THE STATISTICS OF THE SMALLEST SPACE ON A LOTTERY TICKET

Robert E. Kennedy and Curtis N. Cooper
Central Missouri State University, Warrensburg, MO 64093
(Submitted March 1990)

Introduction

A day hardly goes by without hearing that some lucky person has become an "instant millionaire" by winning a lottery. Recently, one of the authors was visiting relatives in Florida when a sequence of winning lottery numbers was announced. (In the Florida state lottery, one chooses six distinct integers from 1 to 49.) Someone suggested that a person might just as well choose 1, 2, 3, 4, 5, and 6 as any other sequence. In fact, why not choose any six consecutive integers . . . what difference does it make? The chances are the same as any other sequence of six distinct integers!

This led to the following analysis of the least interval between consecutive members of a sequence of six integers. Here, we are concerned with the set of possible lottery tickets for the Florida state lottery. That is, the set of all possible six distinct integers from 1 to 49. The calculation given below can be generalized to "r integers from 1 to n are chosen." The generalization will be given at the end of this article. For clarity, however, we will use Florida's lottery as an example of the technique involved.

In what follows, we let L be the set of all possible Florida lottery tickets. That is,

$$L = \{(t_1, t_2, t_3, t_4, t_5, t_6) : 1 \le t_1 < t_2 < t_3 < t_4 < t_5 < t_6 \le 49\}.$$

We also define the function f on L by:

$$f(t_1, t_2, t_3, t_4, t_5, t_6) = \min\{t_{i+1} - t_i : i = 1, 2, 3, 4, 5\}.$$

Thus, if $t \in L$, we can think of $f(t)$ as the "smallest space" on the ticket t. Our purpose is to determine the mean smallest space with respect to the members of L. That is,

$$\frac{\sum_{t \in L} f(t)}{\binom{49}{6}}$$

will be determined.

Determination of the Mean of the Smallest Spaces of L

Consider the set of 5-tuples,

$$D = \{(d_1, d_2, d_3, d_4, d_5) : 5 \le d_1 + d_2 + d_3 + d_4 + d_5 \le 48;\ d_i \ge 1\},$$

and the function $F : L \to D$ defined by

$$F((t_1, t_2, t_3, t_4, t_5, t_6)) = (t_2 - t_1, t_3 - t_2, t_4 - t_3, t_5 - t_4, t_6 - t_5).$$

It is clear that F is a function from L *onto* D. This will enable us to efficiently determine

$$\sum_{t \in L} f(t)$$

by use of a particular partition of D. If $d \in D$, we note that

$$\#\{t \in L : F(t) = d\} = 49 - s,$$

where $\#$ is used to denote the number of elements in a set and where

$$d = (d_1, d_2, d_3, d_4, d_5) \quad \text{and} \quad s = d_1 + d_2 + d_3 + d_4 + d_5.$$

For $d = (d_1, d_2, d_3, d_4, d_5) \in D$, we define

$$s(d) = d_1 + d_2 + d_3 + d_4 + d_5$$
$$a(d) = \min\{d_1, d_2, d_3, d_4, d_5\}$$
$$m(d) = \#\{i : d_i = a(d)\}.$$

When the context is clear, we will just write s, a, or m. We now see that

$$f(t) = a(F(t)),$$

and that

$$5 \le s \le 48,$$
$$1 \le a \le 9,$$

and $\quad 1 \le m \le 5.$

For each triple (i, j, k) with $5 \le i \le 48$, $1 \le j \le 9$, and $1 \le m \le 5$, we define

$$D_{ijk} = \{d \in D : s(d) = i;\ a(d) = j;\ m(d) = k\}$$

and note that

$$\mathscr{D} = \{D_{ijk} : 5 \le i \le 48;\ 1 \le j \le 9;\ 1 \le k \le 5\}$$

is a partition of D. Since

$$(*) \qquad \sum_{t \in D} f(t) = \sum_{D_{ijk} \in \mathscr{D}} (49 - i)j(\#D_{ijk}),$$

we proceed to determine the right side of (*) by first considering each $k = 1$, 2, 3, 4, and 5. For this, we use the following theorem. Its statement and proof are found in [1: Theorem 2.4.3; pp. 145-46].

Theorem: For integers $r, r_1, r_2, \ldots, r_n$, the number of solutions to

$$x_1 + x_2 + \cdots + x_n = r$$
$$x_i \ge r_i \text{ for } i = 1, 2, \ldots, n$$

is

$$\binom{n - 1 + r - r_1 - r_2 - \cdots - r_n}{n - 1}$$

Thus, if we let s and a be given, we use the above theorem to find the number of solutions to

$$d_1 + d_2 + d_3 + d_4 + d_5 = s$$
$$d_i \ge r_i;\ i = 1, 2, 3, 4, 5,$$

for $m = 1, 2, 3, 4$, and 5. For example, if $m = 1$, $d_i = a$ for some i and $d_j \ge a + 1$ for $j \ne i$. Since there are $\binom{5}{1}$ ways to choose the d_i and, by the theorem,

$$x_1 + x_2 + x_3 + x_4 = s - a$$
$$x_i \ge a + 1;\ i = 1, 2, 3, 4$$

has

$$\binom{s - 5a - 1}{3}$$

solutions, it follows that

$$\#D_{s,a,1} = 5\binom{s - 5a - 1}{3}.$$

Similarly, we obtain

$$\#D_{s,a,2} = \binom{5}{2}\binom{s-5a-1}{2} = 10\binom{s-5a-1}{2},$$

$$\#D_{s,a,3} = \binom{5}{3}\binom{s-5a-1}{1} = 10\binom{s-5a-1}{1},$$

$$\#D_{s,a,4} = \binom{5}{4}\binom{s-5a-1}{0} = 5\binom{s-5a-1}{0},$$

and

$$\#D_{s,a,5} = \binom{5}{5}\binom{s-5a-1}{-1} = \binom{s-5a-1}{-1}.$$

It should be noted here that we will use the convention

$$\binom{n}{-1} = \begin{cases} 1, & \text{if } n = -1; \\ 0, & \text{otherwise} \end{cases}$$

and

$$\binom{i}{k} = 0 \quad \text{if } i < k.$$

Thus, for fixed s and a,

$$\sum_{k=1}^{5} \#D_{sak} = \sum_{k=1}^{5} \binom{5}{k}\binom{s-5a-1}{4-k},$$

and by [2: Formula 21; p. 58], we have

$$\sum_{k=1}^{5} \#D_{sak} = \binom{s-5a+4}{4} - \binom{s-5a-1}{4}$$

Hence,

$$\sum_{t \in L} f(t) = \sum_{i=5}^{48} \sum_{j=1}^{9} \sum_{k=1}^{5} (49-i)j(\#D_{ijk}) = \sum_{i=5}^{48} \sum_{j=1}^{9} (49-i)j \sum_{k=1}^{5} \#D_{ijk}$$

which is, by the above,

$$= \sum_{i=5}^{48} (49-i) \sum_{j=1}^{9} j\left[\binom{i-5j+4}{4} - \binom{i-5j-1}{4}\right]$$

and by telescoping the inner sum,

$$= \sum_{i=5}^{48} \sum_{j=1}^{9} (49-i)\binom{i-5j+4}{4} = \sum_{j=1}^{9} \sum_{i=0}^{49} \binom{49-i}{1}\binom{i-5j+4}{4}$$

$$= \sum_{j=1}^{9} \binom{54-5j}{6}$$

by [2: Formula 25; p. 58]. We have, then, that the mean "smallest space" on a (Florida) lottery ticket is

$$\frac{\sum_{t \in L} f(t)}{\binom{49}{6}} = \frac{\sum_{j=1}^{9} \binom{54-5j}{6}}{\binom{49}{6}}$$

which is approximately 1.88.

Distribution

Of interest, also, would be a list of how lottery tickets are distributed with respect to the "smallest space" concept. For example, how many of the $\binom{49}{6}$ Florida state lottery tickets have a "smallest space" of 3?

This can be answered readily by noting that by omitting the j factor in the summand of (*) and summing with a fixed j, we have that the number of Florida lottery tickets with a "smallest space" of a is

$$\sum_{\substack{D_{ijk} \in \mathscr{D} \\ j = a}} (49 - i) \# D_{ijk},$$

which simplifies to

$$\binom{54 - 5a}{6} - \binom{49 - 5a}{6}$$

Hence, we can construct the following list of how the Florida lottery tickets are distributed with respect to the "smallest space" idea.

smallest space	*number of such tickets*
1	6924764
2	3796429
3	1917719
4	869884
5	340424
6	107464
7	24129
8	2919
9	84

Thus, it can be observed that close to 91% of all possible Florida state lottery tickets have a "smallest space" of 1, 2, or 3. It seems, then, that it might be wise to choose a lottery ticket that has a "smallest space" of 1, 2, or 3 and avoid those with a "smallest space" greater than 3.

Conclusion

As stated earlier, the above could be generalized to a lottery where r numbers from the sequence 1, 2, 3, ..., n are chosen. Using the same technique as before, it is easily shown that the mean of the "smallest space" of all possible lottery tickets where r numbers are chosen from 1, 2, 3, ..., n is

$$\frac{\displaystyle\sum_{j=1}^{\left[\frac{n-1}{r-1}\right]} \binom{n - (r - 1)(j - 1)}{r}}{\binom{n}{r}}$$

and that the number of such lottery tickets with a "least space" of a is

$$\binom{n - (r - 1)(a - 1)}{r} - \binom{n - (r - 1)a}{r}.$$

Of course, another approach in investigating lottery tickets might be to analyze the collection of lottery tickets with respect to the "largest space" on a ticket. This should also be of interest, and we encourage the reader to make such an analysis.

References

1. J. L. Mott, A. Kandel, & T. P. Baker. *Discrete Mathematics for Computer Scientists*. New York: Reston Publishing Co., 1983.
2. D. E. Knuth. *The Art of Computer Programming*, Vol. 1. New York: Addison-Wesley Publishing Co., 1969.

*****