# ON THE NUMBER OF PREDECESSORS IN CONSTRAINED RANDOM MAPPINGS

BERNHARD GITTENBERGER

Abstract. We consider random mappings from an $n$–element set into itself with constraints on coalescence as introduced by Arney and Bender. A local limit theorem for the distribution of the number of predecessors of a random point in such a mapping is presented by using a generating function approach and singularity analysis.

## 1. Introduction

For $M_n = \{1, 2, \ldots, n\}$ let $F_n$ denote the set of all functions from $M_n$ into itself equipped with the uniform distribution. Then an element of $F_n$ is called a random mapping. Each random mapping $f$ can be represented by a functional graph, i.e. the graph consisting of the nodes $1, 2, \ldots, n$, and of the edges $(i, f(i))$, $i = 1, \ldots, n$. Various characteristics of random mappings have been studied. See e.g. [1, 3, 4, 6, 8, 9, 10, 12].

Arney and Bender [1] examined a more general model: They considered mappings such that the number of preimages of every point lies in a given set $D$ of nonnegative integers (with $0 \in D$) or, equivalently, the degrees of the nodes of the functional graph have to be in $D$. Let $F_n^D$ denote the set of those mappings on $M_n$. Arney and Bender derived the distributions of many random mapping characteristics for $F_n^D$ mainly by means of combinatorial counting arguments. In this paper we will study the distribution of a further random mapping characteristic using a generating function approach.

Let $x \in M_n$ and $f \in F_n^D$. $y \in M_n$ is called a *predecessor* of $x$ if there exists a $j > 0$ such that the $j$-th iterate of $f$ applied on $y$ yields $x$. Denote the number of predecessors of $x$ by $\omega(x)$. Then we will show

**Theorem 1.** *Let $\phi(x) = \sum_{k\in D} x^k/k!$ and $\beta$ be the positive root of $\beta\phi'(\beta) = \phi(\beta)$. Furthermore define $\lambda := \beta^2\phi''(\beta)/\phi(\beta)$ and $d := \gcd\{k : k \in D\}$. Then for a randomly chosen point of a mapping of $F_n^D$ the expected number of predecessors equals $\sim \sqrt{\frac{\pi n}{2\lambda}}$ and moreover a local limit theorem holds: If $d|r$ then*

$$\mathbf{P}[\omega = r] = \frac{d}{\sqrt{2\pi\lambda r^3 \left(1 - \frac{r}{n}\right)}} (1 + o(1))$$

*for $r \to \infty$ and $n - r \to \infty$. Otherwise $\mathbf{P}[\omega = r] = 0$*

*Remark .* For the special cases $D = \{0, 1, 2, ...\}$ and $D = \{0, k\}$ the above result was obtained by Rubin and Sitgreaves [11].

## 2. Combinatorial Background

The basic concept which our method relies on is that of combinatorial constructions described e.g. in [13]: Let $\mathcal{A}$ be a set of combinatorial objects where each object $\sigma \in \mathcal{A}$ has a size $|\sigma|$. If $a_n$ denotes the number of objects in $\mathcal{A}$ having size $n$, then

$$\hat{A}(z) = \sum_{n\geq 0} a_n \frac{z^n}{n!}$$

*Date*: December 5, 1996.

*Key words and phrases.* random mappings, generating functions.

is called the (exponential) generating function (GF) of $\mathcal{A}$.

The advantage of the generating function approach is the fact that there is a direct correspondence between operations on combinatorial constructions and on GFs (see [13]). Using the graph representation of random mappings it is easy to see that they may be regarded as multisets of cycles of Cayley trees, i.e. labelled rooted trees. Let $\mathcal{A}$ denote the set of Cayley trees. Then $\mathcal{A}$ can be defined recursively by the symbolic equation

$$\mathcal{A} = \circ * \mathtt{mset}(\mathcal{A}),$$

where $\circ$ represents a node. Analogously, we get for the set of random mappings, $\mathcal{F}$:

$$\mathcal{F} = \mathtt{mset}(\mathtt{cycle}(\mathcal{A}))$$

This implies the following equations for the corresponding GFs: The GF of Cayley trees is given by

$$a(z) = ze^{a(z)}$$

and the GF of random mappings

$$f(z) = \frac{1}{1-a(z)}.$$

As we want to study the random mappings with constraints on coalescence, we have to modify $a(z)$ properly and obtain

$$a(z) = z\phi(a(z)) \quad \text{where } \phi(z) = \sum_{k\in D} \frac{z^k}{k!}.$$

and the GF of constrained random mappings is given by

$$f(z) = \frac{1}{1-b(z)} \quad \text{where } b(z) = z\phi'(\bar{a}(z))$$

since the root of each tree has an additional predecessor coming from the cycle.

Let $c_{nk}^{(r)}$ denote the number of mappings in $F_n^D$ which have exactly $k$ points $x$ satisfying $\omega(x) = r$, where $\omega(x)$ is the number of predecessors of the point $x$. The number of all points in all mappings of $F_n^D$ that satisfy this equation is denoted by $b_n^{(r)}$. Hence we have

$$\frac{1}{|F_n^D|n} b_n^{(r)} = \mathbf{P}\{\omega = r\}.$$

Thus for establishing the local limit theorem we have to compute $b_n^{(r)}$. Obviously the relation

$$b_n^{(r)} = \sum_{k\geq 1} k c_{nk}^{(r)}$$

holds. Therefore $b_n^{(r)}$ can be calculated by

$$b_n^{(r)} = \left[\frac{z^n}{n!}\right] (c_r)_u(z,1) \tag{1}$$

where $[z^n]f$ denotes the coefficient of $z^n$ in the power series of $f$ and $(c_r)_u$ denotes the partial derivative with respect to $u$. $c_r(z,u)$ is defined by

$$c_r(z,u) = \sum_{n,k\geq 0} c_{nk}^{(r)} \frac{z^n}{n!} u^k$$

and can be obtained by marking the nodes we are keeping track of in the functional graphs which corresponds to introducing a further variable in the GF. Further examples of these marking techniques in combinatorial constructions can be found in [3, 4].

(1) shows that in order to get the asymptotic behaviour of $b_n^{(r)}$ we have to evaluate the coefficients of certain power series. In order to do this we will use the following theorem by Flajolet and Odlyzko [7].

**Theorem 2.** *Let $f(z)$ be analytic in the domain*

$$\Delta = \{z/|z| \le s + \eta, |arg(z - s)| \ge \phi\},$$

*where $s, \eta$ are positive real numbers and $0 < \phi < \frac{\pi}{2}$. Furthermore let $\sigma(u) = u^\alpha \log^\beta u,\ \alpha \notin \{0, -1, -2, \dots\}$. Then the Taylor coefficients of $f$ satisfy*

$$f(z) \sim \sigma\left(\frac{1}{1 - z/s}\right) \text{ for } z \to s \text{ in } \Delta \implies [z^n]f(z) \sim \frac{\sigma(n)}{s^n n\Gamma(\alpha)}.$$

*Analogous formulas hold for $\mathcal{O}$ and $o$ instead of $\sim$.*

## 3. Proof of Theorem 1

For convenience, define $a_r = [z^r]a(z)$, $b_{mr} = [z^r]b(z)^m$ and assume that $d = 1$. $b_n = [z^n]1/(1-b(z))$. First of all, let us set up the GF $c_r(z, u)$. Therefore we have to mark all nodes that are roots of a tree with size $r$. This leads to the following modification of $a(z)$:

$$t_r(z, u) = z\phi(t_r(z, u)) + (u - 1)a_r z^r.$$

Moreover we have to take into account that the set of predecessors of a node which belongs to a cycle consists of all nodes in the component. Hence in a component of size $r$ all nodes in the cycle have to be marked. For components containing a cycle of length $m$ this yields the GF

$$\frac{(z\phi'(t_r(z, u)))^m}{m} + \frac{1}{m}(u^m - 1)z^r b_{mr}.$$

Thus we have

$$c_r(z, u) = \frac{1}{1 - z\phi'(t_r(z, u))} \exp\left(z^r \sum_{m=1}^{r} \frac{u^m - 1}{m} b_{mr}\right)$$

and consequently

$$(c_r)_u(z, 1) = \frac{a_r z^{r+1}\phi''(a(z))}{(1 - b(z))^3} + \frac{z^r \sum_{m=1}^r b_{mr}}{1 - b(z)}. \tag{2}$$

Note that $b_{mr} = 0$ if $m = 0$ or $m > r$ and thus

$$\sum_{m=1}^{r} b_{mr} = \sum_{m=0}^{\infty} b_{mr} = [z^r]\frac{1}{1 - b(z)} = b_r.$$

Hence the functions $a(z)$ and $b(z)$ contain the information we need. $a(z)$ has a singularity $\rho$ on the circle of convergence which is determined by

$$\begin{aligned} a(\rho) &= \rho\phi(a(\rho)) \\ 1 &= \rho\phi'(a(\rho)). \end{aligned}$$

To proceed we will need the following lemma which is an immediate consequence of [5, Theorem 7.1]:

**Lemma 1.** *Let $F(z, y)$ be continuous in $\{(x, y) : |z - z_0| < r_1,\ |y - y_0| < r_2\}$. Furthermore assume that if one variable is fixed, then $F$ is analytic in the other variable. If $F(z_0, y_0) = F_y(z_0, y_0) = 0,\ F_z(z_0, y_0) \ne 0$ and $F_{yy}(z_0, y_0)$, then there exists a function $y(z)$ which admits the local representation*

$$y(z) \sim y_0 - \sqrt{\frac{2F_z(z_0, y_0)}{F_{yy}(z_0, y_0)}}\sqrt{z_0 - z}$$

*for $z \to z_0$.*

Thus in the vicinity of $\rho$ $a(z)$ admits the following local expansion:

$$a(z) \sim \beta - \sqrt{\frac{2\phi(\beta)}{\phi''(\beta)}}\sqrt{1-\frac{z}{\rho}} = \beta - \frac{\beta\sqrt{2}}{\sqrt{\lambda}}\sqrt{1-\frac{z}{\rho}}, \quad \text{as } z \to \rho$$

where $\beta = a(\rho)$ and $\lambda = \beta^2\phi''(\beta)/\phi(\beta)$. By expanding $\phi(z)$ we immediately obtain

$$b(z) \sim 1 - \rho\sqrt{2\phi(\beta)\phi''(\beta)}\sqrt{1-\frac{z}{\rho}} = 1 - \sqrt{2\lambda}\sqrt{1-\frac{z}{\rho}} \quad \text{as } z \to \rho$$

and thus

$$\frac{1}{1-b(z)} \sim \frac{1}{\sqrt{2\lambda}}\left(\frac{1}{1-z/\rho}\right)^{1/2} \quad \text{as } z \to \rho. \tag{3}$$

Now we are able to apply Theorem 2 and get

$$a_r \sim \frac{\beta}{\rho^r\sqrt{2\pi\lambda r^3}} \quad \text{and} \quad b_r \sim \frac{1}{\rho^r\sqrt{2\pi\lambda r}} \quad \text{as } r \to \infty. \tag{4}$$

Using this we finally get

$$\begin{aligned}
\frac{b_n^{(r)}}{|F_n^D|n} &= \frac{1}{nb_n}[z^n](c_r)_u(z,1) \\
&\sim \frac{\rho^n\sqrt{2\pi\lambda n}}{n}\left(\frac{a_r\lambda}{\beta}[z^{n-r}]\frac{1}{(1-b(z))^3} + [z^{n-r}]\frac{1}{1-b(z)}[z^r]\frac{1}{1-b(z)}\right) \\
&\sim \frac{\rho^n\sqrt{2\pi\lambda n}}{n}\left(\frac{\lambda}{\rho^r\sqrt{2\pi\lambda r^3}}\frac{\sqrt{n-r}}{\rho^{n-r}\sqrt{2\pi\lambda^3}} + \frac{1}{\rho^r\sqrt{2\pi\lambda r}}\frac{1}{\rho^{n-r}\sqrt{2\pi\lambda(n-r)}}\right) \\
&= \frac{1}{\sqrt{2\pi\lambda r^3(1-r/n)}}
\end{aligned}$$

What remains to do is the calculation of the mean value $\mu_n$: By (2) we have

$$\begin{aligned}
\mu_n &= \frac{1}{nb_n}[z^n]\sum_{r\ge 0} r(c_r)_u(z,1) \\
&= \frac{1}{nb_n}[z^n]\left(\frac{\phi''(a(z))}{(1-b(z))^3}z^2a'(z) + \frac{1}{1-b(z)}z\left(\frac{1}{1-b(z)}\right)'\right)
\end{aligned}$$

Using the functional equation of $a(z)$ we immediately get

$$a'(z) = \frac{a(z)}{z(1-b(z))}.$$

This implies

$$\mu_n = \frac{1}{nb_n}[z^n]\left(\frac{2z\phi''(a(z))a(z)}{(1-b(z))^4} + \frac{z\phi'(a(z))}{(1-b(z))^3}\right)$$

Now observe that for $z \to \rho$ we have $z\phi''(a(z))a(z) \sim \lambda$. Thus we get by using (3) and (4) and applying Theorem 2

$$\mu_n = \sqrt{\frac{\pi n}{2\lambda}}\left(1 + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)\right)$$

and the proof is complete.

*Remark 1.* In an analogous way it is possible to reobtain Arney and Bender's [1] results by the method presented here or to derive local limit theorems for other random mapping characteristics. Note, for instance, that recently Baron, Drmota and Mutafchief [2] derived the missing distribution of [1, Table II] by using a generating function approach.

*Remark 2.* It should be mentioned that the coefficient of coalescence $\lambda$ which occurs in the distributions of several random mapping characteristics has a simple probabilistic interpretation, as Aldous pointed out: If we consider the trees which build up the random mappings as representations of Galton-Watson branching processes, then the offspring distribution is determined by the tree function $\phi(z)$ and $\lambda$ is equal to its variance.

## References

[1] J. Arney and E. A. Bender, Random mappings with constraints on coalescence and number of origins, *Pacific J. Math.,* 103 (1982), 269-294.

[2] G. Baron, M. Drmota and L. Mutafchief, Predecessors in random mappings, *Combinatorics, Probability and Computing*, to appear.

[3] M. Drmota and M. Soria, Marking in combinatorial constructions: generating functions and limiting distributions, *Theoretical Comp. Science 144* (1995), 67-99.

[4] M. Drmota and M. Soria, Marking in combinatorial constructions: images and preimages in random mappings, *SIAM J. on Discrete Math.*, to appear.

[5] E. A. Evgrafov, *Analytic Functions,* Dover, New York, 1966.

[6] P. Flajolet and A. M. Odlyzko, Random mapping statistics, In *Advances in Cryptology* (1990), J.-J. Quisquater and J. Vandewalle, Eds., vol. 434 of *Lecture Notes in Computer Science,* Springer Verlag, pp. 329-354. Proceedings of Eurocrypt '89, Houtalen, Belgium, April 1989.

[7] P. Flajolet and A. M. Odlyzko, Singularity analysis of generating functions, *SIAM J. on Discrete Math. 3,* 2 (1990), 216-240.

[8] B. Harris, Probability distributions related to random mappings, *Annals of Mathematical Statistics* 31, 2 (1960), 1045-1062.

[9] V. F. Kolchin, *Random Mappings,* Optimization Software, Inc., New York, 1986.

[10] L. R. Mutafciev, The limit distribution of the number of nodes in low strata of random mappings, *Statistics & Probability Letters* 7 (1989), 247-251.

[11] H. Rubin and R. Sitgreaves, Probability distributions related to random transformations of a finite set, *Technical Report No. 19a,* Applied Mathematics and Statistics Laboratory, Stanford, 1954.

[12] V. E. Stepanov, Limit distributions of certain characteristics of random mappings, *Th. Prob. Appl.* 14 (1969), 612-626.

[13] J. S. Vitter and P. Flajolet, Average-case analysis of algorithms and data structures, in *Handbook of Theoretical Computer Science,* J. van Leeuwen, Ed., vol. A: Algorithms and Complexity. North Holland, 1990, ch. 9, pp. 431-524.

Bernhard Gittenberger
Department of Algebra and Discrete Mathematics
TU Wien
Wiedner Hauptstrasse 8-10/118
A-1040 Wien,
Austria