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 $x_{j+1} = f(x_j)$ $(i \leq j \leq i+l-2)$ 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 $m$th moment of the length of the $i$th 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 $\alpha$ be any characteristic of the cycle structure of a function $f \in F_n$ (e.g. $\alpha = \text{the longest cycle is of length } l$), we first find a formula relating the number of functions with characteristic $\alpha$ to the number of permutations with characteristic $\alpha$. 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 $m'T(n-m, i)$ 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,

$$T(n, m) = C_{n,m} \sum_{i=0}^{m} m'iT(n-m, i),$$

where $T(n, 0) = 0$ for $n > 0$ and $T(0, 0) = 1$. The solution to the recurrence relation is
\( T(n, m) = C_{n-1,m-1} n^{n-m} \) which may be verified inductively by applying the binomial theorem in the induction step. If \( \alpha \) 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 \( \alpha \) and \( N(n, \alpha) \) denote the number of functions in \( F_n \) having cycles with characteristic \( \alpha \). 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 \( \alpha_{i,r} \) to denote the characteristic, "the \( r \)th longest cycle is of length \( l \)", we have

\[
N(n, \alpha_{i,r}) = \sum_{i=1}^{n} T(n, i) P(i, \alpha_{i,r})
\]

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

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

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

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

The number of permutations of \( j \) points is \( j! \); therefore over all the permutations of \( j \) points, \( P_j \), the expected value of the \( m \)th moment of the length, \( l \), of the \( r \)th longest cycle is

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

Therefore,

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

Shepp and Lloyd [2] show that

\[
E_{F_n,r}(l^m) = (G_{r,m} + e_{r,m}) j^m,
\]
where
\[
\lim_{j \to \infty} (\varepsilon_{r,m,i}) = 0,
\]
\[
G_{r,m} = \int_0^\infty x^{m-1} \frac{[E(x)]^{r-1}}{m!} \left[ \frac{1}{(r-1)!} \exp \left\{ -E(x) - x \right\} \right] dx,
\]
and
\[
E(x) = \int_x^\infty \frac{e^{-y}}{y} dy.
\]

Also, Shepp and Lloyd [2] show that
\[
E_{F_i,1}(l) = G_{1,1}(j + \frac{1}{2}) + o(1).
\]

Therefore
\[
E_{F_i,n}(l^m) = \sum_{j=1}^n C_{n-1,j-1} n^{n-j} n^{-m} j^m (G_{r,m} + \varepsilon_{r,m,i})
\]
\[
= G_{r,m} \sum_{j=1}^n \frac{(n-1)! j^{m+1}}{(n-j)! n^j} + \varepsilon_{r,m,n}
\]
where
\[
\varepsilon_{r,m,n} = \sum_{j=1}^n \frac{(n-1)! j^{m+1}}{(n-j)! n^j} \varepsilon_{r,m,j}.
\]

We will let
\[
Q_n(k) = \sum_{j=1}^n \frac{(n-1)! j^k}{(n-j)! n^j}
\]
and show that
\[
(A) \quad \lim_{n \to \infty} \varepsilon_{r,m,n}/Q_n(m+1) = 0.
\]

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

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

Then since all terms in \( Q_n(m+1) \) are positive, the right hand term of (B) is less than \( \delta \). Since the first \( \sqrt{n-1} \) terms of \( Q_n(m+1) \) are always increasing, there exists a \( q \) such that
\[
\left( \sum_{j=1}^k \frac{(n-1)! j^{m+1}}{(n-j)! n^j} / Q_n(m+1) \right) < \delta
\]
when \( n > q \). Letting \( K = \max_{1 \leq i \leq k} |\varepsilon_{r,m,i}| \), then when \( n > q \), the left hand term of (B) is less than \( K\delta \), so that we have (A). Thus,
\[
\lim_{n \to \infty} \frac{E_{F_i,n}(l^m)}{Q_n(m+1)} = G_{r,m}.
\]
To calculate $Q_n(k)$ for $k \geq 1$ note that

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

$$= -\delta_{k,1} + \sum_{j=0}^{k} \frac{(n-1)!(n-j)^{k-1}}{(n-j)!n!}$$

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

$$= -\delta_{k,1} + \sum_{i=0}^{k-1} (-1)^{k-1-i}C_{k-1,i} \sum_{j=1}^{n} \frac{n!j^i}{(n-j)!n!}$$

$$= -\delta_{k,1} + \sum_{i=1}^{k-1} (-1)^{k-1-i}C_{k-1,i}nQ_n(i)$$

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

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

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

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

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

$$a_{n,k} = 1 \cdot 3 \cdot 5 \cdots (k-1)n^{k/2}Q_n(0) \quad \text{if } k \text{ is even},$$

$$= 2 \cdot 4 \cdot 6 \cdots (k-1)n^{(k-1)/2} \quad \text{if } k \text{ is odd},$$

then

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

Collecting together the various results for large $n$ we have:

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

$$E_{F_{n,1}}(l^m) = (1 \cdot 3 \cdot 5 \cdots m) \left( \frac{\pi}{2} \right)^{1/2} G_{r,m}n^{m/2} + o(n^{m/2}) \quad \text{for } m \text{ odd}$$

$$= (2 \cdot 4 \cdot 6 \cdots m) G_{r,m}n^{m/2} + o(n^{m/2}) \quad \text{for } m \text{ even}.$$
Using Shepp and Lloyd's [2] results for moments of shortest cycles one can also show that

\[ E_{F_n}(s) = S_{r,1}Q_n(1, r) + o(Q_n(1, r)) \]

and

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

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

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

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 \( n \leq 5 \) the formula gave too low an answer. For \( 6 \leq n \leq 50 \) 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


University of Wisconsin, Computer Sciences Department, Madison, Wisconsin