CYCLE LENGTH IN A RANDOM FUNCTION

BY

P. W. PURDOM AND J. H. WILLIAMS

Let X be a finite set of n points and F_n be the class of n^n functions from X into X. For any $f \in F_n$ and $x_0 \in X$, the sequence, x_0, $x_1 = f(x_0)$, $x_2 = f(x_1), \ldots$, is eventually cyclic, i.e. there exists J and there exists l such that $j > J$ implies $x_j = x_{j+l}$. We will call l distinct points $x_i, x_{i+1}, \ldots, x_{i+l-1}$ a *cycle* if $x_{j+1} = f(x_j)$ $(i \leqq j \leqq i+l-2)$ and $f(x_{i+l-1}) = x_i$. Clearly different choices of the starting value, x_0, may lead to different cycles.

The length of the longest cycle in a function is of interest in the generation of pseudo-random numbers [1]. We consider the expected value of the length and the mth moment of the length of the ith longest cycle where the function $f \in F_n$ is selected at random.

Given $f \in F_n$, let Y be the subset of X consisting of all the points in cycles; then f restricted to Y is a permutation. Letting α be any characteristic of the cycle structure of a function $f \in F_n$ (e.g. $\alpha \equiv$ the longest cycle is of length l), we first find a formula relating the number of functions with characteristic α to the number of permutations with characteristic α. We then use the results of Shepp and Lloyd [2] giving asymptotic expressions for the expected values of the various moments of cycle lengths in permutations to find the asymptotic expressions for these values for functions.

We say that a function f *directly connects* x_i to x_j if $x_j = f(x_i)$ and that f *connects* x_i to x_j if there is a sequence of directly connected points starting with x_i going to x_j. Then the subset, Y, consists of just those points that are connected to themselves. We say that a subset $Z \subset X$ is a tree rooted on a point x_m if: (1) $x_m \in X - Z$, (2) $x \in Z$ implies x is connected to x_m, and (3) no point in $X - Z$ is connected to a point in Z. Clearly any $f \in F_n$ connects some of the points into cycles and the remainder of the points into trees rooted on points in cycles.

Let $T(n, m)$ denote the number of ways of connecting n points into trees rooted on m of the n points. Since $C_{n,m}$ is the number of ways the root points may be chosen and $m^i T(n-m, i)$ is the number of ways the remaining points may be connected if exactly i of them are directly connected to the m roots, we have the recurrence relation,

$$T(n, m) = C_{n,m} \sum_{i=0}^{n-m} m^i T(n-m, i),$$

where $T(n, 0) = 0$ for $n > 0$ and $T(0, 0) = 1$. The solution to the recurrence relation is

Received by the editors May 18, 1967.

$T(n, m) = C_{n-1,m-1} n^{n-m}$ which may be verified inductively by applying the binomial theorem in the induction step. If α is some characteristic of the cycle structure of a function, let $P(n, \alpha)$ denote the number of permutations of n points having cycles with characteristic α and $N(n, \alpha)$ denote the number of functions in F_n having cycles with characteristic α. Then,

$$N(n, \alpha) = \sum_{i=1}^{n} T(n, i) P(i, \alpha)$$

since there are $T(n, i)P(i, \alpha)$ ways of having i of the points in cycles, and there may be from 1 to n points in cycles.

Using $\alpha_{l,r}$ to denote the characteristic, "the rth longest cycle is of length l", we have

$$N(n, \alpha_{l,r}) = \sum_{i=l}^{n} T(n, i) P(i, \alpha_{l,r})$$

since $P(i, \alpha_{l,r}) = 0$ for $i < l$. Then, over F_n the expected value of the mth moment of the length, l, of the rth longest cycle is

$$E_{F_n,r}(l^m) = \frac{\sum_{l=1}^{n} l^m N(n, \alpha_{l,r})}{\sum_{l=0}^{n} N(n, \alpha_{l,r})},$$

since the $l=0$ term in the numerator contributes nothing for $m>0$. But $\sum_{l=0}^{n} N(n, \alpha_{l,r}) = n^n$ since this is just the number of functions in F_n. Therefore,

$$E_{F_n,r}(l^m) = \frac{\sum_{l=1}^{n} l^m \sum_{j=l}^{n} T(n, j) P(j, \alpha_{l,r})}{n^n}$$

$$= \frac{\sum_{j=1}^{n} T(n, j) \sum_{l=1}^{j} l^m P(j, \alpha_{l,r})}{n^n}.$$

The number of permutations of j points is $j!$; therefore over all the permutations of j points, P_j, the expected value of the mth moment of the length, l, of the rth longest cycle is

$$E_{P_j,r}(l^m) = \frac{\sum_{l=1}^{j} l^m P(j, \alpha_{l,r})}{j!}.$$

Therefore,

$$E_{F_n,r}(l^m) = \frac{\sum_{j=1}^{n} T(n, j) j! E_{P_j,r}(l^m)}{n^n}.$$

Shepp and Lloyd [2] show that

$$E_{P_j,r}(l^m) = (G_{r,m} + \varepsilon_{r,m,j}) j^m,$$

where

$$\lim_{j\to\infty} (\varepsilon_{r,m,j}) = 0,$$

$$G_{r,m} = \int_0^\infty \frac{x^{m-1}}{m!}\frac{[E(x)]^{r-1}}{(r-1)!}\exp[-E(x)-x]\,dx,$$

and

$$E(x) = \int_x^\infty \frac{e^{-y}}{y}\,dy.$$

Also, Shepp and Lloyd [2] show that

$$E_{P_j,1}(l) = G_{1,1}(j+\tfrac{1}{2})+o(1).$$

Therefore

$$\begin{aligned} E_{F_n,r}(l^m) &= \sum_{j=1}^{n} C_{n-1,j-1}n^{n-j}n^{-n}j!\,j^m(G_{r,m}+\varepsilon_{r,m,j}) \\ &= G_{r,m}\sum_{j=1}^{n}\frac{(n-1)!\,j^{m+1}}{(n-j)!\,n^j}+\bar{\varepsilon}_{r,m,n} \end{aligned}$$

where

$$\bar{\varepsilon}_{r,m,n} = \sum_{j=1}^{n}\frac{(n-1)!\,j^{m+1}}{(n-j)!\,n^j}\varepsilon_{r,m,j}.$$

We will let

$$Q_n(k) = \sum_{j=1}^{n}\frac{(n-1)!\,j^k}{(n-j)!\,n^j}$$

and show that

$$\text{(A)}\qquad \lim_{n\to\infty}\bar{\varepsilon}_{r,m,n}/Q_n(m+1) = 0.$$

Given $\delta > 0$ we first pick a k such that $|\varepsilon_{r,m,i}| < \delta$ for $i > k$ (Shepp and Lloyd [2]) and then rewrite our limit as

$$\text{(B)}\qquad \lim_{n\to\infty}\frac{\sum_{j=1}^{k}\frac{(n-1)!\,j^{m+1}}{(n-j)!\,n^j}\varepsilon_{n,m,j}}{Q_n(m+1)}+\frac{\sum_{j=k+1}^{n}\frac{(n-1)!\,j^{m+1}}{(n-j)!\,n^j}\varepsilon_{r,m,j}}{Q_n(m+1)}.$$

Then since all terms in $Q_n(m+1)$ are positive, the right hand term of (B) is less than δ. Since the first $\sqrt{n}-1$ terms of $Q_n(m+1)$ are always increasing, there exists a q such that

$$\left(\sum_{j=1}^{k}\frac{(n-1)!\,j^{m+1}}{(n-j)!\,n^j}\bigg/Q_n(m+1)\right) < \delta$$

when $n > q$. Letting $K = \max_{1\le i\le k}|\varepsilon_{r,m,i}|$, then when $n > q$, the left hand term of (B) is less than $K\delta$, so that we have (A). Thus,

$$\lim_{n\to\infty}\frac{E_{F_n,r}(l^m)}{Q_n(m+1)} = G_{r,m}.$$

To calculate $Q_n(k)$ for $k \geqq 1$ note that

$$\begin{aligned} nQ_n(k-1)-Q_n(k) &= \sum_{j=1}^{n} \frac{(n-1)!(n-j)j^{k-1}}{(n-j)!n^j} \\ &= -\delta_{k,1}+\sum_{j=0}^{n} \frac{(n-1)!(n-j)j^{k-1}}{(n-j)!n^j} \\ &= -\delta_{k,1}+\sum_{j=1}^{n} \frac{n!(j-1)^{k-1}}{(n-j)!n^j} \\ &= -\delta_{k,1}+\sum_{i=0}^{k-1} (-1)^{k-1-i}C_{k-1,i}\sum_{j=1}^{n} \frac{n!j^i}{(n-j)!n^j} \\ &= -\delta_{k,1}+\sum_{i=1}^{k-1} (-1)^{k-1-i}C_{k-1,i}nQ_n(i) \end{aligned}$$

where $\delta_{k,1}$ is the Kronecker delta. Therefore

$$Q_n(k) = n\left[\sum_{i=0}^{k-2} (-1)^{k-i}C_{k-1,i}Q_n(i)\right]+\delta_{k,1}.$$

The value of $Q_m(0)$ is $(n!e^n/n^{n+1})[1-\gamma(n,n)/(n-1)!]$ where $\gamma(n,n)$ is the incomplete gamma function and can be approximated by (Knuth [3])

$$\frac{1}{n}\left[\left(\frac{\pi n}{2}\right)^{1/2}-\frac{1}{3}+\frac{1}{12}\left(\frac{\pi}{2n}\right)^{1/2}-\frac{91}{540n}+\frac{1}{288}\left(\frac{\pi}{2n^3}\right)^{1/2}+O(n^{-2})\right].$$

Further, we have from the recurrence relation that $Q_n(1)=\delta_{1,1}=1$ and $Q_n(2) = nQ_n(0)$.

Now $Q_n(k)$ is a polynomial in n plus $Q_n(0)$ times a polynomial in n. For large n, $Q_n(k)$ can be approximated by its leading term; i.e. letting

$$\begin{aligned} a_{n,k} &= 1\cdot 3\cdot 5\cdots(k-1)n^{k/2}Q_n(0) \quad \text{if } k \text{ is even,} \\ &= 2\cdot 4\cdot 6\cdots(k-1)n^{(k-1)/2} \quad \text{if } k \text{ is odd,} \end{aligned}$$

then

$$Q_n(k) = a_{n,k}+o(n^{(k-1)/2}).$$

Collecting together the various results for large n we have:

$$E_{F_n,1}(l) = G_{1,1}\left(\frac{\pi n}{2}\right)^{1/2}+\frac{1}{6}+o(1)$$

$$\begin{aligned} E_{F_n,r}(l^m) &= (1\cdot 3\cdot 5\cdots m)\left(\frac{\pi}{2}\right)^{1/2} G_{r,m}n^{m/2}+o(n^{m/2}) \quad \text{for } m \text{ odd} \\ &= (2\cdot 4\cdot 6\cdots m)G_{r,m}n^{m/2}+o(n^{m/2}) \quad \text{for } m \text{ even.} \end{aligned}$$

Using Shepp and Lloyd's [2] results for moments of shortest cycles one can also show that

$$E_{F_n,r}(s) = S_{r,1}Q_n(1, r)+o(Q_n(1, r))$$

and

$$E_{F_n,r}(s^m) = S_{r,m}Q_n(m, r-1)+o(Q_n(m, r-1)) \quad \text{for } m > 2,$$

where $E_{F_n,r}(s^m)$ is the expected value of the mth moment of the rth shortest cycle, $S_{r,m}$ is a coefficient given in Shepp and Lloyd [2], and

$$Q_n(m, r) = \sum_{j=1}^{n} \frac{(n-1)!j^m(\log j)^r}{(n-j)!n^j}.$$

We have not found any asymptotic results for this sum when $r \neq 0$.

For values of n from 1 to 50 we compared the actual average length of the longest cycle to that predicted by the formula

$$l_{\text{ave}} = .7824816n^{1/2}+.104055+.0652068n^{-1/2}-.1052117n^{-1}+.0416667n^{-3/2}$$

obtained by taking the first five terms in the expansion of $G_{1,1}Q_n(2)+1/6$. For $n \leqq 5$ the formula gave too low an answer. For $6 \leqq n \leqq 50$ the formula gave too high an answer, with the maximum error of .00895 at $n=24$. Above $n=24$ the error slowly decreased to .00808 at $n=50$.

References

1. D. E. Knuth, *The art of computer programming*, Vol. 2, Addison-Wesley, Reading, Mass. (to appear).

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

3. D. E. Knuth, *The art of computer programming*, Vol. 1, Addison-Wesley, Reading, Mass., 1968, p. 117.

University of Wisconsin, Computer Sciences Department,
Madison, Wisconsin