THE DISTRIBUTION OF SPACES ON LOTTERY TICKETS

Norbert Henze
Institut für Mathematische Stochastik, Universität Karlsruhe,
Englerstr. 2, 76128 Karlsruhe, Germany
(Submitted March 1994)

1. INTRODUCTION

In many lotteries (e.g., Florida State, Canadian, German) people choose six distinct integers from 1 to 49 so that the set of all lottery tickets is given by

$$T = \{t = (t_1, t_2, \dots, t_6) : 1 \le t_1 < t_2 < \cdots < t_6 \le 49\}.$$

Assuming a uniform distribution over all $\binom{49}{6}$ tickets, Kennedy and Cooper [1] obtained the expectation and the distribution of the "smallest space" random variable

$$S(t) = \min\{t_{j+1} - t_j : j = 1, 2, 3, 4, 5\}$$

and asked for the distribution of the "largest spacing"

$$L(t) = \max\{t_{j+1} - t_j : j = 1, 2, 3, 4, 5\}.$$

By means of a certain "shrinking procedure," we provide a simple derivation of the results of Kennedy and Cooper. Moreover, we use this idea to obtain the distribution (and expectation) of L as well as the (joint) distribution of the individual "spacing" random variables given by

$$X_j(t) = t_{j+1} - t_j, \quad j = 1, \dots, 5. \tag{1.1}$$

A generalized lottery will be treated in the final section. As a bit of convenient but nonstandard notation, let

$$\binom{m}{n}^+ = \begin{cases} 0, & \text{if } m < 0, \\ \binom{m}{n}, & \text{otherwise,} \end{cases}$$

denote a slight modification of the binomial coefficient $\binom{m}{n}$.

2. DISTRIBUTION OF A SINGLE SPACING

We first consider the distribution of the j^{th} spacing random variable X_j defined in (1.1). The crucial observation is that a 6-tuple $t = (t_1, \dots, t_6)$ from T satisfying $t_{j+1} - t_j \ge k$, where $k \in \{1, 2, \dots, 44\}$ may be "shrunk" into a 6-tuple $u = (u_1, \dots, u_6)$, where

$$\begin{aligned} u_\nu &= t_\nu, & \nu &= 1, 2, \dots, j, \\ u_\nu &= t_\nu - (k-1), & \nu &= j+1, \dots, 6. \end{aligned}$$

Obviously, this "shrinking procedure" is a one-to-one mapping from $\{t \in T : t_{j+1} - t_j \ge k\}$ onto the set $M = \{(u_1, \dots, u_6) : 1 \le u_1 < u_2 < \cdots < u_6 \le 49 - (k-1)\}$ which has cardinality $\binom{50-k}{6}$. We therefore obtain

$$P(X_j \ge k) = \binom{50-k}{6}^+ \bigg/ \binom{49}{6}, \quad k \ge 1,$$

and thus

$$P(X_j = k) = P(X_j \ge k) - P(X_j \ge k+1)$$
$$= \binom{49}{6}^{-1}\left[\binom{50-k}{6}^{+} - \binom{49-k}{6}^{+}\right] = \binom{49-k}{5}^{+}\Big/\binom{49}{6}, \quad k \ge 1.$$

Using the general fact that, for an integer-valued random variable N, expectation and variance may be computed from

$$E(N) = \sum_{k\ge1} P(N \ge k) \tag{2.1}$$

and

$$\mathrm{Var}(N) = 2\sum_{k\ge1} kP(N\ge k) - E(N) - (E(N))^2 \tag{2.2}$$

(this is readily seen upon writing

$$E(N) = \sum_{j\ge1} jP(N=j) = \sum_{j\ge1}\left(\sum_{k=1}^{j} 1\right)P(N=j),$$
$$E(N(N+1)) = \sum_{j\ge1} j(j+1)P(N=j) = 2\sum_{j\ge1}\left(\sum_{k=1}^{j} k\right)P(N=j),$$

and then interchanging the order of summation); it follows that

$$E(X_j) = \binom{49}{6}^{-1}\sum_{k=1}^{44}\binom{50-k}{6} = \frac{50}{7} = 7.1428\ldots$$

and

$$\mathrm{Var}(X_j) = 2\cdot\binom{49}{6}^{-1}\sum_{k=1}^{44} k\cdot\binom{50-k}{6} - E(X_j) - (E(X_j))^2 = \frac{3225}{98} = 32.9081\ldots.$$

Note that the distribution of X_j does not depend on j, which is intuitively obvious.

3. JOINT DISTRIBUTION OF SPACINGS

For the sake of lucidity, we first consider the joint distribution of two spacings X_i and X_j, where $1 \le i < j \le 5$. Here the idea is to "shrink" a ticket $(t_1, \ldots, t_6) \in T$ satisfying $t_{i+1} - t_i \ge k$, $t_{j+1} - t_j \ge \ell$, where $k, \ell \ge 1$, $k + \ell \le 45$, into the 6-tuple $(u_1, \ldots, u_6)$, where

$$\begin{aligned} u_\nu &= t_\nu, & \nu &= 1, \ldots, i, \\ u_\nu &= t_\nu - (k-1), & \nu &= i+1, \ldots, j, \\ u_\nu &= t_\nu - (k-1) - (\ell-1), & \nu &= j+1, \ldots, 6. \end{aligned}$$

Since the "shrinking mapping" is now one-to-one from $\{t \in T : t_{i+1} - t_i \ge k,\ t_{j+1} - t_j \ge \ell\}$ onto $\{(u_1, \ldots, u_6) : 1 \le u_1 < \cdots < u_6 \le 49 - (k-1) - (\ell-1)\}$, we obtain

$$P(X_i \ge k, X_j \ge \ell) = \binom{51-k-\ell}{6}^{+}\Big/\binom{49}{6}, \quad k, \ell \ge 1,$$

and thus, by the inclusion-exclusion principle

$$\begin{aligned} P(X_i = k, X_j = \ell) &= P(X_i \geq k, X_j \geq \ell) - P(X_i \geq k, X_j \geq \ell + 1) \\ &\quad - P(X_i \geq k+1, X_j \geq \ell) + P(X_i \geq k+1, X_j \geq \ell+1) \qquad (3.1) \\ &= \binom{49}{6}^{-1}\left[\binom{51-k-\ell}{6}^{+} - 2\binom{50-k-\ell}{6}^{+} + \binom{49-k-\ell}{6}^{+}\right] = \binom{49-k-\ell}{4}^{+} \Big/ \binom{49}{6}, \end{aligned}$$

$(k, \ell \geq 1)$. From this and

$$\begin{aligned} E(X_iX_j) &= \sum_{k\geq 1}\sum_{\ell \geq 1} k\ell P(X_i = k, X_j = \ell) = \sum_{k\geq 1}\sum_{\ell\geq 1} P(X_i \geq k, X_j \geq \ell) \\ &= \binom{49}{6}^{-1} \sum_{k\geq 1}\sum_{\ell\geq 1}\binom{51-k-\ell}{6}^{+} = \frac{1275}{28} = 45.535\ldots \end{aligned}$$

the correlation coefficient between X_i and X_j is given by

$$\rho(X_i, X_j) = \frac{E(X_iX_j) - E(X_i)E(X_j)}{(\mathrm{Var}(X_i)\mathrm{Var}(X_j))^{1/2}} = -\frac{1}{6}. \qquad (3.2)$$

The fact that $\rho(X_i, X_j)$ is negative is also intuitively obvious since large values of X_i tend to produce small values of X_j and vice versa.

It should now be clear how to obtain the joint distribution of more than two spacings. For example, a ticket $(t_1, \ldots, t_6)$ satisfying

$$t_{i+1} - t_i \geq k_i, \quad i = 1, \ldots, 5, \qquad (3.3)$$

where $k_1 + \cdots + k_5 \leq 48$, may be "shrunk" into the ticket $(u_1, \ldots, u_6)$, where

$$u_1 = t_1, \quad u_j = t_j - \sum_{\nu=1}^{j-1}(k_\nu - 1), \quad 2 \leq j \leq 6.$$

This shrinking mapping is one-to-one from the set of tickets satisfying (3.3) onto the set of ordered 6-tuples from 1 to $54 - \sum_{\nu=1}^{5} k_\nu$. We therefore have

$$P(X_j \geq k_j \text{ for } j = 1, 2, \ldots, 5) = \binom{54-k_1-k_2-k_3-k_4-k_5}{6}^{+} \Big/ \binom{49}{6} \qquad (3.4)$$

$(k_1 \geq 1, \ldots, k_5 \geq 1)$, and probabilities of the type $P(X_j = \ell_j, j = 1, 2, \ldots, 5)$ may be obtained from (3.4) and the method of inclusion and exclusion by analogy with (3.1). Note that the joint distribution of $(X_1, X_2, X_3, X_4, X_5)$ is invariant with respect to permutations of the X_j.

4. THE DISTRIBUTION OF THE SMALLEST SPACING

The idea of "ticket shrinking" yields the following simple derivation of the results of Kennedy and Cooper [1] concerning the minimum spacing $S = \min(X_1, X_2, X_3, X_4, X_5)$.

Since $S \geq k$ if and only if each of the X_j is not smaller than k, (3.4) entails

$$P(S \geq k) = \binom{54-5k}{6}^{+} \Big/ \binom{49}{6}, \quad k \geq 1,$$

and thus

$$P(S=k)=P(S\ge k)-P(S\ge k+1)$$
$$=\binom{49}{6}^{-1}\left[\binom{54-5k}{6}^{+}-\binom{49-5k}{6}^{+}\right],\quad k\ge 1.$$

From (2.1) the expectation of S is

$$E(S)=\binom{49}{6}^{-1}\sum_{k=1}^{9}\binom{54-5k}{6}=\frac{4381705}{2330636}=1.88004\ldots,$$

and, in addition to Kennedy and Cooper, the variance of S [computed from (2.2)] is given by

$$\operatorname{Var}(S)=\frac{6842931587015}{5431864164496}=1.25977\ldots.$$

5. THE DISTRIBUTION OF THE LARGEST SPACING

We now answer the question posed by Kennedy and Cooper [1] concerning the distribution of the largest spacing $L=\max(X_1, X_2, X_3, X_4, X_5)$.

Noting that $L\ge k$ if and only if at least one of the X_j is not smaller than k, the reasoning of section 3 and the inclusion-exclusion formula yield

$$P(L\ge k)=P(X_1\ge k \text{ or } X_2\ge k \text{ or } \cdots \text{ or } X_5\ge k)$$
$$=5P(X_1\ge k)-\binom{5}{2}P(X_1\ge k,\ X_2\ge k)+\binom{5}{3}P(X_1\ge k,\ X_2\ge k,\ X_3\ge k)$$
$$-\binom{5}{4}P(X_j\ge k;\ j=1,\ldots,4)+5P(X_j\ge k;\ j=1,\ldots,5)$$
$$=\binom{49}{6}^{-1}\sum_{j=1}^{5}(-1)^{j-1}\binom{5}{j}\binom{49-j(k-1)}{6}^{+}$$

[$k\ge 1$; note that $P(L\ge k)=0$ if $k\ge 45$] and thus

$$P(L=k)=P(L\ge k)-P(L\ge k+1)$$
$$=\binom{49}{6}^{-1}\sum_{j=1}^{5}(-1)^{j-1}\binom{5}{j}\left[\binom{49-j(k-1)}{6}^{+}-\binom{49-jk}{6}^{+}\right]\quad (k=1,2,\ldots,44).$$

Figure 5.1 shows a bar chart of the probability distribution of the maximum spacing L.

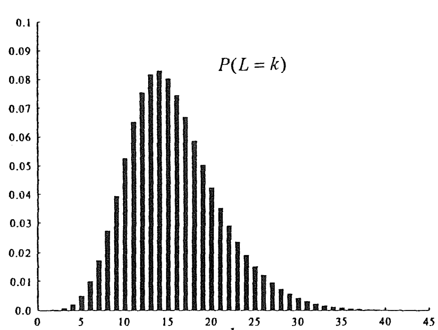

FIGURE 5.1. Distribution of the Largest Spacing on a "6/49" Lottery Ticket

Note that the distribution is skewed to the right. The mode is 14 and has a probability of 0.0828..., whereas the mean "largest space" is given by

$$E(L) = \sum_{k=1}^{44} P(L \geq k) = \frac{109376345}{6991908} = 15.643\ldots.$$

6. THE GENERAL CASE

It is clear that the reasoning given above carries over nearly literally to the case of a generalized lottery where r numbers from the sequence 1, 2, ..., n are chosen. For a ticket $t = (t_1, \ldots, t_r)$ with $1 \leq t_1 < \cdots < t_r \leq n$ let, as above, $X_j(t) = t_{j+1} - t_j, 1 \leq j \leq r -$, denote a single spacing, and write $S(t) = \min_{1 \leq j \leq r-1} X_j(t)$, $L(t) = \max_{1 \leq j \leq r-1} X_j(t)$ for the smallest resp. the largest spacing.

As a simple consequence of the idea of "ticket shrinking," we have

$$P(X_{j_1} \geq k_1, X_{j_2} \geq k_2, \ldots, X_{j_m} \geq k_m) = \binom{n - \sum_{\nu=1}^{m}(k_\nu - 1)}{r}^{+} \Big/ \binom{n}{r} \tag{6.1}$$

($1 \leq m \leq r-1$; $1 \leq j_1 < j_2 < \cdots < j_m \leq r-1$; $k_1 \geq 1, \ldots, k_m \geq 1$) which entails that the individual spacings are exchangeable, i.e., the joint distribution of any subset of $X_1, \ldots, X_{r-1}$ depends only on the cardinality of this subset.

For a single spacing X_j, it follows that

$$\begin{aligned}
P(X_j \geq k) &= \binom{n+1-k}{r}^{+} \Big/ \binom{n}{r}, \quad k \geq 1, \\
P(X_j = k) &= \binom{n}{r}^{-1}\left[\binom{n+1-k}{r}^{+} - \binom{n-k}{r}^{+}\right] = \binom{n-k}{r-1}^{+} \Big/ \binom{n}{r}, \quad k \geq 1, \\
E(X_j) &= \sum_{k=1}^{n+1-r} P(X_j \geq k) = \frac{n+1}{r+1}, \\
\operatorname{Var}(X_j) &= 2 \cdot \sum_{k=1}^{n+1-r} kP(X_j \geq k) - \frac{n+1}{r+1} - \left(\frac{n+1}{r+1}\right)^2 = \frac{(n+1)r(n-r)}{(r+1)^2(r+2)}.
\end{aligned} \tag{6.2}$$

Note that $P(X_j = k) = 0$ if $k > n+1-r$.

For the smallest spacing S, we have

$$\begin{aligned}
P(S \geq k) &= \binom{n-(r-1)(k-1)}{r}^{+} \Big/ \binom{n}{r}, \quad k \geq 1, \\
P(S = k) &= \left[\binom{n-(r-1)(k-1)}{r}^{+} - \binom{n-(r-1)k}{r}^{+}\right] \Big/ \binom{n}{r}, \quad k \geq 1, \\
E(S) &= \binom{n}{r}^{-1} \sum_{k=1}^{\left[\frac{n-1}{r-1}\right]} \binom{n-(r-1)(k-1)}{r}^{+}
\end{aligned}$$

(see also Kennedy and Cooper [1]).

Finally,

$$P(L \geq k) = \binom{n}{r}^{-1} \sum_{\nu=1}^{r-1} (-1)^{\nu-1} \binom{r-1}{\nu} \binom{n-\nu(k-1)}{r}^{+}, \quad k \geq 1,$$

$$P(L = k) = \binom{n}{r}^{-1} \sum_{\nu=1}^{r-1} (-1)^{\nu-1} \binom{r-1}{\nu} \left[\binom{n-\nu(k-1)}{r}^{+} - \binom{n-\nu k}{r}^{+} \right], \quad k \geq 1,$$

$$E(L) = \binom{n}{r}^{-1} \sum_{k=1}^{n-(r-1)} \sum_{\nu=1}^{r-1} (-1)^{\nu-1} \binom{r-1}{\nu} \binom{n-\nu(k-1)}{r}^{+}.$$

Note that $P(L = k) = 0$ if $k > n - r + 1$ and $P(S = k) = 0$ if $k > (n-1)/(r-1)$.

Remark: In addition to $X_1(t), \ldots, X_{r-1}(t)$, one could introduce the spacings $X_0(t) = t_1$ and $X_r(t) = n + 1 - t_r$. By an obvious modification of the "shrinking argument," it is readily seen that (6.1) remains valid for the larger range $1 \leq m \leq r+1$, $0 \leq j_1 < j_2 < \cdots < j_m \leq r$ which entails the exchangeability of $X_0, X_1, \ldots, X_r$.

Since $\sum_{j=0}^{r} X_j = n+1$, it follows that

$$n+1 = E\left(\sum_{j=0}^{r} X_j \right) = \sum_{j=0}^{r} E(X_j) = (r+1) \cdot E(X_j)$$

which gives a second derivation of (6.2). Moreover, from the equality

$$0 = \operatorname{Var}\left(\sum_{j=0}^{r} X_j \right) = \sum_{j=0}^{r} \operatorname{Var}(X_j) + \sum_{\substack{j=0 \\ j \neq k}}^{r} \sum_{k=0}^{r} \operatorname{Cov}(X_j, X_k)$$

and exchangeability, we obtain the covariance

$$\operatorname{Cov}(X_j, X_k) = -\frac{1}{r} \operatorname{Var}(X_j), \quad 0 \leq j \neq k \leq r,$$

and thus the correlation coefficient

$$\rho(X_j, X_k) = -\frac{1}{r}, \quad 0 \leq j \neq k \leq r,$$

which is a generalization of (3.2).

Finally, redefining S and L as to include the spacings X_0 and X_r, the expressions for the distribution and expectation of S resp. L continue to hold if each "$r-1$" is replaced by "$r+1$" [of course, $\binom{n}{r}$ remains unchanged].

REFERENCE

1. R. E. Kennedy & C. N. Cooper. "The Statistics of the Smallest Space on a Lottery Ticket." *The Fibonacci Quarterly* **29.4** (1991):367-70.

AMS Classification Numbers: 60C05, 05A99, 60E05

❖❖❖