A Mean Paradox

by

John Konvalina, Jack Heidel, and Jim Rogers

Department of Mathematics University of Nebraska at Omaha Omaha, NE 68182-0243 USA

Phone: (402) 554-2836 Fax: (402) 554-2975 E-mail: johnkon@unomaha.edu

A Mean Paradox

In our department we have a mathematical biology group that includes student researchers studying mathematical models (known as Boolean networks) of complex biological systems [4]. Some of these models can be represented as permutations on a finite set whose size grows exponentially. One objective is to compute cycle structure statistics of the permutations, since these often have some biological relevance. The important cycle statistics include the average or mean number of cycles in a set of permutation and the mean cycle length. Since the space of permutations is so large, one must often resort to simulations and sampling, including randomly generating permutations and studying their cycles. It is well known from group theory that every permutation on a finite set can be written as a cycle or a product of disjoint cycles [2]. A classical result from combinatorics and discrete probability is that the mean cycle length of permutations on *n* symbols is approximately $n/\log n$. However, if we randomly generate a permutation of degree *n*, perform the decomposition and average the cycle lengths, and repeat this process many times, we find that the overall average or mean cycle length differs nontrivially from the classical result. Why should the mean cycle length arising from random permutation generation of an applied modeling problem differ from the classical mean cycle length? In this note we will show there is a surprisingly nontrivial resolution of this *mean* paradox.

First, we prove the classical result that the mean cycle length is approximately $n/\log n$. The origin of this proof can be traced back at least 50 years to a paper that appeared in this *Monthly* by R.E. Greenwood [3]. Let c_n denote the mean cycle length in permutations of size *n*. The *Stirling numbers of the first kind*, denoted by $\begin{bmatrix} n \\ k \end{bmatrix}$, arise as coefficients of the following polynomial :

$$F(x) = x(x+1)(x+2)\cdots(x+n-1) = \sum_{k=1}^{n} {n \brack k} x^{k}$$
(1)

The Stirling number $\begin{bmatrix} n \\ k \end{bmatrix}$ has a combinatorial interpretation as the number of permutations of the numbers 1 to *n* with exactly *k* cycles in the cycle decomposition **[5]**. Thus, summing over *k* yields the total number of cycles. But this is just the derivative of (1) evaluated at x = 1. That is,

$$F'(1) = \sum_{k=1}^{n} k \begin{bmatrix} n \\ k \end{bmatrix} .$$

If we divide F'(1) by F(1) = n!, the total number of permutations, we obtain the mean number of cycles in a permutation :

$$\frac{F'(1)}{F(1)} = \sum_{k=1}^{n} \frac{1}{k} = H_n$$

where H_n is the *n*th Harmonic number. It is well known that $H_n \sim \log n$. Thus, the total number of cycles over all permutations is $n!H_n$. The total of all cycle lengths is n!n since the sum of the lengths of the cycles in a given permutation is *n*. So the mean cycle length is

$$c_n = \frac{n!n}{n!H_n} = \frac{n}{H_n} \sim \frac{n}{\log n}.$$

Observe that the mean cycle length in the classical case was actually obtained by averaging over all the *cycles*. But, in our simulations we randomly generate permutations and then averaged the resulting cycle lengths. What is the mean cycle length when averaging over all *permutations*? Let p_n denote the mean cycle length over all permutations. There are $\begin{bmatrix} n \\ k \end{bmatrix}$ permutations with exactly *k* cycles in the cycle decomposition. The average cycle length for these permutations is n/k since the total number of elements is *n* and there are *k* cycles. Thus, summing over *k* and dividing by the total number of permutations, we have

$$p_n = \frac{\sum_{k=1}^n \frac{n}{k} \begin{bmatrix} n \\ k \end{bmatrix}}{n!}$$

Using (1) we obtain an integral representation for the mean cycle length:

$$p_{n} = \frac{\sum_{k=1}^{n} \frac{1}{k} {n \brack k}}{(n-1)!} = \frac{\int_{0}^{1} \frac{F(x)}{x} dx}{(n-1)!}$$

Or, using the gamma function $\Gamma(x)$:

$$p_n = \frac{\int_{0}^{1} (x+1)(x+2)\cdots(x+n-1)dx}{(n-1)!} = \frac{\int_{0}^{1} \frac{\Gamma(x+n)}{\Gamma(x+1)}dx}{\Gamma(n)}$$

Next, we will derive a recurrence relation for p_n . Observe that

$$p_{n+1} = \frac{\int_{0}^{1} (x+1)(x+2)\cdots(x+n-1)(x+n)dx}{n!} =$$

$$=\frac{\int_{0}^{1} (x+1)(x+2)\cdots(x+n-1)\,dx}{(n-1)!} + \frac{\int_{0}^{1} x\,(x+1)(x+2)\cdots(x+n-1)\,dx}{n!} =$$

$$= p_n + \frac{\int_0^1 x(x+1)(x+2)\cdots(x+n-1)dx}{n!}.$$

To evaluate the integral we use a classical result from the Calculus of Finite Differences (**[6]** pp.130, 182):

$$\int_{0}^{1} x(x+1)(x+2)\cdots(x+n-1)dx = \mathbf{B}_{n}^{(n)}(n) = (-1)^{n} \mathbf{B}_{n}^{(n)}$$

where $B_n^{(n)}$ is the *n*th Bernoulli number of order *n*. The exponential generating function for these numbers is ([6] p.135):

$$\sum_{n=0}^{\infty} (-1)^n B_n^{(n)} \frac{t^n}{n!} = \frac{-t}{(1-t)\log(1-t)} \quad .$$
⁽²⁾

So the recurrence relation for p_n is

$$p_{n+1} = p_n + \frac{(-1)^n \underline{B}_n^{(n)}}{n!}$$
(3)

with $p_0 = 0$, $\mathbf{B}_0^{(0)} = 1$. Using the generating function (2) together with (3) we obtain the generating function G(t) for p_n :

$$G(t) = \sum_{n=0}^{\infty} p_n t^n = \frac{-t^2}{(1-t)^2 \log(1-t)}$$
(4)

A generalization of the generating function (4) was studied by Flajolet and Odlyzko [1]. Using Cauchy's integral formula and Hankel type contours, they derived asymptotic expansions for the coefficients. Applying their Theorem 3A we obtain the following asymptotic expansion for the mean cycle lengths over the permutations:

$$p_{n+1} \sim \frac{n}{\log n} \left(1 + \sum_{k \ge 1} \frac{a_k}{\log^k n} \right)$$

where
$$a_k = \frac{d^k}{dx^k} \left(\frac{1}{\Gamma(-x)} \right)_{x=-2}$$
.

.

So the resolution of the mean paradox has led us to a nontrivial asymptotic expansion of the mean cycle length p_n from a closed-form expression of the associated generating function .

Acknowledgement

This research has been partially supported by NIH grant #1R016M672-01

References

- 1. Flajolet, P. and Odlyzko, A., Singularity analysis of generating functions. *SIAM J. Discrete Math.* **3** (1990) 216-240.
- 2. Gallian, J., Contemporary Abstract Algebra (5th ed.), Houghton-Mifflin, Boston, 2002.
- 3. Greenwood, RE., The number of cycles associated with the elements of a permutation group, *Amer. Math. Monthly*, **60** (1953) 407-409.
- Heidel, J., Maloney, J., Farlow, C., and Rogers, J., Finding cycles in synchronous Boolean networks with applications to biochemical sytems, *Int. J. Bif. and Chaos*, 13 (2003) 535-552.
- 5. Konvalina, J., A unified interpretation of the binomial coefficients, the Stirling numbers, and the Gaussian coefficients, *Amer. Math. Monthly*, **107** (2000),901-910.
- 6. Milne-Thomson, L., The Calculus of Finite Differences, Chelsea, New York, 1981.