient test revisited:

$$\begin{aligned} F_m &\longrightarrow \mathbb{Z}/n!\mathbb{Z} = F/K \\ F_m &\longrightarrow \mathrm{Sym}_n = F/S \\ & \qquad\qquad m = \frac{n(n-1)}{2} \end{aligned}$$

$$\begin{aligned} \mu(K) &= \frac{1}{n!} + \mu_1(K)s + \cdots \\ \mu(S) &= \frac{1}{n!} + \mu_1(S)s + \cdots \end{aligned}$$

Character theory yields:

$$\mu_1(K) \approx \frac{n!}{\pi^2}$$

$$\mu_1(S) < 10$$

Context-free and regular sets

- **Chomski and Schutzenberger**:
  If R is a context free language then $\mu(R)$ is an algebraic function.
- We can uniformise $\mu(R)$ at $s = 0$:
  $$\mu(R) = m_0 + m_1 s^{1/n} + m_2 s^{2/n} + \cdots$$
- If R is a regular set (produced by a finite automaton) then $\mu(R)$ is a rational function.

- If R is rational then
  - either R is thick,

$$\mu(R) = \mu_0(R) + O(s), \quad \mu_0(R) > 0$$

  - or R is sparse,

$$\mu(R) = \mu_1(R)s + O(s^2)$$

Regular sets

If A and B are regular subsets in F,

- $A \cup B$, $A \cap B$ and $A \smallsetminus B$ are regular.
- The prefix closure $\bar{A}$ of A is regular.

  Here, the *prefix closure* $\bar{A}$ is the set of all initial segments of all words in A.
- The product and the inverse set

  $$AB = \{\, ab \mid a \in A,\, b \in B \,\}$$

  $$A^{-1} = \{\, a^{-1} \mid a \in A \,\}$$

  are regular.

- Every finite subset is regular.
- If $\phi : F \longrightarrow F$ is an endomorphism then $\phi(A)$ is regular.

Thick regular sets

A *cone* $C = C(w)$ is a set of all elements in F containing the given word w as initial segment.

Obviously, cones are thick regular sets:

$$\mu(C(w)) = \frac{1}{|S_{|w|}|} + O(s).$$

Theorem A regular subset $R \subset F$ is thick

if and only if

its prefix closure $\bar{R}$ contains a cone.