J. Appl. Prob. **45,** 271–278 (2008)
Printed in England

THE NECKLACE PROCESS

COLIN MALLOWS,* *Avaya Labs*
LARRY SHEPP,** *Rutgers University*

Abstract

Start with a necklace consisting of one white bead and one black bead, and add new beads one at a time by inserting each new bead between a randomly chosen adjacent pair of old beads, with the proviso that the new bead will be white if and only if both beads of the adjacent pair are black. Let W_n denote the number of white beads when the total number of beads is n. We show that $\mathrm{E}\, W_n = n/3$ and, with $c^2 = \frac{2}{45}$, that $(W_n - n/3)/c\sqrt{n}$ is asymptotically standard normal. We find that, for all $r \geq 1$ and $n > 2r$, the rth cumulant of the distribution of W_n is of the form nh_r. We find the expected numbers of gaps of given length between white beads, and examine the asymptotics of the longest gaps.

Keywords: Special nonlinear recurrence; Markov chain; urn model

2000 Mathematics Subject Classification: Primary 60C05; 60D05
Secondary 62M05

1. The necklace process

We start with a necklace consisting of one white bead and one black bead. We add beads one at a time, putting each one into a gap (between beads) that is chosen at random, i.e. with probability $1/n$ for each gap when there are n beads, with the proviso that the new bead is white if and only if both adjacent beads are black. So it is impossible for two white beads to be adjacent to each other. Our study of this process was stimulated by consideration of a simple model of a communications network in which we have a cycle of active nodes, with new nodes added in random positions. We call a node 'white' if it is still connected to the same neighbors as when it first entered the system. We do this so that white nodes can be removed, backtracking the construction process. The resulting 'cycle' process is not quite the same as our 'necklace' process, since in the 'cycle' process, when a new node is added next to a 'white' node, the new node is white while the old white node becomes black. But the two processes are equivalent; the effect of adding a node next to an existing white node in the 'cycle' process is the same as adding a black node on the other side of this white node in the 'necklace' process. And these two positions for the new node are equally likely.

Suppose that when there are n beads in the necklace, the number of white beads is W_n. We will show that W_n is a Markov chain, and that the mean and variance of W_n are exactly $n/3$ (for all $n \geq 3$) and $2n/45$ (for all $n \geq 4$), respectively. The distribution of $(W_n - n/3)/\sqrt{2n/45}$ is asymptotically standard normal.

We are unable to find formulae for the distribution of W_n, but we will show that there are constants $h_1, h_2, \dots$ $(h_1 = \frac{1}{3})$ such that the rth cumulant of the distribution of W_n is of the form

Received 23 April 2007; revision received 27 August 2007.
* Postal address: Avaya Labs, 233 Mt. Airy Road, Basking Ridge, NJ 07920, USA.
Email address: colinm@research.avayalabs.com
** Postal address: Department of Statistics, 501 Hill Center, Busch Campus, Rutgers University, Piscataway, NJ 08854-8019, USA. Email address: shepp@stat.rutgers.edu

https://doi.org/10.1239/jap/1208358967 Published online by Cambridge University Press

$k_r(W_n) = nh_r$ for $n > 2r$ (there are anomalous values for $n \leq 2r$). This suggests that perhaps the distribution of W_n could be approximated by the distribution of a sum of n independent copies of a random variable with cumulants h_r, $r = 1, 2, \ldots$; but we show that there is no such random variable.

We also derive, for each $j \geq 2$, the expected number of gaps (between white beads) of length exactly j in the necklace, and examine the asymptotics of the longest gap.

We show that our process is very different from (i) a process with a random permutation of $n/3$ white beads and $2n/3$ black beads, subject to the condition that no two white beads are adjacent, and (ii) an urn model in which $n/3$ black beads are thrown randomly into $n/3$ gaps, where each gap is bounded by a white–black pair. For example, in our process the expected number of gaps (between white beads) of length 2 is $2n/15$, whereas in the random-permutation model this expected number is about $n/6$ and in the urn model it is about $n/3\mathrm{e}$.

2. The number of white beads

When there are n beads in the necklace and a new bead is added in a random position, if it is adjacent to an existing white bead then the number of white beads does not change (because the new bead must be black). The number of such positions is $2W_n$, one on each side of each white bead. If the new bead is added between two black beads then the new bead is white, so the number of white beads increases by 1. So, we have the following Markovian structure:

$$\mathrm{P}(W_{n+1} = W_n) = \frac{2W_n}{n},$$
$$\mathrm{P}(W_{n+1} = W_n + 1) = 1 - \frac{2W_n}{n}. \tag{1}$$

The conditional expectation of W_{n+1}, given W_n, is therefore

$$\mathrm{E}(W_{n+1} \mid W_n) = \frac{n-2}{n} W_n + 1,$$

and we have the recurrence relation

$$\mathrm{E}(W_{n+1}) = \frac{n-2}{n}\,\mathrm{E}(W_n) + 1.$$

Since $W_3 = 1$, this implies that $\mathrm{E}(W_n) = n/3$ for all $n \geq 3$.

Note that we could have chosen to start with a necklace consisting of a single bead (of either color). Then the second bead would have to be of the opposite color, and we have our two-bead starting point. An alternative formulation would be to require that the necklace has a unit circumference, starting with a single bead at $x = 0$ and adding the nth bead in position X_n, where these Xs are independent, continuous random variables on $(0, 1)$. For this model, the distributions of the variables we are interested in (W_n and the numbers of beads in the gaps between the white beads) are the same as for our model.

3. Moments

On seeing (1), one of the authors was of the opinion that a productive way to obtain asymptotic results would be by setting up one or more martingales (functions of W_n and arbitrary parameters). However, after much effort, this approach did not seem to yield useful results. This and all our more pedestrian attacks, attempting to obtain formulae or generating functions for the probability distribution of W_n, were hindered by the complexity of the problem. We find it remarkable that underlying this complexity are some very simple relations involving the moments.

From (1), using the result $\mathrm{E}(W_n) = n/3$, we can derive a recurrence relation for the second moment of W_n, namely

$$\mathrm{E}(W_{n+1}^2) = \frac{n-4}{n}\,\mathrm{E}(W_n^2) + \frac{2n+1}{3},$$

so that, for $n > 4$, $\mathrm{E}(W_n^2) = n^2/9 + 2n/45$ and $\operatorname{var}(W_n) = 2n/45$. Similar calculations for moments of orders 3, 4, 5, and 6 show that in each case, for sufficiently large n, each of the corresponding cumulants is exactly a multiple of n. We have the following result.

Theorem 1. *For process (1), there are constants $h_1, h_2, \dots$ such that, for all $r \geq 1$ and all $n > 2r$, the rth cumulant of the distribution of W_n is nh_r.*

We prove this result in Section 8. Note that this implies that the distribution of $(W_n - n/3)/\sqrt{n}$ is asymptotically Gaussian (with zero mean and variance $\frac{2}{45}$). The variance $2n/45$ is one fifth of the variance of a binomial distribution $\mathrm{B}(n, \frac{1}{3})$.

The form of the cumulants of W_n suggests that there might be a random variable Z, say, perhaps with support $(0, \frac{1}{2})$, so that the variable W_n would be distributed approximately as the sum of n independent and identically distributed copies of Z. The possibility that W_n has an (approximate) additive structure is plausible, because the evolution of the necklace between any pair of white beads is independent of what happens elsewhere. However, we have the following result (which we prove in Section 9).

Theorem 2. *The constants $h_1, h_2, \dots$ are not the cumulants of a proper distribution.*

4. Gaps

There are simple relations involving the lengths of the gaps between white beads. Suppose that when there are n beads altogether, there are $G_2(n)$ gaps of length 2, $G_3(n)$ gaps of length 3, and so on. (We have $G_1(n) = 0$ because no two white beads can be next to each other.) Then we must have

$$\begin{aligned} G_2(n) + G_3(n) + G_4(n) + \cdots &= W_n, \\ 2G_2(n) + 3G_3(n) + 4G_4(n) + \cdots &= n, \end{aligned}$$

since the first sum is equal to the total number of gaps, which equals the number of white beads, and the second sum equals the total number of beads.

When a new bead is added, several things may happen. If the new bead is adjacent to an existing white bead, the gap on that side of that bead becomes longer by 1. If the new bead is between two black beads, which lie in a gap of length j, say, (where $j \geq 3$) then this gap is deleted and is replaced by two shorter gaps with lengths summing to $j + 1$. In Section 10 we describe an examination of the possible cases and show that the expected numbers of counts satisfy recurrences similar to the one for $\mathrm{E}(W_n)$, above; namely, for $j \geq 2$ and $n > j + 2$,

$$\mathrm{E}(G_j(n+1)) = \frac{n-j-2}{n}\,\mathrm{E}(G_j(n)) + (j+3)b_j,$$

where $b_j = (j-1)(j+2)2^j/(j+3)!$. This leads to the following exact result (which we prove in Section 10).

https://doi.org/10.1239/jap/1208358967 Published online by Cambridge University Press

Theorem 3. *In the necklace process, for $j \geq 2$ and $n \geq j+3$, the expected number of gaps of length j is* $\mathrm{E}(G_j(n)) = nb_j$, *where* $b_j = (j-1)(j+2)2^j/(j+3)!$.

Also, we find that $\mathrm{E}(G_j(j+2)) = \mathrm{E}(G_j(j+3))$ (this value of $\mathrm{E}(G_j(j+2))$ does not conform to the formula in Theorem 3). The only other nonzero values are $\mathrm{E}(G_j(j)) = 2^{j-2}/(j-1)!$.

We present two more results on gaps, leaving the proofs as exercises for the reader. First, let $L_1(n)$ be the length of the gap between the original white bead and its closest neighbor (clockwise). Then, for $2 \leq k \leq n-2$, we have

$$\mathrm{P}(L_1(n) = k) = \frac{2^{k-1}(k-1)}{(k+1)!}$$

and $\mathrm{P}(L_1(n) = n) = 2^{k-2}/(k-1)!$. Hence, $\mathrm{E}(L_1(n)) \to (\mathrm{e}^2 - 1)/2 = 3.195$, a little larger then the overall average length, which is 3.

Next, let $L_{\text{last}}(n)$ be the length of the gap between the last white bead to enter and its clockwise closest neighbor. Then

$$\mathrm{P}(L_{\text{last}}(n) = k) \to \frac{3}{n} \sum_{j=k+1}^{\infty} G_j(n);$$

whence, $\mathrm{E}(L_{\text{last}}(n)) \to (3\mathrm{e}^2 - 17)/2 = 2.584$, a little smaller than 3.

5. Random permutations

It is interesting to compare these results with those of the model that arranges black and white beads at random, subject to having no two white beads adjacent. When n is large, W_n is close to $n/3$, so it makes sense to compare the expected number of gaps of various lengths in our necklace process with $n = 3m$ to those in the random-permutation process with m white beads and $2m$ black beads. We can view this latter process as randomly permuting (in a ring) m black beads and m white–black pairs. It is easy to derive the result, in this process, that, for $m > 1$, the expected number of gaps of length j is

$$\mathrm{E}(G_j(n)) = \frac{m^{(2)} m^{(j-2)}}{(2m-1)^{(j-1)}},$$

where $k^{(i)} = k(k-1)(k-2)\cdots(k-i+1) = k!/(k-i)!$. Thus, for large m, the expected number of gaps of length j is asymptotically $m/2^{j-1}$, which is not the same as our result for our necklace process.

6. Random urns

Another comparison is with the model in which m black beads are thrown at random into m urns, each of which already contains one black bead. Here the urns are defined as the gaps between the white beads in a ring that starts (with $n = 2m$) with m white–black pairs. Again we take $m = n/3$. For this model, the expected number of urns that end up with j black beads is

$$\mathrm{E}(G_j(n)) = \binom{m}{j-1} \frac{(m-1)^{m+1-j}}{m^m}.$$

In Table 1 we compare the results for the three processes we have discussed for the case in which $m = 1000$.

TABLE 1: Expected numbers of gaps of various lengths for three processes, where $n = 3000$.

Length of gap	Necklace process	Random permutations	Urns process
2	400	500	368
3	333	250	368
4	171	125	184
5	67	63	61
6	21	31	15
7	6	16	3
8	1	8	1
9	0	4	0
10	0	2	0
11	0	1	0
12	0	0	0

7. Asymptotics

We have presented formulae for the expected number of gaps of length j each for three different processes. To compare the lengths of the longest gaps, we use standard asymptotic techniques to derive, for each process, the length j_{longest} for which $\mathrm{E}(G_j(n))$ is approximately equal to 1. We find that, for the necklace process, and also for the urns process,

$$j_{\text{longest}} \sim \frac{\ln n}{\ln \ln n},$$

while, for the random-permutation process,

$$j_{\text{longest}} \sim \frac{\ln n}{\ln 2}.$$

Apart from these asymptotic results, we have nothing to say about the distribution of the longest gap.

8. Proof of Theorem 1

The cumulant-generating function of W_n (which must exist and have a convergent Taylor series for small t, because W_n has finite support) is

$$\begin{aligned} f_n(t) &= \log(\mathrm{E}(\exp(tW_n))) \\ &= k_1(n)t + \frac{k_2(n)t^2}{2} + \frac{k_3(n)t^3}{3!} + \cdots . \end{aligned}$$

The basic recurrence (1) gives

$$f_{n+1}(t) = f_n(t) + \log\biggl(\mathrm{e}^t - \frac{2}{n}(\mathrm{e}^t - 1) f_n'(t)\biggr), \tag{2}$$

and, using a Taylor series expansion, we can show that $k_1(n) = n/3$, $k_2(n) = 2n/45$, and so on, for $n > 4$. We need to prove this 'and so on' for all $n > 2r$.

https://doi.org/10.1239/jap/1208358967 Published online by Cambridge University Press

We define the function $h(t)$ that solves the differential equation

$$\mathrm{e}^{h(t)} = \mathrm{e}^t - 2(\mathrm{e}^t - 1)h'(t) \tag{3}$$

and $h(0) = 0$, which is found to be

$$h(t) = \log\left(\frac{\sqrt{y}}{\arctan(\sqrt{y})}\right),$$

where $y = \mathrm{e}^t - 1$. Note that the function $h(t)$ is

$$h(t) = -\log \int_0^1 \frac{\mathrm{d}u}{1 + yu^2},$$

and has a Taylor series expansion

$$h(t) = h_1 t + \frac{h_2 t^2}{2} + \frac{h_3 t^3}{3!} + \cdots,$$

which converges for $|t| < \ln 2$. Also, $h_1 = \frac{1}{3}$ and $h_2 = \frac{2}{45}$.

We will show that, for all $r \geq 1$, $k_r(n) = nh_r$, provided that $n > 2r$. We already know that this is true for $r = 1$. Suppose that we have shown this for all $j \leq r - 1$. From (2) we have, as $t \to 0$,

$$f_n(t) = nh(t) + (k_r(n) - nh_r)\frac{t^r}{r!} + O(t^{r+1}).$$

Using $[t^r/r!]g(t)$ to denote the coefficient of $t^r/r!$ in $g(t)$, from (3) we have

$$\begin{aligned} k_r(n+1) &= k_r(n) + \left[\frac{t^r}{r!}\right]\log\left(\mathrm{e}^t - \frac{2}{n}(\mathrm{e}^t - 1)f_n'(t)\right) \\ &= k_r(n) + \left[\frac{t^r}{r!}\right]\log\left(\mathrm{e}^t - \frac{2}{n}(\mathrm{e}^t - 1)\left(nh'(t) + (k_r(n) - nh_r)\frac{t^{r-1}}{(r-1)!}\right)\right). \end{aligned}$$

But, from (3), this is

$$\begin{aligned} k_r(n) &+ \left[\frac{t^r}{r!}\right]\left(h(t) + \log\left(1 - 2(\mathrm{e}^t - 1)\mathrm{e}^{-h(t)}\frac{t^{r-1}}{(r-1)!}\frac{k_r(n) - nh_r}{n}\right)\right) \\ &= k_r(n) + h_r - 2r\frac{k_r(n) - nh_r}{n} \\ &= \frac{n - 2r}{n}k_r(n) + (2r + 1)h_r, \end{aligned}$$

and it follows that no matter what $k_r(2r)$ is, for all $n > 2r$, we have $k_r(n) = nh_r$.

9. Proof of Theorem 2

The constants h_r for $r = 1, 2, 3$, and 4 are $\frac{1}{3}$, $\frac{2}{45}$, $-\frac{2}{945}$, and $-\frac{22}{4725}$, respectively. (These can be derived from the distribution of W_9, which is easily found to be $\mathrm{P}(W_9 = (1, 2, 3, 4)) = (1, 60, 192, 62)/315$.) Hence, the first four moments of the random variable $Z - \frac{1}{3}$ (if it exists) must be $\mu_1 = 0$, $\mu_2 = \frac{2}{45}$, $\mu_3 = -\frac{2}{945}$, and $\mu_4 = \frac{2}{1575}$. But a standard condition for the existence of a random variable with these moments is that the determinant of the 3×3 matrix $\boldsymbol{M}$ with $M_{ij} = \mu_{i+j}$, $i, j = 0, 1, 2$, should be nonnegative. But here the determinant is $-\frac{32}{893\,025}$.

10. Proof of Theorem 3

We derive the expectations $\mathrm{E}(G_j(n))$. Suppose that when the necklace contains n beads, the number of gaps of length j is $G_j(n)$ for all j. Then $\sum G_j(n) = W_n$ and $\sum jG_j(n) = n$. We have $G_3(3) = 1$ and $G_j(3) = 0$ for all $j \neq 3$. Also, $G_j(n) = 0$ for $j > n$. In the following, for clarity we write G_j for $G_j(n)$, the probabilities are conditional on the state of the necklace at stage n. On examining the possibilities when a new bead is added, we find that

$$\mathrm{P}(G_2(n+1) = G_2 - 1) = \frac{2G_2}{n},$$
$$\mathrm{P}(G_2(n+1) = G_2) = \frac{2G_3 + 2G_4 + 3G_5 + 4G_6 + 5G_7 + \cdots}{n},$$
$$\mathrm{P}(G_2(n+1) = G_2 + 1) = \frac{2G_4 + 2G_5 + 2G_6 + 2G_7 + \cdots}{n},$$
$$\mathrm{P}(G_2(n+1) = G_2 + 2) = \frac{G_3}{n},$$

so that

$$\begin{aligned}\mathrm{E}(G_2(n+1)) &= G_2 + \frac{-2G_2 + 2G_3 + 2G_4 + 2G_5 + \cdots}{n} \\ &= \frac{n-4}{n}G_2 + \frac{2}{n}W_n.\end{aligned}$$

Similarly, we find that

$$\begin{aligned}\mathrm{E}(G_3(n+1)) &= G_3 + \frac{2G_2 - 3G_3 + 2G_4 + 2G_5 + \cdots}{n} \\ &= \frac{n-5}{n}G_3 + \frac{2}{n}W_n, \\ \mathrm{E}(G_4(n+1)) &= G_4 + \frac{2G_3 - 4G_3 + 2G_4 + 2G_5 + \cdots}{n} \\ &= \frac{n-6}{n}G_4 + \frac{2}{n}(W_n - G_2), \\ \mathrm{E}(G_5(n+1)) &= G_5 + \frac{2G_4 - 4G_5 + 2G_6 + 2G_7 + \cdots}{n} \\ &= \frac{n-7}{n}G_5 + \frac{2}{n}(W_n - G_2 - G_3),\end{aligned}$$

and, generally, for $j \geq 4$, we have

$$\mathrm{E}(G_j(n+1)) = \frac{n-j-2}{n}G_j + \frac{2}{n}\left(W_n - \sum_{i=2}^{j-2} G_i\right).$$

Hence, it is easy to show that, for $n \geq j+3$, we have the unconditional probabilities

$$\mathrm{E}(G_j(n)) = \frac{(j-1)(j+2)2^j}{(j+3)!}n.$$

There are anomalous values for $n \leq j+2$. We find that $G_j(n)$ is 0 for $n < j$, and also for $n = j+1$. Also, $G_j(j) = 2^{j-2}/(j-1)!$, since at each stage the new bead must be adjacent

https://doi.org/10.1239/jap/1208358967 Published online by Cambridge University Press

to the single existing white bead. We will show that

$$G_j(j+2) = G_j(j+3) = \frac{(j-1)2^j}{(j+1)!}.$$

To see this, note that $G_j(j+2)$ is 0 except when $W_{j+2} = 2$, and that the necklace contains exactly one gap of length 2 and one gap of length j. The second white bead could of been added when the number of beads was any of $3, 4, \ldots, j-1$. Each of these possibilities has the same probability, namely $2^j/(j+1)!$, so the total probability is $(j-1)2^j/(j+1)!$, as we claim.