Distribution of image points in random mappings

Michèle Soria
Université Paris VI

November 10, 1996

[summary by Pierre Nicodème]

Abstract

This talk presents a general theorem which can be used to identify the limiting distribution for a class of combinatorial schemata. For example, many parameters in random mappings can be covered in this way.

1. Methods

We consider the general working scheme "Symbolic Structures $\mathcal{A}$ or $\{\mathcal{A}, \omega\} \to$ Generating Functions $a(z)$ or $a(u,z) \to a_n$ or $a_{n,k}$". Then by Cauchy's formula, we get for structures $\mathcal{A}$

$$a(z) = \sum_{\alpha \in \mathcal{A}} \frac{z^{|\alpha|}}{|\alpha|!} = \sum_{n \geq 0} a_n \frac{z^n}{n!} \Longrightarrow \frac{a_n}{n!} = \frac{1}{2i\pi} \oint a(z) \frac{dz}{z^{n+1}}.$$

When considering marked structures with parameters $\{\mathcal{A}, \omega\}$, (ω is a mapping $\mathcal{A} \to \mathbb{N}$), we have

$$a(u,z) = \sum_{\alpha \in \mathcal{A}} u^{\omega(\alpha)} \frac{z^{|\alpha|}}{|\alpha|!} = \sum_{n,k} a_{n,k} u^k \frac{z^n}{n!}.$$

In this case, $a_{n,k}$ can be obtained by double Cauchy inversion, or by Cauchy inversion and Continuity Theorem. Table 1 gives some examples of translation of marked combinatorial structures to generating functions. The mark is represented by character "$\bullet$" and translated to parameter u.

By a classical theorem about characteristic functions (X_n) converges weakly to Y if and only if $\phi_{X_n}(\theta)$ converges to $\phi_Y(\theta)$ for all θ, with $\phi_Z = E(e^{i\theta Z})$. We also have $a(u,z) = \sum_{n,k} u^k \frac{z^n}{n!} = \sum_n p_n(u) \frac{z^n}{n!}$, which gives the probability generating function of X_n as $p_n(u)/p_n(1) = \sum_n \Pr(X_n = k) u^k$. We refer to [2] for the concept of (labelled) combinatorial structures and their translation to generating functions.

Description	Structure	Generating Function
Degree at the root in Cayley trees	$\mathcal{A} = \text{Node} \times \text{Set}(\bullet\mathcal{A})$	$a(u,z) = z\exp(ua(z))$
Random Mappings	$\mathcal{G} = \text{Set}(\text{Cycle}(\mathcal{A}))$	$g(z) = \frac{1}{1-a(z)}$
— by number of cycles	$\mathcal{G} = \text{Set}(\bullet\, \text{Cycle}(\mathcal{A}))$	$g(u,z) = \exp\left(u \log \frac{1}{1-a(z)}\right)$
— by number of trees	$\mathcal{G} = \text{Set}(\text{Cycle}(\bullet\mathcal{A}))$	$g(u,z) = \frac{1}{1-ua(z)}$

Table 1. Some examples of generating functions

2. Trees and Random Mappings

A random mapping is an arbitrary mapping $\phi : \{1, \ldots, n\} \to \{1, \ldots, n\}$ such that every mapping has probability n^{-n}. A mapping ϕ can be identified to its functional graph G_ϕ with vertices $\{1, \ldots, n\}$ and edges $(i, \phi(i))$, for $1 \geq i \geq n$. Each component of G_ϕ consists of a cycle and every cyclic point is the root of a tree.

The basic property for analysis is that solutions of functional equations usually have algebraic singularity of square-root type. For trees, we get $a(u,z) = t(u,z) - h(u,z)\sqrt{1 - z/\rho(u)}$. For sequences of trees, we get an expression of the form $1/(1 - a(u,z))$, and for random mappings an expression of the form

$$s(u,z) = \frac{1}{1 - T(u,z) + h(u,z)\sqrt{1 - z/\rho(u)}}.$$

We recall that when we get an expression of the form $1/(1 - uC(z))$, the asymptotic distribution of the corresponding random variable depends on the value $C(\rho_c)$, where ρ_c is the only singularity on the circle of convergence of $C(z)$. If $C(\rho_c) > 1$, the limit law is normal; if $C(\rho_c) < 1$, the limit law is derivative of geometric, and if $C(\rho_c) = 1$ the limit law is Rayleigh.

3. Examples

Leaves. For Cayley trees, we have $a(u,z) = ze^{a(u,z)} + z(u-1)$, for sequences of trees, $s(u,z) = 1/(1 - a(u,z))$, and for functional graphs

$$g(u,z) = \frac{1}{1 - ze^{a(u,z)}} = \frac{1}{1 - a(u,z) + z(u-1)}.$$

Nodes of arity r. For trees,

$$a(u,z) = z\left(\sum_{m \neq r} \frac{a^m(u,z)}{m!} + u\frac{a^r(u,z)}{r!}\right) = ze^{a(u,z)} + z(u-1)\frac{a^r(u,z)}{r!}.$$

For sequences of trees, we have $s(u,z) = 1/(1 - a(u,z))$, and for functional graphs,

$$g(u,z) = \frac{1}{1 - a(u,z) + z(u-1)\left(\frac{a^{r-1}}{(r-1)!} - \frac{a^r}{r!}\right)}.$$

Nodes at distance d from a cycle. We have the recurrence

$$a_0(u,z) = ua(z), \qquad a_{d+1}(u,z) = ze^{a_d(u,z)}.$$

For functional graphs, this gives $g(u,z) = 1/(1 - a_d(u,z))$.

Nodes with r pre-images in total. For trees, we have $a(u,z) = ze^{a(u,z)} + (u-1)\alpha_{r+1}z^{r+1}$, where $\alpha_{r+1} = (r+1)^r$ is the number of trees of size $r+1$. For functional graphs, we have $\mathcal{G} = \text{Set}(\text{Cycle}(\mathcal{A}))$, which translates to $g(z) = \exp\left(\sum_{p \geq 0} \frac{a^p(z)}{p}\right)$. This gives

$$g(u,z) = \frac{1}{1 - ze^{a(u,z)}} \exp\left(\frac{z^r}{r}\sum_p (u^p - 1)\frac{r^{r-p}}{(r-p)!}\right) = \frac{K(u,z)}{1 - a(u,z) + (u-1)\alpha_{r+1}z^{r+1}}.$$

Nodes d iterated. (These nodes are at distance $\geq d$ from a leaf.) For trees, we have

$$a_d(u,z) = xue^{a_d(u,z)} - (u-1)l_d(z) \qquad \text{with} \quad l_0(z)=0, \quad l_{d+1}(z) = ze^{l_d(z)}.$$

For functional graphs, we have, for nodes at distance $\geq d$ of a leaf of their sub-tree, $s_d(u,z) = 1/(1-a_d(u,z))$. For nodes at distance $\geq d$ of a leaf, we have

$$g_d(u,z) = \frac{1}{1-uze^{a_d(u,z)}} = \frac{1}{1-a_d(u,z)-(u-1)l_d(z)}.$$

4. A classification for limit laws of random mappings parameters

We begin with a proposition which applies to functional equations of trees.

Proposition 1. *Let $F(u,z,a(u,z))$ be a power series in three variables with non-negative coefficients and $F(0,0,0)=0$. Suppose that the system of equations $\{\tau = F(1,\rho,\tau), 1 = F'_a(1,\rho,\tau)\}$ has positive solutions ρ and τ such that $F'_z(1,\rho,\tau) \neq 0$ and $F''_{aa}(1,\rho,\tau) \neq 0$. Then, $F(u,z,a)=0$ has for solution*

$$a(u,z) = t(u,z) - h(u,z)\sqrt{1-z/\rho(u)},$$

with t,h,ρ analytic,

$$t(1,\rho(1)) = \tau(1) \equiv \tau, \quad \rho(1) = \rho \quad \text{and} \quad h(1,\rho(1)) = \sqrt{\frac{2\rho F'_z(1,\rho,\tau)}{F''_{aa}(1,\rho,\tau)}}.$$

We arrive to a general theorem which seems to be the proper theorem to discuss random mappings. We consider a generating function $g(u,z) = \sum_{n,k} g_{n,k}u^k z^n$ corresponding for variables X_n to a probability distribution $\Pr(X_n = k) = g_{n,k}/g_n$. We consider a local expansion in the neighbourhood of $u=1, z=\rho(u)$, of the form

$$g(u,z) = \frac{1}{1-T(u,z)+h(u,z)\sqrt{1-z/\rho(u)}}.$$

T, h and ρ are analytic and $T(1,\rho)=1$.

Theorem 1. *With these hypotheses (T,h,ρ analytic and $T(1,\rho)=1$),*

1. *If $\rho'(u) = 0$ and $T'_u(1,\rho) > 0$, then $X_n/\sqrt{n} \to \mathcal{R}(\lambda)$, where $\lambda = \frac{1}{2}\left(\frac{h(\rho,1)}{T'_u(\rho,1)}\right)^2$ and $\mathcal{R}(\lambda)$ is the Rayleigh distribution of density $\lambda x \exp\left(-\frac{\lambda}{2}x^2\right)$. Moreover $E(X_n) \approx \sqrt{\frac{\pi n}{2\lambda}}$ and* $\mathrm{Var}(X_n) \approx \left(2-\frac{\pi}{2}\right)\frac{n}{\lambda}$.
2. *If $\rho'(1) \neq 0$ and $T'_u(1,\rho) = 0$, then $\frac{X_n - \mu n}{\sqrt{\sigma^2 n}} \to \mathcal{N}(0,1)$, where $\mu = -\rho'(1)/\rho(1)$ and $\sigma^2 = \mu^2 + \mu - \rho''(1)/\rho(1)$. Moreover $E(X_n) \approx \mu n$ and* $\mathrm{Var}(X_n) \approx \sigma^2 n$.
3. *If $\rho'(1) \neq 0$ and $T'_u(1,\rho) \neq 0$, then $\frac{X_n - \mu n}{\sqrt{\sigma^2 n}} \to \mathcal{N}(0,1) \star \mathcal{R}(\sigma^2\lambda)$, where μ and σ are defined as in (2), λ is defined as in (1) and the star operator represents the convolution operation.*

Remark . If $T(1,\rho) \neq 1$, then $\frac{X_n-\mu n}{\sqrt{\sigma^2 n}} \to \mathcal{N}(0,1)$, (except if $\rho'(u)=0$ and $T(1,\rho) < 1$, in which case $X_n \to \delta\mathcal{G}$, derivative of a geometric law).

The density and characteristic functions in these different cases are as follows.

1. $\mathcal{R}$ (Rayleigh) $f_{\mathcal{R}(\lambda)}(x) = \lambda x e^{-\lambda x^2/2}$, and $\phi_{\mathcal{R}}(\theta) = 1 + i\theta\sqrt{\frac{\pi}{2}}e^{-\theta^2/2}(1 - i\,\mathrm{erf}(\theta/\sqrt{2}))$.
2. $\mathcal{N}$ (Normal) $f_{\mathcal{N}}(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ and $\phi_{\mathcal{N}}(\theta) = e^{-\theta^2/2}$.

3. $\mathcal{N} \star \mathcal{R}$ (Normal conv. Rayleigh) $f_{\mathcal{N}\star\mathcal{R}}(x) = (e^{-x^2/4} - e^{-x^2/2})/\sqrt{2\pi} + \frac{xe^{-x^2/4}}{2\sqrt{2}}\mathrm{erf}(x/2)$ and $\phi_{\mathcal{N}\star\mathcal{R}}(\theta) = \phi_{\mathcal{N}}(\theta) \times \phi_{\mathcal{R}}(\theta)$.

Proof. (Sketch) Let $g(u,z) = \sum_{n\geq 0} p_n(u)z^n/n!$ with $p_n(1) = g_n$. The proof rests on the convergence of the corresponding characteristic functions to (1) $\phi_{\mathcal{R}}(\theta)$, (2) $e^{-\theta^2/2}$, (3) $e^{-\theta^2/2} \times \phi_{\mathcal{R}}(\theta)$. For instance, in case (1), the characteristic function $p_n(e^{i\theta/\sqrt{n}})/g_n$ converges to $\phi_{\mathcal{R}}(\theta)$. The proofs in the different cases make use of Cauchy inversions along suitable contours of the complex plane [1]. □

5. Applications

We note Ξ_n the law of $\frac{X_n - \mu n}{\sqrt{\sigma^2 n}}$.

Leaves. We have $a(z) = t(u,z) - h(u,z)\sqrt{1 - z/\rho(u)}$. This gives $\{\tau = \rho e^{\tau} + (u-1)\rho, 1 = \rho e^{\tau}\}$, which gives $\{t(1,\rho) \equiv \tau(1) = 1, \rho(1) = \rho\}$, and also by differentiation wrt u $\{\tau' = (\rho e^{\tau})' + \rho + (u-1)\rho', 0 = (\rho e^{\tau})'\}$, these two last equations give $\{\tau'(1) = \rho, \rho'(1) = -\rho^2 \neq 0\}$. This gives for sequences of trees $t(1,\rho) = 1, \rho'(1) \neq 0, t'_u(1,\rho) \neq 0$, and therefore $\Xi_n \to \mathcal{N}\star\mathcal{R}$. This also gives for functional graphs, with $T(u,z) = t(u,z) - (u-1)z$, $T(1,\rho) = 1, \rho'(1) \neq 0, T'_u(1,\rho) = \tau'(1) - \rho = 0$, and therefore $\Xi_n \to \mathcal{N}$.

Nodes with in-degree r. As before, $a(z) = t(u,z) - h(u,z)\sqrt{1 - z/\rho(u)}$. We have $\{\tau = \rho e^{\tau} + \rho(u-1)\frac{\tau^r}{r!}, 1 = \rho e^{\tau} + \rho(u-1)\frac{\tau^{r-1}}{(r-1)!}\}$. This gives $\tau(1) = 1$ and $\rho(1) = \rho$. By differentiation wrt u, we obtain $\tau'(1) = \rho\left(\frac{1}{r!} - \frac{1}{(r-1)!}\right)$ and $\rho'(1) = \frac{-\rho^2}{r!} \neq 0$. For sequences of trees, we get $t(1,\rho) = 1, \rho'(1) \neq 0$ and, if $r \geq 2$, $t'_u(1,\rho) \neq 0$, which implies $\Xi_n \to \mathcal{N}\star\mathcal{R}$. If $r = 1$, the limit law is normal. For functional graphs, we have $T(u,z) = t(u,z) - z(u-1)\left(\frac{a^{r-1}}{(r-1)!} - \frac{a^r}{r!}\right)$. We get $T(1,\rho) = 1, \rho'(1) \neq 0$, and $T'_u(1,\rho) = 0$, which implies that $\Xi_n \to \mathcal{N}$.

Nodes at distance d from a cycle. We have $a_d(u,z) = t_d(u,z) - c_d(u,z)\sqrt{1 - ez}$, with $t_0(z) = ug(z)$, $t_d(u,z) = ze^{t_{d-1}(u,z)}$, $c_0(z) = uk(z)$, $c_d(u,z) = t_d(u,z)c_{d-1}(u,z)$. This gives $\rho' = 0, t_d(1,\rho) = 1, t'_d(1,\rho) = 1$. Applying this results to $g(u,z) = 1/(1 - a_d(u,z))$, we get $T(1,\rho) = 1, T'_u(1,\rho) \neq 0, \rho = Cst$, which implies that $\Xi_n \to \mathcal{R}$.

Nodes with in-degree r. (Same method.) We have for sequences of Cayley trees $\xi_n \to \mathcal{N}\star\mathcal{R}$, and for functional graphs $\Xi_n \to \mathcal{N}$.

Nodes at distance $\geq d$ from a leaf. (Same method.) From a leaf of their own subtree (sequences of Cayley trees), $\Xi_n \to \mathcal{N}\star\mathcal{R}$. In the general case, $\Xi_n \to \mathcal{N}$.

Nodes at distance d from a leaf. (Same method.) If the path contains no cyclic edge, $\Xi_n \to \mathcal{R}\star\mathcal{N}$ (except if $d = 1$, in which case $\Xi_n \to \mathcal{N}$). If cyclic edges are allowed, for $d \leq 2$, we have $\Xi_n \to \mathcal{N}$. (Conjecture: this last result is true for all d.)

References

[1] Drmota (Michael) and Soria (Michèle). – Images and preimages in random mappings. *SIAM Journal on Discrete Mathematics*, vol. 10, n° 2, May 1997, pp. 246–269.

[2] Vitter (Jeffrey Scott) and Flajolet (Philippe). – Analysis of algorithms and data structures. In van Leeuwen (J.) (editor), *Handbook of Theoretical Computer Science*, Chapter 9, pp. 431–524. – North Holland, 1990.