J. Appl. Prob. **45,** 1060–1070 (2008)
Printed in England

A RANDOM PERMUTATION MODEL ARISING IN CHEMISTRY

MARK BROWN,* *The City College of New York*
EROL A. PEKÖZ,** *Boston University*
SHELDON M. ROSS,*** *University of Southern California*

Abstract

We study a model arising in chemistry where n elements numbered $1, 2, \dots, n$ are randomly permuted and if i is immediately to the left of $i+1$ then they become stuck together to form a cluster. The resulting clusters are then numbered and considered as elements, and this process keeps repeating until only a single cluster is remaining. In this article we study properties of the distribution of the number of permutations required.

Keywords: Random permutation; hat-check problem

2000 Mathematics Subject Classification: Primary 60C05
Secondary 92E20

1. Introduction

For the classic hat-check problem first proposed in 1708 by Montmort [2], the following variation appears in [6, p. 93]. Each member of a group of n individuals throws his or her hat in a pile. The hats are shuffled, each person chooses a random hat, and the people who receive their own hat depart. Then the process repeats with the remaining people until everybody has departed; let N be the number of shuffles required. With X_i representing the total number of people who have departed after shuffle number i, it is easy to show that $X_i - i$ is a martingale and, thus, by the optional sampling theorem we elegantly see that $\mathrm{E}[N] = n$.

Someone getting their own hat can also be thought of as corresponding to a cycle of length one in a random permutation. Properties of cycles of various lengths in random permutations have been studied extensively; see [1] and [3] for entry points to this literature. A variation of this problem was presented in [5], where it was given as a model for a chemical bonding process. Below we discuss this variation and study its properties. We quote the following description of the chemistry application from [5], where a recursive formula was given to numerically compute the mean.

> There are 10 molecules in some hierarchical order operating in a system. A catalyst is added to the system and a chemical reaction sets in. The molecules line up. In the line-up from left to right molecules in consecutive increasing hierarchical order bond together and become one. A new hierarchical order sets among the fused molecules. The catalyst is added again

Received 30 September 2008; revision received 7 November 2008.
* Postal address: Department of Mathematics, The City College of New York, New York, NY 10031-9100, USA. Email address: cybergarf@aol.com
Supported by the National Security Agency under grant H98230-06-01-0149.
** Postal address: Department of Operations and Technology Management, Boston University, 595 Commonwealth Avenue, Boston, MA 02215, USA. Email address: pekoz@bu.edu
*** Postal address: Department of Industrial and System Engineering, University of Southern California, Los Angeles, CA 90089, USA. Email address: smross@usc.edu

to the system and the whole process starts all over again. The question raised is how many times catalysts are expected to be added in order to get a single lump of all molecules.

This variation presented in [5] can be abstractly stated as follows. Suppose that we have n elements numbered $1, 2, \dots, n$. These elements are randomly permuted, and if i is immediately to the left of $i+1$ then i and $i+1$ become stuck together to form (possibly with other adjacently numbered elements) a cluster. These clusters are then randomly permuted and if a cluster ending with i immediately precedes one starting with $i+1$ then those two clusters join together to form a new cluster. This continues until there is only one cluster, and we are interested in $N(n)$, the number of permutations that are needed. For instance, suppose that $n = 7$ and that the first permutation is

$$3, 4, 5, 1, 2, 7, 6,$$

which results in the clusters $\{3, 4, 5\}$, $\{1, 2\}$, $\{6\}$, and $\{7\}$. If a random permutation of these four clusters gives the ordering

$$\{6\}, \{7\}, \{3, 4, 5\}, \{1, 2\}$$

then the new sets of clusters are $\{6, 7\}$, $\{3, 4, 5\}$, and $\{1, 2\}$. If a random permutation of these three clusters gives the ordering

$$\{3, 4, 5\}, \{6, 7\}, \{1, 2\}$$

then the new sets of clusters are $\{3, 4, 5, 6, 7\}$ and $\{1, 2\}$. If a random permutation of these two clusters gives the ordering

$$\{1, 2\}, \{3, 4, 5, 6, 7\}$$

then there is now a single cluster $\{1, 2, 3, 4, 5, 6, 7\}$ and $N(7) = 4$.

The random variable $N(n)$ can be analyzed as a first passage time from state n to state 1 of a Markov chain whose state is the current number of clusters. When the state of this chain is i, we will designate the clusters as $1, \dots, i$, with 1 being the cluster whose elements are smallest, 2 being the cluster whose elements are the next smallest, and so on. For instance, in the preceding $n = 7$ case, the state after the first transition is 4, with 1 being the cluster $\{1, 2\}$, 2 being the cluster $\{3, 4, 5\}$, 3 being the cluster $\{6\}$, and 4 being the cluster $\{7\}$. With this convention, the transitions from state i are exactly the same as if the problem began with the i elements, $1, \dots, i$.

In Section 2 we compute the transition probabilities of this Markov chain and use them to obtain some stochastic inequalities. In Section 3 we obtain upper and lower bounds on $\mathrm{E}[N(n)]$, as well as bounds on its distribution. In Section 4 we give results for a circular version of the problem.

2. The transition probabilities

With the above definitions, let D_n be the decrease in the number of clusters starting from state n. Then we have the following proposition.

Proposition 1. *For* $0 \le k < n$,

$$\mathrm{P}(D_n = k) = \frac{n-k+1}{nk!} \sum_{i=0}^{n-k+1} \frac{(-1)^i}{i!}.$$

Proof. Letting A_i be the event that i immediately precedes $i+1$ in the random permutation, then D_n is the number of events $A_1, \dots, A_{n-1}$ that occur. Then, with

$$S_j = \sum_{0<i_1<\cdots<i_j<n} \mathrm{P}(A_{i_1} \cdots A_{i_j}),$$

the inclusion/exclusion identity (see [4, p. 106]) gives

$$\mathrm{P}(D_n = k) = \sum_{j=k}^{n-1} S_j \binom{j}{k} (-1)^{j+k}.$$

Now consider $\mathrm{P}(A_{i_1} \cdots A_{i_j})$. If we think of a permutation of n elements as having n degrees of freedom then, for each event A_i in the intersection, one degree of freedom in the permutation is dropped. For instance, suppose that we want $\mathrm{P}(A_2 A_3 A_6)$. Then, in order for these three events to occur, 2, 3, and 4 must be consecutive values of the permutation, as must be 6 and 7. Because there are $n-5$ other values, there are thus $(n-3)!$ such permutations. Similarly, for the event $A_2 A_4 A_6$ to occur, 2 and 3 must be consecutive values of the permutation, as must be 4, 5 and 6, 7. As there are $n-6$ other values, there are $(n-3)!$ such permutations. Consequently, for $0 < i_1 < \cdots < i_j < n$,

$$\mathrm{P}(A_{i_1} \cdots A_{i_j}) = \frac{(n-j)!}{n!}.$$

As a result,

$$S_j = \binom{n-1}{j} \frac{(n-j)!}{n!} = \frac{n-j}{nj!},$$

which yields

$$\begin{aligned}
\mathrm{P}(D_n = k) &= \sum_{j=k}^{n-1} \binom{j}{k} (-1)^{j+k} \frac{n-j}{nj!} \\
&= \sum_{i=0}^{n-k-1} (-1)^i \binom{k+i}{k} \frac{n-k-i}{n(k+i)!} \\
&= \sum_{i=0}^{n-k-1} (-1)^i \frac{n-k-i}{nk!\,i!} \\
&= \frac{1}{nk!} \left((n-k) \sum_{i=0}^{n-k-1} \frac{(-1)^i}{i!} - \sum_{i=1}^{n-k-1} \frac{(-1)^i}{(i-1)!} \right) \\
&= \frac{1}{nk!} \left((n-k+1) \sum_{i=1}^{n-k-2} \frac{(-1)^i}{i!} + (n-k) \frac{(-1)^{n-k-1}}{(n-k-1)!} \right).
\end{aligned}$$

Thus, the result follows once we show that

$$(n-k) \frac{(-1)^{n-k-1}}{(n-k-1)!} = (n-k+1) \left(\frac{(-1)^{n-k-1}}{(n-k-1)!} + \frac{(-1)^{n-k}}{(n-k)!} + \frac{(-1)^{n-k+1}}{(n-k+1)!} \right)$$

or, equivalently, that

$$\frac{(-1)^{n-k}}{(n-k-1)!} = (n-k+1) \left(\frac{(-1)^{n-k}}{(n-k)!} + \frac{(-1)^{n-k+1}}{(n-k+1)!} \right)$$

or

$$1 = (n-k+1)\left(\frac{1}{n-k} - \frac{1}{(n-k+1)(n-k)}\right),$$

which is immediate.

Remark 1. A recursive expression for $\mathrm{P}(D_n = k)$, though not in closed form, was given in [5].

From Proposition 1 we immediately conclude that D_n converges in distribution to a Poisson random variable with mean 1.

Corollary 1. *We have* $\lim_{n\to\infty} \mathrm{P}(D_n = k) = \mathrm{e}^{-1}/k!$.

We now present two results that will be used in the next section. Recall from [6, p. 133] that a discrete random variable X is said to be likelihood ratio smaller than Y if $\mathrm{P}(X = k)/\mathrm{P}(Y = k)$ is nonincreasing in k.

Corollary 2. *With the above definitions, D_n is likelihood ratio smaller than a Poisson random variable with mean* 1.

Proof. We need to show that $k!\,\mathrm{P}(D_n = k)$ is nonincreasing in k. But, with $B_k = nk!\,\mathrm{P}(D_n = k)$ we have

$$\begin{aligned} B_{k-1} - B_k &= \sum_{i=0}^{n-k+1} \frac{(-1)^i}{i!} + (n-k+2)\frac{(-1)^{n-k+2}}{(n-k+2)!} \\ &= \sum_{i=0}^{n-k} \frac{(-1)^i}{i!} \\ &> 0, \end{aligned}$$

which proves the result.

Corollary 3. *The state of the Markov chain after a transition from state n, $n - D_n$, is likelihood ratio increasing in n.*

Proof. From Proposition 1,

$$\mathrm{P}(n - D_n = k) = \frac{k+1}{n(n-k)!} \sum_{i=0}^{k+1} \frac{(-1)^i}{i!}.$$

Consequently,

$$\frac{\mathrm{P}(n+1-D_{n+1} = k)}{\mathrm{P}(n - D_n = k)} = \frac{n}{(n+1)(n+1-k)}.$$

As the preceding is increasing in k, the result follows.

3. The random variable $N(n)$

Let X_i be the ith decrease in the number of clusters, so that

$$S_k \equiv n - \sum_{i=1}^{k} X_i$$

is the state of the Markov chain, starting in state n, after k transitions, $k \geq 1$.

Proposition 2. *We have*

$$\mathrm{P}(N(n) > k) \geq \sum_{i=0}^{n-1} \frac{\mathrm{e}^{-k} k^i}{i!}.$$

Proof. Let the Y_i, $i = 1, \dots, k$, be independent Poisson random variables, each with mean 1. Now, because likelihood ratio is a stronger ordering than stochastic order (see Proposition 4.20 of [6]), it follows by Corollary 2 that X_i, conditional on $X_1, \dots, X_{i-1}$, is stochastically smaller than a Poisson random variable with mean 1. Consequently, the random vector $X_1, \dots, X_k$ can be generated in such a manner that $X_i \leq Y_i$ for each $i = 1, \dots, k$. But this implies that

$$\begin{aligned}\mathrm{P}(N(n) > k) &= \mathrm{P}(X_1 + \cdots + X_k < n) \\ &\geq \mathrm{P}(Y_1 + \cdots + Y_k < n) \\ &= \sum_{i=0}^{n-1} \frac{\mathrm{e}^{-k} k^i}{i!}.\end{aligned}$$

We now consider bounds on $\mathrm{E}[N(n)]$.

Proposition 3. *We have*

$$\mathrm{E}[N(n)] \leq n - 1 + \sum_{i=1}^{n-1} \frac{1}{i}.$$

Proof. First note that

$$\mathrm{E}[D_n] = \sum_{i=1}^{n-1} \mathrm{P}(i \text{ immediately precedes } i+1) = \frac{n-1}{n}. \tag{1}$$

Because the Markov chain cannot make a transition from a state into a higher state and $\mathrm{E}[D_n]$ is nondecreasing in n, it follows from Proposition 5.23 of [6] that

$$\mathrm{E}[N(n)] \leq \sum_{i=2}^{n} \frac{1}{\mathrm{E}[D_i]} = n - 1 + \sum_{i=1}^{n-1} \frac{1}{i}.$$

Proposition 4. *We have*

$$\mathrm{E}[N(n)] \geq n - 1 + \frac{\mathrm{e}}{n(\mathrm{e}-1)} + \frac{\mathrm{e}}{(\mathrm{e}-1)^2} \sum_{j=2}^{n-1} \frac{1}{j}.$$

Proof. To begin, note that

$$Z_k = \sum_{i=1}^{k} (X_i - \mathrm{E}[X_i \mid X_1, \dots, X_{i-1}]), \qquad k \geq 1, \tag{2}$$

is a zero-mean martingale. Hence, by the martingale stopping theorem,

$$\mathrm{E}[Z_{N(n)}] = 0. \tag{3}$$

Now, because $\mathrm{E}[X_i \mid X_1, \dots, X_{i-1}]$ is the expected decrease from state S_{i-1}, it follows from (1) that

$$\mathrm{E}[X_i \mid X_1, \dots, X_{i-1}] = \mathrm{E}[X_i \mid S_{i-1}] = \mathrm{E}[D_{S_{i-1}} \mid S_{i-1}] = 1 - \frac{1}{S_{i-1}}.$$

Using this, and the fact that $\sum_{i=1}^{N(n)} X_i = n - 1$, we obtain, from (2) and (3),

$$n - 1 - \mathrm{E}[N(n)] + \mathrm{E}\biggl[\sum_{i=1}^{N(n)} \frac{1}{S_{i-1}}\biggr] = 0.$$

Now (notationally suppressing its dependence on the initial state n), let T_j denote the amount of time that the Markov chain spends in state j, $j > 1$. Then

$$\sum_{i=1}^{N(n)} \frac{1}{S_{i-1}} = \sum_{j=2}^{n} \frac{T_j}{j}.$$

Hence,

$$\mathrm{E}[N(n)] = n - 1 + \sum_{j=2}^{n} \frac{1}{j} \mathrm{E}[T_j] \geq n - 1 + \frac{\mathrm{e}}{(\mathrm{e}-1)^2} \sum_{j=2}^{n-1} \frac{1}{j} + \frac{\mathrm{e}}{n(\mathrm{e}-1)},$$

where, for the inequality, we made use of the following proposition.

Proposition 5. *We have*

$$\mathrm{E}[T_n] = \frac{1}{\mathrm{P}(D_n > 0)} \geq \frac{\mathrm{e}}{\mathrm{e}-1},$$
$$\mathrm{E}[T_j] = \frac{\mathrm{P}(T_j > 0)}{\mathrm{P}(D_j > 0)} \geq \frac{\mathrm{e}}{(\mathrm{e}-1)^2}.$$

To prove Proposition 5, we will need a series of lemmas.

Lemma 1. *Let W_j, $2 \leq j < n$, denote the state of the Markov chain from which the first transition to a state less than or equal to j occurs. Then, for $r > j$,*

$$\mathrm{P}(T_j > 0 \mid W_j = r) \geq \mathrm{P}(T_j > 0 \mid W_j = j+1) = \mathrm{P}(D_{j+1} = 1 \mid D_{j+1} \geq 1).$$

Proof. Let $Y_r = r - D_r$. Then,

$$\begin{aligned} \mathrm{P}(T_j > 0 \mid W_j = r) &= \mathrm{P}(D_r = r - j \mid D_r \geq r - j) \\ &= \mathrm{P}(Y_r = j \mid Y_r \leq j) \\ &= \frac{\mathrm{P}(Y_r = j)}{\sum_{i=1}^{j} \mathrm{P}(Y_r = i)} \\ &= \frac{1}{\sum_{i=1}^{j} \mathrm{P}(Y_r = i)/\mathrm{P}(Y_r = j)}. \end{aligned} \tag{4}$$

But, for $i \leq j$, it follows from Corollary 4 that

$$\frac{\mathrm{P}(Y_{r+1} = j)}{\mathrm{P}(Y_r = j)} \geq \frac{\mathrm{P}(Y_{r+1} = i)}{\mathrm{P}(Y_r = i)}$$

or, equivalently, that

$$\frac{\mathrm{P}(Y_{r+1}=i)}{\mathrm{P}(Y_{r+1}=j)} \le \frac{\mathrm{P}(Y_r=i)}{\mathrm{P}(Y_r=j)}.$$

Thus, by (4), $\mathrm{P}(T_j > 0 \mid W_j = r)$ is nondecreasing in r.

Lemma 2. *For all* $j \ge 2$,

$$\mathrm{P}(D_{j+1} = 1 \mid D_{j+1} \ge 1) \ge \frac{\mathrm{e}^{-1}}{1-\mathrm{e}^{-1}}.$$

Proof. Let $M_k = \sum_{i=0}^{k}(-1)^i/i!$. By Proposition 1 we need to show that

$$\frac{M_{j+1}}{1-(j+2)M_{j+2}/(j+1)} \ge \frac{\mathrm{e}^{-1}}{1-\mathrm{e}^{-1}}.$$

That is, we need to show that, for all $n \ge 3$,

$$M_n(1-\mathrm{e}^{-1}) - \mathrm{e}^{-1}\biggl(1-\frac{n+1}{n}\,M_{n+1}\biggr) \ge 0.$$

Case 1. Suppose that n is even and that $n > 2$. Then,

$$\begin{aligned}
M_n&(1-\mathrm{e}^{-1}) - \mathrm{e}^{-1}\biggl(1-\frac{n+1}{n}\,M_{n+1}\biggr)\\
&= M_n(1-\mathrm{e}^{-1}) - \mathrm{e}^{-1}\biggl[1-\frac{n+1}{n}\biggl(M_n - \frac{1}{(n+1)!}\biggr)\biggr]\\
&= M_n\biggl(1+\frac{\mathrm{e}^{-1}}{n}\biggr) - \mathrm{e}^{-1}\biggl(1+\frac{1}{nn!}\biggr)\\
&\ge \mathrm{e}^{-1}\biggl(1+\frac{\mathrm{e}^{-1}}{n}\biggr) - \mathrm{e}^{-1}\biggl(1+\frac{1}{nn!}\biggr)\\
&= \frac{\mathrm{e}^{-1}}{n}\biggl(\mathrm{e}^{-1} - \frac{1}{n!}\biggr)\\
&> 0,
\end{aligned}$$

where we used the fact that $M_n > \mathrm{e}^{-1}$.

Case 2. Suppose that n is odd. In this case,

$$\begin{aligned}
M_n&(1-\mathrm{e}^{-1}) - \mathrm{e}^{-1}\biggl(1-\frac{n+1}{n}\,M_{n+1}\biggr)\\
&= \biggl(M_{n+1} - \frac{1}{(n+1)!}\biggr)(1-\mathrm{e}^{-1}) - \mathrm{e}^{-1}\biggl(1-\frac{n+1}{n}\,M_{n+1}\biggr)\\
&= M_{n+1}\biggl(1+\frac{\mathrm{e}^{-1}}{n}\biggr) - \frac{1-\mathrm{e}^{-1}}{(n+1)!} - \mathrm{e}^{-1}\\
&\ge \mathrm{e}^{-1}\biggl(1+\frac{\mathrm{e}^{-1}}{n}\biggr) - \frac{1-\mathrm{e}^{-1}}{(n+1)!} - \mathrm{e}^{-1}\\
&= \mathrm{e}^{-1}\biggl(\frac{\mathrm{e}^{-1}}{n} + \frac{1}{(n+1)!}\biggr) - \frac{1}{(n+1)!},
\end{aligned}$$

which will be nonnegative provided that

$$\mathrm{e}^{-2} \geq \frac{n}{(n+1)!}(1-\mathrm{e}^{-1})$$

or, equivalently, that

$$\mathrm{e}(\mathrm{e}-1) \leq \frac{(n+1)!}{n},$$

which is easily seen to be true when $n \geq 3$. This completes the proof of Lemma 2.

We need one additional lemma.

Lemma 3. *As* $n \to \infty$, $\mathrm{P}(D_n = 0) \downarrow \mathrm{e}^{-1}$.

Proof. By Proposition 1,

$$\mathrm{P}(D_n = 0) = \frac{n+1}{n} M_{n+1},$$

yielding $\lim_n \mathrm{P}(D_n = 0) = \mathrm{e}^{-1}$. To show that the convergence is monotone, note that

$$\begin{aligned}\frac{n+1}{n} M_{n+1} - \frac{n+2}{n+1} M_{n+2} &= \frac{n+1}{n} M_{n+1} - \frac{n+2}{n+1}\left(M_{n+1} + \frac{(-1)^n}{(n+2)!}\right)\\ &= \frac{M_{n+1}}{n(n+1)} + \frac{(-1)^{n+1}}{(n+1)(n+1)!}.\end{aligned}$$

When n is odd, the preceding is clearly positive. When n is even, $M_{n+1} = M_n - 1/(n+1)!$, and, thus, we must show that

$$M_n - \frac{1}{(n+1)!} \geq \frac{1}{(n+1)(n-1)!}$$

or, equivalently, that

$$M_n \geq \frac{1}{n!},$$

which follows since, for n even,

$$M_n = M_{n-1} + \frac{1}{n!} \geq \frac{1}{n!}.$$

Proof of Proposition 5. Given that state j is entered, the time spent in that state will have a geometric distribution with parameter $\mathrm{P}(D_j > 0)$. Hence,

$$\mathrm{E}[T_j] = \frac{\mathrm{P}(T_j > 0)}{\mathrm{P}(D_j > 0)}.$$

Now, $\mathrm{P}(T_n > 0) = 1$, and, by Lemma 3, $\mathrm{P}(D_n > 0) \leq 1 - \mathrm{e}^{-1}$, which verifies the first part of Proposition 5. Also, for $2 \leq j < n$, Lemmas 1 and 2 yield

$$\mathrm{P}(T_j > 0) \geq \mathrm{P}(D_{j+1} = 1 \mid D_{j+1} \geq 1) \geq \frac{\mathrm{e}^{-1}}{1-\mathrm{e}^{-1}}.$$

Hence, by Lemma 3,

$$\mathrm{E}[T_j] \geq \frac{\mathrm{e}^{-1}}{(1-\mathrm{e}^{-1})^2} = \frac{\mathrm{e}}{(\mathrm{e}-1)^2},$$

which completes the proof of Proposition 5.

TABLE 1.

n	Lower bound	Upper bound
100	102.62	104.19
1000	1004.72	1006.50
1 000 000	1 000 011.08	1 000 013.41

Corollary 4. *We have*

$$n-1+\frac{\mathrm{e}}{n(\mathrm{e}-1)}+\frac{\mathrm{e}}{(\mathrm{e}-1)^2}\ln\left(\frac{n}{2}\right)\le \mathrm{E}[N(n)]\le n+\ln\left(\frac{2n-1}{3}\right).$$

Proof. Let X be uniformly distributed between $j-\frac{1}{2}$ and $j+\frac{1}{2}$. Then,

$$\ln\left(\frac{j+1/2}{j-1/2}\right)=\int_{j-1/2}^{j+1/2}\frac{1}{x}\,\mathrm{d}x=\mathrm{E}\left[\frac{1}{X}\right]\ge\frac{1}{\mathrm{E}[X]}=\frac{1}{j},$$

where the inequality used Jensen's inequality. Hence,

$$\sum_{j=2}^{n-1}\frac{1}{j}\le\ln\left(\frac{n-1/2}{3/2}\right)=\ln\left(\frac{2n-1}{3}\right),$$

and the upper bound follows from Proposition 3. To obtain the lower bound, we use Proposition 4 along with the inequality

$$\ln\left(\frac{j+1}{j}\right)=\int_{j}^{j+1}\frac{1}{x}\,\mathrm{d}x\le\frac{1}{j}.$$

Remarks. 1. Corollary 4 yields the results given in Table 1.

2. It follows from Corollary 3, using a coupling argument, that $N(n)$ is stochastically increasing in n.

4. The circular case

Whereas we have previously assumed that at each stage the clusters are randomly arranged in a linear order, in this section we suppose that they are randomly arranged around a circle, again with all possibilities being equally likely. We suppose that if a cluster ending with i is immediately counterclockwise to a cluster beginning with $i+1$ then these clusters merge. Let $N^*(n)$ denote the number of stages needed until all n elements are in a single cluster, and let D_n^* denote the decrease in the number of clusters from state n.

Lemma 4. *For $n\ge 2$,*

$$\mathrm{E}[D_n^*]=\mathrm{var}(D_n^*)=1.$$

Proof. If B_i is the event that i is the counterclockwise neighbor of $i+1$ then

$$D_n^*=\sum_{i=1}^{n-1}\mathbf{1}_{B_i}.$$

Now,

$$P(B_i) = \frac{(n-2)!}{(n-1)!} = \frac{1}{n-1}, \qquad i = 1, \dots, n-1,$$

and, for $i \neq j$,

$$P(B_i B_j) = \frac{(n-3)!}{(n-1)!}.$$

Hence,

$$E[D_n^*] = \sum_{i=1}^{n-1} \frac{1}{n-1} = 1$$

and

$$\begin{aligned} \operatorname{var}(D_n^*) &= \sum_{i=1}^{n-1} \frac{1}{n-1}\left(1 - \frac{1}{n-1}\right) + 2\binom{n-1}{2}\left(\frac{(n-3)!}{(n-1)!} - \frac{1}{(n-1)^2}\right) \\ &= \frac{n-2}{n-1} + 1 - \frac{n-2}{n-1} \\ &= 1. \end{aligned}$$

Proposition 6. *We have*

$$E[N^*(n)] = n - 1.$$

Proof. The proof is by induction on n. Because $P(N^*(2) = 1) = 1$, it is true when $n = 2$, and so assume that $E[N_k^*] = k - 1$ for all $k = 2, \dots, n-1$. Then,

$$E[N^*(n) \mid D_n^*] = 1 + E[N^*(n - D_n^*) \mid D_n^*], \tag{5}$$

yielding

$$\begin{aligned} E[N^*(n)] &= 1 + \sum_{i=0}^{n-1} E[N^*(n-i)]\, P(D_n^* = i) \\ &= 1 + E[N^*(n)]\, P(D_n^* = 0) + \sum_{i=1}^{n-1} E[N^*(n-i)]\, P(D_n^* = i) \\ &= 1 + E[N^*(n)]\, P(D_n^* = 0) + \sum_{i=1}^{n-1} (n - i - 1)\, P(D_n^* = i) \\ &= 1 + E[N^*(n)]\, P(D_n^* = 0) + (n-1)(1 - P(D_n^* = 0)) - E[D_n^*] \\ &= 1 + E[N^*(n)]\, P(D_n^* = 0) + (n-1)(1 - P(D_n^* = 0)) - 1, \end{aligned}$$

which proves the result.

Remark. Proposition 6 could also have been proved by using a martingale stopping argument, as in the proof of Proposition 4.

Proposition 7. *For $n > 2$,*

$$\operatorname{var}(N^*(n)) = n - 1.$$

Proof. Let $V(n) = \operatorname{var}(N^*(n))$. The proof is by induction on n. As it is true for $n = 3$, since $N^*(3)$ is geometric with parameter $\frac{1}{2}$, assume it is true for all values between 2 and n. Now,

$$\operatorname{var}(N^*(n) \mid D_n^*) = \operatorname{var}(N^*(n - D_n^*) \mid D_n^*)$$

and, from (5) and Proposition 6,

$$\mathrm{E}[N^*(n) \mid D_n^*] = n - D_n^*.$$

Hence, by the conditional variance formula,

$$\begin{aligned} V(n) &= \sum_{i=0}^{n-1} V(n-i)\,\mathrm{P}(D_n^* = i) + \operatorname{var}(D_n^*) \\ &= V(n)\,\mathrm{P}(D_n^* = 0) + \sum_{i=1}^{n-1} V(n-i)\,\mathrm{P}(D_n^* = i) + 1. \end{aligned} \tag{6}$$

Now, because $\mathrm{P}(D_n^* = n-2) = 0$ and $V(1) = 0$, the induction hypothesis yields

$$\sum_{i=1}^{n-1} V(n-i)\,\mathrm{P}(D_n^* = i) = \sum_{i=1}^{n-1} (n-i-1)\,\mathrm{P}(D_n^* = i).$$

Hence, from (6),

$$\begin{aligned} V(n) &= V(n)\,\mathrm{P}(D_n^* = 0) + (n-1)(1 - \mathrm{P}(D_n^* = 0)) - \mathrm{E}[D_n^*] + 1 \\ &= V(n)\,\mathrm{P}(D_n^* = 0) + (n-1)(1 - \mathrm{P}(D_n^* = 0)), \end{aligned}$$

which proves the result.

References

[1] Arratia, R. and Tavaré, S. (1992). The cycle structure of random permutations. *Ann. Prob.* **20,** 1567–1591.

[2] De Montmort, P. R. (1708). *Essay d'Analyse sur le Jeux de Hazard*. Quillau, Paris.

[3] Diaconis, P., Fulman, J. and Guralnick, R. (2008). On fixed points of permutations. *J. Algebraic Combin.* **28,** 189–218.

[4] Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, Vol. I, 3rd edn. John Wiley, New York.

[5] Rao, M. B. and Kasala, S. (2008). A discrete probability problem in chemical bonding. Preprint, Universite de Technologie de Compiegne, France.

[6] Ross, S. M. and Peköz, E. (2007). *A Second Course in Probability*. ProbabilityBookstore.com, Boston, MA.