An introduction to asymptotic enumeration and large random objects

Mireille Bousquet-Mélou, CNRS, Bordeaux, France

http://www.labri.fr/~bousquet

Asymptotic enumeration

Let A be a set of discrete objects equipped with a size:

$$size: \mathcal{A} \rightarrow \mathbb{N}$$
 $a \mapsto |a|$

Assume that for all n,

$$A_n := \{a \in A : |a| = n\}$$
 is finite.

Let $a(n) = |\mathcal{A}_n|$.

Exact enumeration: determine the sequence a(n)

$$a(n) = \cdots$$
 or $A(t) := \sum_{n \ge 0} a(n)t^n = \cdots$

Asymptotic enumeration: estimate the numbers a(n), as $n \to \infty$

$$a(n) \sim \cdots$$
 or $a(n) \leq \cdots$ or $a(n) = O(\cdots)$ or $a(n)^{1/n} \to \cdots$

Landau's unforgettable notation

As $n \to \infty$,

$$a(n) \sim b(n) \quad \iff \quad \frac{a(n)}{b(n)} o 1$$

$$a(n) = O(b(n)) \iff |a(n)| \le Cb(n)$$
 for some constant C

$$a(n) = o(b(n)) \iff \frac{a(n)}{b(n)} \to 0$$

An introduction to asymptotic enumeration

Part 1. Why are asymptotic results interesting?

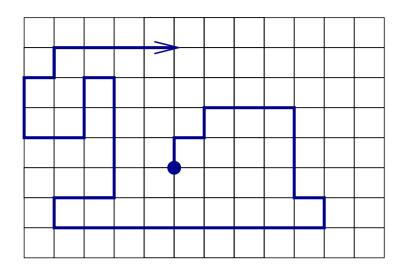
Typical results

• Large random objects

Why are asymptotic results interesting?

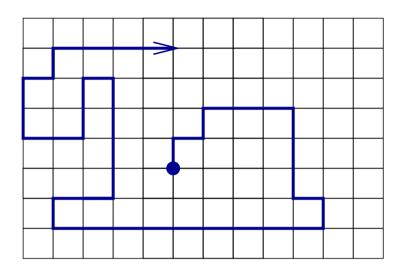
• They are less demanding than exact results

Self-avoiding walks in \mathbb{Z}^d



Theorem [Hara-Slade 92]: for $d \geq$ 5, there exists μ and κ such that $a(n) \sim \kappa \, \mu^n.$

Self-avoiding walks in \mathbb{Z}^d



Theorem [Hara-Slade 92]: for $d \geq 5$, there exists μ and κ such that $a(n) \sim \kappa \, \mu^n$.

Conjecture: on the square lattice (d = 2)

$$a(n) \sim \kappa \,\mu^n \,n^{11/32}.$$

Theorem [Hammersley-Welsh 62]: on the square lattice,

$$\mu^n \le a(n) \le \mu^n \, \beta^{\sqrt{n}}$$

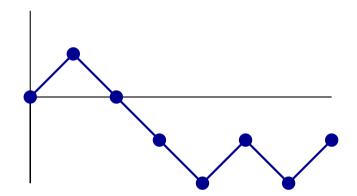
Why are asymptotic results interesting ... even when exact results are known?

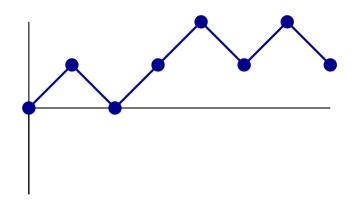
Why are asymptotic results interesting?

- They are less demanding than exact results
- They help deciding how restrictive a constraint is

Decide how restrictive a constraint is Ex: 1D lattice walks with steps ± 1

- Total number of n-step walks: 2^n
- Number of non-negative walks [exercise]: $\binom{n}{\lfloor n/2 \rfloor}$

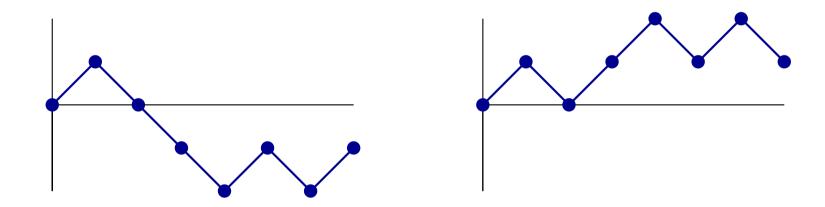




Decide how restrictive a constraint is Ex: 1D lattice walks with steps ± 1

- Total number of n-step walks: 2^n
- Number of non-negative walks [exercise]:

$$\binom{n}{\lfloor n/2 \rfloor} \sim \frac{\sqrt{2}}{\sqrt{\pi}} \frac{2^n}{\sqrt{n}}$$

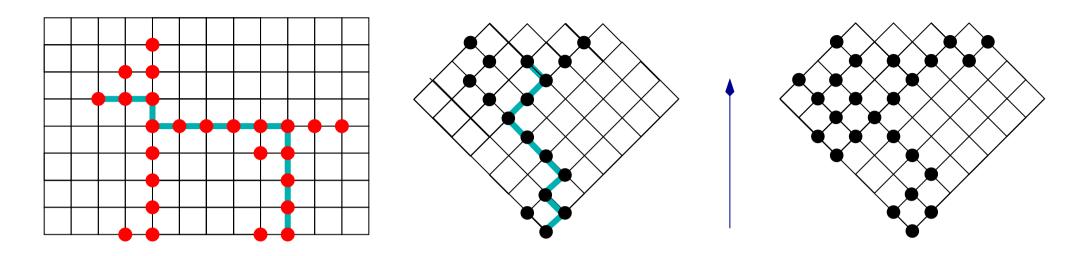


"The probability that a random walk remains non-negative decays like $\sqrt{2}/\sqrt{\pi n}$ as n grows"

Much more about random objects later

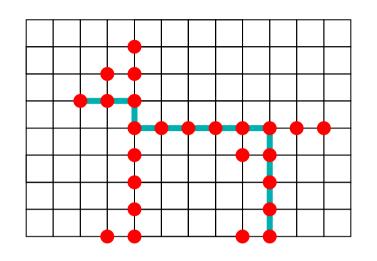
Why are asymptotic results interesting?

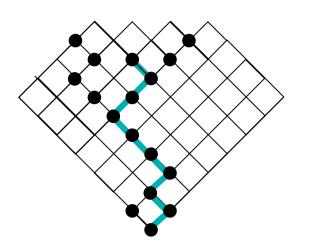
- They are less demanding than exact results
- They help deciding how restrictive a constraint is
- They allow us to compare classes of objects in the same scale

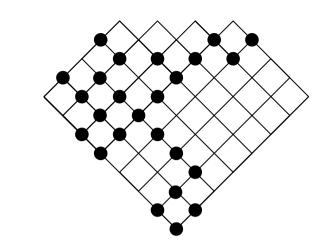


General animals: ???

Directed [Dhar 83] Multi-directed [MBM-R]







General animals: ???

Directed [Dhar 83] Multi-directed [MBM-R]

$$D(t) = \frac{Q}{1 - Q}$$

$$D(t) = \frac{Q}{1 - Q} \quad \text{while} \quad M(t) = \frac{Q}{(1 - Q) \left[1 - \sum_{k \ge 1} \frac{Q^{k+1}}{1 - Q^k (1 + Q)} \right]},$$

with

$$Q = \frac{1 - t - \sqrt{(1+t)(1-3t)}}{2t}.$$

Ex: Directed animals [Dhar 83] vs. multi-directed animals [MBM-Rechnitzer]:

$$D(t) = \frac{Q}{1 - Q} \quad \text{while} \quad M(t) = \frac{Q}{(1 - Q) \left[1 - \sum_{k \ge 1} \frac{Q^{k+1}}{1 - Q^k (1 + Q)} \right]},$$

with

$$Q = \frac{1 - t - \sqrt{(1+t)(1-3t)}}{2t}.$$

Asymptotics:

$$d(n) \sim \kappa \frac{3^n}{\sqrt{n}}$$
, while $m(n) \sim \kappa \mu^n$ with $\mu = 3.58789436...$

Multi-directed animals are exponentially more numerous. $\stackrel{\circ}{\smile}$

Ex: Directed animals [Dhar 83] vs. multi-directed animals [MBM-Rechnitzer]:

$$D(t) = \frac{Q}{1 - Q} \quad \text{while} \quad M(t) = \frac{Q}{(1 - Q) \left[1 - \sum_{k \ge 1} \frac{Q^{k+1}}{1 - Q^k (1 + Q)} \right]},$$

with

$$Q = \frac{1 - t - \sqrt{(1+t)(1-3t)}}{2t}.$$

Asymptotics:

$$d(n) \sim \kappa \frac{3^n}{\sqrt{n}}$$
, while $m(n) \sim \kappa \mu^n$ with $\mu = 3.58789436...$

Multi-directed animals are exponentially more numerous. $\overset{\circ}{\smile}$



BUT General animals: $\kappa \mu^n/n$ with $\mu \simeq 4.06...$

Standard scales

Typically, asymptotic results look like this:

$$a(n) \sim n^{\alpha n} \mu^{n^{\beta}} n^{\gamma} (\log n)^{\eta} \kappa$$

Examples

$$n! \sim (n/e)^n \sqrt{2\pi n}$$

$$w(n) \sim \kappa \mu^n (\log n)^{1/4}$$

$$p(n) \sim \frac{1}{4n\sqrt{3}} e^{\pi\sqrt{2n/3}}$$

$$g(n) \sim \kappa n! (27.2...)^n n^{-7/2}$$

permutations

SAW in D = 4 (conj.)

Ferrers diagrams with n cells (partitions)

labelled planar graphs on n vertices [Gimenez-Noy 05]

Why are asymptotic results interesting?

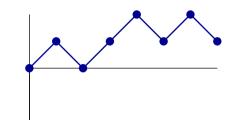
- They are less demanding than exact results
- They help deciding how restrictive a constraint is
- They allow us to compare classes of objects in the same scale
- They tell us about large random objects

Large random objects

Ex: 1D walks with steps ± 1

The number of non-negative n-step walks is

$$\binom{n}{\lfloor n/2 \rfloor} \sim \sqrt{2} \; \frac{2^n}{\sqrt{\pi n}}$$



 \Rightarrow The probability that an n-step random walk remains non-negative is equivalent to κ/\sqrt{n} .

Uniform probabilistic distribution: each object of A_n occurs with the same probability 1/a(n), with $a(n) = |A_n|$.

Comparison of classes: if $\mathcal{B}_n \subset \mathcal{A}_n$ and O is a random object of \mathcal{A}_n

$$\operatorname{prob}(O \text{ belongs to } \mathcal{B}_n) = \frac{b(n)}{a(n)}.$$

The study of additional statistics

$$size: \mathcal{A} \rightarrow \mathbb{N}$$
 $s: \mathcal{A} \rightarrow \mathbb{Z}$ $a \mapsto |a|$ $a \mapsto s(a)$

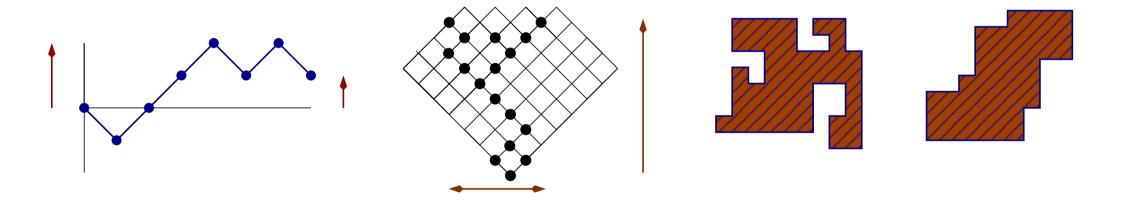
When objects are taken uniformly in A_n , the statistic s becomes a random variable S_n :

$$\mathbb{P}(\mathsf{S}_n = k) = \frac{a(n,k)}{a(n)}$$

where a(n,k) is the number of objects of size n for which the additional statistic s equals k.

Examples of additional statistics

- ullet the final height of a 1D walk with n steps,
- the maximal height of such a walk,
- \bullet the width/height of a directed animal with n cells,
- ullet the area enclosed by a SAP of perimeter 2n,
- \bullet the area enclosed by a staircase polygon of perimeter 2n...



What do we want to know about this additional statistic?

• Its average value (and its behaviour as $n \to \infty$)

$$\mathbb{E}(\mathsf{S}_n) = \frac{1}{a(n)} \sum_{|a|=n} s(a) = \frac{1}{a(n)} \sum_{k} ka(n,k)$$

where a(n,k) is the number of objects of size n and statistic k.

More general moments

$$\mathbb{E}((S_n)^i) = \frac{1}{a(n)} \sum_{|a|=n} s(a)^i = \frac{1}{a(n)} \sum_k k^i a(n,k)$$

Examples

 \bullet 1D random walks: take s to be the final height of the walk. Then

$$\mathbb{E}(S_n) = 0$$
 (symmetry!) $\mathbb{E}(S_n^2) = n$

Examples

 \bullet 1D random walks: take s to be the final height of the walk. Then

$$\mathbb{E}(S_n) = 0$$
 (symmetry!) $\mathbb{E}(S_n^2) = n$

Area of staircase polygons

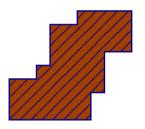
$$\mathbb{E}\left(\frac{\mathsf{S}_n}{n^{3/2}}\right) \sim \frac{\sqrt{\pi}}{4}, \quad \mathbb{E}\left(\left(\frac{\mathsf{S}_n}{n^{3/2}}\right)^2\right) \sim \frac{5}{24}$$

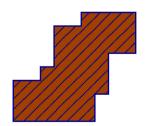
More generally,

$$\mathbb{E}\left(\left(\frac{\mathsf{S}_n}{n^{3/2}}\right)^j\right) \sim \frac{c_j}{c_0} \; \frac{\Gamma(-1/2)}{\Gamma((3j-1)/2)}$$

with

$$c_0 = -1/2, \quad c_j = \frac{j(3j-4)}{8}c_{j-1} + \sum_{i=1}^{j-1} {j \choose i}c_i c_{j-i}$$





Interlude: The Gamma function

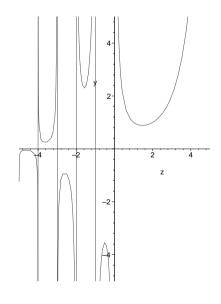
$$\Gamma: \mathbb{C}\setminus\{0,-1,-2,-3,\cdots\}\to \mathbb{C}$$

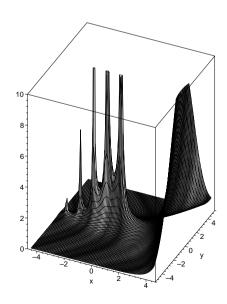
Euler's definition: For $\Re(z) > 0$,

$$\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt \quad \Rightarrow \Gamma(n+1) = n!$$

Functional equation

$$\Gamma(z+1) = z\Gamma(z)$$





Interlude: The Gamma function

Complement formula:

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin \pi z} \Rightarrow \Gamma(1/2) = \sqrt{\pi}$$

Asymptotics [Stirling's formula]: as $z \to \infty$ with $|Arg(z)| < \pi - \delta$,

$$\Gamma(z+1) = z\Gamma(z) = \left(\frac{z}{e}\right)^z \sqrt{2\pi z} \left(1 + \frac{1}{12z} + O(1/z^2)\right)$$

What do we want to know about this additional statistic?

• Its moments $\mathbb{E}((S_n)^j)$ (and their behaviour as $n \to \infty$)

$$\triangleleft$$
 \triangleleft \Diamond \triangleright \triangleright

- Its law, described
 - either by its distribution function

$$F_n(x) := \mathbb{P}(S_n \le x)$$

– or by its probability generating function:

$$G_n(u) := \mathbb{E}(u^{\mathsf{S}_n}) = \sum_k \mathbb{P}(\mathsf{S}_n = k)u^k = \sum_k \frac{a(n,k)}{a(n)}u^k,$$

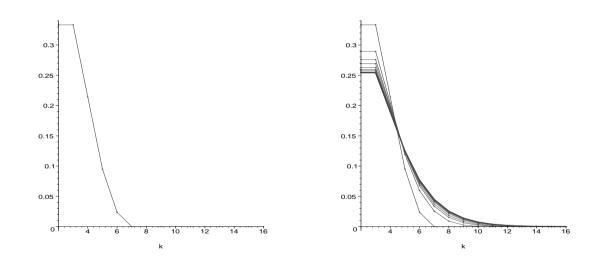
and the behaviour of this law as $n \to \infty$.

Ex. 1: The number of contacts in a Dyck path

Let A(t, u) count Dyck paths by the half-length (t) and contacts (u) [ex.]:

$$A(t,u) = \frac{u}{1 - uP(t)}$$
 with $P(t) = \frac{1 - \sqrt{1 - 4t}}{2}$.

 \Rightarrow plot of a(n,k)/a(n), the probability that a 2n-step paths has k contacts (n fixed)



Convergence of the probability distribution to a discrete law:

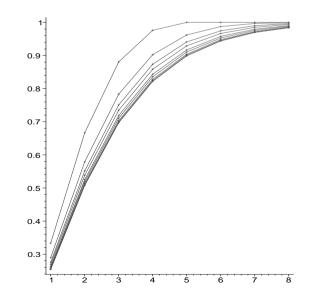
$$\mathbb{P}(\mathsf{S}_n = k+1) \to \frac{k}{2^{k+1}}$$

Ex.1 (contd.): The number of contacts in a Dyck path

$$\mathbb{P}(\mathsf{S}_n = k+1) \to \frac{k}{2^{k+1}}$$

⇒ Convergence of the distribution function:

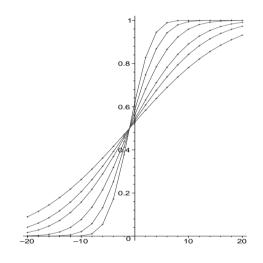
$$\mathbb{P}(\mathsf{S}_n \leq \ell+1) \to \sum_{k=1}^\ell \frac{k}{2^{k+1}}$$



Ex. 2: The final height of a random walk

For $k \equiv n \mod 2$,

$$\mathbb{P}(\mathsf{S}_n = k) = \frac{1}{2^n} {n \choose (n-k)/2} \quad \Rightarrow \quad \mathbb{P}(\mathsf{S}_n \le \ell) = \frac{1}{2^n} \sum_{k=-n}^{\ell} {n \choose (n-k)/2}$$



What happens?

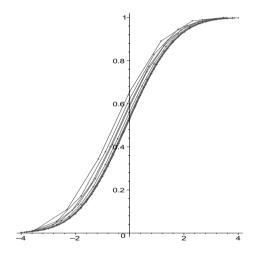
$$\mathbb{E}(\mathsf{S}_n^2) = n$$

 \Rightarrow Consider instead the random variable S_n/\sqrt{n}

Ex. 2 (contd.): The final height of a random walk

The distribution function of S_n/\sqrt{n} :

$$\mathbb{P}\left(\frac{\mathsf{S}_n}{\sqrt{n}} \le x\right)$$



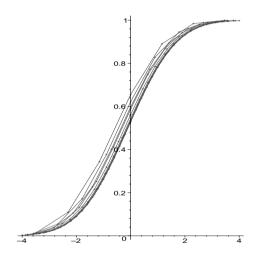
$$\mathbb{P}\left(\frac{\mathsf{S}_n}{\sqrt{n}} \le x\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$$

Distribution function of the Gaussian law

Convergence in law (or: in distribution)

Def. The random variables S_n converge in law to the random variable S_n having distribution function F(x), if for all x [where F is continuous],

$$\mathbb{P}(S_n \leq x) = F_n(x) \to F(x) = \mathbb{P}(S \leq x).$$



Ex. The final height of a random walk of length n, normalized by \sqrt{n} , converges in law to a centered Gaussian of variance 1.

$$rac{\mathsf{S}_n}{\sqrt{n}} \stackrel{d}{ o} \mathcal{N}(\mathsf{0},\mathsf{1})$$

Ex. 3: The height of binary trees

Let $F_n(x)$ be the probability that a random binary tree of size n has height at most $x\sqrt{n}$.

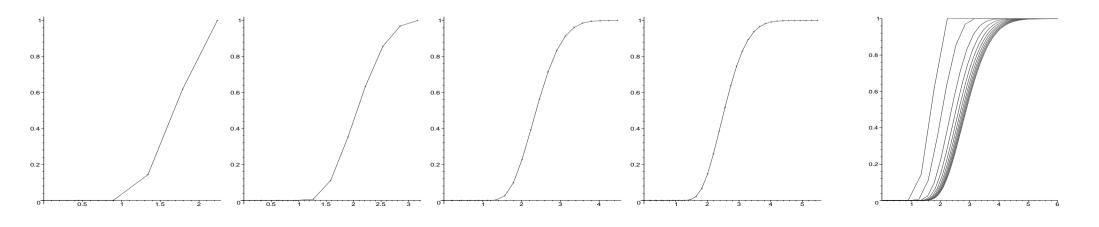
Convergence in law of the (normalized) height [Kemp 79], [Flajolet-Odlyzko 82]

$$F_n(x) = \mathbb{P}\left(\frac{\mathsf{H}_n}{\sqrt{n}} \le x\right) \to F(x/2)$$

where

$$F(x) = \frac{2\pi^{5/2}}{x^3} \sum_{k=-\infty}^{\infty} k^2 e^{-\pi^2 k^2/x^2}$$

Distribution function of a theta law.



Ex. 4, 5, 6: Staircase polygons

For random staircase polygons of perimeter 2n:

• The normalized diameter converges to a theta distribution:

$$\mathbb{P}\left(\frac{2\mathsf{D}_n}{\sqrt{n}} \le x\right) \to \frac{2\pi^{5/2}}{x^3} \sum_{k=-\infty}^{\infty} k^2 e^{-\pi^2 k^2/x^2}$$

• The width concentrates at n/2, with gaussian fluctuations:

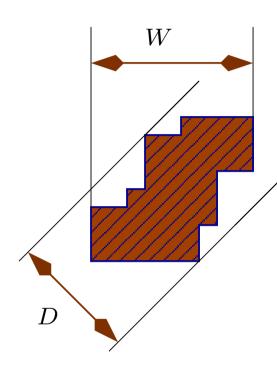
$$\mathbb{P}\left(\frac{\mathsf{W}_n - n/2}{\sqrt{n}} \le x\right) \to \sqrt{2/\pi} \int_{-\infty}^x e^{-2t^2} dt$$

• The normalized area converges to an Airy distribution:

$$\frac{\mathsf{4A}_n}{n\sqrt{n}} \stackrel{d}{\to} \mathcal{A}$$

where the Airy distribution A is characterized by its moments.

 \Rightarrow Connections with properties of the Brownian excursion (height, area).



Where will we find the information on the statistic s?

Bivariate generating functions:

$$A(t,u) = \sum_{a \in \mathcal{A}} t^{|a|} u^{s(a)} = \sum_{n,k} a(n,k) t^n u^k$$

Specialization u = 1:

$$A(t,1) = \sum_{a \in A} t^{|a|} = \sum_{n} a(n)t^{n}$$

• Average value of S_n :

$$u\frac{\partial}{\partial u}A(t,u) = \sum_{n} t^{n} \left(\sum_{k} ka(n,k)u^{k} \right) \Rightarrow u\frac{\partial}{\partial u}A(t,u) \Big|_{u=1} = \sum_{n} t^{n}a(n)\mathbb{E}(S_{n})$$

Where will we find the information on the statistic s?

Bivariate generating functions:

$$A(t,u) = \sum_{a \in \mathcal{A}} t^{|a|} u^{s(a)} = \sum_{n,k} a(n,k) t^n u^k$$

Specialization u = 1:

$$A(t,1) = \sum_{a \in A} t^{|a|} = \sum_{n} a(n)t^{n}$$

• Average value of S_n :

$$u\frac{\partial}{\partial u}A(t,u) = \sum_{n} t^{n} \left(\sum_{k} ka(n,k)u^{k} \right) \Rightarrow u\frac{\partial}{\partial u}A(t,u) \Big|_{u=1} = \sum_{n} t^{n}a(n)\mathbb{E}(S_{n})$$

• Further moments:

$$\left(u\frac{\partial}{\partial u}\right)^{i}A(t,u) = \sum_{n} t^{n} \left(\sum_{k} k^{i}a(n,k)u^{k}\right) \Rightarrow \left(u\frac{\partial}{\partial u}\right)^{i}A(t,u)\Big|_{u=1} = \sum_{n} t^{n}a(n)\mathbb{E}(\mathsf{S}_{n}^{i})$$

Where will we find the information on the statistic s?

Bivariate generating functions:

$$A(t,u) = \sum_{a \in \mathcal{A}} t^{|a|} u^{s(a)} = \sum_{n,k} a(n,k) t^n u^k$$

Specialization u = 1:

$$A(t,1) = \sum_{a \in A} t^{|a|} = \sum_{n} a(n)t^{n}$$

• Average value of S_n :

$$u\frac{\partial}{\partial u}A(t,u) = \sum_{n} t^{n} \left(\sum_{k} ka(n,k)u^{k} \right) \Rightarrow u\frac{\partial}{\partial u}A(t,u) \Big|_{u=1} = \sum_{n} t^{n}a(n)\mathbb{E}(S_{n})$$

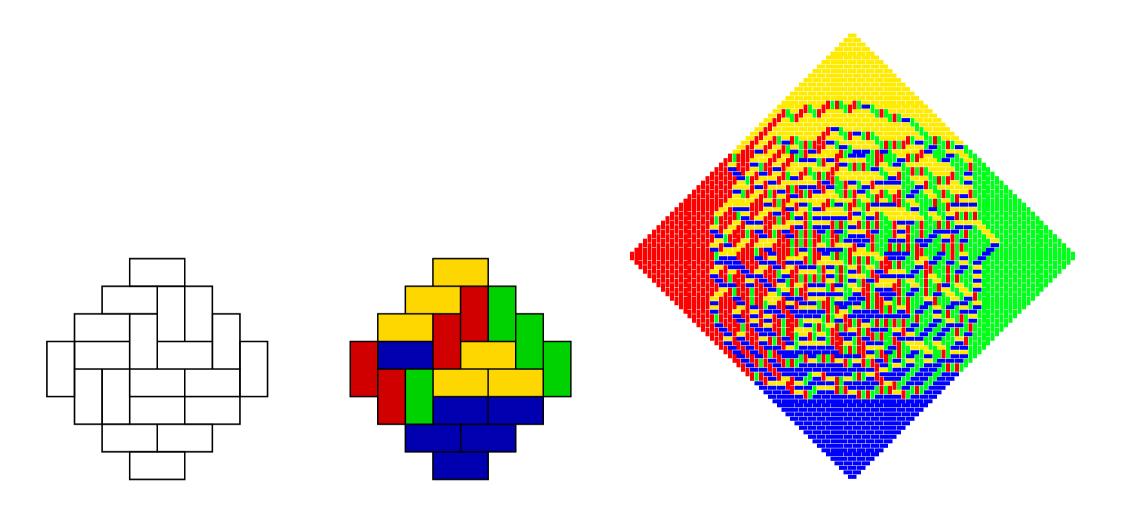
• Further moments:

$$\left(u\frac{\partial}{\partial u}\right)^{i}A(t,u) = \sum_{n} t^{n} \left(\sum_{k} k^{i}a(n,k)u^{k}\right) \Rightarrow \left(u\frac{\partial}{\partial u}\right)^{i}A(t,u)\Big|_{u=1} = \sum_{n} t^{n}a(n)\mathbb{E}(S_{n}^{i})$$

• Probability generating function:

$$A(t,u) = \sum_{n} t^{n} \left(\sum_{k} a(n,k) u^{k} \right) = \sum_{n} t^{n} a(n) \mathbb{E}(u^{S_{n}})$$

More complex statistics: tilings of the aztec diamond



The arctic circle phenomenon [Propp et al.]



Why do we want to know about large objects?

• To study the complexity of algorithms

Complexity of algorithms

Problem: generate a non-negative 1D random walk of length n with uniform probability

Solution: generate a general random walk (by tossing a coin n times) and only keep it when it is non-negative!

Recall that the probability that an n-step random walk is non-negative is $\sim \kappa/\sqrt{n}$

Complexity: about \sqrt{n} attempts $\times n$ tosses per attempt $\simeq n^{3/2}$

Complexity of algorithms

Problem: generate a non-negative 1D random walk of length n with uniform probability

Solution: generate a general random walk (by tossing a coin n times) and only keep it when it is non-negative!

Recall that the probability that an n-step random walk is non-negative is $\sim \kappa/\sqrt{n}$

Complexity: about \sqrt{n} attempts $\times n$ tosses per attempt $\simeq n^{3/2}$

This is silly! One can reject a bad walk as soon as it reaches -1.

Analysis: only $(1+\sqrt{2})n$ tosses are needed on average

This rejection algorithm is linear on average.

Why do we want to know about large objects?

• To study the complexity of algorithms

• To study phase transitions in statistical mechanics

Phase transitions in statistical mechanics

"As one parameter (temperature, fugacity...) varies, the behaviour of a large random object changes drastically"

Phase transitions in statistical mechanics

"As one parameter (temperature, fugacity...) varies, the behaviour of a large random object changes drastically"

Ex: Self-avoiding polygons of perimeter 2n with non-uniform probability:

$$\operatorname{prob}(p) = \frac{1}{Z_n(u)} u^{a(p)}$$

where a(p) is the area of the polygon p and

$$Z_n(u) = \sum_{|p|=2n} u^{a(p)}$$

counts SAPs of perimeter 2n by area (the partition function).

The average area is

$$\mathbb{E}(A_n) = \frac{1}{Z_n(u)} \sum_{|p|=2n} a(p) u^{a(p)} = \frac{u Z_n'(u)}{Z_n(u)}$$

Does the behaviour of $\mathbb{E}(A_n)$ undergo a drastic change as u varies?

Phase transitions in statistical mechanics

Draw a polygon of perimeter 2n with probability

$$\operatorname{prob}(p) = \frac{1}{Z_n(u)} u^{a(p)}$$

A toy example: rectangular polygons [exercise!]

$$\mathbb{E}(A_n) \sim \begin{cases} n & \text{for } u < 1, \\ n^2/6 & \text{for } u = 1, \\ n^2/4 & \text{for } u > 1. \end{cases}$$

 \Rightarrow Phase transition at u = 1.

General SAP

$$\mathbb{E}(A_n) \sim \left\{ egin{array}{ll} \kappa n & ext{for } u < 1, \ \kappa n^{3/2} & ext{for } u = 1, \ n^2/4 & ext{for } u > 1 \end{array}
ight. \ ext{[Fisher-Guttmann-Whittington 91]}
ight.$$

Why do we want to know about large objects?

- To study the complexity of algorithms
- To study phase transitions in statistical mechanics
- Because of Universality, which makes asymptotic results meaningful

Universality: Asymptotic results are meaningful

"Certain quantities, or laws, describing the behaviour of large random objects, do not depend on the details of the problem, but only on its dimension"

Universality: Asymptotic results are meaningful

"Certain quantities, or laws, describing the behaviour of large random objects, do not depend on the details of the problem, but only on its dimension"

Universal exponents

Self-avoiding walks

Conjecture: For any 2D lattice (square, triangular, hexagonal...), the number of n-step SAW is asymptotic to

$$\kappa \mu^n n^{11/32}$$

where κ and μ depend on the lattice, but not the exponent

• 1D random walks

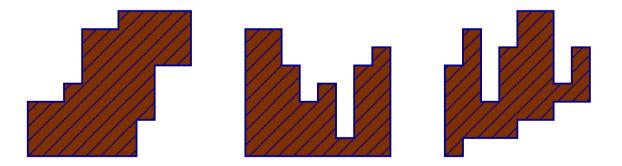
The probability that a walk taking step i with probability p_i ($i \le K$) ends at 0 after n steps is asymptotic to κ/\sqrt{n} as soon as the average of the steps is 0 ($\sum_i i p_i = 0$).

Universality: Asymptotic results are meaningful

Universal laws

- Plane trees The height of plane trees, suitably normalized, converges to a theta distribution, whether we take plane trees, binary trees, ternary trees...
- Self-avoiding polygons

The area of a random staircase polygon/Directed and convex polygon/Bargraphs, suitably normalized, converges to an Airy distribution [Duchon], [Richard]



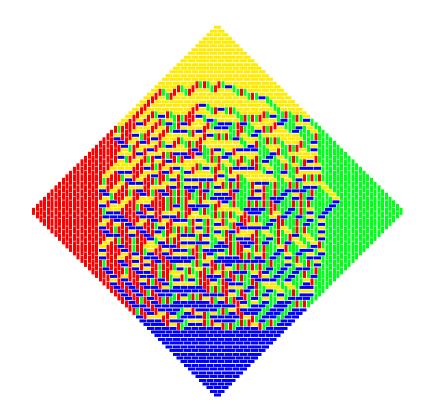
Conjecture: also true for general SAPs!

• ICM 2006: Deift's plenary lecture on the universality of the Tracy-Widom distribution

To conclude...

Asymptotic results are interesting!

- they explain and quantify what you see
- they allow you to talk to other people (probabilists, physicists...)
- they give you an idea of The Big Picture



Overview

Part 1. Asymptotic results are interesting

Part 2. Asymptotics of sequences a(n)

- 2.1. Three techniques...
 - Bare-hand asymptotics on sums
 - Singularity analysis
 - Saddle point asymptotics

2.2 ... and their applications

- Automated asymptotics for algebraic series
- (Almost) automated asymptotics for D-finite series

Part 3. Limit laws

Bivariate series and perturbation analysis