RANDOM MAPPINGS WITH CONSTRAINTS ON COALESCENCE AND NUMBER OF ORIGINS

James Arney and Edward A. Bender

In § 2 we tabulate for easy reference probability distributions associated with some functions of random mappings on large sets (e.g., number of points on cycles, size of the component containing x) when the number of immediate predecessors of each point is required to lie in some set $\mathscr{D}$.

Our results allow the number of origins to be restricted, a useful constraint in some shift register situations. Although limiting the number of immediate predecessors to $\{0, 1, 2\}$ and constraining the number of origins is in some ways a poor model for random shift registers, we show in §§ 3 and 4 that most of the tabulated results fit shift register data quite well.

Derivations of our results are given in §§ 5 through 9.

1. **Introduction and terminology.** Let X be an n element set and let φ be a mapping from X to X. We can picture φ as a directed graph with vertices X and edges $(x, \varphi(x))$, $x \in X$. We use the graph theory terminology *component* and *cycle*. The component containing x is denoted $K(x)$. For sufficiently large m, $\varphi^m(x)$ lies on a cycle denoted $C(x)$. The least m such that $\varphi^m(x) \in C(x)$ is called the *tail length of* x and written $t(x)$. The *six length of* x, written $s(x)$, is the length of the path from x to the first repeat; i.e., $t(x)$ plus the size of $C(x)$. If x lies on a cycle, the set of $u \in X$ such that $\varphi^{t(u)}(u) = x$ is called the *tree of* x and is written $T(x)$. It consists precisely of those elements which first "hit" a cycle at x when φ is iterated. If $x \in T(u)$, we define $T(x) = T(u)$. The elements of $\varphi^{-1}(x)$ are called the *immediate predecessors* of x. The number of points with r immediate predecessors is denoted by $n_r(\varphi)$. The elements of $\bigcup \varphi^{-i}(x)$, where the union ranges over all $i \geq 0$ are called the *predecessors* of x. The number of points with r predecessors is denoted by $N_r(\varphi)$. A point without predecessors is called an *origin*. The number of origins is $n_0(\varphi) = N_1(\varphi)$. Let $p_i(\varphi) = n_i(\varphi)/n$, the probability that a point chosen at random has exactly i immediate predecessors, then we call $\sum_i (i - 1)p_i(\varphi) = \lambda(\varphi)$ *the coefficient of coalescence of* φ. Many authors have employed this concept. Among the equivalent definitions are

(a) the variance of the number of predecessors of a random point,

(b) $(n - 1)$ times the probability that two distinct elements have the same immediate successor,

    (c)   the expected number of other immediate predecessors to the immediate successor of a randomly selected point.

The equivalence of these four definitions is easily established by simple counting arguments. We use $\#(S)$ to denote the cardinality of S. For example, $\#(C(x))$ is the length of the cycle of x.

    Let $\mathscr{D}$ be a set of nonnegative integers including zero and at least one integer greater than one. Let τ be the set of mappings from X to X such that the vertices of the associated directed graphs all have indegrees lying in $\mathscr{D}$. In other words, if $\varphi \in \tau$ and $n_i(\varphi) \neq 0$, then $i \in \mathscr{D}$. Three special types of $\mathscr{D}$ have been considered in the literature:

    (a)   all nonnegative integers,
    (b)   the integers 0 and k,
    (c)   the integers between 0 and k inclusive.

Rubin and Sitgreaves [9] and Harris [6] discuss (a). Rubin and Sitgreaves also discuss (b) and (c), and Harris also discusses (a) when 1-cycles are forbidden. We establish results for all $\mathscr{D}$. The total number of vertices in the graph of φ can be computed by counting the immediate predecessors of each point. This gives us $n = \sum_i n_i(\varphi)$. It follows that n must be a multiple of $gcd(\mathscr{D})$, the greatest common divisor of the elements in $\mathscr{D}$.

    We can ask for the distributions of various quantities over τ; however, we must be clear just what we are asking for. One possibility is to define a function $f(\varphi)$ and ask for its distribution assuming that all $\varphi \in \tau$ are equally likely. This gives us information about the overall appearance of a random map. The number of cycles of a map is an example of this. Another possibility is to define a function $g(x, \varphi)$ and ask for its distribution assuming that all $(x, \varphi) \in X \times \tau$ are equally likely. This gives us information about the appearance of a random map when viewed from a radom point. The tail length of a point is an example of this. We usually write $g(x)$ instead of $g(x, \varphi)$ to emphasize the fact that we are choosing a random point. When f or g can be obtained simply by counting (e.g., the cycles of φ, or the number of points between x and a cyclic point), it is quite likely that the distributions, or at least the means, can be computed asymptotically. See Table II in the next section. Another type of function is

$$f(\varphi) = \max_{x \in X} g(x, \varphi) .$$

Since this requires more than simple counting, it is much harder to deal with. See § 4.

    Of particular interest is the case $n = 2^l$ and $\mathscr{D} = \{0, 1, 2\}$. This has some of the properties of shift registers of length l because in

such divices a state has at most two predecessors. This model fits shift register data fairly well, but predicts $.293n$ origins whereas the average number of origins for a random shift register of length l is $n/4 = 2^{l-2}$. (This is easily proved; see below.) The number of origins for a shift register might be restricted for other reasons. For example, when the recursion is written in the form

$$(1.1) \qquad x_l = f(x_1, \cdots, x_{l-1}) + x_0 g(x_1, \cdots, x_{l-1}) \ ,$$

then the number of origins is precisely the number of solutions to $g = 0$. (The expected number of zeros of g is $2^{l-1}/2$, proving the earlier claim that the expected number of origins in a random shift register is $n/4$.) If there is no further information about f and g, we might look at data for random functions with this many origins. Thus, it is of interest to restrict the mappings in τ to have a specified number of origins. This can sometimes be done by writing down explicit sums. We are unable to specify the number of origins when using generating functions; however, a compromise is possible: assign to φ a probability proportional to y^{n_0}, where y is chosen so that the expected value of $n_0 = n_0(\varphi)$ has a particular value (e.g., $1/4$ for random shift registers). Setting $y = 1$ gives the unconstrained case. This is done for two reasons. First, this is a natural approach when using generating functions. (See [2] for some applications.) Second, assigning this probability to φ leads to a distribution for which the variance of $n_0(\varphi)$ is relatively small. Table II includes both the constrained and unconstrained cases.

We would like to thank L. Liporace, G. Soules, and C. Terry for their help.

2. **Summary of results.** The following notation is used in the tables.

$\mathscr{D}$ the set of possible values for the number of immediate predecessors of points;

φ a mapping such that the number of immediate predecessors of each point is in $\mathscr{D}$ and, possibly, such that the number of origins is specified;

d the greatest common divisor of the elements of $\mathscr{D}$;

n the number of points being mapped (always a multiple of d);

$\sum'$ sums ranging over those nonzero $m \in \mathscr{D}$ for which the factorials in the sum make sense;

$\rho = 1/(\sum' \beta^{m-1}/(m-1)!)$ (β is defined below);

$\lambda = \rho \sum' \beta^{m-1}/(m-2)!$ (β is defined below).

When there is no constraint on the number of origins:

β is the positive root of $\sum' (m-1)\beta^m/m! = 1$,

$$G(r) = 1 + (r - 1)^2/\lambda,$$
$$y = 1.$$

When the number of origins is asymptotic to nf:

β is the positive root of $\sum'(m - mf - 1)\beta^m/m! = 0$,
$$G(r) = (\lambda + (r - 1)^2 - r^2 f)/(\lambda - f - \lambda f),$$
$$y = f\beta/\rho;$$

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt \; ;$$

j given a density function f, $\Pr\{\text{variable} = j\} = f(j)$;

$\#(S)$ the size of the set S;

$n_r(\varphi)$ the number of points with exactly r immediate predecessors under φ;

$N_r(\varphi)$ the number of points x having a total of r predecessors, including x, under φ;

$\lambda(\varphi) = \sum i(i - 1)n_i(\varphi)/n$, the coefficient of coalescence of φ;

$c(\varphi)$ the number of cyclic elements of φ;

$m(\varphi)$ the number of cycles of φ; i.e., the number of components of φ;

$C(x)$ the cycle that x leads into;

$K(x)$ the component containing x;

$k_t(x)$ the number of points in a set of t randomly chosen points that lie in the same component as x; $\Pr\{k_t(x) = t\}$ is the probability that $t + 1$ randomly chosen points lie in the same component;

$T(x)$ the tree of tails at the cycle point x leads into;

$t(x)$ the tail length of x;

$s(x)$ the six length of x; i.e., $t(x) + \#(C(x))$.

The parameters needed in Table II for some cases of particular interest are given in Table I. If $f(j)$ is given as a "probability density function" entry in Table II, it means that the maximum difference between the true cumulative distribution function and the one computed using the table entry approaches zero. (For $\#(K(x))$ it is necessary to restrict the maximum to $j \ll n$.) A somewhat stronger statement in terms of density functions is

$$\lim_{n \to 0} \max_j (|\Pr\{\text{item} = j\} - f(j)|/\max_k f(k)) = 0 \; ,$$

TABLE I. Parameters for Cases of Particular Interest

$\mathscr{D}$	$n_0(\varphi)$	β	ρ	λ
$[0, \infty)$	free	1	e^{-1}	1
$\{0, k\}$	free	$(k!/(k - 1))^{1/k}$	$\beta(k - 1)/k$	$k - 1$
$[0, k]$	free			$k + \beta - k\beta$
$\{0, 1, 2\}$	free	$\sqrt{2}$	$\sqrt{2} - 1$	$2 - \sqrt{2}$
$\{0, 1, 2\}$	fn	$2f/(1 - 2f)$	$1/(1 - 2f)$	$2f$

TABLE II. Asymptotic Means and Probability Densities

item	mean	probability density function	Sec.
immed. pred. $n_r(\varphi)$, $r \in \mathscr{D}$	$r = 0$: $\mu = \rho y n / \beta$ $r \neq 0$: $\mu = n \rho \beta^{r-1} / r!$	normal with variance $\mu - G(r)\mu^2/n$	9
total pred. $N_r(\varphi)$ $1 \ll r \ll n$	$nd/\sqrt{2\pi\lambda r^3}$ if $d \mid (r-1)$	distribution unknown	8
coalescence $\lambda(\varphi)$	λ	normal: see Section 9	9
cyclic elts. $c(\varphi)$	$\sqrt{\pi n/2\lambda}$	$(j\lambda/n) \exp(-j^2\lambda/2n)$	5
no. cycles $m(\varphi)$	$\frac{1}{2} \ln n$	normal: mean = variance	5
cycle length $\#(C(x))$	$\frac{1}{2}\sqrt{\pi n/2\lambda}$	$\sqrt{2\pi\lambda/n}\phi(j\sqrt{\lambda/n})$	5
comp. size $\#(K(x))$	$2n/3$	$1/(2n\sqrt{1-j/n})$ for $d \mid j$ and $j \ll n$	7
same comp. $k_t(x)$	$2t/3$	$\binom{t}{j} \Big/ (2t+1)\binom{t-1/2}{j}$	7
tree size $\#(T(x))$	$n/3$	conjecture: $1/2\sqrt{nj}$ $j \gg 1$	7
tail length $t(x)$	same mean and distribution as $\#(C(x))$		5
six length $s(x)$	same mean and distribution as $c(\varphi)$		6

where the domain of j in the maximum is restricted for $\#(K(x))$. The last column in Table II gives the section where the results for that entry are proved. The variance for $n_r(\varphi)$, the first entry in Table II, should be set to zero if the value of $n_r(\varphi)$ is forced; i.e., if $\#(\mathscr{D}) = 2$ or if the number of origins is fixed and either $r = 0$ or $\#(\mathscr{D}) = 3$. The variance of λ is not given because of its complexity. See § 9 for its computation.

3. Comparison with shift register data. We generated shift registers of lenth ten at random subject to the constraint that $n_0(\varphi)$ take on some preassigned value. For each value of n_0 we produced 2000 different registers. We choose 2000 to reduce sampling error. Since the amount of work grows like 2^l where l is register length, it is necessary to keep l reasonably small. For $l = 10$, our crude program took about two minutes per value of n_0. We were quite surprised at the accuracy of the fit with $n_0 = 4$ since Table II contains asymptotic results that are based on $n_0 \to \infty$.

The means of various items in Table II for these data sets are shown in Table III. The largest difference between the theoretical

TABLE III. Data from Ten Long Shift Registers

item	property	the number of origins		
		256	32	4
$c(\varphi)$	data mean	55.78	152.58	418.40
	Table II mean	56.72	160.42	453.75
	max. diff.	.03	.04	.08
$m(\varphi)$	data mean	4.096	4.989	6.071
	Table II mean	3.466	3.466	3.466
	(3.1) mean	4.043	5.083	6.122
	max. diff.	.15	.15	.13
$\#(C(x))$	data mean	28.93	78.63	212.73
	Table II mean	28.36	80.21	226.87
	max. diff.	.04	.02	.03
$\#(K(x))$	data mean	693.33	695.94	654.58
	Table II mean	682.67	682.67	682.67
	max. diff.	.03	.03	.07
$\#(T(x))$	data mean	342.69	343.52	272.20
	Table II mean	341.33	341.33	341.33
	max. diff.	.01	.11	.37

and the observed cumulative distributions, denoted by "max. diff.", is a finer measure of how well Table II fits shift register data. Earlier runs with 500 random shift registers and the Monte Carlo rule that variance due to sampling is inversely proportional to the number of examples indicates that a max. diff. of about .04 is to be expected in Table III because of sampling variation.

The values of $n_r(\varphi)$ and $\lambda(\varphi)$ are not studied because they are completely determined. We have not included $t(x)$ or $s(x)$ because they are very closely connected with $\#(C(x))$ and $c(\varphi)$ respectively, as discussed in the next section. We have excluded $N_r(\varphi)$ and $k_t(x)$ because determining them seems to be quite time consuming. The distribution for the latter should be good since it is derived from that for $\#(K(x))$ in § 7.

Most of the predictions fit the data rather well even when there are as few as four origins; however, the predicted number of cycles is significantly low and the fit for $\#(T(x))$ is not very good. The poor fit for the $\#(T(x))$ distribution suggests that either the distribution conjectured in Table III is incorrect or the structure of shift registers is important for this statistic.

The $m(\varphi)$ discrepancy appears to be due to the fact that $\ln 2^{10}$ is rather small. We can obtain the improved estimate

(3.1) $$\overline{m(\varphi)} \approx \frac{1}{2} \ln (n/2\lambda) + \gamma$$

by studying the generating function $-(1 - R)^{-1}\ln(1 - R)$ in the paragraph leading to (7.10) more carefully. Here $\gamma = .5772156649\cdots$ is Euler's constant. This estimate is shown in Table III. The fit to the $m(\varphi)$ cumulative distribution can probably be improved by obtaining a more accurate estimate for the variance as we did for the mean. It appears that we should add the term $\pi^2/6$ to the asymptotic value given in Table II.

A discussion of estimates for average maxima and a comparison with the same shift registers used in Table III appears in § 4. Comments on models of shift registers appear in the last section.

4. **The average of a maximum.** As remarked earlier, studying the distribution, or even just the mean, of a maximum is usually quite difficult. In these remarkable paper [10] Shepp and Lloyd obtain the asymptotic distribution for the length of the rth longest and rth shortest cycles of a random permutation. The basic idea is that asymptotically the cycle lengths are independently distributed Poisson random variables. It is quite likely that this idea can be used to obtain distributions of various maxima associated with random mappings. Our results are much more modest: We use Shepp and Lloyd's result to deduce cycle length information for random mappings and give crude methods for estimating other maxima.

Shepp and Lloyd have shown that the expected value of the maximum cycle length of a random permutation on $[1, c]$ is $.62432965c$. Using Lemma 1 in § 6 and averaging over all φ we obtain

$$(4.1) \qquad \text{average maximum cycle length} \sim .62432965\overline{c(\varphi)}\ .$$

We now turn our attention to $K(x)$ and $T(x)$. If a_i denotes a sequence of nonnegative numbers with maximum value a_0, then

$$(\textstyle\sum a_i^2)/(\sum a_i) \leqq a_0 \leqq \sqrt{\sum a_i^2}\ .$$

Suppose the a_i's are either $\#(K(x))$ where x runs through a representative of each component, or $\#(T(x))$ where x runs through the cyclic points. Then $\sum a_i = n$ and $\sum a_i^2$ is n times the average given in Table II. This leads to the bounds

$$(4.2\,\mathrm{a}) \qquad 2n/3 \leqq \overline{\max \#(K(x))} \leqq \sqrt{2/3}n = .816n$$

$$(4.2\,\mathrm{b}) \qquad n/3 \leqq \overline{\max \#(T(x))} \leqq n/\sqrt{3} = .577n\ .$$

The question arises: What should we do with these estimates? One possibility is to average them, but then we must decide which average to use. Arbitrarily choosing the arithmetic mean we get

$$(4.3) \qquad \overline{\max \#(K(x))} \approx .7416n \qquad \text{and} \qquad \overline{\max \#(T(x))} \approx .4553n\ .$$

TABLE IV. Maxima for Ten Long Shift Registers

item	quantity	the number of origins		
		256	32	4
comp.	data max.	781.9	787.4	758.2
	(4.3) est.	759.4	759.4	759.4
	(4.4) est.	819.0	819.0	819.0
tree	data max.	493.4	501.1	447.4
	(4.3) est.	466.2	466.2	466.2
	(4.4) est.	512.0	512.0	512.0
cycle	data max.	35.2	96.8	263.2
	Sh.-L. theory	35.4	100.2	283.3
	Sh.-L. data	34.8	95.3	261.2

Another method for estimating a maximum uses the idea that, once the component associated with the maximum is removed, what is left can be treated as a new graph which has a maximum, and so on.   Let the fraction of points which belong to the maximum component of a graph be r, then we have

$$(4.4) \qquad \sum \left(r(r-1)^k n \right)^2 \approx \overline{\#(K(x))} n = 2n^2/3 .$$

The left hand side is a geometric series. We have, after summing and manipulating, $r/(2-r) \approx 2/3$ and so $r \approx 4/5$. Using this approach for tree we obtain $r/(2-r) \approx 1/3$ and so $r \approx 1/2$. Table IV compares the predictions with data from the shift registers used in the previous section. The last two lines refer to predictions based on Shepp and Lloyd's result. The "theory" case uses the predicted value for $\overline{c(\varphi)}$ and the "data" case uses the observed value.

5. Relations between six, tail, and cycle lengths. In this section we show that if the asymptotic density function for one of $s(x)$, $t(x)$, $c(\varphi)$ and $\#(C(x))$ is known, then the asymptotic density functions of the other three are easily obtained. At the end of the section we derive the asymptotic density function for $m(\varphi)$ using the density for $c(\varphi)$ and the corresponding result for permutations.

We begin with

LEMMA 1. *Let a set X and numbers n_r and c be given. Let $\mathscr{S}$ be the set of all mappings φ from X to X with $c(\varphi) = c$ and $n_r(\varphi) = n_r$ for all r. Then the distribution of cycle lengths of the mappings in $\mathscr{S}$ is independent of X and the n_r's. In particular, it is the same as the distribution for permutations on a set of c elements.*

Proof. It suffices to show the last assertion in the lemma. Let $\varphi \in \mathscr{S}$. Break all edges of the associated graph that lie on cycles. The c resulting trees were equally likely to arise from any other recombination of them that forms a permutation of their roots. $\square$

We will establish the following simple relationship between $c(\varphi)$ and $\#(C(x))$:

$$(5.1) \qquad \Pr\{\#(C(x)) = j\} = \sum_{k \geq j} \Pr\{c(\varphi) = k\}/k \ .$$

Before proceeding with the derivation we observe that a point in a random permutation of $[1, k]$ belongs to a cycle of length $j \leq k$ with probability $1/k$. To see this arrange $[1, k]$ in a one line permutation at random, create cycles by breaking the permutation before π_i if $\pi_j < \pi_i$ for all $j < i$, and look at the length of the cycle containing k. The derivation of (5.1) is now straightforward: the left side of (5.1) can be rewritten as

$$\sum_{k \geq j} \Pr\{\#(C(x)) = j \text{ and } c(\varphi) = k\}$$
$$= \sum_{k \geq j} \Pr\{\#(C(x)) = j \,|\, c(\varphi) = k\} \Pr\{c(\varphi) = k\} \ .$$

The conditional probability equals $1/k$ by Lemma 1 and our observation about random permutations of $[1, k]$.

Let $F(j)$ and $f(j)$ be the density functions for $\#(C(x))$ and $c(\varphi)$ respectively. We can rewrite (5.1) as $F(j) = \sum f(k)/k$ or, equivalently, $f(j) = j(F(j) - F(j + 1))$. These lead to the asymptotic results

$$(5.2) \qquad F(j) = \int_j^\infty \frac{f(v)}{v} dv \quad \text{and} \quad f(j) = -jF'(j) \ .$$

We now show that $t(x)$ and $\#(C(x))$ have asymptotically the same distribution and that a relationship like (5.1) holds asymptotically for $s(x)$ and $t(x)$. Choose a point x at random and construct its six, except for the final closure. Even allowing for indegree restrictions and constraints on the number of origins, it is easy to see that all closures, except possibly the one which makes x cyclic are equally likely. For asymptotic purposes, this cyclic closure is unimportant since $\mathscr{D}$ contains a number exceeding 1. It follows that $t(x)$ and $\#(C(x))$ have asymptotically the same distribution and that

$$\Pr\{t(x) = j \,|\, s(x) = k\} \sim 1/k \quad \text{for} \quad 0 < j < k \ .$$

(We do not know if this holds for $j = 0$.) This is enough to determine the distribution of $s(x)$ just as we proved (5.1): for $j > 1$ we have

$$\mathrm{Pr}\,\{t(x) = j\} = \sum_{k>j} \mathrm{Pr}\,\{t(x) = j\,|\,s(x) = k\}\,\mathrm{Pr}\,\{s(x) = k\}$$

$$\sim \sum_{k>j} \mathrm{Pr}\,\{s(x) = k\}/k \quad \text{as} \quad j \longrightarrow \infty\,.$$

Since $t(x)$ and $\#(C(x))$ have asymptotically the same distribution,

$$\mathrm{Pr}\,\{\#(C(x)) = j\} \sim \sum_{k \geqq j} \mathrm{Pr}\,\{s(x) = k\}/k \quad \text{as} \quad j \longrightarrow \infty\,.$$

Comparing this with (5.1) we see that $\mathrm{Pr}\,\{s(x) = j\} \sim \mathrm{Pr}\,\{c(\varphi) = j\}$. For a fixed j we obtain the desired result trivially since both the $s(x)$ and $c(\varphi)$ density functions approach zero rapidly. The asymptotic density function for $s(x)$ is determined in the next section.

We now assume that the asymptotic density function for $c(\varphi)$ is given and derive the asymptotic density function for $m(\varphi)$. We need the result of Gončarov [5] that the number of cycles in a random permutation on a k element set is asymptotically normally distributed with mean and variance $\ln k$. Combining this result with Lemma 1 we have

$$\mathrm{Pr}\,\{m(\varphi) = t\} \sim \sum_{k \geqq 1} \mathrm{Pr}\,\{c(\varphi) = k\}\,\mathrm{Pr}\,\{t\,|\,\Psi(\ln k,\,\ln k)\}\,,$$

where $\Psi(\mu,\,v)$ denotes the normal distribution with mean μ and variance v. The table entry for $m(\varphi)$ now follows easily from the table entry for $c(\varphi)$. (All that is needed about the distribution of $c(\varphi)$ is its mean, the fact the mean is unbounded as $n \to \infty$, and the fact that asymptotically the distribution is clumped near its mean.)

6. **The density function for six length.** Rubin and Sitgreaves obtain simple expressions for $\mathrm{Pr}\,\{s(x) = s\}$ when $\mathscr{D}$ contains all non-negative integers and when $\mathscr{D} = \{0,\,k\}$. See their formulas (7.2) and (13.24). We determine the asymptotic density function for $s(x)$ in general by essentially the same method that they use for the case $\mathscr{D} = [0,\,k]$. Let $\mathscr{N}_i \subseteqq X$ be those points with indegree i. Then $n_i = \#(\mathscr{N}_i)$. Let $\mathrm{Ch}\,(A\,|\,B)$ denote the number of ways to choose A subject to the condition B. By a simple counting argument

$$(6.1) \qquad\qquad \mathrm{Ch}\,((\varphi,\,x)\,|\,\mathscr{N}) = \frac{n \cdot n!}{\prod_i i!^{n_i}}\,.$$

Let $S(x) = \{\varphi^k(x),\,k \geq 1\}$, $\mathscr{S}_i = \mathscr{N}_i \cap S(x)$, and $s_i = \#(\mathscr{S}_i)$. We have

$$\mathrm{Ch}\,((\varphi,\,x)\,|\,\mathscr{N},\,s(x) = s,\,\#(C(x)) = c < s,\,s,\,x \in \mathscr{N}_j,\,\varphi^s(x) \in \mathscr{S}_k)$$

$$= \prod_i \mathrm{Ch}\,(\mathscr{S}_i) \cdot \mathrm{Ch}\,(x) \cdot \mathrm{Ch}(\varphi^s(x)) \cdot \mathrm{Ch}\,(\text{order for the six})$$

$$\times\,\mathrm{Ch}\,(\text{successors for points not in the six})$$

$$= \prod_i \binom{n_i}{s_i} \cdot (n_j - s_j) \cdot s_k \cdot (s - 2)!$$

$$\times (n - s)! / (\prod_i i!^{n_i - s_i})(\prod_{i \neq k} (i - 1)!^{s_i})(k - 1)!^{s_k - 1}(k - 2)! \ .$$

Rearranging, summing over $c < s$, j, and k, and using $\sum s_j = s - 1$, we obtain

Ch $((\varphi, x) \,|\, \mathscr{N}, s(x) = s, \#(C(x)) < s, s)$

$$= (n + 1 - s)!(s - 1)!(\sum_{k \neq 0} (k - 1)s_k) \prod_i \binom{n_i}{s_i} i^{s_i} \Big/ i!^{n_i} \ .$$

Dividing this by (6.1) we obtain

(6.2)
$$\Pr \{s(x) = s, \ t(x) \neq 0, \ \text{and} \ s \,|\, \mathscr{N}\}$$
$$= ((n + 1 - s)!(s - 1)! / n \cdot n!)(\sum_{k \,l\, 0} (k - 1)s_k) \prod_i \binom{n_i}{s_i} i^{s_i} \ .$$

The modifications for the case $c = s$ are straightforward. They arise from three differences; first, the start of the six is now no different from any of the other points on the six, second, $\#(S(x)) = s$, one larger than it was before, and, third, there is no need to sum over c and j. We obtain

(6.3)
$$\Pr \{s(x) = s, \ t(x) = 0, \ \text{and} \ s \,|\, \mathscr{N}\}$$
$$= ((n - s)! \ s! / n \cdot n!) \prod_i \binom{n_i}{s_i} i^{s_i} \ .$$

We now express the sum of (6.2) and (6.3) over all s as the coefficient of an analytic function. The coefficient of z^i in $(1 + kz)^r$ is $\binom{r}{i} k^i$ and so the coefficient of z^i in $k(k - 1)rz(1 + kz)^{r-1}$ is $(k - 1)i\binom{r}{i}k^i$. Recalling that in (6.2) $\sum s_i = s - 1$, we obtain

(6.4)
$$\Pr \{s(x) = s \,|\, \mathscr{N}\}$$
$$= \frac{(n + 1 - s)! \ (s - 1)!}{n \cdot n!}$$
$$\times \operatorname*{Coef}_{z^s} \left[\prod_i (1 + iz)^{n_i} \left(\frac{s}{n + 1 - s} + \sum_k \frac{k(k - 1)n_k z^2}{1 + kz} \right) \right] \ .$$

The coefficient of z^s can be determind by the saddle point method. A description of this technique is given in de Bruijn [4]. Let $f(z)$ be the function from which we wish to extract the coefficient. By the Cauchy Residue theorem, the coefficient of z^s is $\oint (f(z)dz)/(2\pi i z^{s+1})$ where the contour is $|z| = r = s/n$. We have

$$|1 + kz|^2 = (1 + kr)^2 - 2kr(1 - \cos \theta) \sim (1 + kr)^2 \exp (-kr\theta^2/(1 + kr)^2)$$

$$\arg (1 + kz) = \arctan ((kr \sin \theta)/(1 + kr \cos \theta)) \sim kr\theta/(1 + kr) \ .$$

From this we can conclude that only very small values of θ contribute significantly to the contour integral and in this range

$$f(z) \sim f(r) \exp (i\theta s \sum kn_k/n(1 + kr) - s\theta^2 \sum kn_k/2n(1 + kr)^2) \ .$$

Note that

$$\sum kn_k/n(1 + kr) = \sum kn_k/n + O(\sum rk(k - 1)n_k/n) \quad \text{and}$$
$$\sum kn_k/2n(1 + kr)^2 = \sum kn_k/2n + O(\sum rk(k - 1)n_k/n) \ .$$

When n_k is near its expected value, $n\rho\beta^{k-1}/k!$ for $k \in \mathscr{D}$ and $k \neq 0$, we may use this in place of n_i to simplify the above equation:

$$\cdot \quad f(z) \sim f(r) \exp (-\theta^2 s/2 + i\theta s) \ .$$

Using this in the contour integral we obtain the estimate

(6.5 a) $$\underset{z^s}{\operatorname{Coef}} f(z) \sim f(r)/r^s\sqrt{2\pi s} \ .$$

Using Stirling's formula

$$k! \sim \sqrt{2\pi k}(k/e)^k \ ,$$

and [1, (3.8)] we have

(6.5 b) $$(n + 1 - s)! \ (s - 1)!/n \cdot n! \sim \sqrt{2\pi/s}(s/en)^s \exp (s^2/2n) \ .$$

We also have

(6.5 c)
$$f(r) \sim \prod (1 + kr)^{n_k}(s/n + (s^2/n) \sum k(k - 1)n_k/n(1 + kr))$$
$$\sim \exp (\sum n_k(kr - k^2r^2/2))(s/n + \lambda s^2/n)$$
$$\sim \exp (s - (1 + \lambda)s^2/2n)(\lambda s^2/n) \ .$$

Combining (6.4) and (6.5) we obtain

(6.6) $$\Pr \{s(x) = s \,|\, \mathscr{N}\} \sim (\lambda s/n) \exp (-\lambda s^2/2n) \ ,$$

where the asymptotic result requires $s \to \infty$ and $s/n^{2/3} \to 0$. Outside this range the density function is very small. Because the variance of $n_r(\varphi)$ is proportional to it mean, it follows that $\#(\mathscr{N}_r) \sim \overline{n_r(\varphi)}$ with probability 1. Consequently, we may remove the conditioning on $\mathscr{N}$ in (6.6).

7. **Generating functions.** In this section we develop generating functions and use them to study the means of $n_0(\varphi)$, $m(\varphi)$, $\#(K(x))$, $\#(C(x))$, $\#(T(x))$, $n_r(\varphi)$, and $\lambda(\varphi)$, respectively. This illustrates the ease with which generating functions can often be used to study means. We also determine the asymptotic density function for

$\sharp(K(x))$ and $k_k(x)$. Until the last paragraph of the section, we assume that $gcd(\mathscr{D}) = 1$. Recall the convention introduced in § 2 that $\sum'$ indicates a sum over all nonzero $m \in \mathscr{D}$ for which the factorials in the sum make sense. The quantities we compute are weighted by y^{n_0} as noted at the end of § 1. We use the variables x, y, and z to keep track of quantities as follows:

x the number of points in the graph ,

y the number of origins , and

z the number of cycles .

The variable x is exponential since the elements of X are labeled. For example, the coefficient of $x^{20}y^5z^4/20!$ is a number associated with mappings on twenty points having five origins and four cycles.

The methods in this paragraph are fairly standard but few descriptions exist. Bender and Goldman [3] discuss a more general situation. There are $(c - 1)!$ possible c-cycles on $[1, c]$. By the exponential formula [3, Thm. 2], the generating function for mappings is given by

$$(7.1) \qquad M = \exp\left(\sum_{c \geq 1} z(c - 1)! \, R^c/c!\right) = (1 - R)^{-z} ,$$

where R is the generating function for objects making up the cycles. These objects are rooted directed trees where all paths lead to the root, the in-degrees of all nonroot vertices belong to $\mathscr{D}$, and one plus the degree of the root belongs to $\mathscr{D}$. We have

$$(7.2) \qquad R = x \sum' B^{m-1}/(m - 1)! ,$$

where B counts trees like those R counts, without the exception for the degree of the root. Thus

$$(7.3) \qquad B = xy + x \sum' B^m/m! .$$

The derivations of (7.2) and (7.3) are similar. In the latter case, a tree consists of either an origin or a root where exactly m roots are joined. The origin contributes the term xy. The union of m trees contributes $B^m/m!$, the $m!$ arising because the order of the trees is irrelevant, and the new root contributes x.

The simplest approach is to solve (7.3) for B and combine this with (7.2) and (7.1); however, this can be done for very few $\mathscr{D}$. Since we intend to study various derivatives of (7.1) and of related generating functions, we can avoid this problem by using implicit differentiation and the following asymptotic estimate.

LEMMA 2. *Suppose $gcd(\mathscr{D}) = 1$ and let $a_n(k)$ denote the coefficient*

of x^n in $(1 - R)^{-k}$. Then

$$a_n(k) \sim n^{k/2-1}/((2\lambda)^{k/2}\rho^n\Gamma(k/2)) ,$$

where β is the positive root of $\sum'(m - 1)\beta^m/m! = y$,

$$\rho = 1/(\sum'\beta^{m-1}/(m - 1)!) , \quad and \quad \lambda = \rho\sum'\beta^{m-1}/(m - 2)! .$$

Proof. Throughout the proof, $A_i(x)$ denotes a function which is analytic near the point ρ defined below. Let $F(x, B) = xy + x\sum' B^m/m! - B$. By the proof of Theorem 5 of [1, p. 505] and (7.3),

(7.5) $$B = B(x) = A_1(x) + A_2(x)(1 - x/\rho)^{1/2} ,$$

where $A_2(\rho) \neq 0$, and $\rho > 0$ and $\beta > 0$ are the solution of the pair of equations $F(\rho, \beta) = 0$, $F_B(\rho, \beta) = 0$. Manipulating these we obtain

(7.6a) $$\sum'(m - 1)\beta^m/m! = y$$

(7.6b) $$\rho\sum'\beta^{m-1}/(m - 1)! = 1 .$$

We can find β from (7.6a) and then ρ from (7.6b). (Note that $gcd(\mathscr{D}) = 1$ implies (7.6a) has a unique root of minimum modulus.) Rewriting (7.6b) we obtain $R(\rho) = 1$. By (7.2) and (7.5),

$$R(x) = A_3(x) + A_4(x)(1 - x/\rho)^{1/2} ,$$

where $A_4(\rho) \neq 0$. Since $R(\rho) = 1$,

$$1 - R(x) = A_5(x)(1 - x/\rho) - A_4(x)(1 - x/\rho)^{1/2} .$$

Using this and $A_4(\rho) \neq 0$, we have

$$\begin{aligned}(1 - R(x))^{-1} &= (1 - x/\rho)^{-1/2}(-A_4(x) + A_5(x)(1 - x/\rho)^{1/2})^{-1}\\ &= (1 - x/\rho)^{-1/2}(A_6(x) + A_7(x)(1 - x/\rho)^{1/2})\\ &= A_6(x)(1 - x/\rho)^{-1/2} + A_7(x) .\end{aligned}$$

By the parenthetic note after (7.6) and Theorem 4 of [1, p. 498], the coefficients of $(1 - R(x))^{-k}$ are asymptotic to $b^k n^{k/2-1}(1/\rho)^n/\Gamma(k/2)$, where $b = A_6(\rho)$. We find $\lambda = b^{-2}/2$ as follows:

$$b^{-2} = \lim_{x \to \rho}(1 - R(x))^2/(1 - x/\rho) = -\rho\frac{d(R(x) - 1)^2}{dx}\bigg|_{x=\rho} .$$

Differentiating (7.2) directly and (7.3) implicitly we have

(7.7a) $$R'(x) = \sum' B^{m-1}/(m - 1)! + xB'\sum' B^{m-2}/(m - 2)!$$

(7.7b) $$B'(x) = \frac{y + \sum' B^m/m!}{1 - R(x)} ,$$

which we substitute in the above to obtain

$$\lambda = b^{-2}/2 = \rho^2(\sum' \beta^{m-2}/(m-2)!)(y + \sum' \beta^m/m!)$$
$$= \rho\beta \sum' \beta^{m-2}/(m-2)! \quad \text{by (7.6)}. \qquad \square$$

The (weighted) total number of mappings is simply the coefficient of $x^n/n!$ in M when $z = 1$. We have $M = (1 - R)^{-1}$. By Lemma 2

$$(7.8) \qquad \text{(number of mappings)} \sim n!/(\rho^n\sqrt{2n\lambda}\,\Gamma(1/2)).$$

Since y keeps track of the number of origins, $y\,\partial M/\partial y$ provides information about the average number of origins. Differentiating (7.1) and (7.2) directly and differentiating (7.3) implicitly, we obtain

$$y\,\partial M/\partial y = x^2 y(\sum' B^{m-2}/(m-2)!)(1-R)^{-3}.$$

Applying Lemma 2 and dividing by (7.8) we obtain the (weighted) average number of origins:

$$(7.9) \qquad \overline{n_0(\varphi)} \sim \rho^2 y(\sum' \beta^{m-2}/(m-2)!)n\Gamma(1/2)/(\lambda\Gamma(3/2)) = ny\rho/\beta.$$

We must set this equal to some value and solve for y, ρ, and β. Substituting from Lemma 2 and rearranging we have the rule:

If $\overline{n_0(\varphi)}/n = f$ is given, let β be the positive
root of $\sum'(m - mf - 1)\beta^m/m! = 0$ and define
$\rho = 1/(\sum' \beta^{m-1}/(m-1)!)$ and $y = f\beta/\rho$.

Since $n_0(\varphi)$ is normally distributed with variance proportional to n, most of the contribution to our various weighted averages come from mappings with $n_0(\varphi)$ near fn.

We now study the average number of cycles. Using the operator $z\,\partial/\partial z$ and setting $z = 1$, we have $-(1-R)^{-1}\ln(1-R)$. Since $\ln(1-R)$ differs from $\ln\sqrt{1 - x/\rho}$ by a function which is analytic near ρ, we need only consider $-(1-R)^{-1}\ln\sqrt{1 - x/\rho}$. Using (7.8), Lemma 2, and either the Taylor series expansion for ln or Theorems 1 and A of [7],

$$(7.10) \qquad \overline{m(\varphi)} \sim \frac{1}{2}\sum_{k=1}^{n-1}\frac{1}{k}\left(\frac{n}{n-k}\right)^{1/2} \sim \frac{1}{2}\ln n.$$

We now study the density function for $\sharp(K(x))$, the size of the component containing x. Rubin and Sitgreaves derive explicit sums when $\mathscr{D}$ contains all nonnegative integers and when $\mathscr{D} = \{0, k\}$ in their equations (6.2) and (13.20) respectively. We begin by determining the number of mappings with connected graphs. The generating function is $\sum R^c/c = -\ln(1-R)$. As in the previous paragraph, this differs from $-\ln\sqrt{1 - x/\rho}$ by a function which is analytic near

ρ, and so

(7.11)   (number of one component mappings) $\sim n!/2n\rho^n$.

Letting $\mathrm{Ch}(\cdots)$ denote the number of ways to choose $\cdots$, we have

$$\mathrm{Pr}\{\sharp(K(x)) = j\} = \mathrm{Ch}(j - 1 \text{ vertices})\cdot\mathrm{Ch}(\text{size } j \text{ component})$$
$$\times \mathrm{Ch}(\text{size } n - j \text{ graph})/\mathrm{Ch}(\text{size } n \text{ graph})$$
$$\sim \binom{n-1}{j-1}(j!/2j\rho^j)((n-j)!/\sqrt{2\pi\lambda(n-j)}\rho^{n-j})$$
$$\times (\sqrt{2\pi\lambda n}\rho^n/n!)$$
$$= 1/2\sqrt{n(n-j)} ,$$

where we have used (7.8) and (7.11) to obtain the asymptotic formula and so require that $j \to \infty$ and $(n - j) \to \infty$.

Using the result for $\sharp(K(x))$, we easily obtain the result for $k_t(x)$:

$$\mathrm{Pr}\{k_t(x) = i\} = \sum_{j=0}^{n} \binom{t}{i} \mathrm{Pr}\{\sharp(K(x)) = j\}(j/n)^i(1 - j/n)^{t-i}$$

$$\sim \sum_{j=0}^{n} \binom{t}{i}(j/n)^i(1 - j/n)^{t-i-1/2}/2n$$

$$\sim \frac{1}{2}\binom{t}{i}\int_0^1 x^i(1 - x)^{t-i-1/2}dx = \frac{1}{2t+1}\frac{\binom{t}{i}}{\binom{t-1/2}{i}} .$$

We used the standard formula $\int x^{r-1}(1 - x)^{s-1}dx = \Gamma(r)\Gamma(s)/\Gamma(r + s)$. When $i = t$, the value of the probability is $(2^t t!)^2/(2t + 1)!$. To compute the expected value, we introduce a factor of i into the original sum and proceed as before.

We now average over points. To do this we distinguish a special element $x_0 \in X$, let G be the generating function for the "part" containing x_0, and let H be the generating function for the remainder. If F is the total generating function, then $F_x = G_x H$. The partial derivative selects the distinguished point x_0. On the right hand side it is applied only to G rather than to GH because x_0 lies in the part associated with G. The following observation is useful. Near $x = \rho$, the dominant part of R' is

$$x(\sum' B^{m-2}/(m - 2)!)(y + \sum' B^m/m!)/(1 - R) .$$

Near $x = \rho$ we can replace x by ρ and B by β in the above for asymptotic purposes. Thus

(7.12)   as $x \longrightarrow \rho$ $R' \sim (\lambda/\rho)(1 - R)^{-1}$.

Although we have already determined the asymptotic probability density function for $\#(K(x))$, we determine its mean because it provides a simple illustration of the method. Since we wish to count the number of vertices in the component containing x_0, we replace x^k by kx^k in the generating function of a component. Thus

$$G = x\frac{\partial}{\partial x}\sum_{c\geq 1} R^c/c = xR'/(1 - R) \, .$$

By (7.12), $G \sim \lambda(1 - R)^{-2}$. Thus

$$G_x \sim \lambda 2R'(1 - R)^{-3} \sim (2\lambda^2/\rho)(1 - R)^{-4} \, .$$

Since a general graph without the component containing x_0 is still a general graph, $H = M = (1 - R)^{-1}$ and

$$F_x \sim (2\lambda^2/\rho)(1 - R)^{-5} \, .$$

By Lemma 2, and the fact that $n^{3/2}$ is asymptotic to $(n - 1)^{3/2}$, the coefficient of $x^{n-1}/(n - 1)!$ is asymptotic to

$$(n - 1)! \, (2\lambda^2/\rho)n^{3/2}/(\rho^{n-1}(2\lambda)^{5/2}\Gamma(5/2)) = 2n! \, \sqrt{n}/(3\rho^n\sqrt{2\lambda}\Gamma(1/2)) \, .$$

Combining this with (7.8) we obtain $\overline{\#(K(x))} \sim 2n/3$.

For $\#(C(x))$, the size of the cycle that x leads into,

$$G = \sum_{c\geq 1} c(R^c/c) = R/(1 - R)$$

and, again, $H = M$. Using (7.12) we obtain

$$F_x \sim (\lambda/\rho)(1 - R)^{-4} \, .$$

By Lemma 2, and the fact that n^2 is asymptotic to $(n - 1)^2$, the coefficient of $x^{n-1}/(n - 1)!$ is asymptotic to

$$(n - 1)! \, (\lambda/\rho)^n n/((2\lambda)^2\rho^{n-1}\Gamma(2)) = n!/4\lambda\rho^n \, .$$

Combining this with (7.8) we obtain

$$\overline{\#(C(x))} \sim \frac{1}{2}\sqrt{n/2\lambda}\Gamma(1/2) \, .$$

For $\#(T(x))$ we set $G = xR'$. The rest of the cycle for the component containing x_0 has generating function $\sum R^{c-1} = (1 - R)^{-1}$, and the generating function for the remainder of the graph is M. Thus $H = (1 - R)^{-2}$. By (7.12) $G_x \sim (\lambda^2/\rho)(1 - R)^{-3}$ and since $R(\rho) = 1$,

$$F_x \sim (\lambda^2/\rho)(1 - R)^{-5} \, .$$

Since this is half the function studied for $K(x)$,

$$\overline{\#(T(x))} \sim \overline{\#(K(x))}/2 \sim n/3 \ .$$

The conjectured distribution for $\#(T(x))$ is based on an interesting new result of Meir and Moon [8]. Consider the set of k vertex rooted trees where each vertex has indegree in $\mathscr{D}$ and each tree is weighted by y^{n_0} as usual. Meir and Moon show that the probability that a point is at a distance j from the root is

$$(\lambda j/k) \exp(-\lambda j^2/2k) \quad \text{when} \quad j/\sqrt{k} \ \text{is bounded} \ .$$

Replacing j by $j + 1$ and pretending that $T(x)$ is obtained from (7.3) rather than (7.2), we have

$$\Pr\{t(x) = j \mid \#(T(x)) = k\} \sim (\lambda j/k) \exp(-\lambda j^2/2k) \ .$$

We will ignore the restriction on j. It follows that

$$\Pr\{t(x) = j\} \sim \sum_{k \geq j} \Pr\{\#(T(x)) = k\}(\lambda j/k) \exp(-\lambda j^2/2k) \ .$$

The left hand side of this relationship appears in Table II. Define

$$k = un \quad j = v\sqrt{n/\lambda} \quad p(u) = \lim_{x \to \infty} n \Pr\{\#(T(x)) = k\} \ .$$

Passing to the limit in the asymptotic relation we obtain the integral equation

$$\int_v^\infty e^{-u^2/2} \, du = v \int_0^1 \frac{p(u)}{u} e^{-v^2/2u} \, du \ .$$

It is easily verified that $p(u) = 1/2\sqrt{u}$ is the solution. This gives the result in Table II. We suspect that this proof can be made rigorous.

Although $n_r(\varphi)$, the number of points with indegree r, is studied in detail in §9, we would like to point out the ease with which the mean of $n_r(\varphi)$ can be found. Since we have studied $n_0(\varphi)$ in (7.9), we assume that $r > 0$. Introduce a variable w to keep track of the points counted by $n_r(\varphi)$:

$$R = x \sum' B^{m-1}/(m - 1)! + x(w - 1)B^{r-1}/(r - 1)! \quad \text{and}$$
$$B = xy + x \sum' B^m/m! + x(w - 1)B^r/r! \ .$$

Differentiating as usual and evaluating at $w = 1$ we obtain

$$R_w \sim \rho B_w \sum' (m - 1)\beta^{m-2}/(m - 1)! = \lambda B_w/\beta$$
$$B_w \sim (\rho\beta^r/r!)/(1 - x \sum' mB^{m-1}/m!) = \rho\beta^r/r!(1 - R)$$
$$M_w \sim (\rho\lambda\beta^{r-1}/r!)/(1 - R)^{-3} \ .$$

By Lemma 2, the coefficient of $x^n/n!$ is asymptotic to

$$n! \, \rho\beta^{r-1}\sqrt{n/2\lambda}/(r! \, \rho^n \Gamma(1/2)) \ .$$

By (7.8), the expected value of n_r is $(\rho\beta^{r-1}/r!)n$. Since

$$\lambda(\varphi) = \sum{}' m(m-1)n_m(\varphi)/n \ ,$$

the expected value of $\lambda(\varphi)$ is $\rho \sum{}' m(m-1)\beta^{m-1}/m! = \lambda$.

We can study the total number of predecessors of noncyclic points by introducing another variable u as follows

$$B(x, u) = uxy + ux \sum{}' B(ux, u)^m/m!$$
$$R(x, u) = x \sum{}' B(x, u)^{m-1}/(m-1)! \ ,$$

and we can consider all predecessors by also replacing M with

$$M(x, u) = \exp\left(\sum_{c \geq 1} R(u^c x, u)/c\right) .$$

If we wish to study only the predecessors of cyclic points, we can use (7.2) and (7.3) to define R and B and define M as above with $R(u^c x, u)$ replaced by $R(u^c x)$. We do not pursue this here. See § 8.

Suppose $gcd(\mathscr{D}) = d > 1$, then the powers of x in B/x and R are multiples of d. If we replace x^d by a new variable, the above methods continue to apply. In this case, all estimates of power series coefficients end up being multiplied by d. Since averages and distributions involve quotients, they are unchanged.

8. **The total number of predecessors.** The mean total number of predecessors of a point is easily obtained. Since x is a predecessor of $\#(C(x))$ cyclic points, the total number of predecessors of cyclic points is precisely the sum of $\#(C(x))$. Since there are $c(\varphi)$ cyclic points, it is easy to see that the average number of total predecessors of a cyclic point is asymptotically $n/2$. Since x is a predecessor precisely of those points counted in $s(x)$, the mean number of total predecessors of a point is the mean of $s(x)$ which is asymptotically $\sqrt{\pi n/2\lambda}$. Also, the total number of predecessors of noncyclic points is the sum of $t(x)$, and so the mean number of total predecessors of noncyclic points is asymptotically the same as $\overline{t(x)} \sim \sqrt{\pi n/8\lambda}$. See the next to the last paragraph of § 7 for a generating function approach.

It is possible to obtain fairly simple expressions for the expected value of $N_r(\varphi)$ in certain cases. Rubin and Sitgreaves show that it equals

$$\frac{n! \, r^{r-1}(n-r)^{n-r}}{(n-r)! \, r! \, n^{n-1}}$$

when $\mathscr{D}$ consists of all nonnegative integers and obtain some rather messy expressions when $\mathscr{D} = \{0, k\}$. The references are their formulae

(8.2), (13.26), and (13.27). It appears that simple sums are impossible to obtain in the general case. We treat it asymptotically by using a standard branching theory technique that was also employed by Rubin and Sitgreaves.

Note that except for cyclic points, the number of predecessors of a point must be one more than a multiple of d. Since the number of cyclic points grows like $\sqrt{n}$, we can ignore them provided $r \ll n$.

Let Ψ_r be the expected value of the probability $N_r(\varphi)/n$, and let $p_r = \overline{p_r(\varphi)}/n$, the expected value of the probability that a point has exactly r immediate predecessors. Then

$$(8.1) \qquad \Psi_r = \sum{}' p_m \sum \Psi_{s_1} \cdots \Psi_{s_m} \quad \text{for} \quad r > 1 \,,$$

where the inner sum extends over all m-tuples $s_1, \cdots, s_m$ such that the sum of the s_i equals $r - 1$ and we have assume that the probabilities involved are independent. To see this, let x be a random point and note that the factor Ψ_{s_i} is associated with the ith immediate predecessor of x and that this point has a total of s_i predecessors. Define $A(z) = \sum \Psi_r z^r$, the sum ranging over $r \geqq 1$, and rewrite (8.1) as

$$(8.2) \qquad A(z) = zp_0 + z \sum{}' p_m A(z)^m \,.$$

Since (8.1) was based on the assumption of independence, the coefficients of $A(z)$ are a good approximation provided $r \ll n$. We do not know p_m exactly, but we do have the asymptotic information of Table II for $\overline{n_m(\varphi)} = np_m$. Thus $p_m \sim \rho\beta^{m-1}/m!$ for $m \neq 0$ and $p_0 \sim y\rho/\beta$. Insert this in (8.2) and multiply by β to obtain

$$(8.3) \qquad \beta A(z) \approx (\rho z)y + (\rho z) \sum{}' (\beta A(z))^m/m! \,.$$

We can treat this as an equality and analyze it. Rubin and Sitgreaves use Lagrange inversion. This generally yields unwieldy results; however, it can be used to supplement the following method when r is small.

Since (8.3) has the same form as (7.3), it can be analyzed in the same way: If $d = 1$, Theorem 5 [1, p. 502] can be used with

$$F(z, w) = \rho y z + \rho z \sum{}' (\beta w)^m/m! - \beta w \,.$$

After a bit of calculation we obtain $\Psi_r \sim 1/\sqrt{2\pi\lambda r^3}$, asymptotically in r. As noted at the end of § 7, this must be multiplied by d if $d > 1$.

9. **Number of immediate predecessors.** In this section we show that every finite set of the $n_r(\varphi)$'s is asymptotically jointly normally distributed and obtain the variances and covariances. From

this it follows that λ is normally distributed with a computable variance.

Let $\mathscr{E}$ and $\mathscr{F}$ be disjoint finite subsets of $\mathscr{D}$ where $\#(\mathscr{D} - \mathscr{F}) \geqq 3$ and $\mathscr{E} \neq \varnothing$. We show that the probability that a map has $n_r(\varphi) = f_r(n)$ for $r \in \mathscr{F}$ and $n_r(\varphi) = \nu_r$ for $r \in \mathscr{E}$ is asymptotically normal in the ν_r with parameters depending on $\mathscr{D}$, $\mathscr{E}$, $\mathscr{F}$, and $f_r(n)$. It is assumed that the $f_r(n)$ are such that a large number of choices exist for n_r, $r \in \mathscr{E}$. When $\mathscr{D}$ is finite it suffices to consider the case $\mathscr{E} = \mathscr{D} - \mathscr{F}$. The modifications for infinite $\mathscr{D}$ are discussed after the finite case. The entry in Table II corresponds to $\mathscr{E} = \{i\}$ and either $\mathscr{F} = \varnothing$ or $\mathscr{F} = \{0\}$, but any set $\mathscr{E}$ containing i can be used to find the variance of i. We will use $\mathscr{E} = \{i, j\}$ so that we can compute covariances, too.

Some new notation is needed. Let β be the positive root of

$$(9.1) \qquad (1 - \sum_{r \in \mathscr{F}} f_r(n)/n)/(\sum_{r \notin \mathscr{F}} \beta^{r-1}/r!)$$
$$= (1 - \sum_{r \in \mathscr{F}} rf_r(n)/n)/(\sum_{r \notin \mathscr{F}} \beta^{r-1}/(r-1)!) \,.$$

(It is to be understood throughout that a sum over $r \notin \mathscr{F}$ is limited to $\mathscr{D}$.) Define

$$(9.2\,\mathrm{a}) \qquad \rho = (1 - \sum_{r \in \mathscr{F}} rf_r(n)/n)/(\sum_{r \notin \mathscr{F}} \beta^{r-1}/(r-1)!) \,,$$

$$(9.2\,\mathrm{b}) \qquad \mu_r = n\rho\beta^{r-1}/r! \quad \text{and} \quad \tau_r = n_r - \mu_r \,.$$

Manipulating (9.1) and (9.2) we obtain

$$(9.3\,\mathrm{a}) \qquad \sum_{r \notin \mathscr{F}} \mu_r = n - \sum_{r \in \mathscr{F}} f_r(n)$$

$$(9.3\,\mathrm{b}) \qquad \sum_{r \notin \mathscr{F}} r\mu_r = n - \sum_{r \in \mathscr{F}} rf_r(n) \,.$$

Note that the constraints $\sum' n_r = n$ and $\sum' rn_r = n$ with the sums over all r lead to the constraints

$$(9.4) \qquad \sum_{r \notin \mathscr{F}} \tau_r = 0 \quad \text{and} \quad \sum_{r \notin \mathscr{F}} r\tau_r = 0 \,.$$

By a simple counting argument, the number of maps such that the n_r's have some preassigned values is

$$(9.5) \qquad N = \frac{(n!)^2}{\prod_{r \in \mathscr{D}} n_r! \, r!^{n_r}} \,.$$

By Taylor's theorem

$$\ln(1 + z) = z - z^2/2 + O(z^3)$$

and Stirling's formula

$$n! \sim \sqrt{2\pi n}(n/e)^n \,,$$

it follows that when $\tau_r = o(n^{2/3})$

(9.6)
$$n_r! \, r!^{n_r}/\mu_r! \, r!^{\mu_r} \sim (n\rho\beta^{r-1}/e)^{\tau_r}(1 + \tau_r/\mu_r)^{n_r}$$
$$\sim (n\rho/\beta)^{\tau_r}\beta^{r\tau_r} \exp{(\tau_r^2/2\mu_r)} \,.$$

Multiply (9.5) by the product over $r \notin \mathscr{F}$ of $\mu_r! \, r!^{\mu_r}$ and use (9.6) and (9.4) to obtain

(9.7)
$$N \sim G(n, f) \exp{(-\sum_{r\in} \tau_r^2/2\mu_r)} \,,$$

for $\tau_r = o(n^{2/3})$, where G is some function independent of the τ_r's which we are not particularly interested in. If any of the τ_r's are large compared to $\sqrt{n}$, then the value of N is small compared to the value of N when the τ_r's are zero. By this and (9.7), the τ_r's are asymptotically jointly normally distributed with means zero. It follows that the n_r's are asymptotically jointly normally distributed with means μ_r. They are not independent because of the constraints on the τ_r's. We can obtain a nonsingular distribution by selecting two distinct indices $i, j \notin \mathscr{F}$, solving for τ_i and τ_j in (9.4), and substituting the result into (9.7). The result is a quadratic form $\tau'A\tau$ inside the exponential function. The inverse of the matrix A is the covariance matrix for τ. Rubin and Sitgreaves use this to derive the covariance when $\mathscr{F} = \varnothing$ and $\mathscr{D} = [0, k]$ in their (14.22). Unfortunately, (9.7) does not apply to infinite $\mathscr{D}$.

We now turn our attention to (possibly) infinite $\mathscr{D}$. Let $\mathscr{R}$ be those elements of $\mathscr{D}$ which are not in either $\mathscr{E}$ or $\mathscr{F}$. We assume that $\#(\mathscr{R}) \geqq 2$, so the distribution on $\mathscr{E}$ is nonsingular.

Let b^* and c^* be given and b and c be arbitrary subject to $b - b^* = o(b^{*2/3})$ and $c - c^* = o(c^{*2/3})$. We next show that

(9.8) $$\sum \prod_{r\in\mathscr{R}} (n_r! \, r!^{n_r})^{-1} \sim \frac{B^b C^{-c}}{b!\sqrt{2\pi b\sigma^2}} \exp{(-(c - bc^*/b^*)^2/2b^*\sigma^2)} \,,$$

where the sum ranges over all n_r's such that

$$\sum_{r\in\mathscr{R}} n_r = b \quad \text{and} \quad \sum_{r\in\mathscr{R}} rn_r = c \,,$$

and the positive numbers $C, B,$ and σ are determined by

(9.9 a) $$(b^*/c^*) \sum_{r\in\mathscr{R}} rC^r/r! = \sum_{r\in\mathscr{R}} C^r/r! = B$$

(9.9 b) $$\sigma^2 = (\sum_{r\in\mathscr{R}} r^2C^r/r!)/B - (c^*/b^*)^2 \,.$$

If the left side of (9.8) is multiplied by $x^b y^c$ and summed over b and c, we obtain $\exp{(x \sum y^r/r!)}$, the sum ranging over $\mathscr{R}$. The coeffi-

cient of x^b is clearly $(\sum y^r/r!)^b/b!$.   Set $y = Cz$ where C is defined as above.   An application of the central limit theorem for independent identically distributed random variables as in [2, § 4] gives the desired result.   In this case there are b independent, identically distributed random variables ξ_i with $\Pr\{\xi_i = k\} = C^k/k!\, B$.   As in the derivation of (9.6), we have

$$b! \sim (b^*!/(b^*)^{b^*})(b^*)^b \exp((b - b^*)^2/2b^*) ,$$

which we need in the next paragraph.

We now return to our problem: We wish to sum (9.5) over all choices for n_r such that $n_r = f_r(n)$ for $r \in \mathscr{F}$, $n_r = \mu_r + \tau_r$ for $r \in \mathscr{E}$, and $\sum' n_r = \sum' r n_r = n$. (In other words, n_r is given for $r \in \mathscr{E} \cup \mathscr{F}$ and n_r is allowed to vary for $r \in \mathscr{R}$ subject to the constraints $\sum' n_r = \sum' r n_r = n$.)   The notation introduced in (9.1) and (9.2) still applies.   We now set

$$(9.10) \qquad b^* = \sum_{r \in \mathscr{R}} \mu_r \quad \text{and} \quad c^* = \sum_{r \in \mathscr{R}} r\mu_r .$$

It follows that $C = \beta$ and $B = b^*\beta/n\rho$. Set

$$b = b^* - \sum_{i \in \mathscr{E}} \tau_i \quad \text{and} \quad c = c^* - \sum_{i \in \mathscr{E}} i\tau_i$$

in (9.8).   By (9.4) the range of the sum in (9.8) is the range over which we wish to sum (9.5).   Combining this with (9.6) for $r \in \mathscr{E}$ we can rewrite (9.5) as

$$(9.11) \qquad N \sim H(n, f) \exp(-(S_1 + S_2 + S_3)/2) ,$$

where H is some function independent of the τ_i's,

$$(9.12\,\mathrm{a}) \qquad S_1 = \sum_{i \in \mathscr{E}} \tau_i^2/\mu_i ,$$

$$(9.12\,\mathrm{b}) \qquad S_2 = (\sum_{i \in \mathscr{E}} \tau_i)^2/b^* , \quad \text{and}$$

$$(9.12\,\mathrm{c}) \qquad S_3 = (\sum_{i \in \mathscr{E}} i\tau_i - (c^*/b^*) \sum_{i \in \mathscr{E}} \tau_i)^2/b^*\sigma^2 .$$

We now turn our attention to the covariance matrix for the $n_r(\varphi)$'s.   Covariance does not depend on which (marginal) distribution is used as long as it contains the two variables whose covariance we want.   We use (6.11) with $\mathscr{E} = \{i, j\}$ to obtain $\mathrm{cov}\,(\tau_i, \tau_j)$ and $\mathrm{var}\,(\tau_i)$. No new ideas are involved in doing this; however, we must unravel the notation in (9.1), (9.2), (9.3), (9.9 b), (9.10), and (9.12) and then carry out involved algebraic manipulations without making errors. There should be an easier way to arrive at the desired result, but we have not found it.   It will be convenient to introduce the notation $R_k = \sum r^k\mu_r$, the sum ranging over $r \in \mathscr{R}$ and $T_k = \sum t^k\mu_t$, the

sum ranging over $t \in \mathscr{R} \cup \mathscr{E}$.

We unravel. The values of β and ρ given by (9.1) and (9.2a) are the same as those given in § 2. Equations (9.8) and (9.9b) are

$$b^* = R_0 \qquad c^* = R_1 \qquad (b^*)^2 \sigma^2 = R_0 R_2 - R_1^2$$

equation (9.12c) is

$$S_3 = ((i\tau_i + j\tau_j)R_0 - (\tau_i + \tau_j)R_1)^2 / R_0(R_0 R_2 - R_1^2) .$$

This completes the unravelling of the notation.

The 2 by 2 matrix associated with the quadratic form $S_1 + S_2 + S_3$ turns out to be

$$\frac{1}{\mu_i \mu_j (R_0 R_2 - R_1^2)} \begin{bmatrix} \mu_j A(i) & \mu_i \mu_j B \\ \mu_i \mu_j B & \mu_i A(j) \end{bmatrix} ,$$

where

$$A(i) = R_0 R_2 - R_1^2 + \mu_i (R_2 + i^2 R_0 - 2i R_1)$$
$$B = R_2 + ij R_0 - (i + j) R_1 .$$

The inverse of this gives the variance and covariance. They can be written in a single formula:

$$(9.13) \qquad \text{cov}\,(\tau_i, \tau_j) = \delta_{ij} \mu_i + \mu_i \mu_j \left(\frac{T_2 - (i + j)T_1 + ij T_0}{T_0 T_2 - T_1^2} \right) .$$

To obtain the entry in Table II, consider the two cases $\mathscr{F} = \varnothing$ and $\mathscr{F} = \{0\}$.

Since the elements of every finite set of $n_r(\varphi)$'s have a joint normal distribution and since those with large r contribute very little to the variance and covariance, we can conclude that since $\lambda(\varphi)$ is a linear combination of them it is normally distributed. The mean of $\lambda(\varphi)$ was computed in § 7. Its variance is $\sum ij(i - 1)(j - 1) \times \text{cov}\,(\tau_i, \tau_j)/n^2$. This can be computed asymptotically using (9.13), but it is quite messy.

10. Concluding remarks. Table II states that the density function for $N_r(\varphi)$ is unknown. We have no ideas concerning its value or how it might be found. It also states that the density function for $\#(T(x))$ is uncertain. We presented a heuristic derivation of it in § 7 and data that casts doubt on its validity in Table III. We think the distribution is probably correct and simply does not fit the data so well when the number of origins is small; however, it is a source of some unease.

One can imagine other ways in which the set of mappings could be restricted. For example the variable $\lambda(\varphi)$ could be constrained. We have not explored this area much; however, one other situation

merits some comment. Suppose that $n_r(\varphi)$ is given for $r \in \mathscr{S}$ and the rest are free. (We considered the case in which $n_0(\varphi)$ is given.) We can compute $\overline{\lambda(\varphi)}$ using (9.1) and (9.2):

$$\overline{\lambda(\varphi)} \sim \sum_{r \notin \mathscr{S}} r(r-1)\mu_r + \sum_{r \in \mathscr{S}} r(r-1)f_r(n) \, .$$

If this value is used for λ, it will give the correct distribution for $s(x)$, as you can easily see from (6.6). By § 5, it follows that the entries in Table II are also correct for $c(\varphi)$, $\#(C(x)$, $t(x)$, and $m(\varphi)$. Equation (9.13) gives the covariances for the joint normal distribution of the $n_r(\varphi)$'s. The method of § 8 can still be used to study $N_r(\varphi)$. We conjecture that the remaining entries in Table II ($\#(K(x))$, $k_t(x)$, and $\#(T(x))$) remain unchanged.

It would be very interesting to restrict the number of mappings to mimic shift register data more accurately. We are unable to allow for the fact that shift registers do indeed shift bits. This omission seems unimportant for the cases we studied. It would be interesting to restrict the function g in (1.1) in some other fashion than simply by counting the number of zeros it has. We have no suggestions on how to attack such problems.

Note added in proof. The conjecture for $\#(T(x))$ in Table II has been verified using an argument like that following (7.11) combined with Lemma 2 and (7.8) as follows.

$$\Pr\{\#T(x) = j\} \cdot \mathrm{Ch}(\text{size } n \text{ mapping}) = \mathrm{Ch}(j-1 \text{ vertices}) \cdot \mathrm{Ch}(j \text{ vertex tree})$$
$$\times \sum_p \mathrm{Ch}(p - 1 \text{ trees with } n - j \text{ vertices}) \cdot \mathrm{Ch}(\text{perm. of } [1, p]) \, .$$

The sum is the coefficient of X^{n-j} in

$$\sum (R^{p-1}/(p-1)!)p! = -(1-R)^{-2} \, ,$$

which is obtained from Lemma 2 with $k = 2$. To obtain $\mathrm{Ch}(j$ vertex tree), apply Lemma 2 with $k = -1$.

REFERENCES

1. E. A. Bender, *Asymptotic methods in enumeration*, SIAM Rev., **16** (1974), 485–515.
2. ———, *Central and local limit theorems applied to asymptotic enumeration*, J. Combinatorial Theory—Series A, **15** (1973), 91–111.
3. E. A. Bender and J. R. Goldman, *Enumerative uses of generating functions*, Indiana Univ. Math. J., **20** (1971), 753–765.
4. N. G. de Bruijn, *Asymptotic Methods in Analysis*, North-Holland, 1958.
5. V. Gončarov, *On the field of combinatory analysis*, Amer. Math. Soc. Trans. Ser. 2, **19** (1962), 1–46. The original Russian article appeared in Izv. Akad. Nauk SSSR. Ser. Math., **8** (1944), 3–48.
6. B. Harris, *Probability distributions related to random mappings*, Ann. Math. Stat., **31** (1960), 1045–1062.

7. R. Jungen, *Sur les séries de Taylor n'ayant que des singularitiés algébrico-logarithmiques sur leur cercle de convergence*, Comm. Math. Helv., **3** (1931), 266–306.

8. A. Meir and J. W. Moon, *On the altitude of nodes in random trees*, Canadian J. Math., **30** (1978), 997–1015.

9. 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, January 1954.

10. L. A. Shepp and S. P. Lloyd, *Ordered cycle lengths in a random permutation*, Trans. Amer. Math. Soc., **121** (1966), 340–357.

INSTITUTE FOR DEFENSE ANALYSES
PRINCETON, NJ 08540
AND
UNIVERSITY OF CALIFORNIA, SAN DIEGO
LA JOLLA, CA 08540