Algorithms Seminar 1999–2000,
F. Chyzak (ed.), INRIA, (2000), pp. 17–22.

Available online at the URL
http://algo.inria.fr/seminars/.

Enumeration of Geometric Configurations on a Convex Polygon

Marc Noy
Departament de Matemàtica Aplicada II, Universitat Politècnica de Catalunya

December 16, 1999

Summary by Michel Nguyễn-Thế

Abstract

We survey recent work on the enumeration of non-crossing configurations on the set of vertices of a convex polygon, such as triangulations, trees, and forests. Exact formulæ and limit laws are determined for several parameters of interest. In the second part of the talk we present results on the enumeration of chord diagrams (pairings of $2n$ vertices of a convex polygon by means of n disjoint pairs). We present limit laws for the number of components, the size of the largest component and the number of crossings. The use of generating functions and of a variation of Levy's continuity theorem for characteristic functions enable us to establish that most of the limit laws presented here are Gaussian. (Joint work by Marc Noy with Philippe Flajolet and others.)

1. Analytic Combinatorics of Non-crossing Configurations [3]

1.1. **Connected graphs and general graphs.** Let $\Pi_n = \{v_1, \ldots, v_n\}$ be a fixed set of points in the plane, conventionally ordered counter-clockwise, that are vertices of a regular n-gon K. Define a *non-crossing graph* as a graph with vertex set Π_n whose edges are straight line segments that do not cross. A graph is *connected* if any two vertices can be joined by a path. Parameters of interest are the number of edges of connected graphs and general graphs, and the number of components of general graphs.

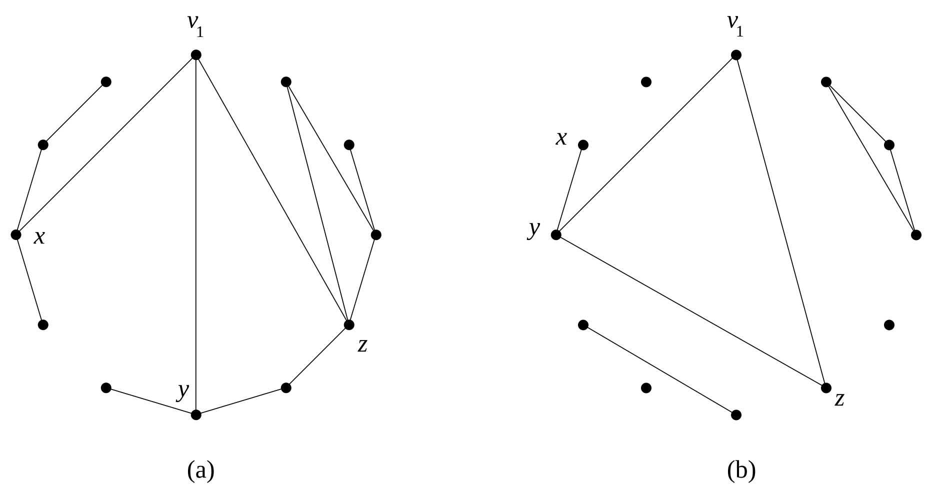

FIGURE 1. (a) A connected non-crossing graph; (b) an arbitrary non-crossing graph.

1.2. **Trees and forests.** A *(general) tree* is a connected acyclic graph and the number of edges in a tree is one less than the number of vertices. The study of trees becomes easier with the introduction of *butterflies* [3], defined to be ordered pairs of trees with a common vertex; a tree appears to be a sequence of butterflies attached to a root. A *forest* is an acyclic graph, in other words a graph whose components are trees.

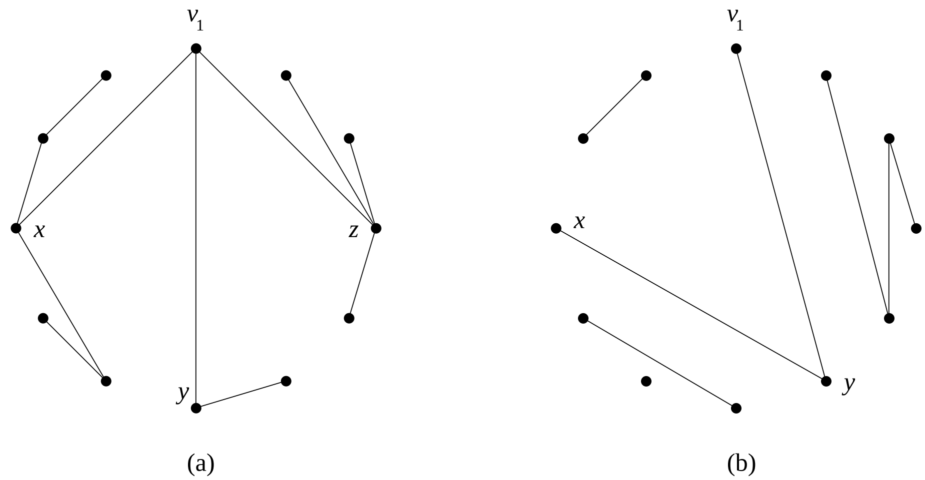

FIGURE 2. (a) A tree; (b) a forest.

1.3. **Triangulations.** A *triangulation* [7] is a set $\mathcal{T}_n$ of $n-3$ non-crossing diagonals v_iv_j which partitions K into $n-2$ triangles. As each triangle corresponds to an internal node of a binary tree (see the generating function of exercise 7.22 of [6]) via a classical bijection due to Euler [11], the number $\widehat{T}_n$ of triangulations is given by the $(n-2)$-th Catalan number $\widehat{T}_n = C_{n-2} = \binom{2n-4}{n-2}/(n-1)$. Let d_i denote the *degree* of the vertex v_i (i.e., the number of diagonals incident with v_i) and $\|v_iv_j\| = \min\big(|i-j|, n-|i-j|\big)$ the length of a diagonal v_iv_j. Define [2]:

$$\Delta_n(\tau) = \max\left\{\, d_i \;\middle|\; i = 0, \ldots, n-1 \,\right\},$$

the *maximal degree* of the vertices, and

$$\lambda_n(\tau) = \max\left\{\, \|v_iv_j\| \;\middle|\; v_iv_j \in \mathcal{T}_n \,\right\},$$

the *length of the longest diagonal* in the triangulation.

Those features are of interest for a triangulation τ because they convey information about the corresponding tree $b(\tau)$: $\Delta_n(\tau)$ measures the *external-node separation* of $b(\tau)$, i.e., the maximal distance between successive external nodes; $\lambda_n(\tau)$ measures its *nearly half* measure, i.e., the size of the largest subtree with not more than half the external nodes.

Using combinatorial bijections and probability lemmas [2], we find:

$$\mathbf{E}[\Delta_n] \sim \log_2 n, \;\text{ and } \mathrm{E}[\lambda_n] \sim \alpha n, \quad \text{where} \quad \alpha = \frac{\sqrt{3}}{\pi} + \frac{1}{3} - \frac{\log\left(2+\sqrt{3}\right)}{\pi} \simeq 0.4654.$$

Let an *ear* of a triangulation τ be a triangle sharing two sides with the polygon, and e_n the number of ears of a triangulation. Let us view triangulations as binary trees and ears as leaves (internal node whose children are external nodes [11]) or roots with at least one child that is an external node, and let $\widehat{B}$ enumerate binary trees by size and number of leaves and $\widehat{T}$ enumerate triangulations by size and number of ears.[1] These generating series satisfy [5]

$$z^2\widehat{T}(z,w) = \big(1+2z(w-1)\big)\widehat{B}(z,w), \quad \text{where} \quad \widehat{B}(z,w) = z\big(w + 2\widehat{B}(z,w) + \widehat{B}(z,w)^2\big),$$

leading to $\mathbf{Var}[e_n] \sim \sqrt{n}/4$ and a Gaussian limit law (see §1.5 below). The expectation

$$\mathbf{E}[e_n] = \frac{n(n-1)}{2(2n-5)} \sim \frac{n}{4}$$

was already known from a combinatorial manipulation of Catalan numbers described in [7].

1.4. **Generating functions.** The combinatorial objects and parameters above, except for extremal ones, lead to univariate and bivariate generating functions, given in Table 1 below.

Configuration	Generating function equation
Connected graphs	$C^3 + C^2 - 3zC + 2z^2 = 0$
—–, edges	$wC^3 + wC^2 - (1+2w)zC + (1+w)z^2 = 0$
Graphs	$G^2 + (2z^2 - 3z - 2)G + 3z + 1 = 0$
—–, edges	$wG^2 + \big((1+w)z^2 - (1+2w)z - 2w\big)G + w + (1+2w)z = 0$
—–, components	$G^3 + (2w^3z^2 - 3w^2z + w - 3)G^2 + (3w^2z - 2w + 3)G + w - 1 = 0$
Trees	$T^3 - zT + z^2 = 0$
—–, leaves	$T^3 + (z^2w - z^2 - z)T + z^2 = 0$
Forests	$F^3 + (z^2 - z - 3)F^2 + (z+3)F - 1 = 0$
—–, components	$F^3 + (w^3z^2 - w^2z - 3)F^2 + (w^3z + 3)F - 1 = 0$
Triangulations	$z^4\widehat{T}^2 + (2z^2 - z)\widehat{T} + 1 = 0$
—–, ears	$z^4\widehat{T}^2 + \big(1+2z(w-1)\big)(2z^2 - z)\widehat{T} + w\big(1+2z(w-1)\big)^2 = 0$

TABLE 1. Generating function equations (z and w mark vertices and the secondary parameter).

A few tricks enable one to make Lagrange inversion applicable and to derive exact formulæ—sometimes involving summations—for all coefficients. For example, the change of variable $T = z+zy$ followed by Lagrange's formula yields:

$$T_n = \frac{1}{2n-1}\binom{3n-3}{n-1} \quad \text{and} \quad T_{n,k} = \frac{1}{n-1}\binom{n-1}{k}\sum_{j=0}^{k-1}\binom{n-1}{j}\binom{n-k-1}{k-1-j}2^{n-2k+j}.$$

Finding $C_{n,k}$ goes through a parameterization of the functional equation of C. To get the coefficients $\widehat{T}$, we use the equality $\widehat{T}_{n,k} = \widehat{B}_{n+2,k-1} + 2\widehat{B}_{n+1,k} - 2\widehat{B}_{n+1,k}$ deduced from $z^2\widehat{T}(z,w) = \big(1+2z(w-1)\big)\widehat{B}(z,w)$.

1.5. **Asymptotics.** All of the univariate generating functions above, and a few others (dissections and partitions of convex polygons) not presented in the talk but available in [3], have a unique dominant singularity ρ in $(0,1)$, and can be written

$$f(z) = c_0 + c_1\left(1-\frac{z}{\rho}\right)^{1/2} + O\left(1-\frac{z}{\rho}\right), \quad \text{entailing} \quad [z^n]f(z) = \frac{c_1}{\Gamma(-1/2)}\left(1+O\left(\frac{1}{n}\right)\right).$$

For example the numbers T_n and F_n of respectively general trees and forests satisfy

$$T_n \asymp (27/4)^n = 6.75^n \quad \text{and} \quad F_n \asymp 8.2246^n, \quad \text{whence} \quad T_n = o(F_n).$$

[1]The expression of $\widehat{T}$, entailing the Gaussian limit of the distribution of ears of triangulations, was established by the author of this summary.

The numbers C_n and G_n of respectively connected and general graphs satisfy

$$C_n \sim \left(\frac{\sqrt{6}}{9} - \frac{\sqrt{2}}{6}\right) \asymp 10.39^n \quad \text{and} \quad G_n \sim \frac{1}{4}\sqrt{99\sqrt{2} - 140} \times \frac{2^n(3 + 2 + \sqrt{2})^n}{\sqrt{\pi}n^{3/2}} \asymp 11.65^n,$$

entailing $C_n/G_n \to 0$ when $n \to \infty$.

The bivariate generating function seen before admits the form

$$f(z, w) = c_0(w) + c_1(w)\left(1 - \frac{z}{\rho(w)}\right)^{1/2} + O\left(1 - \frac{z}{\rho(w)}\right);$$

this leads to

$$f_n(w) = \gamma(w)\left(\frac{1}{\rho(w)}\right)^n\left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right), \text{ or } \frac{f_n(w)}{f_n(1)} = \frac{\gamma(w)}{\gamma(1)}\left(\frac{\rho(1)}{\rho(w)}\right)^n\left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right).$$

From the Quasi-Powers theorem [5, 8], which is a consequence of Levy's continuity theorem for characteristic functions, one deduces that f_n is asymptotically normal. The mean μ_n and variance σ_n satisfy $\mu_n \sim \kappa n$ and $\sigma_n^2 \sim \lambda n$ for algebraic numbers κ and λ.

For instance, for the distribution of the number of edges in the space of connected graphs of given size, we have $\kappa = \left(1 + \sqrt{3}\right)/2 \simeq 1.366$.

2. Analytics Combinatorics of Chord Diagrams [4]

2.1. **Definitions.** Take $2n$ points on a circle, labelled 1, 2, ..., $2n$, and join them in disjoint pairs by n chords. The resulting configuration is called a *chord diagram.* A diagram is *connected* if no set of chords can be separated from the remaining chords by a line. A *component* is a maximal connected subdiagram.

2.2. Components.

2.2.1. *Number of components.* Let $C(z) = \sum_{n \geqslant 0} C_n z^n$ be the generating function of connected diagrams of size n. The bivariate generating function $I(z, w) = \sum_{n,k \geqslant 0} I_{n,k} w^k z^n$ of diagrams of size n and k components satisfies $I(z, w) = 1 + wC\left(zI(z, w)^2\right)$.

We have the following result:

Theorem 1. *Let X_n be the number of components in a random diagram of size n.*

1. *For $k \geqslant 1$, one has* $\mathbf{P}[X_n = k] \underset{n\to\infty}{=} \dfrac{e^{-1}}{(k-1)!}\left(1 + o(1)\right)$.
2. *The mean μ_n and the variance σ_n of the distribution satisfy* $\mu_n \underset{n\to\infty}{\sim} 2$ *and* $\sigma_n^2 \underset{n\to\infty}{\sim} 1$.

Sketch of proof. The proof of the first point makes use of "monoliths," or "monolithic diagrams," where a diagram is said to be monolithic if: *(i)* it consists solely of the connected component that contains 1 (called the root component) and of isolated edges; *(ii)* for any two such isolated edges (a, b) and (c, d), one never has $a < c < d < b$ or $c < a < b < d$ (in other words, two isolated chords are never in a dominance relation).

The ordinary generating function of monoliths reads $M(z) = C\left(z/(1 - z)^2\right)$, and according to Stein and Everett [12] $C_n/I_n = e^{-1} + o(1)$, so one can deduce the relation $M_n \sim I_n$, i.e., that almost every diagram is a monolith. The number $M_{n,k}$ of monoliths of size n with k components is given by

$$M_{n,k} = \binom{2n - k}{k - 1} C_{n-k+1} \sim \frac{e^{-1}}{(k - 1)!} I_n.$$

As to the second point, using $2zC(z)C'(z) = C(z)^2 + C(z) - z$, which is deduced from

$$C_n = (n-1)\sum_{j=1}^{n-1} C_j C_{n-j}$$

and $C_1 = 1$ [9, 13], one finds

$$\mu_n = \frac{\partial I}{\partial w}(z,w)\Big|_{w=1} = \frac{1}{z}\big(I(z) + h(z) - 2\big), \quad \text{where} \quad h(z) = I(z)^{-1}.$$

Hence, letting $g_n = h_n/I_n$, one obtains

$$g_n = 1 - \sum_{k=1}^{n-1} g_k \binom{n}{k}\binom{2n}{2k}^{-1} = 1 - \frac{1}{n} + \frac{3}{4n^2} + O(n^{-3}),$$

and $\mu_n = \dfrac{I_{n+1} + h_{n+1}}{I_n} = \dfrac{2n+1}{n+1} + O(n^{-1}) \sim 2$. Similar computations yield the variance. □

2.2.2. *Largest connected component.*

Theorem 2. *Let L_n be the size of the largest connected component in a random diagram of size n. Then, as $n \to \infty$, the mean μ_n and the variance σ_n of the distribution of L_n are*

$$\mathbf{E}[L_n] = n - 1 + o(1), \qquad \mathbf{Var}[L_n] = 1 + o(1),$$

and for any fixed $k \geqslant 1$, one has $\mathbf{P}[n - L_n = k] = \dfrac{e^{-1}}{k!}\big(1 + o(1)\big)$. In other words, thre random variable $n - L_n$ follows a Poisson law of parameter 1.

The proof relies on the analysis of the largest component in a monolith, namely, the root component with probability $1 - o(1)$, the other components being only edges. The number $M_{n,k}$ of monoliths of size n with root component of size $n-k$ is given by:

$$M_{n,k} = \binom{2n-k-1}{k} C_{n-k} \underset{n\to\infty}{\sim} \frac{e^{-1}}{(k-1)!} I_n.$$

2.3. **Crossings.** Let κ denote the number of chord crossings in a chord diagram, and let $\mathcal{I}_n$ be the set of all diagrams of size n. Flajolet and Noy proved the following result:

Theorem 3. *Let X_n be the random variable equal to the value of κ taken over the set of chord diagrams $\mathcal{I}_n$ of size n endowed with the uniform probability distribution.*

1. *The mean μ_n and the variance σ_n of the distribution of X_n are given by*

$$\mu_n = \mathbf{E}[X_n] = \frac{n(n-1)}{6} \quad \textit{and} \quad \sigma_n^2 = \mathbf{Var}[X_n] = \frac{n(n-1)(n+3)}{45}, \quad \textit{respectively.}$$

2. *The distribution of X_n is Gaussian in the asymptotic limit: for all real x, one has*

$$\lim_{n\to\infty} \mathbf{P}\left[\frac{X_n - \mu_n}{\sigma_n} \leqslant x\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy.$$

Sketch of proof. Flajolet and Noy prove a stronger result by computing the moments of any order. They use the *exact formula* discovered by Touchard [14] and Riordan [10], namely that the series

$$\phi_n(q) = \sum_{w\in\mathcal{I}_n} q^{\kappa(w)} \quad \text{equals} \quad \frac{1}{(1-q)^n} \sum_{k=-n}^{n} (-1)^k q^{k(k-1)/2} \binom{2n}{n+k}.$$

Using the equality $e^{a^2/2} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2} e^{ax}\, dx$ for $a = k\sqrt{t}$, one obtains:

$$\phi_n(e^t) = \frac{1}{(1-e^t)^n} \sum_{k=-n}^{n} (-1)^k e^{-kt/2} \binom{2n}{n+k} e^{k^2 t/2} = \frac{1}{2\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2} x^{2n} H(x,t)^n\, dx,$$

where $H(x,t) = \dfrac{2\sinh^2(x\sqrt{t}/2 - t/4)}{x^2 \exp(t/2)\sinh(t/2)}$.

Taking derivatives with respect to t and taking the limit when $t \to 0$ yields the moments of any order; this proves the first point of the claim.

The Laplace method delivers the asymptotic relation

$$e^{-u\mu_n/\sigma_n} \frac{\phi_n(u/\sigma_n)}{\phi_n(1)} = e^{u^2/2}\big(1 + O(n^{-1/5})\big).$$

From Levy's continuity theorem for Laplace transforms [1], one concludes that $(X_n - \mu_n)/\sigma_n$ converges in distribution towards $\mathcal{N}(0,1)$. $\square$

Bibliography

[1] Billingsley (Patrick). – *Probability and measure.* – John Wiley & Sons Inc., New York, 1995, third edition, xiv+593p. A Wiley-Interscience Publication.

[2] Devroye (L.), Flajolet (P.), Hurtado (F.), Noy (M.), and Steiger (W.). – Properties of random triangulations and trees. *Discrete & Computational Geometry*, vol. 22, n° 1, 1999, pp. 105–117.

[3] Flajolet (Philippe) and Noy (Marc). – Analytic combinatorics of non-crossing configurations. *Discrete Mathematics*, vol. 204, n° 1-3, 1999, pp. 203–229.

[4] Flajolet (Philippe) and Noy (Marc). – Analytic combinatorics of chord diagrams. In Krob (D.), Mikhalev (A. A.), and Mikhalev (A. V.) (editors), *Formal Power Series and Algebraic Combinatorics.* pp. 191–201. – Springer Verlag, 2000. Proceedings of FPSAC'2000, June 2000, Moscow.

[5] Flajolet (Philippe) and Sedgewick (Robert). – *The average case analysis of algorithms: multivariate asymptotics and limit distributions.* – Research Report n° 3162, Institut National de Recherche en Informatique et en Automatique, 1997. 123 pages.

[6] Graham (Ronald L.), Knuth (Donald E.), and Patashnik (Oren). – *Concrete mathematics.* – Addison-Wesley Publishing Co., Reading, MA, 1989, xiv+625p. A foundation for computer science.

[7] Hurtado (F.) and Noy (M.). – Ears of triangulations and Catalan numbers. *Discrete Mathematics*, vol. 149, n° 1-3, 1996, pp. 319–324.

[8] Hwang (Hsien-Kuei). – *Théorèmes limites pour les structures combinatoires et les fonctions arithmétiques.* – Thèse de doctorat, École polytechnique, December 1994.

[9] Nijenhuis (Albert) and Wilf (Herbert S.). – The enumeration of connected graphs and linked diagrams. *Journal of Combinatorial Theory. Series A*, vol. 27, n° 3, 1979, pp. 356–359.

[10] Riordan (John). – The distribution of crossings of chords joining pairs of $2n$ points on a circle. *Mathematics of Computation*, vol. 29, 1975, pp. 215–222. – Collection of articles dedicated to Derrick Henry Lehmer on the occasion of his seventieth birthday.

[11] Sedgewick (Robert) and Flajolet (Philippe). – *An introduction to the analysis of algorithms.* – Addison-Wesley Publishing Co., Reading, MA, 1996.

[12] Stein (P. R.) and Everett (C. J.). – On a class of linked diagrams. II. Asymptotics. *Discrete Mathematics*, vol. 21, n° 3, 1978, pp. 309–318.

[13] Stein (Paul R.). – On a class of linked diagrams. I. Enumeration. *Journal of Combinatorial Theory. Series A*, vol. 24, n° 3, 1978, pp. 357–366.

[14] Touchard (Jacques). – Sur un problème de configurations et sur les fractions continues. *Canadian Journal of Mathematics*, vol. 4, 1952, pp. 2–25.