The necklace process

Colin Mallows and Larry Shepp
Avaya Labs and Statistics Department, Rutgers University
e-mail: colinm@research.avayalabs.com, shepp@stat.rutgers.edu

April 9, 2007

Abstract

.

Start with a necklace consisting of one white bead and one black one, 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 explore the properties of this process using the fact that the process (n, W_n) is an embedded Markov chain. We show $EW_n = n/3$ and that with $c^2 = 2/45$, $(W_n - n/3)/c\sqrt{n}$ is asymptotically standard normal. We find that for all $r >= 1$ and $n > 2r$, the r-th 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.

Exploring tempting conjectures shows that our process is different from that obtained by arranging $n/3$ white beads and $2n/3$ black beads in random order, subject to the "no-two-white-beads-adjacent" requirement, and is also different from the process which attempts to reconstruct the contents of the gaps between $n/3$ white beads by throwing $n/3$ black beads at random into $n/3$ urns, each of which starts with one black bead. The necklace process seems to be new.

1 The necklace process

Stimulated by a problem in the design of a communications network, we start with a necklace consisting of one white bead and one black one. We add beads one at a time, putting each one into a gap (betwen 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. 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 formulas for the distribution of W_n, but we will show that there are constants $h_1, h_2, \cdots$ $(h_1 = 1/3)$ such that the r-th cumulant of the distribution of W_n is of the form $k_r(W_n) = nh_r$ 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 $h_r, r = 1, 2, \cdots$; but we show 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 each of two other processes: first, random permutation of $n/3$ white beads and $2n/3$ black beads, subject to the condition that no two whites are adjacent; second, an urn model in which $n/3$ black beads are thrown randomly into $n/3$ gaps, each gap being 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 $n/3e$.

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, 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, the new bead is white, and the number of whites increases by 1. So we have the Markovian structure

$$\begin{aligned} P(W_{n+1} = W_n) &= 2W_n/n \\ P(W_{n+1} = W_n + 1) &= 1 - 2W_n/n \end{aligned} \tag{1}$$

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

$$E(W_{n+1}|W_n) = ((n-2)/n)W_n + 1$$

and we have the recurrence

$$E(W_{n+1}) = ((n-2)/n)E(W_n) + 1.$$

Since $W_3 = 1$, this implies that $E(W_n) = n/3$ 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 the necklace to have unit circumference, starting with a single bead at $x = 0$, and adding the n-th bead in position X_n, where these X's 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 gaps between white beads) are the same as for our model.

3 Moments

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

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

$$E(W_{n+1}^2) = ((n-4)/n)E(W_n^2) + (2n+1)/3$$

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

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

We prove this at the end of the paper. Note that this implies that the distribution of $(W_n - n/3)/\sqrt{n}$ is asymptotically Gaussian (with zero mean and variance $2/45$). The variance $2n/45$ is one fifth of the variance of a Binomial distribution $B(n, 1/3)$.

The form of the cumulants of W_n suggests that there might be a random variable Z say, perhaps with support $(0, 1/2)$, so that the variable W_n would be distributed approximately as the sum of n i.i.d. 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 show in a later section:

Theorem 2 *The constants $h_1, h_2, \cdots$ 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, etc. ($G_1(n) = 0$ 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 the total number of gaps, which equals the number of white beads, and the second sum is the total number of beads.

When a new white 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 gets 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 $j + 1$. Appendix 2 describes an examination of the possible cases, and shows that the expected numbers of the counts satisfy recurrences similar to the one for $(E(W_n))$ above, namely for $j \geq 2$ and $n > j + 2$

$$E(G_j(n+1)) = ((n - j - 2)/n)E(G_j(n)) + (j+3)b_j$$

where $b_j = (j-1)(j+2)2^j/(j+3)!$. It follows that we have the exact result (which we prove at the end of the paper):

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

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

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 $2 \leq k \leq n - 2$

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

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

Next, let $L_{last}(n)$ be the length of the gap between the last white bead to enter and its clockwise closest neighbor. Then $P(L_{last}(n) = k) \to (3/n)\sum_{j=k+1}^{\infty} G_j(n)$ whence $E(L_{last}(n)) \to (3e^2 - 17)/2 = 2.584$, 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 whites and $2m$ blacks. We can view this latter process as randomly permuting (in a ring) m blacks and m white-black pairs. It is easy to derive the result that for this process, for $m > 1$ the expected number of gaps of length j is

$$E(G_j(n)) = m^{(2)} m^{(j-2)} / (2m-1)^{(j-1)}$$

where $k^{(i)} = k(k-1)(k-2)\cdots(k-i+1) = k!/(k-i)!$. Thus for m large, 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 balls are thrown at random into m urns, each of which already contains one black ball. 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. We again take $m = n/3$. For this model the expected number of urns that end up with j black balls is

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

The following table compares the results for the three processes we have discussed, for the case $m = 1000$.

Expected numbers of gaps of various lengths for three processes, $n = 3000$

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

7 Asymptotics

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

$$j_{longest} \sim \ln n / \ln \ln n$$

while for the "random permutation" process

$$j_{longest} \sim \ln n / \ln 2$$

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

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

The basic recurrence (1) gives

$$f_{n+1}(t) = f_n(t) + \log(e^t - (2/n)(e^t - 1)f_n'(t)) \quad (2)$$

and we can show by expanding in Taylor series that $k_1(n) = n/3$, $k_2(n) = 2n/45$ for $n > 4$, etc. We need to prove this "etc" for all $n > 2r$.

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

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

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

$$h(t) = \log(\sqrt{y}/\arctan(\sqrt{y}))$$

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

$$h(t) = -\log \int_0^1 \frac{du}{1 + yu^2}$$

and has a Taylor series

$$h(t) = h_1 t + h_2 t^2/2 + h_3 t^3/3! ...$$

which converges for $|t| < \ln 2$. Also $h_1 = 1/3, h_2 = 2/45$.

We will show that for all $r \geq 1$, $k_r(n) = nh_r$, provided $n > 2r$. We already know that this is true for $r = 1$. Suppose 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 $\left[\frac{t^r}{r!}\right] g(t)$ to denote the coefficient of $\frac{t^r}{r!}$ in $g(t)$, from (3) we have

$$k_r(n+1) = k_r(n) + \left[\frac{t^r}{r!}\right] \log(e^t - (2/n)(e^t - 1)f_n'(t))$$

$$= k_r(n) + \left[\frac{t^r}{r!}\right] \log(e^t - (2/n)(e^t - 1)(nh'(t) + (k_r(n) - nh_r)\frac{t^{r-1}}{(r-1)!}$$

But from (3) this is

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

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

Proof of Theorem 2

The constants h_r for $r = 1, 2, 3, 4$ are $1/3, 2/45, -2/945, -22/4725$. (These can be derived from the distribution of W_9, which is easily found to be $P(W_9 = (1,2,3,4)) = (1, 60, 192, 62)/315)$). Hence the first four moments of the random variable $Z - 1/3$ (if it exists) must be $\mu_1, \mu_2, \mu_3, \mu_4 = 0, 2/45, -2/945, 2/1575$. But a standard condition for the existence of a random variable with these moments is that the determinant of the 3x3 matrix M with $M_{ij} = \mu_{i+j}$ $i, j = 0, 1, 2$ should be non-negative. But here this determinant is $-32/893025$.

Proof of Theorem 3

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

$$\begin{aligned}
P(G_2(n+1) = G_2 - 1) &= (2G_2)/n \\
P(G_2(n+1) = G_2) &= (2G_3 + 2G_4 + 3G_5 + 4G_6 + 5G_7 + \cdots)/n \\
P(G_2(n+1) = G_2 + 1) &= (2G_4 + 2G_5 + 2G_6 + 2G_7 + \cdots)/n \\
P(G_2(n+1) = G_2 + 2) &= (G_3)/n
\end{aligned}$$

so that

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

Similarly we find

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

and generally for $j \geq 4$

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

Hence it is easy to show that for $n \geq j+3$

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

There are anomalous values for $n \leq j+2$. $G_j(n)$ is zero 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 to the single existing white bead. We will show that

$$G_j(j+2) = G_j(j+3) = (j-1)2^j/(j+1)!$$

To see this, note that $G_j(j+2)$ is zero except when $W_{j+2} = 2$ and the necklace contains exactly one gap of length 2 and one of length j. The second white bead can have been added when the number of beads was any of $3, 4, \cdots j-1$. 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.