On the Evolution of Islands

Peter G. Doyle Colin Mallows Alon Orlitsky
Larry Shepp

October 5, 1998

1 Introduction

Suppose we have n cells arranged in a ring, or alternatively, in a row. We pick a cell at random and mark it; we pick one of the remaining unmarked cells at random and mark it; and so on until after n steps each cell is marked. After the k'th cell has been marked, the configuration of the marked cells defines some number of *islands* separated by *seas* (See Figure 1). An *island* is a maximal set of adjacent marked cells; a *sea* is a maximal set of adjacent unmarked cells. Let ξ_k be the random number of islands after k cells have been marked. Clearly $\xi_1 = \xi_n = 1$, and for a ring of cells $\xi_{n-1} = 1$ as well. We show that for n cells in a ring

$$E_{\text{ring}}\left(\frac{1}{\xi_1\xi_2\cdots\xi_{n-1}}\right) = \frac{1}{n!}\binom{2n-2}{n-1} = \frac{1}{(n-1)!}C_{n-1},$$

where C_k is the k'th Catalan Number

$$C_k = \frac{1}{k+1}\binom{2k}{k}.$$

For n cells in a row, the answer is the same as for $n+1$ cells in a ring. To see this, break the ring at the position of the last cell marked. Hence

$$E_{\text{row}}\left(\frac{1}{\xi_1\xi_2\cdots\xi_{n-1}}\right) = \frac{1}{(n+1)!}\binom{2n}{n} = \frac{1}{n!}C_n.$$

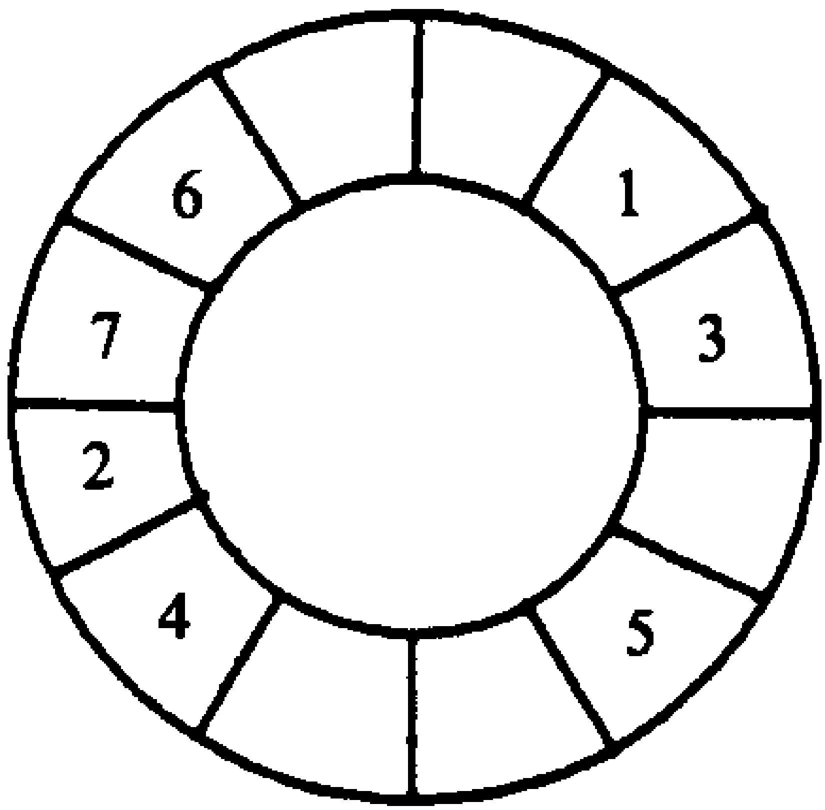

Figure 1: $n = 12$ cells, $k = 7$ marked cells, $\xi_k = 3$ islands. Numbers denote the time a cell was marked.

This latter formula is used in a companion paper, Shepp [1], to show that certain random graphs are disconnected.

From now on we will consider only the ring case, and write E instead of E_{ring} throughout.

The formula

$$E\left(\frac{1}{\xi_1\xi_2\cdots\xi_{n-1}}\right) = \frac{1}{n!}\binom{2n-2}{n-1}$$

is a special case of the following formula, valid for all $1 \le k \le \ell \le n-1$:

$$E\left(\frac{1}{\xi_k\cdots\xi_\ell}\right) = \frac{(k-1)!(n-\ell-1)!}{(n-1)!}\left[\binom{n+\ell-k}{n-k} - \binom{n+\ell-k}{\ell-k}\right] .$$

Another particular case of this general formula is

$$E\left(\frac{1}{\xi_k}\right) = \frac{n! - k!(n-k)!}{(n-1)!k(n-k)}.$$

We will also show that for all $1 \le k \le n-1$

$$E(\xi_k) = \frac{k(n-k)}{n-1}$$

and for all $1 \le k \le \ell \le n-1$

$$E(\xi_k\xi_\ell) = \frac{k(n-\ell)}{n-1} + \frac{k(n-k-1)(\ell-1)(n-\ell)}{(n-1)(n-2)} .$$

2 $E(\frac{1}{\xi_k\cdot\ldots\cdot\xi_\ell})$

We give two proofs that

$$E\left(\frac{1}{\xi_1\cdot\ldots\cdot\xi_{n-1}}\right) = \frac{1}{(n-1)!}C_{n-1} .$$

The first proof is inductive, the second uses a more elegant counting argument. The more general equation can be proved using similar methods.

2.1 An inductive proof

A straightforward inductive attack on this problem would number the cells in order $1, 2, ..., n$, and would define σ_k to be the number of the k'th marked cell. The sequence $\sigma_1, \sigma_2, \ldots, \sigma_n$ gives a complete description of the evolution of the process. This attack is unlikely to succeed, since the number of islands after k cells have been marked is a complicated function of these random variables. The trick in problems like this is to find a convenient *partial* description of the process under study, a description that captures what is of interest and that has simple probability properties. A similar trick is effective in problems in mechanics, where the judicious choice of a coordinate system can make all the difference.

Note that if we are interested only in the number of islands at each stage, then when there are exactly i islands, the *sizes* of these islands are irrelevant to the subsequent development. So we consider the situation where there are i islands and m cells still to be marked. Letting

$$\eta_j = \xi_{n-j}$$

we observe that, conditional on the event $\{\eta_m = i\}$, the random variables $\eta_1, \eta_2, \ldots, \eta_{m-1}$ have a distribution that does not depend on n. So we can define

$$f(m, i) = E\left(\frac{1}{\eta_m \eta_{m-1} \ldots \eta_1} | \eta_m = i\right)$$

and $E\left(\frac{1}{\xi_1 \cdot \ldots \cdot \xi_{n-1}}\right) = f(n-1, 1)$ (we can start the whole process after the first cell has been marked, since this must give just one island). We shall set up and solve a recurrence for f.

With $f(m, i)$ as defined above, we consider what can happen when the next cell is marked. There are m empty cells, and the next cell marked is equally likely to be any one of them. The crucial step in this approach is the observation that conditional on $\{\eta_m = i\}$, all possible *sizes* of the i seas are equally likely: the probability that when there are m cells still to be marked, there are exactly i islands and the sizes of the intervening seas are $\{m_1, m_2, ..., m_i\}$ (where necessarily each m_j is at least 1) is independent of $\{m_1, ...m_i\}$. This can be shown formally by Bayes' theorem.

It is convenient to distinguish two kinds of empty cells. An empty cell that is adjacent clockwise to an marked cell is called a *tied* cell. There are i

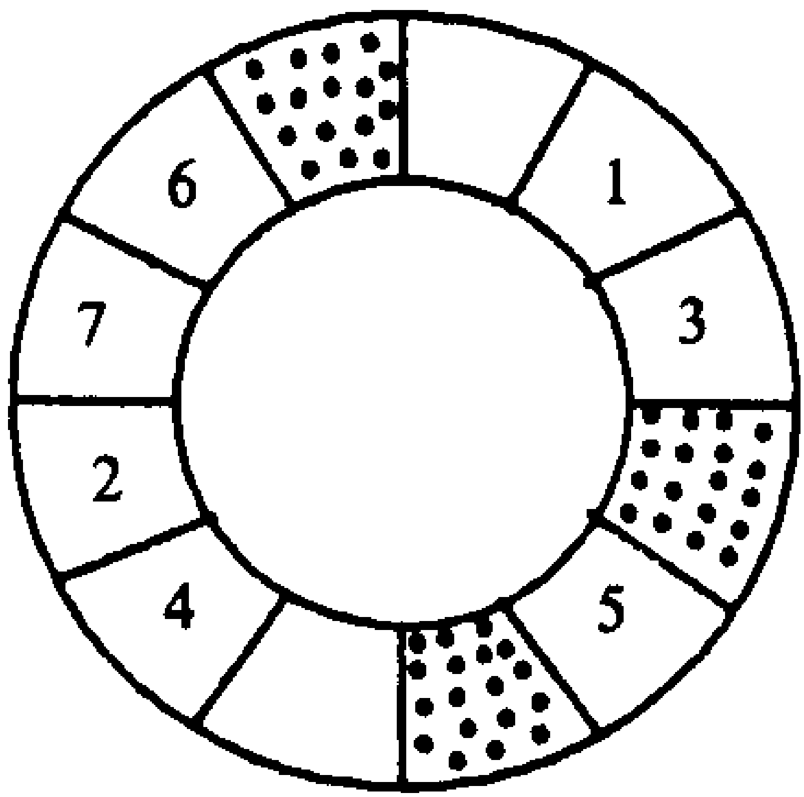

Figure 2: Tied (shaded) and free cells

such tied cells, and $m-i$ remaining *free* cells (See Figure 2). (We do not count an empty cell that is adjacent anticlockwise to an island as being tied to that island.)

With probability i/m the next cell to be marked is a tied cell; and then (using the "crucial observation" above) with probability $(i-1)/(m-1)$ there is an marked cell clockwise from it; with probability $(m-i)/(m-1)$ there is a free cell clockwise from it. On the other hand, with probability $(m-i)/m$ the next cell to be marked is a free cell; and then with probability $i/(m-1)$ the next clockwise cell is marked, and with probability $(m-i-1)/(m-1)$ it is empty. This gives the recurrence

$$\begin{aligned} f(m,i) &= \frac{1}{i}\Big(\frac{i}{m}\Big(\frac{i-1}{m-1}f(m-1,i-1)+\frac{m-i}{m-1}f(m-1,i)\Big) \\ &\quad +\frac{m-i}{m}\Big(\frac{i}{m-1}f(m-1,i)+\frac{m-i-1}{m-1}f(m-1,i+1)\Big)\Big) \end{aligned}$$

valid for $m \geq i$, with the boundary conditions $f(m,m)=1/m!$ since when $m=i$ we must have $\eta_j = j$ for $j=m-1,m-2,...,1$.

To solve this recurrence, put

$$f(m,i)=\frac{(m-i)!(i-1)!}{m!(m-1)!}a(m-i,m)$$

so that

$$a(d,m)=a(d,m-1)+2a(d-1,m-1)+a(d-2,m-1),$$

valid for $d \geq 0, m \geq 1$, with the boundary conditions

$$a(0,m)=1.$$

We recognize this recurrence as being related to binomial coefficients. Working out a few values of $a(d,m)/\binom{2m}{d}$ easily leads to the conjecture

$$a(d,m)=\frac{m-d}{m}\binom{2m}{d}$$

which does indeed satisfy the recurrence above. Thus we have

$$f(m,i)=\frac{i!}{(m+i)!}\binom{2m}{d}$$

so that finally we have

$$E\left(\frac{1}{\xi_1\cdot\ldots\cdot\xi_{n-1}}\right)=f(n-1,1)=\frac{1}{n!}\binom{2n-2}{n-1}=\frac{1}{(n-1)!}C_{n-1}.$$

2.2 A counting-argument proof

Let σ_i be the i'th marked cell. $(\sigma_1,\ldots,\sigma_n)$ is a permutation of $\{1,\ldots,n\}$. Each such permutation gives rise to a sequence $(c_1,c_2,\ldots,c_n)$ where c_i is the number of islands after the i'th cell has been marked. Call a sequence $(c_1,c_2,\ldots,c_{n-1})$ of positive integers *admissible* if $c_1=c_{n-1}=1$ and any two successive entries differ by at most 1. Let $\delta_i=c_{i+1}-c_i$ be the increment in the number of islands when the $i+1$'st cell is marked, and let $\omega=\sum_i 1-|\delta_i|$.

The number of permutations that gives rise to an admissible sequence $(c_1,c_2,\ldots,c_{n-1})$ is:

$$1\cdot 2^{1-|\delta_1|}c_1\cdot 2^{1-|\delta_2|}c_2\cdot\ldots\cdot 2^{1-|\delta_{n-2}|}c_{n-2}\cdot 1\cdot n=n2^{\omega}c_1c_2\cdots c_{n-1}\ .$$

To see this, think of a child assembling a necklace of beads, one bead at a time. The child can be working on more than one string at once; these strings are kept in a more or less circular ring, arranged in the same order as in the finished necklace. As each successive bead is added, it is joined to any bead that it will be adjacent to in the finished necklace. Figure 3 illustrates a possible arrangement after the child placed seven beads, forming three strings. Suppose there are c_i strings after the i'th bead has been added. If $\delta_i=1$ then the $i+1$'st bead creates a new island and there are c_i possible new-island locations. If $\delta_i=-1$, then the $i+1$'st bead connects two islands and there are c_i possible pairs of adjacent islands. If $\delta_i=0$, then the $i+1$'st bead is added to an existing island and there are c_i islands, each with two sides, hence there are $2c_i$ ways to add the bead. Once all the beads have been placed, there are n ways to spin them before obtaining a recipe for marking the cells.

Dividing the number of ways an admissible sequence $c_1,\ldots,c_n$ can arise by $n!$ gives the probability of the sequence:

$$P\left((\xi_1,\ldots,\xi_n)=(c_1,\ldots,c_n)\right)=\frac{2^{\omega}\xi_1\xi_2\cdots\xi_{n-1}}{(n-1)!}.$$

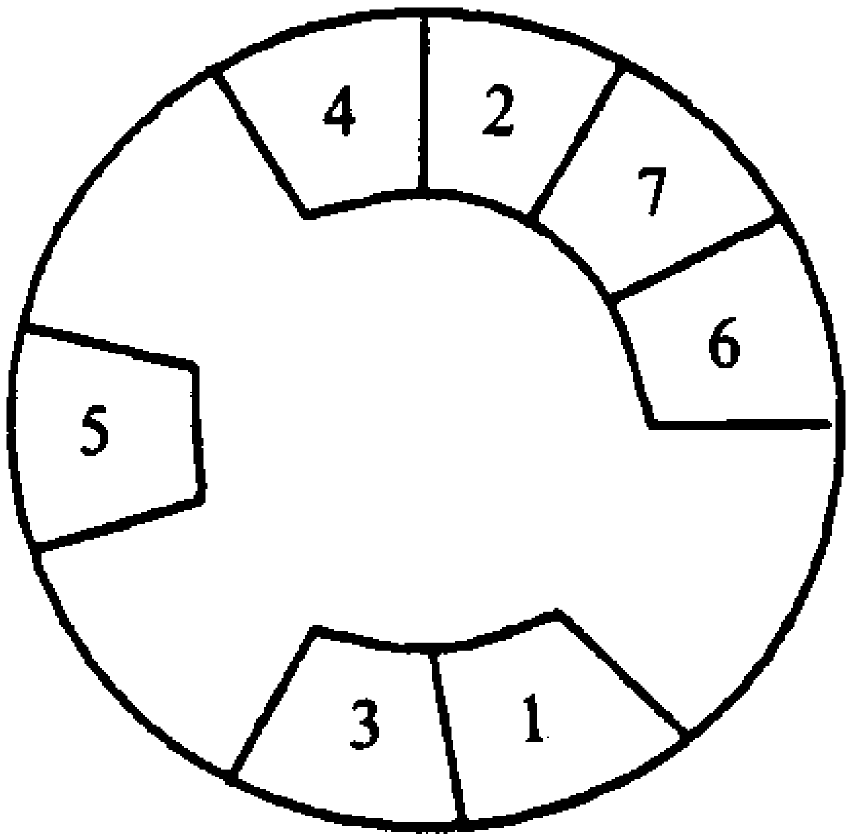

Figure 3: Assembling a necklace of beads.

The expected value that we are interested in is thus

$$E\left(\frac{1}{\xi_1\cdots\xi_{n-1}}\right)=\frac{1}{(n-1)!}\sum_{\substack{(c_1,c_2,\ldots,c_{n-1})\\ \text{admissible}}} 2^{\omega}.$$

So we just need to evaluate this sum.

Consider all possible walks $(x_0=0,x_1,\ldots,x_{2n-1},x_{2n}=0)$ on the non-negative integers that start from 0, go up or down 1 each time, and return to 0 for the first time after the $2n$'th step. The number of such walks is well known to be

$$\frac{1}{2n-1}\binom{2n-1}{n}=\frac{1}{n}\binom{2n-2}{n-1}=C_{n-1}.$$

Given such a walk, the sequence

$$(\frac{x_2}{2},\frac{x_4}{2},\ldots,\frac{x_{2n-2}}{2})$$

is an admissible sequence, and every admissible sequence arises from 2^{ω} different walks. Hence

$$\sum_{\substack{(c_1,c_2,\ldots,c_{n-1})\\ \text{admissible}}} 2^{\omega}=C_{n-1},$$

and

$$E\left(\frac{1}{\xi_1\cdots\xi_{n-1}}\right)=\frac{1}{(n-1)!}C_{n-1}.$$

3 Further results

For any possible sequence $\xi_1,\ldots,\xi_k$ of islands in the ring, the sequence $M_1,\ldots,M_{\xi_k}$ of sea sizes at time k is uniformly distributed: every positive sequence $m_1,\ldots,m_{\xi_k}$ satisfying

$$\sum_{i=1}^{\xi_k} m_i=n-k$$

arises as the value of $M_1, \ldots, M_{\xi_k}$ with the same probability. Therefore, the sequence $\xi_1, \ldots, \xi_{n-1}$ is a Markov Chain.

Using the uniformity of $M_1, \ldots, M_{\xi_k}$, and letting $\xi_0 \stackrel{\text{def}}{=} 0$, it is easy to see that for $1 \leq k \leq n-1$:

$$P\left(\xi_k | \xi_{k-1} = \xi\right) = \begin{cases} \frac{\xi(\xi-1)}{(n-k)(n-k+1)} & \text{if} \quad \xi_k = \xi - 1 \\ \\ \frac{2\xi(n-k+1-\xi)}{(n-k)(n-k+1)} & \text{if} \quad \xi_k = \xi \\ \\ \frac{(n-k-\xi)(n-k+1-\xi)}{(n-k)(n-k+1)} & \text{if} \quad \xi_k = \xi + 1 \ . \end{cases}$$

Hence

$$E(\xi_k | \xi_{k-1} = \xi) = 1 + \frac{n-k-1}{n-k+1}\xi \ .$$

and

$$E(\xi_k) = 1 + \frac{n-k-1}{n-k+1}E(\xi_{k-1}) \ . \tag{1}$$

Solving the recurrence with $E(\xi_0) = 0$ we obtain

$$E(\xi_k) = \frac{k(n-k)}{n-1} \ .$$

Similarly,

$$E(\xi_k^2 | \xi_{k-1} = \xi) = 1 + 2\frac{n-k-1}{n-k}\xi + \frac{(n-k-1)(n-k-2)}{(n-k+1)(n-k)}\xi^2 \ .$$

This, when solved, yields

$$E(\xi_k^2) = \frac{k(n-k)}{(n-1)(n-2)}\left(k(n-k) - 1\right) \ .$$

Equation (1) can also be used to show that for all $1 \leq k \leq \ell \leq n-1$

$$E(\xi_\ell | \xi_k) = \frac{(\ell - k)(n-\ell)}{n-k-1} + \frac{(n-\ell)(n-\ell-1)}{(n-k)(n-k-1)}\xi_k \ .$$

Therefore

$$\begin{aligned} E(\xi_k \cdot \xi_\ell) &= E\left(\xi_k E(\xi_\ell | \xi_k)\right) \\ &= \frac{k(n-\ell)}{n-1} + \frac{k(n-k-1)(\ell-1)(n-\ell)}{(n-1)(n-2)} \ . \end{aligned}$$

An alternative way of proving that

$$E(\xi_k) = \frac{k(n-k)}{n-1}$$

is via the differences $\xi_i - \xi_{i-1}$. They satisfy

$$P\left(\xi_i - \xi_{i-1} = \delta\right) = \begin{cases} \frac{(n-i)(n-i-1)}{(n-1)(n-2)} & \text{if} \quad \delta = 1 \\ \\ \frac{2(n-i)(i-2)}{(n-1)(n-2)} & \text{if} \quad \delta = 0 \\ \\ \frac{(i-1)(i-2)}{(n-1)(n-2)} & \text{if} \quad \delta = -1 \ . \end{cases}$$

To see that, consider the permutation σ that maps i to the cell marked at time i. The number of islands increases, decreases, or remains the same at time i, corresponding to whether i is a local minimum, maximum, or a middle point, of the inverse permutation σ^{-1}. Since σ is distributed uniformly over all permutations of $\{1, \ldots, n\}$, so is σ^{-1}. The integer i is a local minimum, maximum, or a middle point of σ^{-1} with the above probabilities. Therefore

$$E(\xi_i - \xi_{i-1}) = \frac{n-2i+1}{n-1}$$

and the result follows.

Finally, we can write down the value of $E(\xi_k)$ directly if we note that

$$\begin{aligned} & E(\xi_k) \\ = \; & E(\mid \{i | \text{after marking } k \text{ cells, } i \text{ is marked and } i+1 \text{ is unmarked}\} \mid) \\ = \; & \sum_i P(\text{after marking } k \text{ cells, } i \text{ is marked and } i+1 \text{ is unmarked}) \\ = \; & n \cdot \frac{k}{n} \cdot \frac{(n-1)-(k-1)}{n-1} \\ = \; & \frac{k(n-k)}{n-1}. \end{aligned}$$

References

[1] L. A. Shepp. Connectedness of certain random graphs. *Israel J. Math.*, 67, 1989.