REGENERATION POINTS IN RANDOM PERMUTATIONS

A. J. STAM
Mathematisch Instituut der Rijksuniversiteit, 9700 AV Groningen, Netherlands
(Submitted July 1983)

1. INTRODUCTION

In this paper, we study a sequence of positive integers defined by recurrence that have applications in combinatorics and probability theory.

Let σ be a permutation of $\mathbb{N}_n = \{1, \ldots, n\}$, i.e., a bijection $\mathbb{N}_n \to \mathbb{N}_n$. Then $k \in \mathbb{N}_n$ is a regeneration point of σ if $\sigma(\mathbb{N}_k) = \mathbb{N}_k$. Here σ will be a random permutation, i.e., we consider σ to be chosen at random from the set S_n of permutations of $\mathbb{N}_n$. Equivalently, we define a probability measure P_n on the power set of S_n by $P_n(\{\sigma_0\}) = P_n(\sigma = \sigma_0) = 1/n!$, $\sigma_0 \in S_n$. Expectation with respect to P_n will be denoted by E_n.

Let A_k be the event that k is a regeneration point of the random permutation. Then

$$P_n(A_k) = k!(n-k)!/n! = \binom{n}{k}^{-1}, \quad k \in \mathbb{N}_n. \tag{1.1}$$

For the event that $k_1, \ldots, k_r$, with $1 \leqslant k_1 < \cdots < k_r \leqslant n$, are regeneration points, we have

$$P_n(A_{k_1}A_{k_2} \ldots A_{k_r}) = k_1!(k_2 - k_1)! \ldots (k_r - k_{r-1})!(n - k_r)!/n!. \tag{1.2}$$

Let M be the total number of regeneration points in σ. The (factorial) moments of M can be expressed in terms of (1.2), e.g.,

$$E_nM = 1 + Q_n = 1 + \sum_{k=1}^{n-1} P_n(A_k) = 1 + \sum_{k=1}^{n-1} \binom{n}{k}^{-1}. \tag{1.3}$$

Note that n is always a regeneration point.

The theory of regeneration points is dominated by the numbers c_n or $c(n)$, $n = 1, 2, \ldots$, where c_n is the number of elements of S_n that have only one regeneration point, or

$$P_n(M = 1) = c_n/n!, \quad n = 1, 2, \ldots . \tag{1.4}$$

This will be seen in Section 2. Here we mention the relation

$$P_n(k) = P_n(\nu = k) = c_k(n-k)!/n!, \quad k \in \mathbb{N}_n, \tag{1.5}$$

where ν is the first regeneration point of the random permutation σ. Since $P_n(1) + \cdots + P_n(n) = 1$, we have

$$\sum_{k=1}^{n} (n-k)!c_k/n! = 1, \quad n \geqslant 1. \tag{1.6}$$

The c_n can be computed recursively from (1.6). We find

$$c_1 = c_2 = 1, \quad c_3 = 3, \quad c_4 = 13, \quad c_5 = 71, \qquad c_6 = 461, \quad c_7 = 3447 = g \times 383. \tag{1.7}$$

From (1.6) we see, by induction on n, that the c_n are odd. Divisibility of the c_n is considered in Section 4. By (1.4), the principle of inclusion and exclusion, and by (1.2),

$$c_n/n! = 1 - P(A_1 \cup \cdots \cup A_{n-1}) = 1 + \sum_{h=1}^{n-1} (-1)^h T_h = \sum_{h=0}^{n-1} (-1)^h T_h. \tag{1.8}$$

Here $T_0 = 1$ and for $h > 0$,

$$\begin{aligned} T_h &= \Sigma' P_n(A_{i_1} A_{i_2} \ldots A_{i_h}) = \Sigma' i_1!(i_2 - i_1)! \ldots (i_h - i_{h-1})!(n - i_h)!/n! \\ &= \Sigma'' j_1! j_2! \ldots j_{h+1}!/n!, \end{aligned}$$

where Σ' sums over all $i_1, \ldots, i_h$ with $1 \leqslant i_1 < \cdots < i_h \leqslant n - 1$ and Σ'' over all $j_1 \geqslant 1, \ldots, j_{h+1} \geqslant 1$ with $j_1 + \cdots + j_{h+1} = n$. In (1.8) this gives, by putting $h = m - 1$.

$$c_n = \sum_{m=1}^{n} (-1)^{m-1} \Sigma^* j_1! \ldots j_m!, \quad n \geqslant 1. \tag{1.9}$$

where Σ^* sums over all $j_1 \geqslant 1, \ldots, j_m \geqslant 1$ with $j_1 + \cdots + j_m = n$.

In Section 2, an integral equation for the exponential generating function of the c_n will be derived. Section 3 studies the asymptotic behavior of c_n for $n \to \infty$. We have $c_n/n! \to 1$, so M tends to 1 in probability as $n \to \infty$. In Section 5, some applications of the c_n in combinatorial probability theory are given.

2. GENERAL FORMULAS

For the total number M of regeneration points we find, by specifying regeneration points only at $j_1, j_1 + j_2, \ldots, j_1 + \cdots + j_m = n$,

$$P_n(M = m) = \Sigma^* c(j_1)c(j_2) \cdots c(j_m)/n!, \quad m \in \mathbb{N}_n, \tag{2.1}$$

where Σ^* is the same as in (1.9). The event $\{M \geqslant m\}$, with $m \geqslant 2$, means that there are at least $m - 1$ regeneration points in $\{1, \ldots, n - 1\}$. This gives, in the same way as (2.1),

$$\begin{aligned} P_n(M \geqslant m) &= \Sigma' c(j_1) \cdots c(j_{m-1})(n - j_1 - \cdots - j_{m-1})!/n! \\ &= \Sigma^* c(j_1) \cdots c(j_{m-1}) j_m!/n!, \quad m = 2, \ldots, n, \end{aligned} \tag{2.2}$$

where Σ' sums over all $j_1 \geqslant 1, \ldots, j_{m-1} \geqslant 1$ with $j_1 + \cdots + j_{m-1} \leqslant n - 1$ and Σ^* is the same as in (1.9).

For the first regeneration point ν we have, with (1.5),

$$\begin{aligned} E_n \nu &= \sum_{k=1}^{n} k c_k (n - k)!/n! = \sum_{k=1}^{n} (n + 1) P_n(k) - \sum_{k=1}^{n} c_k (n + 1 - k)!/n! \\ &= (n + 1) - (n + 1) \sum_{k=1}^{n} P_{n+1}(k) = (n + 1) P_{n+1}(n + 1) = c_{n+1}/n!. \end{aligned} \tag{2.3}$$

From the relation $k^2 = (n + 2 - k)(n + 1 - k) + (2n + 3)k - (n + 2)(n + 1)$, we find, in a similar way,

$$E_n \nu^2 = \{2(n + 1)c_{n+1} - c_{n+2}\}/n!. \tag{2.4}$$

Let

$$C(z) = \sum_{k=1}^{\infty} z^k c_k/k!, \quad |z| < 1. \tag{2.5}$$

From (1.6),

$$z(1 - z)^{-1} = \sum_{n=1}^{\infty} z^n \sum_{k=1}^{n} c_k(n - k)!/n! = \sum_{k=1}^{\infty} c_k \sum_{n=k}^{\infty} z^n(n - k)!/n!$$
$$= \sum_{k=1}^{\infty} c_k \sum_{j=0}^{\infty} z^{k+j} j!/(k + j)!.$$

With the relation

$$\int_0^z (z - x)^{k-1}(1 - x)^{-1}dx = (k - 1)! \sum_{j=0}^{\infty} z^{k+j} j!/(k + j)!,$$

to be derived by putting $x = zt$ and expanding $(1 - zt)^{-1}$, we see that

$$z(1 - z)^{-1} = \int_0^z C'(z - x)(1 - x)^{-1}dx,$$

and with partial integration, noting that $C(0) = 0$,

$$z(1 - z)^{-1} = C(z) + \int_0^z (1 - x)^{-2}C(z - x)dx, \quad |z| < 1. \tag{2.6}$$

The author was unable to find a solution of (2.6) in closed form. The Neumann series solution gives a series of iterated convolutions which, on expansion into powers of z, leads back to (1.9).

3. ASYMPTOTIC BEHAVIOR

We use the notation for falling factorials

$$(n)_r = n!/(n - r)!, \quad r = 0, \ldots, n, \quad n = 1, 2, \ldots . \tag{3.1}$$

First we consider $Q_n = E_n M - 1$ given by (1.3). Rockett [4] gave an expression for

$$\sum \binom{n}{k}^{-1},$$

but direct use of (1.3) seems better for asymptotic estimates. We have

$$Q_n = 2n^{-1} + 4(n)_2^{-1} + V\,(n_3)^{-1}, \quad n \geqslant 6, \tag{3.2}$$

$$V_n = \sum_{k=3}^{n-3} k!(n - k)!/(n - 3)!, \quad n \geqslant 6. \tag{3.3}$$

<u>Theorem 1</u>. We have

$$V_n \geqslant 12, \ n \geqslant 7; \quad V_n \leqslant 156/7, \ n \geqslant 6; \tag{3.4}$$

$$V_n = 12 + O(n^{-1}), \ n \to \infty; \tag{3.5}$$

$$V_{n+1} < V_n, \ n \geqslant 11. \tag{3.6}$$

<u>Proof</u>: The first inequality in (3.4) follows by considering the terms with $k = 3$ and $k = n - 3$ in (3.3). The relation (3.5) follows by estimating the

terms in (3.3) with $k = 4$, $k = n - 4$, and $5 \leqslant k \leqslant n - 5$. From (3.3), for $n \geqslant 6$,

$$\begin{aligned} V_{n+1} - V_n &= 6 + \sum_{k=3}^{n-3} k!(n-k)!\{(n+1-k)(n-2)^{-1} - 1\}/(n-3)! \\ &= 6 + 4(n-2)^{-1}V_n - \sum_{k=3}^{n-3}(k+1)!(n-k)!/(n-2)! \\ &= 6 + 4(n-2)^{-1}V_n - \sum_{h=4}^{n-2} h!(n+1-h)!/(n-2)! \\ &= 12 + 4(n-2)^{-1}V_n - V_{n+1}, \end{aligned} \tag{3.7}$$

so that $2V_{n+1} = 12 + (n+2)(n-2)^{-1}V_n$. Substituting this into (3.7) shows that $V_{n+1} < V_n$, for $n \geqslant 6$, if and only if

$$V_n > 12 + 48(n-6)^{-1}. \tag{3.8}$$

From the terms in (3.3) with $k \leqslant 5$ and $k \geqslant n - 5$,

$$V_n \geqslant 12 + 48(n-3)^{-1} + 240(n-3)^{-1}(n-4)^{-1}, \quad n \geqslant 11.$$

Applying this to (3.8) we find (3.6). From (3.6) and direct computation of V_n, $n = 6, \ldots, 11$, we see that max $V_n = 156/7$ is reached for $n = 11$. Better bounds for larger n may be obtained from (3.6) by computing some V_n.

For the study of c_n, we introduce the following notation, see (1.5) and (3.1):

$$H_n = 1 - c_n/n! = P_n(\nu \leqslant n-1) = \sum_{k=1}^{n-1} c_k(n-k)!/n!; \tag{3.9}$$

$$D_n = (n)_3\{H_n - 2n^{-1} - (n)_2^{-1}\}, \quad n \geqslant 3. \tag{3.10}$$

We need some numerical values of nH_n and D_n. By means of (1.6), (3.9), and (3.10), the values of nH_n and D_n for $3 \leqslant n \leqslant 200$ were computed for the author at the University of Groningen Computing Centre. Part of the values are given in Tables 1 and 2, but the most important numerical result is

$$D_{n+1} < D_n, \quad n = 13, \ldots, 199. \tag{3.11}$$

Table 1

n	nH_n	n	nH_n	n	nH_n
1	0.000000	4	1.833333	7	2.212500
2	1.000000	5	2.041667	8	2.227579
3	1.500000	6	2.158333	9	2.220660

Theorem 2. With D_n defined by (3.9) and (3.10),

$$D_n > 4, \; n \geqslant 9; \; D_n < 6, \; n \geqslant 20. \tag{3.12}$$

Proof: Since $c_k \leqslant k!$, we see from (3.9), (1.1), (1.3), (3.2), and (3.4) that $nH_n \leqslant nQ_n < 3$, $n \geqslant 9$.

With Table 1, we then extend this to

$$nH_n < 3, \quad n \geq 1. \tag{3.13}$$

From (3.9), for $n \geq 7$,

$$n!H_n \geq (n-1)!c_1 + (n-2)!c_2 + (n-3)!c_3 + 6c(n-3) + 2c(n-2) + c(n-1) \tag{3.14}$$

With (1.7) and (3.13), writing $c_k = k!(1 - H_k)$ for $k \geq n - 3$, this gives

$$H_n > 2n^{-1} + 3(n)_3^{-1} - 18(n)_4^{-1}, \quad n \geq 7. \tag{3.15}$$

From (3.15) we see that $nH_n > 2$, $n > 9$, and then from Table 1,

$$nH_n > 2, \quad n \geq 5. \tag{3.16}$$

From (3.9) for $n \geq 9$, with $c_k \leq k!$,

$$n!H_n \leq \left(\sum_{k=1}^{4} + \sum_{k=n-4}^{n-1}\right) c_k (n-k)! + (n-9)5!(n-5)!. \tag{3.17}$$

With (1.7) and (3.16), writing $c_k = k!(1 - H_k)$ for $k \geq n - 4$, we find

$$H_n \leq 2n^{-1} + (n)_2^{-1} + h(n)(n)_3^{-1}, \quad n \geq 9; \tag{3.18}$$

$$h(n) = 5 + 25(n-3)^{-1} + (120(n-9) - 48)(n-3)^{-1}(n-4)^{-1}. \tag{3.19}$$

Table 2

n	D_n	n	D_n	n	D_n
3	-2.000000	10	6.625992	17	6.687779
4	-3.000000	11	7.376414	18	6.406247
5	-2.500000	12	7.702940	19	6.156020
6	-0.833333	13	7.726892	20	5.939237
7	1.375000	14	7.561317	21	5.754089
8	3.558333	15	7.295355	21	5.596962
9	5.356944	16	6.991231	23	5.463713

By elementary computation we see that $h(n+1) < h(n)$ for $145n > 1876$ or $n \geq 13$ and $h(196) < 6$, so that $h(n) < 6$, $n \geq 196$. Hence, $D_n < 6$, $n \geq 196$, by (3.10). The second inequality in (3.12) then follows from (3.11) and Table 2, and it shows that

$$H_n < 2n^{-1} + 2(n)_2^{-1}, \quad n \geq 20. \tag{3.20}$$

From (3.9), for $n \geq 9$,

$$n!H_n \geq \left(\sum_{k=1}^{4} + \sum_{k=n-4}^{n-1}\right) c_k (n-k)!.$$

Here we apply (1.7) for $k \leq 4$ and write $c_k = k!(1 - H_k)$ for $k \geq n - 4$. Application of (3.20) for $k = n - 3$, $n - 4$, and of (3.10) with $D_n < 6$ then gives

$$H_n > 2n^{-1} + (n)_2^{-1} + 4(n)_3^{-1} + g(n)(n)_4^{-1}, \quad n \geq 24; \tag{3.21}$$

$$g(n) = 17 - 72(n - 4)^{-1} - 48(n - 4)^{-1}(n - 5)^{-1}. \tag{3.22}$$

Since $g(n + 1) > g(n)$, $n \geq 6$, and $g(13) > 0$,

$$H_n > 2n^{-1} + (n)_2^{-1} + 4(n)_3^{-1}, \quad n \geq 24,$$

i.e., $D_n > 4$, $n \geq 24$. The first inequality in (3.12) then follows from Table 2.

Remark 1. By taking into account more terms in (3.9), it was proved in Stam [5] that $D_n = 4 + O(n^{-1})$ and $D_{n+1} < D_n$, $n \geq 13$.

Remark 2. For the conditional probability $P(A_1|M \geq 2)$, we see, using (1.1), (3.10), and (3.12), that

$$P(A_1|M \geq 2) = P(A_1)/P(M \geq 2) = n^{-1}H_n^{-1} \to \frac{1}{2}, \quad n \to \infty,$$

and in the same way,

$$P(A_{n-1}|M \geq 2) \to \frac{1}{2},$$

so that the regeneration points concentrate near the end points of $\mathbb{N}_n$ as $n \to \infty$.

4. DIVISIBILITY

From (1.6) we have, since m divides $h!$ if $h \geq m$, the congruences

$$\sum_{j=0}^{m-1} j!c_{n-j} \equiv 0 \pmod{m}, \quad n \geq m. \tag{4.1}$$

Let $d_n = d_n(m)$ be the remainder of c_n on division by m. Then the recurrence (4.1) also holds for the d_i and determines them completely if $d_1, \ldots, d_{m-1}$ are given. Since $d_n \in \{0, \ldots, m - 1\}$, there are at most m^{m-1} possibilities for the sequence $u_k = (d_k, \ldots, d_{k+m-2})$. One of them is $u_k = (0, \ldots, 0)$ and this would give $d_n = 0$, $n \geq 1$, which is excluded because $c_1 = 1$. So we must have $u_k = u_{k+p}$ for some k and some minimal $p \leq m^{m-1} - 1$. Since any u_k determines all d_n, $n \geq 1$, with (4.1) and the coefficients in (4.1) do not depend on n, it follows that the sequence d_n, $n \geq 1$ is periodic with period p.

If $m = 3$, then (4.1) becomes

$$c_n + c_{n-1} + 2c_{n-2} \equiv 0 \quad \text{or} \quad c_n + c_{n-1} - c_{n-2} \equiv 0, \bmod 3, \ n \geq 3,$$

so that $(-1)^n c_n$ satisfies the same recurrence mod 3 as the Fibonacci numbers, but the initial conditions are different. We find

$$c_n \equiv 1, 1, 0, 1, 2, 2, 0, 2, 1, 1, \bmod 3, \ n = 1, \ldots, 10.$$

So p has its maximal value 8.

If $m = 4$, then (4.1) gives

$$c_n + c_{n-1} + 2c_{n-2} + 6c_{n-3} \equiv c_n + c_{n-1} + 2c_{n-2} + 2c_{n-3} \equiv 0, \bmod 4, \ n \geq 4.$$

Since the c_i are odd, this gives

$$c_n + c_{n-1} \equiv 0, \bmod 4, \ n \geq 4.$$

With (1.7) we see that $c_n \equiv 1$, mod 4, if n is even and $c_n \equiv 3$, mod 4, if $n \geq 3$ is odd.

Since the c_n are odd, we have $c_n \equiv 1$, $c_n \equiv 3$, $c_n \equiv 5$, mod 6, if $c_n \equiv 1$, $c_n \equiv 0$, $c_n \equiv 2$, mod 3, respectively.

The Computing Centre of the University of Groningen computed the sequences for $m = 5$, $m = 7$, and part of the sequence for $m = 11$. For $m = 5$ the period is 62, whereas $5^4 - 1 = 624$. For $m = 7$ the period is 684, whereas $7^6 - 1 = 117649$. For $m = 5$, 7, and 11 all possible values of $c_n \pmod{m}$ occur. It is conjectured that this holds for all prime m.

We note that for m prime the last two coefficients in (4.1) are 1 and -1 (mod m) by Wilson's theorem (see Grosswald [2, Ch. 4.3]).

5. APPLICATIONS IN COMBINATORIAL PROBABILITY THEORY

If σ and τ are independent stochastic elements of S_n and one of them has uniform distribution P_n, then the points $k \in \mathbb{N}_n$ such that $\sigma(\mathbb{N}_k) = \tau(\mathbb{N}_k)$ have the same joint distribution as the regeneration points of a random permutation, since $\sigma(\mathbb{N}_k) = \tau(\mathbb{N}_k)$ if and only if $\sigma^{-1}\tau(\mathbb{N}_k) = \mathbb{N}_k$ and $\sigma^{-1}\tau$ has probability distribution P_n.

Let $X_1, \ldots, X_n$ be independent random variables with common continuous distribution function and $Y_1, \ldots, Y_n$ their increasing order statistics, i.e., the value of Y_k is the k^{th} smallest of the values of $X_1, \ldots, X_k$. Then the stochastic points k in $\mathbb{N}_n$ such that $X_1 + \cdots + X_k = Y_1 + \cdots + Y_k$ have the same joint probability distribution as the regeneration points of a random permutation of $\mathbb{N}_n$. We have $Y_1 < Y_2 < \cdots < Y_n$ with probability 1 and the conditional distribution of $X_1, \ldots, X_n$ given $Y_i = y_i$, $i = 1, \ldots, n$ is the same as the distribution of $\sigma(y_1), \ldots, \sigma(y_n)$, where σ is a random element of S_n (see Rényi [3]). Furthermore, $\sigma(y_1) + \cdots + \sigma(y_k) = y_1 + \cdots + y_k$ if and only if

$$\sigma(\{y_1, \ldots, y_k\}) = \{y_1, \ldots, y_k\}.$$

A deeper application is the following. Let σ and τ be independent random elements of S_n. Dixon [1] defined t_n as the probability that the subgroup $\langle \tau, \sigma \rangle$ of S_n generated by σ and τ is transitive, i.e., has $\mathbb{N}_n$ as the only orbit. This occurs if and only if $\sigma(A) = \tau(A) = A$ for no proper subset A of $\mathbb{N}_n$. Using formal power series, Dixon [1] proved that

$$\sum_{k=1}^{n} (n-k)!k!kt_k = n!n, \quad n \geqslant 1. \tag{5.1}$$

A slightly shorter proof starts from U_1, the orbit of $\langle \sigma, \tau \rangle$ that contains 1. By the definition of t_k, we have

$$P(U_1 = A) = (k!)^2 t_k ((n-k)!)^2/(n!)^2,$$

if $A \subset \mathbb{N}_n$, $1 \in A$, and $|A| = k$. So

$$P(|U_1| = k) = \binom{n-1}{k-1} P(U_1 = A) = (n-k)!k!kt_k(n!n)^{-1}.$$

Equation (5.1) states that these probabilities sum to 1. From (1.6) and (1.7),

$$\sum_{k=1}^{n} (n-k)!c_{k+1} = \sum_{j=2}^{n+1} (n+1-j)!c_j = (n+1)! - n!c_1 = n!n, \quad n \geqslant 1.$$

Comparing this with (5.1) we see that the sequences c_{n+1} and $n!nt_n$, $n \geqslant 1$, are determined (uniquely) by the same recurrence. So

$$n!nt_n = c_{n+1}, \quad n \geqslant 1.$$

The author was unable to find a direct combinatorial proof of this result.

Let $X = (X_1, \ldots, X_n)$ be a random sample with replacement from $\mathbb{N}_m$, or a random function $\mathbb{N}_n \to \mathbb{N}_m$. If $X(\mathbb{N}_k) = \mathbb{N}_k$, then $X_1, \ldots, X_k$ defines a bijection $\mathbb{N}_k \to \mathbb{N}_k$. So the probability that h is the first k with $X(\mathbb{N}_k) = \mathbb{N}_k$ is

$$c_h m^{n-h} m^{-n} = c_h m^{-h}$$

and the probability that there is at least one such k is

$$\sum_{h=1}^{m \wedge n} m^{-h} c_h.$$

When the sample is drawn without replacement, so that $n \leqslant m$, the corresponding probability is

$$\sum_{h=1}^{n} c_h (m - h)!/m!.$$

REFERENCES

1. J. D. Dixon. "The Probability of Generating the Symmetric Group." *Math. Zeitschr.* 110 (1969):199-205.
2. E. Grosswald. *Topics from the Theory of Numbers*. New York: Macmillan, 1966.
3. A. Rényi. "Théorie des éléments saillants d'une suite d'observations." In *Colloquium on Combinatorial Methods in Probability Theory*. Matematisk Institut Aarhus Universitetet Danmark, 1962.
4. A. M. Rockett. "Sums of the Inverses of Binomial Coefficients." *The Fibonacci Quarterly* 19, no. 5 (1981):433-37.
5. A. J. Stam. "Asymptotics for Regeneration Points of Random Permutations." Report T.W. 242, Mathematisch Instituut Rijksuniversiteit Groningen.

◆◇◆◇◆