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 schemas. For example, many parameters in random mappings can be covered in this way.

1. Methods

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

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

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

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

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 “•” 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 $\theta$, 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.

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

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 \( \phi \) can be identified to its functional graph \( G_\phi \) with vertices \( \{1, \ldots, n\} \) and edges \( (i, \phi(i)) \), for \( 1 \leq i \leq n \). Each component of \( G_\phi \) 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 \( \rho_c \) 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

\[
a_0(u, z) = u a(z), \quad a_{d+1}(u, z) = z e^{a_d(u, z)}.
\]

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 \( G = \text{Set}(\text{Cycle}(A)) \), which translates to \( g(z) = \exp \left( \sum_{p \geq 0} \frac{\alpha_{p+1}}{p} \right) \). This gives

\[
g(u, z) = \frac{1}{1 - ze^{a(u, z)}} \exp \left( \frac{z^r}{r} \sum_{p} \frac{u^p - 1}{(r - p)!} r^{r-p} \right) = \frac{K(u, z)}{1 - a(u, z) + (u - 1) \alpha_{r+1} z^{r+1}}.
\]
Nodes $d$ iterated. (These nodes are at distance $\geq d$ from a leaf.) For trees, we have

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

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

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

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 $\rho$ and $\tau$ such that $F'_a(1, \rho, \tau) \neq 0$ 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 $t, h, \rho$ analytic,

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

We arrive to a general theorem which seems to be the proper theorem to discuss random mappings. We consider a generating function $g(u, z) = \sum_{n,k} g_{n,k} w^k z^n$ 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 $\rho$ are analytic and $T(1, \rho) = 1$.

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

1. If $\rho'(u) = 0$ and $T'_u(1, \rho) > 0$, then $X_n/\sqrt{n} \rightarrow \mathcal{R}(\lambda)$, where $\lambda = \frac{1}{2} \left( \frac{\mu(1)}{\tau_a(1)} \right)^2$ and $\mathcal{R}(\lambda)$ is the Rayleigh distribution of density $\lambda x \exp \left(-\frac{1}{2} x^2\right)$. Moreover $E(X_n) \approx \sqrt{\frac{2\pi}{\lambda}}$ and $\text{Var}(X_n) \approx \frac{\lambda}{2}.$

2. If $\rho'(1) \neq 0$ and $T'_u(1, \rho) = 0$, then $\frac{X_n - \mu n}{\sqrt{\sigma^2_n}} \rightarrow \mathcal{N}(0, 1)$, where $\mu = -\rho'(1)/\rho(1)$ and $\sigma^2 = \mu^2 + \mu - \rho''(1)/\rho(1)$. Moreover $E(X_n) \approx \mu n$ and $\text{Var}(X_n) \approx \sigma^2 n$.

3. If $\rho'(1) \neq 0$ and $T'_u(1, \rho) \neq 0$, then $\frac{X_n - \mu n}{\sqrt{\sigma^2_n}} \rightarrow \mathcal{N}(0, 1) \star \mathcal{R}(\sigma^2\lambda)$, where $\mu$ and $\sigma$ are defined as in (2), $\lambda$ is defined as in (1) and the star operator represents the convolution operation.

Remark. If $T(1, \rho) \neq 1$, then $\frac{X_n - \mu n}{\sqrt{\sigma^2_n}} \rightarrow \mathcal{N}(0, 1)$, (except if $\rho'(u) = 0$ and $T(1, \rho) < 1$, in which case $X_n \rightarrow \delta \mathcal{G}$, 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 $\phi_\mathcal{R}(\theta) = 1 + i\theta\sqrt{\frac{\pi}{2}}e^{-\theta^2/2}(1 - i\text{erf}(\theta/\sqrt{2}))$.

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) $f_{\mathcal{N} \star \mathcal{R}}(x) = \left( e^{-x^2/4} - e^{-x^2/2} \right) / \sqrt{2\pi} + \frac{e^{-x^2/4}}{2\sqrt{2}} \text{erf}(x/2)$ 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 $\Xi_n$ the law of $\frac{X_n - \mu_n}{\sqrt{\sigma_n^2}}$.

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 $t(1, \rho) = 1, \rho(1) = 1$ and also by differentiation wrt $u$ $\tau' = (\rho e^\tau)' + \rho + (u - 1) \rho', 0 = (\rho e^\tau)'$, and these 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 $\Xi_n \rightarrow \mathcal{N} \star \mathcal{R}$. This also gives for functional graphs, with $T(u, z) = t(u, z) - \frac{1}{2}(u - 1)z, T(1, \rho) = 1, \rho'(1) \neq 0, T'_u(1, \rho) = \tau'(1) - \rho = 0$, and therefore $\Xi_n \rightarrow \mathcal{N}$.

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

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 $t_0(u) = u$, $t_d(u, z) = ze^{t_{d-1}(u, z)}, c_0(u) = u, c_d(u, z) = t_d(u, z) c_{d-1}(u, z)$. 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 $\Xi_n \rightarrow \mathcal{R}$.

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

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

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

References