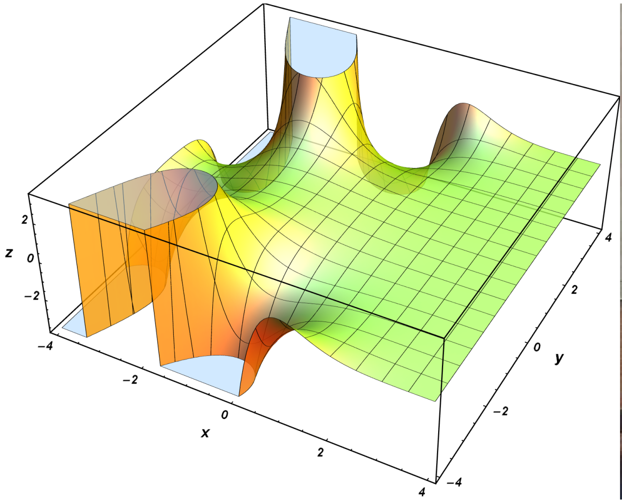
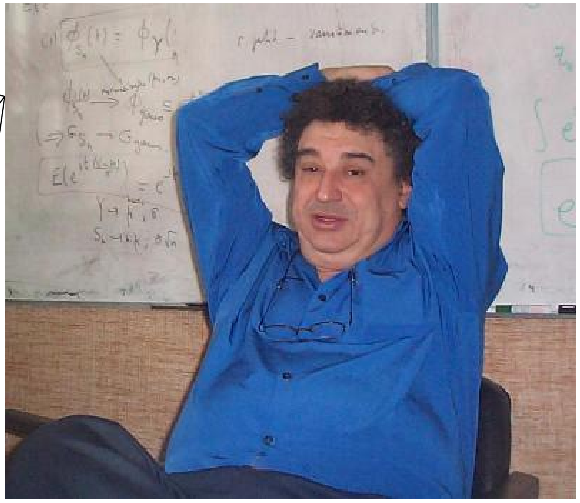
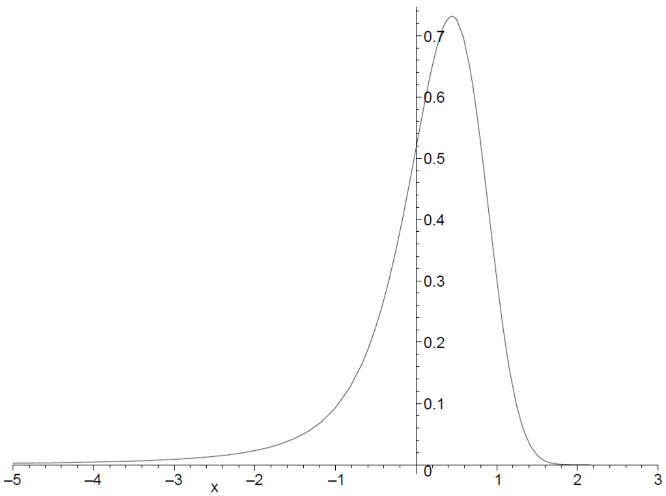
The Airy function and its modern avatars in combinatorics, physics, probability theory...

Cyril Banderier (CNRS / Univ. Paris Nord) [$$$: ANR PhysComb]

March 30, 2012, Séminaire Lotharingien #68 [32 years!]
(talk based on "Philippe Flajolet and Algebraic Combinatorics")







Which one is Philippe's grandfather?

Which one is Philippe's grandfather?

Sir George Biddell Airy
(1801-1892)
Royal astronomer
Cambridge

Philippe Flajolet
(1885-1948)
Republican astronomer
Observatory of Lyon

The Airy function


Sir George Biddel Airy:
On the intensity of Light in the neighbourhood of a Caustic.
Trans. Camb. Phil. Soc. v. 6 (1838)

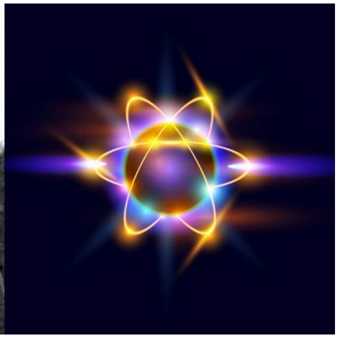
Airy Function $y'' - zy = 0$

$$\mathrm{Ai}(z) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{i(zt+t^3/3)}\,dt$$

$$= \frac{1}{\pi 3^{2/3}}\sum_{n=0}^{\infty}\frac{\Gamma(\frac{n+1}{3})}{n!}\sin\left(\frac{2(n+1)\pi}{3}\right)\left(3^{1/3}z\right)^n$$

Airy $\approx$ hypergeometric series ${}_2F_0(z)$
$\approx$ integral representation $\approx$ Bessel functions $I_\nu(z), K_\nu(z)$ at $\nu = 1/3$
physics: optics, quantum mechanics, electromagnetics, radiative transfer
combinatorics : in some limit laws of discrete structures

The Airy function: a special function ubiquitous in physics



Schrödinger equation:

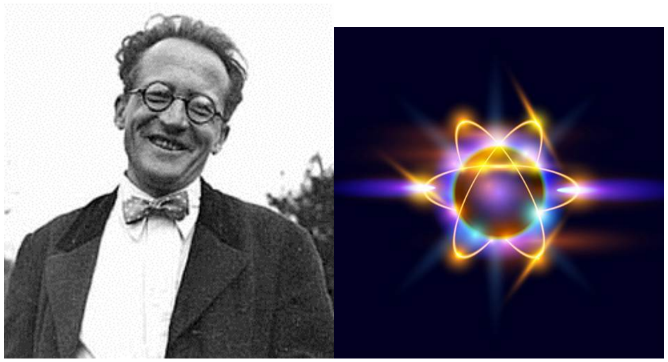
$-\frac{\hbar}{2w}\psi''(x) + gx\psi(x) = E\psi(x)$

[If you don't know about physics, just see this as a differential equation for a classical function ψ from $\mathbb{C}$ to $\mathbb{C}$.]

For which E do we have $\psi(0) = 0$?

Hint: Airy Function $y'' - zy = 0$

The Airy function: a special function ubiquitous in physics



Schrödinger equation:

$$-\frac{\hbar}{2w}\psi''(x) + gx\psi(x) = E\psi(x)$$

[If you don't know about physics, just see this as a differential equation for a classical function ψ from $\mathbb{C}$ to $\mathbb{C}$.]

For which E do we have $\psi(0) = 0$?

Hint: Airy Function $y'' - zy = 0$

Answer: A change of variable gives that E has to be a zero of the Airy function!

Recently proven phenomenon in physics: Quantum states of neutrons in the Earth's gravitational field, at energy levels being nothing else than quotients of Airy zeroes! (up to 4 significative digits!)

[Thanks to the polish gang "Combinatorial Physics" for this nice example!]

Some precisions about the previous example, as its sketchy presentation sounds to contradict the Cauchy–Lipschitz theorem (aka Picard–Lindelöf theorem).

It is true that for a given N, and some initial conditions, the equation $-a.y'' + b.x.y = N.y$ has a unique solution $y(z) = A.\,\mathrm{AiryAi}(blah(z)) + B.\,\mathrm{AiryBi}(blah(z))$ where $blah(z)$ is the appropriate linear change of variable.

If we stop here, there will always be a solution with $y(0) = 0$, but now, the quantum mechanics comes into the game: the physics of the Schrödinger equation implies that $y(\pm\infty) < \infty$ and thus $B = 0$.

The initial condition $y(0) = 0$ then implies
either $A = 0$ (but then $y(z) = 0$, which is not a physical solution)
either $blah(0)$ is a zero α_k of the Airy function, which in turn constraints N to belong to a discrete set of values: $N = -a^{1/3} b^{2/3} \alpha_k$. □

The Airy function

George Biddel Airy:
On the intensity of Light in the neighbourhood of a Caustic.
Trans. Camb. Phil. Soc. v. 6 (1838)

Airy Function $y'' - zy = 0$

$$\mathrm{Ai}(z) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{i(zt+t^3/3)}\,dt$$

$$= \frac{1}{\pi 3^{2/3}} \sum_{n=0}^{\infty} \frac{\Gamma(\frac{n+1}{3})}{n!} \sin\left(\frac{2(n+1)\pi}{3}\right)\left(3^{1/3}x\right)^n$$

Airy $\approx$ hypergeometric series ${}_2F_0(z)$
$\approx$ integral representation $\approx$ Bessel functions $I_\nu(z), K_\nu(z)$ at $\nu = 1/3$
physics: optics, quantum mechanics, electromagnetics, radiative transfer
COMBINATORICS: in some limit laws of discrete structures

Airy limit lawSSSSSSSS

There are 3 families of limit laws related to the Airy function:
• Tracy-Widom distribution (distribution of spectra of random matrices; zeroes of ζ; patience sorting)
• Area-Airy distribution (area under Brownian motion)
• Map-Airy distribution (largest connected component in a graph, stable law 3/2)

All of them can be obtained via analytic combinatorics (c) [Flajolet–Sedgewick 2009].

The Airy function in statistical mechanics

The Airy function is related to the distribution F_2 of the largest eigenvalue of some random hermitian matrices [Tracy & Widom 1993-2002]:

$$F_2(s) := \exp\left(-\int_s^{\infty}(x-s)q(x)^2dx\right)$$

where q is a solution of the Painlevé II equation $q'' = sq + 2q^3$ with $q(s) \sim \mathrm{Ai}(s)$ as $s \to \infty$.

Key object: Airy Kernel:$=\frac{\mathrm{Ai}(x)\mathrm{Ai}'(y)-\mathrm{Ai}'(x)\mathrm{Ai}(y)}{x-y}$

Recent experiments: $F_2(s)$ as a law for gaps in superconductors!

The Airy function in statistical mechanics

The Airy function is related to the distribution F_2 of the largest eigenvalue of some random hermitian matrices [Tracy & Widom 1993-2002]:

$$F_2(s) := \exp\left(-\int_s^\infty (x-s)q(x)^2 dx\right)$$

where q is a solution of the Painlevé II equation $q'' = sq + 2q^3$ with $q(s) \sim \text{Ai}(s)$ as $s \to \infty$.

Key object: Airy Kernel:$=\frac{\mathrm{Ai}(x)\mathrm{Ai}'(y)-\mathrm{Ai}'(x)\mathrm{Ai}(y)}{x-y}$

Recent experiments: $F_2(s)$ as a law for gaps in superconductors!

Length of the longest increasing subsequence in a permutation

E.g.: 2 3 8 6 4 5 1 7 its longest increasing subsequence has length 5.

(computable in linear time via Young Tableaux).

[Ulam 61, Erdős 35, Logan & Shepp 77, Vershik & Kerov 77-85, Odlyzko-Rains 93, Kim 96, Baik–Deift–Johansson 99, Aldous–Diaconis 99]:

$$\mathbb{P}\left(\frac{L_n - 2\sqrt{n}}{n^{1/6}} \leq x\right) \to F_2(x)$$

THE FIRST CYCLES IN AN EVOLVING GRAPH

Philippe FLAJOLET,* Donald E. KNUTH,** and Boris PITTEL†
Computer Science Department, Stanford University, Stanford CA 94305, U.S.A.

* Permanent address: *INRIA, Rocquencourt, 78150 Le Chesnay (France).*
** Permanent address: *Computer Science Department, Stanford University, Stanford, CA 94305 (U.S.A.).*
† Permanent address: *Mathematics Department, Ohio State University, Columbus, OH 43210 (U.S.A.).*

Revised November 1988

If successive connections are added at random to an initially disconnected set of n points, the expected length of the first cycle that appears will be proportional to $n^{\frac{1}{6}}$, with a standard deviation proportional to $n^{\frac{1}{4}}$. The size of the component containing this cycle will be of order $n^{\frac{1}{2}}$, on the average, with standard deviation of order $n^{\frac{7}{12}}$. The average length of the kth cycle is proportional to $n^{\frac{1}{6}}(\log n)^{k-1}$. Furthermore, the probability is $\sqrt{\frac{2}{3}} + O(n^{-\frac{1}{3}})$ that the graph has no components with more than one cycle at the moment when the number of edges passes $\frac{1}{2}n$. These results can be proved with analytical methods based on combinatorial enumeration with multivariate generating functions, followed by contour integration to derive asymptotic formulas for the quantities of interest.

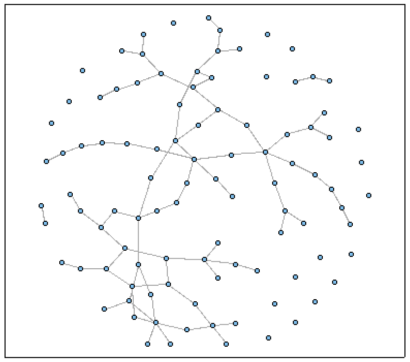
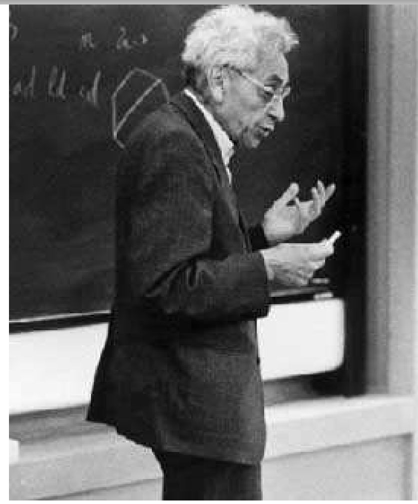
A classic paper by Erdős and Rényi [6] inaugurated the study of the random graph process, in which we begin with a totally disconnected graph and enrich it by successively adding edges. Algorithms that deal with graphs often mimic such a process, inputting a sequence of edges until some stopping criterion occurs, based on the configuration of edges seen so far. To analyze such algorithms, we wish to estimate relevant characteristics of the resulting graph. For example, we might stop when the graph first contains a particular kind of subgraph, and we might ask how large that subgraph is.

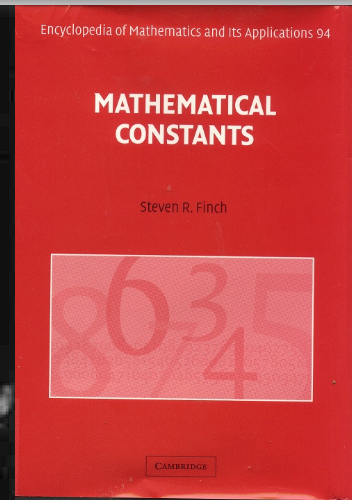
The purpose of this paper is to introduce analytical methods by which such questions can be answered systematically. In particular, we will apply the ideas to an interesting question posed by Paul Erdős and communicated by Edgar Palmer to the 1985 Seminar on Random Graphs in Posnań: "What is the expected length of the first cycle in an evolving graph?" The answer turns out to be rather surprising: The first cycle has length $Kn^{\frac{1}{6}} + O(n^{\frac{3}{22}})$ on the average, where

$$K = \frac{1}{\sqrt{8\pi}\,\mathrm{i}} \int_{-\infty}^{\infty} \int_{\Gamma} \mathrm{e}^{(\mu+2s)(\mu-s)^2/6} \frac{\mathrm{d}s}{s}\, \mathrm{d}\mu \approx 2.0337$$

for a certain contour Γ. The form of this result suggests that the expected

This research was supported in part by the National Science Foundation under grant CCR-86-10181, and by Office of Naval Research contract N00014-87-K-0502.



Erdős–Rényi model of random graphs (1959):

G(n,p): graphs with n vertices, and probability p to get an edge

Lot of transition phases when the proportion of edges is increasing.

The first cycle has length $\sim Kn^{1/6}$ where

$$K = \frac{1}{\sqrt{8\pi i}} \int_{-\infty}^{\infty} \int \exp\left((\mu + 2s) \frac{(\mu - s)^2}{6} \right) \frac{ds}{s} d\mu \approx 2.0337$$

PF: Give me an integral!

Further works: the Janson–Knuth–Łuczak–Pittel 1993 "giant paper of the giant component" ... story still goes on in the 2000's, e.g. Ravelomanana, Wormald, Noy, Drmota... +Bollobás' probabilist school

Airy Phenomena and Analytic Combinatorics of Connected Graphs

Philippe Flajolet
Algorithms Project, INRIA Rocquencourt, 78153 Le Chesnay (France).
Philippe.Flajolet@inria.fr.

Bruno Salvy
Algorithms Project, INRIA Rocquencourt, 78153 Le Chesnay (France).
Bruno.Salvy@inria.fr.

Gilles Schaeffer
LIX – CNRS, École polytechnique, 91128 Palaiseau (France).
Gilles.Schaeffer@lix.polytechnique.fr

Submitted: Oct 11, 2002 & Mar 12, 2004; Accepted: Apr 7, 2004; Published: May 27, 2004.
MR Subject Classifications: 05A15, 05A16, 05C30, 05C40, 05C80

Abstract

Until now, the enumeration of connected graphs has been dealt with by probabilistic methods, by special combinatorial decompositions or by somewhat indirect formal series manipulations. We show here that it is possible to make analytic sense of the divergent series that expresses the generating function of connected graphs. As a consequence, it becomes possible to derive analytically known enumeration results using only first principles of combinatorial analysis and straight asymptotic analysis—specifically, the saddle-point method. In this perspective, the enumeration of connected graphs by excess (of number of edges over number of vertices) derives from a simple saddle-point analysis. Furthermore, a refined analysis based on coalescent saddle points yields complete asymptotic expansions for the number of graphs of fixed excess, through an explicit connection with Airy functions.

Introduction

E. M. Wright, of Hardy and Wright fame, initiated the enumeration of labelled connected graphs by number of vertices and edges in a well-known series of articles [34, 35, 36]. In particular, he discovered that the generating functions of graphs with a fixed excess of number of edges over number of vertices has a rational expression in terms of the tree function $T(z)$. Wright's approach is based on the fact that deletion of an edge in a

Connected Graphs

Philippe Flajolet, Bruno Salvy, and Gilles Schaeffer.
Airy phenomena and analytic combinatorics of connected graphs.
Electronic Journal of Combinatorics, 2004.

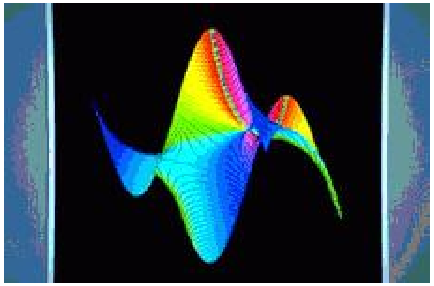
It is possible to make analytic sense of the divergent series that expresses the generating function of connected graphs $C(z) = \ln\left(\sum 2^{\binom{n}{2}} \frac{z^n}{n!}\right)$.

The enumeration of connected graphs by excess (of number of edges over number of vertices) derives from a simple saddle-point analysis. Furthermore, a refined analysis based on coalescent saddle points yields complete asymptotic expansions for the number of graphs of fixed excess, through an explicit connection with Airy functions.

build on works by E.M. Wright 70's, Knuth–Flajolet–Pittel 1989, Janson–Knuth–Łuczak–Pittel 1993 "Giant paper on the giant component"

Algorithms Project's Logo

Our logo shows the behaviour in the complex plane of the generating function of connected graphs counted according to number of nodes and edges. In critical regions, two saddle points coalesce giving rise to a so-called "monkey saddle" (a saddle that you'd use if you had three legs!)



The fine analysis of this coalescence is crucial to the understanding of connectivity in random graphs. This problem has applications in the design of communication networks and it relates to a famous series of problems initiated by Erdös and Renyi in the late 1950's. See the paper [Janson, Knuth, Luczak, Pittel: The birth of the giant component. Random Structures Algorithms 4 (1993), no. 3, 231--358]. As said by Alan Frieze in his review [**MR**94h:05070]:

> *"This paper and its predecessor [**MR**90d:05184] mark the entry of generating functions into the general theory of random graphs in a significant way. Previously, their use had mainly been restricted to the study of random trees and mappings. Most of the major results in the area, starting with the pioneering papers of P. Erdös and A. Renyi [**MR**22#10924] have been proved without significant use of generating functions. However, at the early stages of the evolution of a random graph we find that it is usually not too far from being a forest, and this allows them an entry..."*

The icon was generated by Maple code like this:

```
f:=u^3/3-3*u+1;
plots[complexplot3d]([Re(f),argument(f)],
                  u=-4-3*I..4+3*I,style=patchcontour,
                           contours=30,numpoints=50*50);
```

Last modified: September 17, 2003. For problems or questions involving this site, please contact [Virginie Collette](#).

Algorithmica (1998) 22: 490–515

Algorithmica

On the Analysis of Linear Probing Hashing[1]

P. Flajolet,[2] P. Poblete,[3] and A. Viola[4]

Dedicated to Don Knuth on the occasion of the 35*th anniversary of his first analysis of an algorithm in* 1962–1963.

Abstract. This paper presents moment analyses and characterizations of limit distributions for the construction cost of hash tables under the linear probing strategy. Two models are considered, that of full tables and that of sparse tables with a fixed filling ratio strictly smaller than one. For full tables, the construction cost has expectation $O(n^{3/2})$, the standard deviation is of the same order, and a limit law of the Airy type holds. (The Airy distribution is a semiclassical distribution that is defined in terms of the usual Airy functions or equivalently in terms of Bessel functions of indices $-\frac{1}{3}, \frac{2}{3}$.) For sparse tables, the construction cost has expectation $O(n)$, standard deviation $O(\sqrt{n})$, and a limit law of the Gaussian type. Combinatorial relations with other problems leading to Airy phenomena (like graph connectivity, tree inversions, tree path length, or area under excursions) are also briefly discussed.

Key Words. Analysis of algorithms, Hashing, Linear probing, Parking problem, Airy functions.

Introduction. *Linear probing hashing*, defined below, is certainly the simplest "in place" hashing algorithm [14], [23], [38].

> A table of length m, $T[1\,..\,m]$ is set up, as well as a hash function h that maps keys from some domain to the interval $[1\,..\,m]$ of table addresses. A collection of n elements with $n \le m$ are entered sequentially into the table according to the following rule: Each element x is placed at the first unoccupied location starting from $h(x)$ in cyclic order, namely the first of $h(x), h(x)+1, \ldots, m, 1, 2, \ldots, h(x)-1$.

For each element x that gets placed at some location y, the circular distance between y and $h(x)$ (that is, $y - h(x)$ if $h(x) \le y$, and $m + h(x) - y$ otherwise) is called its *displacement*. Displacement is both a measure of the cost of inserting x and of the cost of searching x in the table. *Total displacement* corresponding to a sequence of hashed values is the sum of the individual displacements of elements. As it determines the *construction cost* of the table, we use both terms interchangeably.

We analyze here the total displacement $d_{m,n}$ of a table of length m (the number of table locations) and size n (the number of keys), under the assumption that all m^n hash

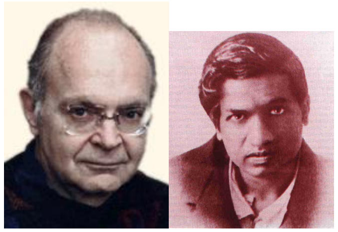
[1] The work of Philippe Flajolet was supported in part by the Long Term Research Project *Alcom-IT* (# 20244) of the European Union. The work of Patricio Poblete was supported in part by FONDECYT(Chile) under Grant 1960881. The work of Alfredo Viola was supported in part by proyecto BID-CONICYT 140/94 and proyecto CONICYT fondo Clemente Estable 2078/96.

[2] Algorithms Project, INRIA, Rocquencourt, 78150 Le Chesnay, France. Philippe.Flajolet@inria.fr.
[3] Department of Computer Science, University of Chile, Casilla 2777, Santiago, Chile. ppoblete@dcc.uchile.cl.
[4] Pedeciba Informatica, Casilla de Correo 16120, Distrito 6, Montevideo, Uruguay. viola@fing.edu.uy.

Received October 5, 1997; revised January 15, 1998. Communicated by H. Prodinger and W. Szpankowski.

Linear Probing Hashing

Philippe Flajolet, Patricio Poblete, and Alfredo Viola.
On the analysis of linear probing hashing.
Algorithmica, 1998.



is solving an old problem, first studied by Knuth, 1962
Knuth then decided to create the field "analysis of algorithms" and to write *The Art of Computer Programming*

$$\delta_z F(z,q) = \frac{F(z,q) - F(qz,q)}{1-q} F(z,q)$$

Displacement in parking functions, law of $D_{n+1,n}/n^{3/2}$ converges to the Airy-area distribution.

Reversing the time: Linear probing hashing ≈ fragmentation process (Bertoin, Pitman "additive coalescent", Rényi "parking functions", ...)

NOTES ON "OPEN" ADDRESSING.

D. Knuth 7/22/63

1. Introduction and Definitions.

Open addressing is a widely-used technique for keeping "symbol tables." The method was first used in 1954 by Samuel, Amdahl, and Boehme in an assembly program for the IBM 701. An extensive discussion of the method was given by Peterson in 1957 [1], and frequent references have been made to it ever since (e.g. Schay and Spruth [2], Iverson [3]). However, the timing characteristics have apparently never been exactly established, and indeed the author has heard reports of several reputable mathematicians who failed to find the solution after some trial. Therefore it is the purpose of this note to indicate one way by which the solution can be obtained.

We will use the following abstract model to describe the method: N is a positive integer, and we have an array of N variables $x_1, x_2, \ldots, x_N$. At the beginning, $x_i = 0$, for $1 \leqq i \leqq N$.

To "enter the k-th item in the table," we mean that an integer a_k is calculated, $1 \leqq a_k \leqq N$, depending only on the item, and the following process is carried out:

1. Set $j = a_k$.
2. "The comparison step." If $x_j = 0$, set $x_j = 1$ and stop; we say "the k-th item has fallen into position x_j."
3. If $j = N$, go to step 5.
4. Increase j by 1 and return to step 2.
5. "The overflow step." If this step is entered twice, the table is full, i.e. $x_i = 1$ for $1 \leqq i \leqq N$. Otherwise set j to 1 and return to step 2.

Observe the cyclic character of this algorithm.

We are concerned with the statistics of this method, with respect to the number of times the comparison step must be executed. More precisely, we consider all of the N^k possible sequences $a_1, a_2, \ldots a_k$ to be equally probable, and we ask, "What is the probability that the comparison step is used precisely m times when the k-th item is placed?"

2. Non-overflow (self-contained) sequences.

Let $\left[{n \atop k}\right]$ denote the number of sequences $a_1, a_2, \ldots a_k$ $(1 \leqq a_i \leqq n)$ in which no overflow step occurs during the entire process of placing k items, if the algorithm is used for $N = n$. (By convention, we set

$$\left[{0 \atop 0}\right] = 1.)$$

Lemma 1: If $0 \leq k \leq n+1$, then $\left[{n \atop k}\right] = (n+1)^k - k(n+1)^{k-1}$.

Proof: This proof is based on the fact that $\left[{n \atop k}\right]$ is precisely the number of sequences $b_1, b_2, \ldots b_k$ $(1 \leqq b_i \leqq n+1)$ in which, if the algorithm is carried out for $N = n+1$, then $x_{n+1} = 0$ at the end of the operation. This follows because every sequence of the former type is one of the latter, and conversely the condition implies in particular that $1 \leqq b_i \leqq n$, and that no overflow step occurs.

But sequences of the latter type are easily enumerated, because the algorithm has circular symmetry; of the $(n+1)^k$ possible sequences $b_1, b_2, \ldots b_k$, exactly $k/(n+1)$ of these leave $x_{n+1} \neq 0$. This shows that

$$\left[{n \atop k}\right] = (n+1)^k \left(1 - \frac{k}{n+1}\right).$$

MATHÉMATIQUES ET INFORMATIQUE

Hachage, arbres, chemins & graphes

Philippe Chassaing† et Philippe Flajolet‡

Mathématiques discrètes et continues se rencontrent et se complètent volontiers harmonieusement. C'est cette thèse que nous voudrions illustrer en discutant un problème classique aux ramifications nombreuses—l'analyse du hachage avec essais linéaires. L'exemple est issu de l'analyse d'algorithmes, domaine fondé par Knuth et qui se situe lui-même « à cheval » entre l'informatique, l'analyse combinatoire, et la théorie des probabilités. Lors de son traitement se croisent au fil du temps des approches très diverses, et l'on rencontrera des questions posées par Ramanujan à Hardy en 1913, un travail d'été de Knuth datant de 1962 et qui est à l'origine de l'analyse d'algorithmes en informatique, des recherches en analyse combinatoire du statisticien Kreweras, diverses rencontres avec les modèles de graphes aléatoires au sens d'Erdös et Rényi, un peu d'analyse complexe et d'analyse asymptotique, des arbres qu'on peut voir comme issus de processus de Galton-Watson particuliers, et, pour finir, un peu de processus, dont l'ineffable mouvement Brownien ! Tout ceci contribuant *in fine* à une compréhension très précise d'un modèle simple d'aléa discret.

1. Hachage

Nous ferons commencer l'histoire par Knuth en 1962 ; Knuth a alors 24 ans et hésite entre une carrière en mathématiques discrètes et une passion pour l'informatique très concrète. L'un des tous premiers problèmes quantitatifs de la science informatique naissante consiste à comprendre comment se comporte une certaine méthode d'accès à des données qui apparaît comme présentant, au vu des simulations, de très bonnes caractéristiques de complexité. L'École de Feller est aussi « sur le coup » (dixit Knuth). Knuth apportera très vite une solution[1] à ces questions (à partir de 1962) et, comme il le dit lui-même, c'est ce succès scientifique initial qui déterminera très largement la suite de sa carrière. Pour

†P. Chassaing : Institut Elie Cartan, Université Henri Poincaré B.P. 239, 54506 Vandoeuvre les Nancy Cedex
`Philippe.Chassaing@iecn.u-nancy.fr`
‡P. Flajolet : Projet ALGORITHMES, INRIA Rocquencourt, F-78153 Le Chesnay
`Philippe.Flajolet@inria.fr`

[1]Le manuscrit est disponible en `http://algo.inria.fr/AofA/Research/11-97.html`.

(ii) Des fonctions de parking aux partitions. Une fonction parking f de n voitures sur $n+1$ places est également décrite par une partition ordonnée de l'ensemble $\{1, 2, \ldots, n\}$ en n morceaux, notée $(B_k(f))_{1 \leq k \leq n}$, avec

$$B_k(f) = f^{-1}(k) \ ;$$

$B_k(f)$ est l'ensemble des voitures, de cardinal noté $x_k(f)$, dont la place k est la première tentative. Définissons $y_k(f)$ de manière analogue à (11) :

$$y_k(f) = x_1(f) + x_2(f) + \cdots + x_k(f) - k + 1; \tag{13}$$

$y_k(f)$ représente alors le nombre de voitures ayant tenté, avec ou sans succès, de se garer sur la $k^{\text{ème}}$ place. Le déplacement total coïncidant avec le nombre total de tentatives infructueuses, on a

$$y_1(f) + y_2(f) + \cdots + y_n(f) - n = D_{n,n+1}(f),$$

et, naturellement, le moment factoriel du déplacement total s'exprime par

$$E\left[\binom{D_{n,n+1}}{k}\right] = \frac{1}{(n+1)^{n-1}} \sum_f \binom{y_1(f) + y_2(f) + \cdots + y_n(f) - n}{k}, \tag{14}$$

où f parcourt l'ensemble des fonctions de parking. (Rappel : on a vu à la section 2 que $(n+1)^{n-1}$ dénombre les fonctions de parking.)

(iii) L'équivalence. Le point clé est que les ensembles de partitions « admissibles », d'une part l'ensemble des partitions $(A_k)_{1 \leq k \leq n}$ issues du parcours d'un arbre (ou plus généralement du parcours d'un graphe connexe) et d'autre part l'ensemble des partitions $(B_k)_{1 \leq k \leq n}$ issues d'une fonction parking, sont confondus. De fait, la condition pour qu'une partition soit admissible dans l'un ou l'autre des sens du terme, est que

$$y_k \geq 1, \qquad 1 \leq k \leq n. \tag{15}$$

En effet, pour un parcours d'arbre ou de graphe, y_k représente la longueur de la file d'attente après le $k^{\text{ème}}$ pas, et l'inégalité (15) traduit la contrainte de connexité sur le graphe ou l'arbre; pour une fonction parking, l'inégalité (15) traduit que la seule place vide est la place $n+1$ (ou 0, indifféremment).

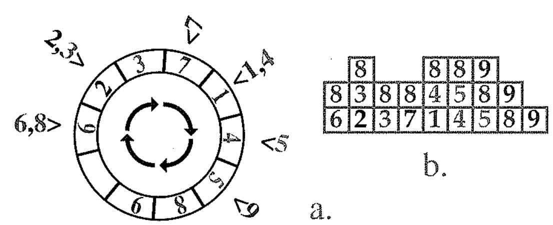


FIG. 3. Le parking induit par la partition ordonnée $(\{6,8\}, \{2,3\}, \varnothing, \{7\}, \{1,4\}, \{5\}, \{9\}, \varnothing, \varnothing)$, apparaissant déjà Figure 2 pour l'arbre γ (a), et les piles des voitures ayant essayé de se garer sur les places 1, 2, etc .., dans l'ordre chronologique (b).

ALGORITHME DU PARCOURS. On part du sommet 0. À tout instant, un sommet peut être dans deux états : l'état « inconnu » ou l'état « exploré ». Initialement, tout sommet (sauf la source) est « inconnu ». L'algorithme opère avec une file d'attente qui à chaque instant contient une liste de sommets déjà visités. Soit $\mathcal{F}_t$ l'état de la file à l'instant t. Initialement $t = 0$ et $\mathcal{F}_0 := \{0\}$. La « boucle » principale de l'algorithme consiste alors à répéter jusqu'à épuisement l'opération suivante :

> Au temps t ($t = 1, 2, \ldots$), choisir un sommet $s_t \in \mathcal{F}_{t-1}$; l'enlever de $\mathcal{F}_{t-1}$. Soit A_t l'ensemble des voisins de s_t (selon l'adjacence du graphe) qui sont encore « inconnus » ; on remplace alors dans $\mathcal{F}_{t-1}$ l'élément s_t par les éléments de A_t, de sorte que
>
> $$\mathcal{F}_t = (\mathcal{F}_{t-1} \setminus \{s_t\}) \cup A_t.$$
>
> Au moment où les nouveaux sommets sont insérés dans $\mathcal{F}_t$, leur état passe du statut « inconnu » au statut « exploré ».

Clairement, ce schéma permet de parcourir tous les sommets d'un graphe connexe en leur rendant visite une fois et une seule. De fait, ce schéma associe à un graphe connexe γ un *arbre couvrant* τ, l'arbre dont les arêtes relient s_t aux éléments de A_t, les arêtes de s_t vers les autres voisins étant inutilisées.

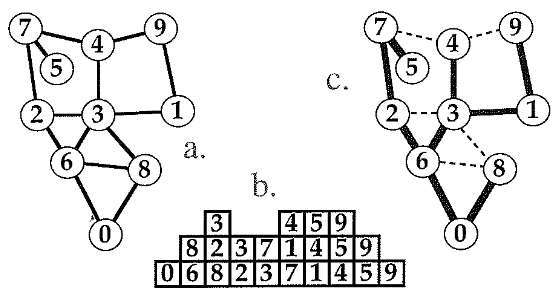


FIG. 2. Un graphe γ (a), les états $(\mathcal{F}_t)_{0 \le t \le 9}$ de la file (b) et l'arbre couvrant (c) associés au parcours en largeur de γ.

L'algorithme est complètement spécifié dès qu'une politique $\mathbf{\Pi}$ fixe le principe de sélection de s_t ainsi que l'ordre d'insertion de ses successeurs $s' \in A_t$. Les principes « LIFO » (*Last-In-First-Out* : on choisit le $s \in \mathcal{F}$ le plus récent, ce qui se gère par une pile) ou « FIFO » (*First-In-First-Out* : on sert le plus ancien dans la file) sont par exemple des politiques correspondant aux parcours appelés « en profondeur d'abord » et « en largeur d'abord ». Le principe par « priorité » (choix du plus petit ou plus grand numéro d'abord) est une autre politique particulièrement intéressante du point de vue combinatoire.

la suite des tailles et positions des blocs, complétée par une suite infinie de couples $(0,0)$. On considère les suites renormalisées

$$Z_\beta^{(m)}(t) = \frac{y_{\lceil mt\rceil}^{m,\lceil\beta\sqrt{m}\rceil}}{\sqrt{m}}, \qquad \Theta^{(m)}(\beta) = \frac{B^{m,\lceil\beta\sqrt{m}\rceil}}{m},$$
$$Z^{(m)} = \left(Z_\beta^{(m)}(t)\right)_{0\le\beta,\,0\le t\le 1}, \qquad \Theta^{(m)} = \left(\Theta^{(m)}(\beta)\right)_{\beta\ge 0}.$$

Ces suites constituent respectivement les « profils de visite » et « profils de placement » du parking.

Soit par ailleurs, comme précédemment, $(e(t))$ l'excursion Brownienne. On définit alors « l'excursion élaguée » construite selon la règle

$$\begin{aligned} Z_\beta(t) &= e(t) - \beta t + \sup_{t-1\leqslant s\leqslant t} (\beta s - e(s)), \\ Z &= (Z_\beta(t))_{0\le\beta,\,0\le t\le 1}, \end{aligned}$$

où β est un paramètre de contrôle. On note alors $\Theta(\beta)$ la suite des largeurs et positions des excursions du processus $t \to Z_\beta(t)$.

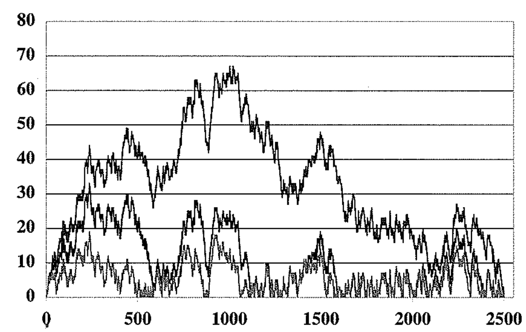


FIG. 5. $t \longrightarrow Z_\beta^{(m)}(t)$ pour $m = 2500$ places
et $\beta = 0, 2, 4$ successivement (1 puis 100 et 200 places vides)

Avec ces notations, en utilisant un couplage *parking–processus empirique*, on peut construire, sur un espace de probabilité approprié, des copies de $Z^{(m)}$ et de Z, telles que soit vérifié le principe suivant (qui généralise (16)) :

Convergence du profil. *Dans la région critique, le profil des visites du parking converge vers l'excursion Brownienne élaguée,*

$$\Pr\left(Z^{(m)} \to Z\right) = 1,$$

où $Z^{(m)}$ converge vers Z au sens de la convergence uniforme sur tout compact de $[0,+\infty) \times [0,1]$. En conséquence, les distributions finies-dimensionnelles du profil des placements du parking, $\Theta^{(m)}$,

Random Maps, Coalescing Saddles, Singularity Analysis, and Airy Phenomena

Cyril Banderier,[1] Philippe Flajolet,[1] Gilles Schaeffer,[2] Michèle Soria[3]
[1] *Algorithms Project, INRIA, Rocquencourt, 78150 Le Chesnay, France*
[2] *Loria, CNRS, Campus Sciences, B.P. 239, 54506 Vandœuvre-lès-Nancy, France*
[3] *Lip6, Université Paris 6, 8 rue du Capitaine Scott, 75005 Paris, France*

Received 10 March 2001; accepted 1 August 2001

ABSTRACT: A considerable number of asymptotic distributions arising in random combinatorics and analysis of algorithms are of the exponential-quadratic type, that is, Gaussian. We exhibit a class of "universal" phenomena that are of the exponential-cubic type, corresponding to distributions that involve the Airy function. In this article, such Airy phenomena are related to the coalescence of saddle points and the confluence of singularities of generating functions. For about a dozen types of random planar maps, a common Airy distribution (equivalently, a stable law of exponent $\frac{3}{2}$) describes the sizes of cores and of largest (multi)connected components. Consequences include the analysis and fine optimization of random generation algorithms for a multiple connected planar graphs. Based on an extension of the singularity analysis framework suggested by the Airy case, the article also presents a general classification of compositional schemas in analytic combinatorics. © 2001 John Wiley & Sons, Inc. Random Struct. Alg., 19, 194–246, 2001

Key Words: Airy function; analytic combinatorics; coalescing saddle points; multiconnectivity; planar map; random graph; random generation; singularity analysis; stable law

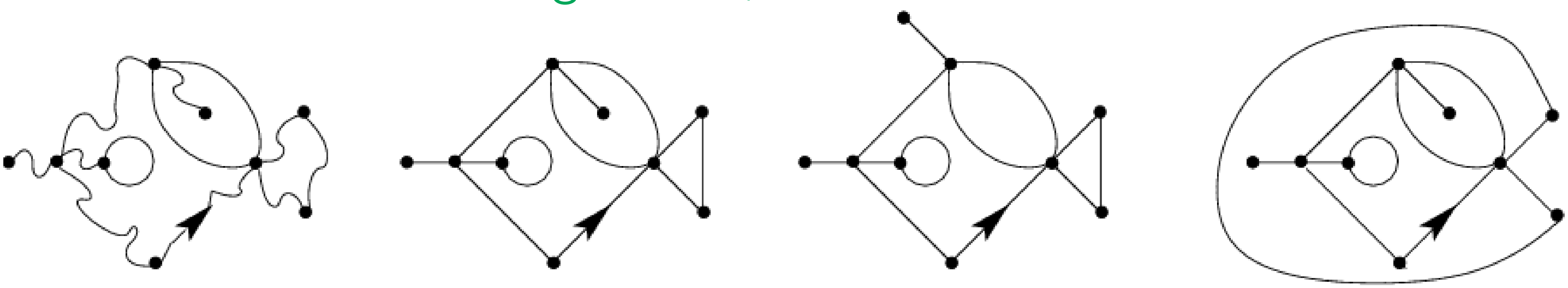
---

Correspondence to: Philippe Flajolet; e-mail: Philippe.Flajolet@inria.fr; http://algo.inria.fr/flajolet.
Contract grant sponsor: IST of EU.
Contract grant number: IST-1999-14186 (ALCOM-FT).

DOI 10.1002/rsa.10021

Random maps, coalescing saddles, and Airy

Cyril Banderier, Philippe Flajolet, Gilles Schaeffer, and Michèle Soria.
Random maps, coalescing saddles, singularity analysis, and Airy phenomena.
Random Structures & Algorithms, 2001.



map = planar graph on the sphere.
Tutte wanted to refute the 4 color theorem, not succeeded but found a way to enumerate maps (∼1960).

Random maps

Our problem:

how to generate (uniformly at random) a connected-map?

Random maps

Our problem:

how to generate (uniformly at random) a connected-map?

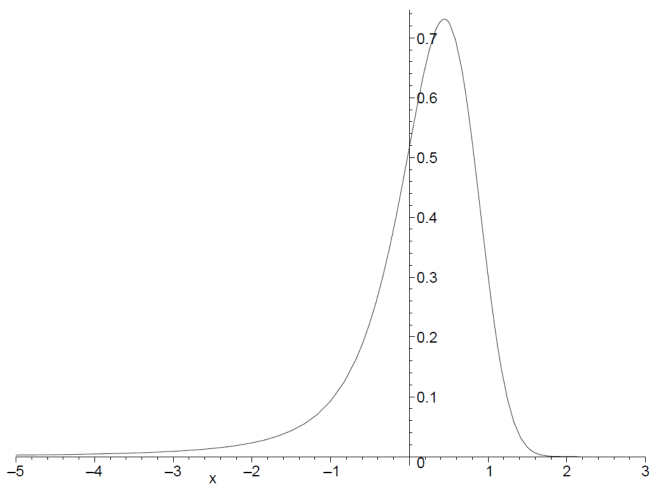
Solution: Use a rejection algorithm to reach size n
1. generate a map of size $f(n)$ 2. extract largest connected component C 3. if its size $\neq n$, then goto 1 4. output C.

Random maps

Our problem:

how to generate (uniformly at random) a connected-map?

Solution: Use a rejection algorithm to reach size n
1. generate a map of size $f(n)$ 2. extract largest connected component C 3. if its size $\neq n$, then goto 1 4. output C.



Theorem The fastest algorithm consists in choosing $f(n) = 3n - (3n)^{2/3}x_0$.
$x_0 := \arg\max(\mathcal{A}(x))$
$=$ peak of the Map-Airy distribution:
$\mathcal{A}(x) = 2\exp\left(-\frac{2}{3}x^3\right)\left(x\mathrm{Ai}(\mathrm{x}^2) - \mathrm{Ai}'(\mathrm{x}^2)\right)$
$\int_{-\infty}^{+\infty} \mathcal{A}(x)dx = 1$

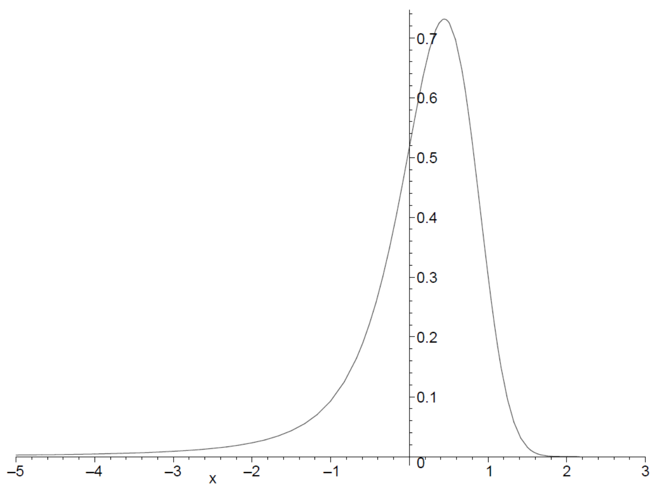
Random maps

Our problem:

how to generate (uniformly at random) a connected-map?

Solution: Use a rejection algorithm to reach size n

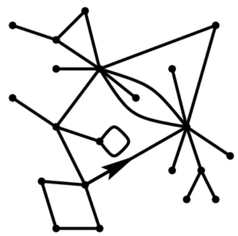
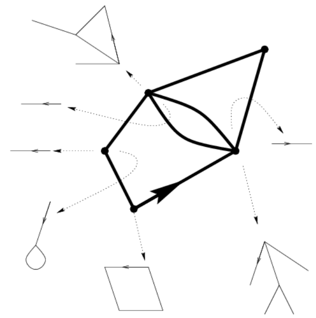
1. generate a map of size $f(n)$
2. extract largest connected component C
3. if its size $\neq n$, then goto 1
4. output C.



Theorem The fastest algorithm consists in choosing $f(n) = 3n - (3n)^{2/3} x_0$.
$x_0 := \arg\max(\mathcal{A}(x))$
$=$ peak of the Map-Airy distribution:
$\mathcal{A}(x) = 2\exp\left(-\frac{2}{3}x^3\right)\left(x\mathrm{Ai}(x^2) - \mathrm{Ai}'(x^2)\right)$
$\int_{-\infty}^{+\infty} \mathcal{A}(x)dx = 1$

Rejection with optimization of the "parameter" to reach size $\approx n$:
Flajolet's "Boltzmann method"

Proof of the rejection algorithms

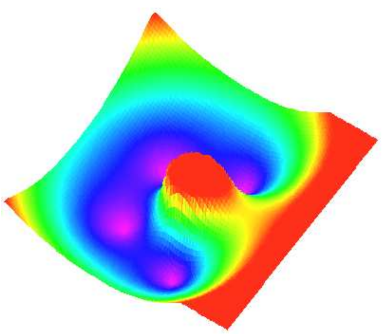




largest 2-connected component

functional equation (Tutte) via the "rootface decomposition"

$Mn, k :=$ # maps with n edges and face of degree k.

$$M(z,u) = \sum M_{n,k} z^n u^k = 1 + u^2 z M(z,u)^2 + uz \frac{M(z,1) - uM(z,u)}{1-u}$$



prob that a map of size n has a kernel of size k =

$$Pr(X_n = k) = \frac{M_{n,k}}{M_n} = C_k \frac{1}{2i\pi} \int_\gamma F(z) G(z)^k H(z)^n dz$$

$$\int_\Gamma \exp(n(a_0 + a_1(z-\tau) + a_2(z-\tau)^2 + a_3(z-\tau)^3 + \dots))dz$$

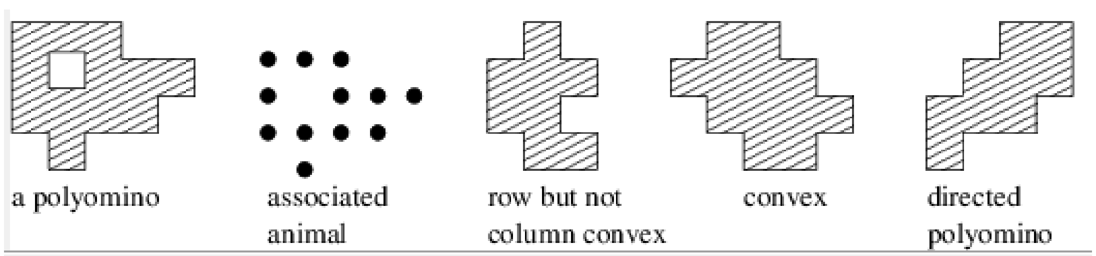
double saddle $\Leftrightarrow a_1 = 0$ et $a_2 = 0$, so one gets Airy.

Bonus: universal stable law for critical compositions, transition phase for 2-SAT?

The Area-Airy distribution

The moments of which are given by $\ln' \mathrm{Ai}(z)$,
density given by a sum of Airy zeroes weighted by an hypergeometric.

- area under a Brownian excursion [Louchard 1984, Takács 1993]
- cost of linear probing hashing [Flajolet–Poblete–Viola 1998]
- connexity of random graphs [Knuth/Flajolet/Spencer]
- area of polyominos [Prellberg 1995, Duchon 1999]
- statistical physics, self avoiding walks [Guttmann/Jensen 00's, Majumdar/Comtet 2005-07, Richard 2001-07, Rambeau/Schehr 2006, 2009. . .]
- additive parameters of context-free grammars (conjecture) , OK for Q-grammars [Duchon 1999, Fill–Kapur 2004]
- Area below discrete lattice paths [Banderier & Gittenberger, Schwerdtfeger 2011, Janson 2007]
- Cost of solving quadratic boolean systems, Gröbner basis computations [Bardet–Faugère–Salvy–Spaenlehauer 2005, 2011]



Algorithmica (2001) 31: 361–377
DOI: 10.1007/s00453-001-0056-0

Algorithmica

Analytic Variations on the Airy Distribution[1]

P. Flajolet[2] and G. Louchard[3]

Abstract. The Airy distribution (of the "area" type) occurs as a limit distribution of cumulative parameters in a number of combinatorial structures, like path length in trees, area below walks, displacement in parking sequences, and it is also related to basic graph and polyomino enumeration. We obtain curious explicit evaluations for certain moments of the Airy distribution, including moments of orders -1, -3, -5, etc., as well as $+\frac{1}{3}$, $-\frac{5}{3}$, $-\frac{11}{3}$, etc. and $-\frac{7}{3}$, $-\frac{13}{3}$, $-\frac{19}{3}$, etc. Our proofs are based on integral transforms of the Laplace and Mellin type and they rely essentially on "non-probabilistic" arguments like analytic continuation. A by-product of this approach is the existence of relations between moments of the Airy distribution, the asymptotic expansion of the Airy function $\mathrm{Ai}(z)$ at $+\infty$, and power symmetric functions of the zeros $-\alpha_k$ of $\mathrm{Ai}(z)$.

Key Words. Brownian excursion area, Airy function, Parking problem, Linear probing hashing.

Introduction. For probabilists, the *Airy distribution* considered here is nothing but the distribution of the *area* under the Brownian excursion. The name is derived from the connection between Brownian motion and the Airy function, a fact discovered around 1980 by several authors; see [16] and [20]. For combinatorialists and theoretical computer scientists, this Airy distribution (of the "area type") arises in a surprising diversity of contexts like parking allocations, hashing tables, trees, discrete random walks, merge-sorting, etc.

The most straightforward description of the Airy distribution is by its moments themselves defined by a simple nonlinear recurrence. We follow here the notations and the normalization of [11].

Definition 1. The Airy distribution (of the "area" type) is the distribution of a random variable $\mathcal{A}$ whose moments are

$$\mu_r \equiv E(\mathcal{A}^r) = \frac{-\Gamma(-\frac{1}{2})}{\Gamma((3r-1)/2)}\Omega_r, \qquad r \geq 1, \tag{1}$$

where the "Airy constants" Ω_r are determined by the quadratic recurrence

$$\Omega_0 = -1, \qquad 2\Omega_r = (3r-4)r\Omega_{r-1} + \sum_{j=1}^{r-1}\binom{r}{j}\Omega_j\Omega_{r-j} \qquad (r \geq 1). \tag{2}$$

[1] This research was partially supported by the IST Programme of the EU under Contract Number IST-1999-14186 (ALCOM-FT).

[2] INRIA, Domaine de Voluceau-Rocquencourt, BP 105, F-78153 Le Chesnay Cedex, France. Philippe.Flajolet@inria.fr.
[3] Département d'Informatique, Université Libre de Bruxelles, CP 212, Boulevard du Triomphe, B-1050 Bruxelles. louchard@ulb.ac.be.

Received June 6, 2000; revised February 17, 2001. Communicated by R. Kemp and H. Prodinger.
Online publication August 28, 2001.

Area-Airy distribution of the Brownian excursion area

The Area-Airy distribution is the distribution of a random variable $\mathcal{A}$ whose moments are

$$\mu_r \equiv E(\mathcal{A}^r) = \frac{-\Gamma(-1/2)}{\Gamma((3r-1)/2)}\Omega_r, \qquad r \geq 1,$$

where the "Airy constants" Ω_r satisfy the quadratic recurrence (more or less distant echo of some simple combinatorial tree decomposition):

$$\Omega_0 = -1, \qquad 2\Omega_r = (3r-4)r\Omega_{r-1} + \sum_{j=1}^{r-1} \binom{r}{j} \Omega_j \Omega_{r-j} \ \ (r \geq 1).$$

The normalized random variable $\frac{\mathcal{A}}{\sqrt{8}}$ is called Brownian excursion area.

r	0	1	2	3	4	5	6	7
Ω_r	-1	$\frac{1}{2}$	$\frac{5}{4}$	$\frac{45}{4}$	$\frac{3315}{16}$	$\frac{25425}{4}$	$\frac{18635625}{64}$	$\frac{18592875}{1}$
μ_r	1	$\sqrt{\pi}$	$\frac{10}{3}$	$\frac{15}{4}\sqrt{\pi}$	$\frac{884}{63}$	$\frac{565}{32}\sqrt{\pi}$	$\frac{662600}{9009}$	$\frac{19675}{192}\sqrt{\pi}$

Table: A table of the Airy constants Ω_r and of the Airy moments μ_r.

Area-Airy distribution: generating functions

The Airy constants Ω_r are characterized by any of the following expansions:

$$\frac{\mathrm{Ai}'(z)}{\mathrm{Ai}(z)} \underset{z\to+\infty}{\sim} \sum_{r=0}^{\infty} \frac{\Omega_r}{2^r} \frac{(-1)^r z^{-(3r-1)/2}}{r!}$$

We also have a linear recurrence on the Airy coefficients Ω_r:

$$18^r\Omega_r = \frac{12r}{6r-1}\frac{\Gamma(3r+1/2)}{\Gamma(r+1/2)} - \sum_{k=1}^{r-1}\binom{r}{k}\frac{\Gamma(3k+1/2)}{\Gamma(k+1/2)}18^{r-k}\Omega_{r-k}. \quad (1)$$

Area-Airy distribution: Laplace transform and density

Let $w(x)$ be the density function of the Area-Airy distribution:

$$w(x) = \frac{d}{dx}\mathbb{P}\{\mathcal{A} \leq x\},$$

the corresponding moment generating function is

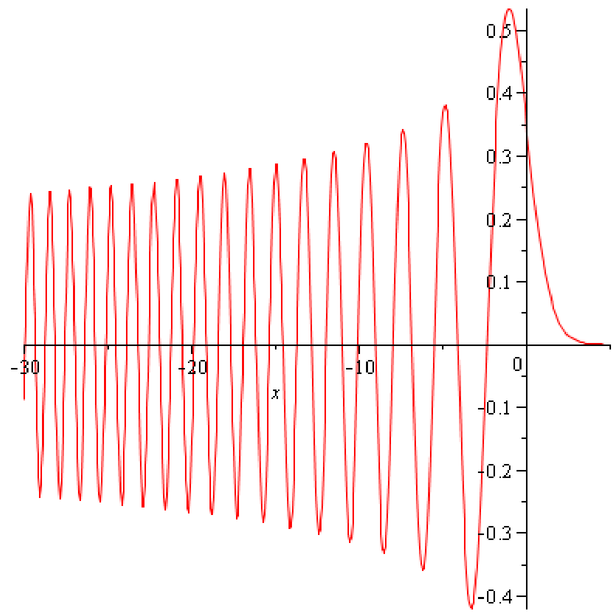
$$E\left[e^{-y\mathcal{A}}\right] = \sum_{r\geq 0} \mu_r \frac{(-y)^r}{r!} = \int_0^\infty e^{-yt}\, w(t)dt.$$

We know that the Area-Airy distribution function satisfies the double Laplace transform relation:

$$\frac{1}{\sqrt{2\pi}} \int_0^\infty \left(e^{-zy} - 1\right) E\left[e^{-y^{2/3}\frac{\mathcal{A}}{\sqrt{8}}}\right] \frac{dy}{y^{3/2}} = 2^{1/3}\left(\frac{\mathrm{Ai}'(2^{1/3}z)}{\mathrm{Ai}(2^{1/3}z)} - \frac{\mathrm{Ai}'(0)}{\mathrm{Ai}(0)}\right).$$

C:\P13\Cheetah\airyzeroes.mw - [Server 1] - Maple 14

```
> plot(AiryAi(x),x=-30..5);
```




```
> evalf(seq(AiryAiZeros(k),k=1..10));
```

$-2.338107410, -4.087949444, -5.520559828, -6.786708090, -7.944133587, -9.022650853, -10.04017434, -11.00852430, -11.93601556, -12.82877675$ **(1)**

```
>
```

Prêt   Mémoire: 41.42M Temps: 2.37s Mode Texte

The moment generating function and the density of the Area-Airy distribution are given by

$$E\left[e^{-y\frac{\mathcal{A}}{\sqrt{8}}}\right] = \sqrt{2\pi}y\sum_{k=0}^{\infty}\exp\left(-\alpha_k y 2^{-1/3}\right)$$

$$w(x) = \frac{8\sqrt{3}}{x^2}\sum_{k=1}^{\infty}e^{-v_k}\, v_k^{2/3}\, U\left(\frac{-5}{6},\frac{4}{3};v_k\right) \qquad v_k = \frac{16\alpha_k^3}{27x^2}.$$

There, the quantities $-\alpha_k$ are the zeros of the Airy function $\mathrm{Ai}(z)$ and $U(a,b;z)$ is the confluent hypergeometric function ($\approx$ the Map-Airy density, strange miracle!, no combinatorial explanation!).

The Airy zeros admit an asymptotic expansion of the form:

$$\alpha_k \sim \rho k^{2/3}\left(1+\sum_{j=1}^{\infty}\frac{a_j}{k^j}\right) \qquad \text{with} \quad \rho = \left(\frac{3\pi}{2}\right)^{\frac{2}{3}}$$

and the expansion starts as

$$\alpha_k \sim \rho k^{2/3}\left(1-\frac{1}{6k}-\frac{\rho^3-15}{144\rho^3k^2}-\frac{\rho^3-45}{1296\rho^3k^3}-\cdots\right).$$

The moments

In what follows, an essential rôle is played by what may be called the "root zeta function" of the Airy function. This function is defined by

$$\Lambda(s) := \sum_{k=1}^{\infty} (\alpha_k)^{-s} \qquad (\Re(s) > \frac{3}{2}),$$

where the sum is defined for $\Re(s) > 3/2$, given the growth of the α_k.

But via the convergent-divergent trick, we can get an analytic continuation of $\Lambda(s)$ on $\mathbb{C}$:

$$\Lambda(s) = \sum_{k\geq 1} (3\pi k/2)^{-2s/3} + \sum_{k\geq 1} \left((\alpha_k)^{-s} - (3\pi k/2)^{-2s/3} \right)$$

now extends by analytic continuation to $\Re(s) > 0$ (and so on for $s \in \mathbb{C}$).

Using Mellin transforms, we get that the moments of the Area-Airy distribution exist for any $s \in \mathbb{C}$ and satisfy

$$\mathbf{E}\left[\left(\frac{\mathcal{A}}{\sqrt{8}}\right)^s\right] = 3\sqrt{\pi}2^{-s/2}\frac{\Gamma(\frac{3}{2}(1-s))}{\Gamma(-s)}\Lambda\left(\frac{3}{2}(1-s)\right),$$

Moments of negative order $\approx$ expansion of $\mathrm{Ai}(z)$ at 0.
Moments of positive order $\approx$ expansion of $\mathrm{Ai}(z)$ at $+\infty$, or $-\infty$.

$$E\left[\left(\frac{\mathcal{A}}{\sqrt{8}}\right)^{-5/3}\right] = \frac{9\sqrt{\pi}2^{5/6}}{\Gamma(1/3)^7}\left(3^{1/3}\Gamma(1/3)^6 - 8\ 3^{5/6}\pi^3\right).$$

(...and complex analysis gives full access to a lot of informations on this Airy distribution: tales behaviour, large deviations, etc.)

Contents

- Airy function. Links with Physics (Schrödinger)
- Airy distributionS
- Tracy-Widom "Airy"-distribution and statistical mechanics
- The first cycle in random graphs
- Connected graphs
- The analysis of linear probing hashing (Knuth-Flajolet)
- Map-Airy distribution (Random maps, coalescing saddles)
- Area-Airy distribution (Brownian, q-grammars)
- Airy root zeta function

Crossing of many fields, crossing of different methods:
still a lot of mysterious phenomena!
(complex analysis (saddles) versus probability theory versus replica method, Airy root zeta function..., asymptotics of unsolvable functional equations from combinatorics)