SOME COMBINATORICS OF LOTTERIES

Shayle R. Searle

Biometrics Unit, New York State College of Agriculture and Life Sciences

Cornell University, Ithaca, N. Y.

BU-1169-M July 1992

ABSTRACT

Details are given of some of the combinatorics associated with lotteries where tickets have r numbers chosen from the first n integers.

INTRODUCTION

Recent years have seen an increasing spread of various forms of lotto throughout the individual states of the U.S.A.; and in overseas countries, too. Almost all versions of these lotteries are adaptations of the Genoese lottery discussed by Bellhouse (1991). Earlier origins, back some two thousand years, can be traced to the the Han Dynasty in China, suggests Morton (1990). Whatever the origins, the current forms ask a player to pick r (usually 6) numbers from the first n (often 54) integers. Prizes are determined by the r numbers shown on the r ping pong (or other) balls selected at random from an urn containing such balls numbered 1 through n. We call those r numbers the draw, and refer to such a lottery as an r/n lottery. Then a player's ticket wins the Jackpot if its r numbers are the same as the draw. (If several players are winners they share the prize equally.)

In addition, for at least some values of $w < r$, tickets having exactly w of their r numbers coinciding with w of the numbers in the draw also win prizes. These prizes are less (usually much less) than the Jackpot won by having all r numbers the same as those of the draw. And in many lotteries, some of these prizes also depend on some other random drawing as well; e.g., a "supplementary" or $(r + 1)^{th}$ number or a drawing from a pack of 52 playing cards.

What is interesting about current lotteries is that quite a range of values are used for n, from 39-54. Although the probability of winning a Jackpot is, in all cases, extremely small (or at least so it appears to me), the relative magnitude of the probabilities can be greatly affected by n. For example, the probability of winning in a 6/54 lottery is 12.8% of that in a 6/39 lottery: put another way, the probability of a ticket winning the Jackpot is almost 8 times as large in a 6/39 lottery as in a 6/54 lottery. Of course, both probabilities are small: 1 out of 3,262,623 in a 6/39 lottery and 1 out of 25,827,165 in a 6/54 lottery.

THE r/n LOTTERY

The number of possible draws in a r/n lottery is the number of combinations of r things chosen from n, which we represent by

$$ {}^{n}C_{r} = \frac{n!}{r!(n-r)!} \, . \tag{1}$$

In most lotteries, a ticket consists of r different numbers, the same number of numbers as in the draw. In the U.S.A., \$1 buys one such ticket in some states, and two such tickets in others. In the latter, the published odds are usually given on a \$1 (i.e., two-ticket) basis, and in most such lotteries every ticket that is the same as the draw shares equally in the Jackpot. To simplify discussion, and particularly for comparing different lotteries, we shall consider all probabilities on a per-ticket basis, thus ignoring the costs of a ticket and the possibility of multiple sales of tickets bearing the same numbers. Then from (1) the probability of a ticket containing the draw, which we shall denote by $P_J(r, n)$, is

$$ P_J(r, n) = \text{probability of a ticket winning the Jackpot} = 1 \Big/ {}^{n}C_{r} = r!(n-r)! \Big/ n! \, . \tag{2}$$

A ticket having the same r numbers as the draw wins the big prize, the Jackpot. But tickets having exactly $w(< r)$ of their r numbers occurring in the r numbers of the draw also win prizes, for a limited number of values of w, such as $r-1$, $r-2$ and $r-3$. And the prizes are much smaller than the Jackpot. For a given w, the number of possible tickets having exactly w numbers occurring in the draw shall be denoted $N_w(r, n)$. Each ticket has ${}^{r}C_{w}$ sets of w numbers. For a given value of w, a ticket will be a winning ticket if the draw contains exactly w of the numbers that are on the ticket, together with $r-w$ numbers that are *not* on the ticket, chosen from the $n-r$ such numbers available.

The number of such possible choices is ${}^{n-r}C_{r-w}$. Therefore, as is implicit in Bellhouse (1991),

$$N_w(r, n) = {}^rC_w \; {}^{n-r}C_{r-w} \; ; \tag{3}$$

and clearly $N_r(r, n) = 1$. Thus, for $w \le r$, the probability of a ticket having exactly w numbers occurring in the draw is, for $w = r, r-1, \cdots, 2, 1, 0$,

$$P_w(r, n) = N_w(r, n) \Big/ {}^nC_r = {}^rC_w \; {}^{n-r}C_{r-w} \Big/ {}^nC_r \, . \tag{4}$$

Notice that for $w = r$ this is $P_J(r, n)$ of (2). And, of course, we can observe the well known result that

$$\sum_{w=0}^{r} N_w(r, n) = \sum_{w=0}^{r} {}^rC_w \; {}^{n-r}C_{r-w} = {}^nC_r \, , \tag{5}$$

so that, as one would expect,

$$\sum_{w=0}^{r} P_w(r, n) = 1 \, .$$

An algebraic proof of (5), in contrast to the familiar combinatoric proof, is that ${}^rC_w$ is the coefficient of x^w in $(1 + x)^r$ and ${}^{n-r}C_{r-w}$ is the coefficient of x^{r-w} in $(1 + x)^{n-w}$. Therefore $N_w(r, n)$ is the coefficient of x^r in $(1 + x)^n$ when the latter is factored as $(1 + x)^w \, (1 + x)^{n-w}$. Hence summing over all values of w gives ${}^nC_r$, the coefficient of x^r in $(1 + x)^n$. An extension of this method of proof is shown in the Appendix.

A recurrence relationship between $N_w(r, n)$ and $N_{w-1}(r, n)$ is, from (3)

$$N_{w-1}(r, n) = \frac{w(n-2r+w)}{(r-w+1)^2} N_w(r, n) \qquad \text{for } w = r, r-1, \cdots, 1, 0 \, . \tag{6}$$

This, along with $N_r(r, n) = 1$, provides convenient computing formulae for $P_w(r, n)$. Thus for $r = 6$ we have $N_w(6, n)$ for $w = 6, 5, \cdots, 0$ as

$$\begin{aligned}
&N_6(6, n) = 1 && N_3(6, n) = \tfrac{4}{9}(n - 8)[N_4(6, n)] \\
&N_5(6, n) = 6(n - 6) && N_2(6, n) = \tfrac{3}{16}(n - 9)[N_3(6, n)] \\
&N_4(6, n) = 1.25(n - 7)[N_5(6, n)] && N_1(6, n) = .08\,(n - 10)[N_2(6, n)] \\
& && N_0(6, n) = \tfrac{1}{36}(n - 11)[N_1(6, n)] \, .
\end{aligned} \tag{7}$$

And for r = 5 we have $N_w(5, n)$ as

$$N_5(5, n) = 1 \qquad N_2(5, n) = \tfrac{1}{3}(n - 7)[N_3(5, n)]$$

$$N_4(5, n) = 5(n - 5) \qquad N_1(5, n) = \tfrac{1}{8}(n - 8)[N_2(5, n)] \tag{8}$$

$$N_3(5, n) = (n - 6)[N_4(5, n)] \qquad N_0(5, n) = .04(n - 9)[N_1(5, n)]$$

All this is the basis of the probabilities for most (if not all) of the state lotto games in the U.S.A. They all have the characteristic that a player's ticket must have exactly the same number of numbers on it as does the draw. A variation on this is to allow tickets to have fewer numbers.

TICKETS HAVING FEWER NUMBERS THAN THE DRAW

In the preceding discussion of the r/n lottery the players' tickets always contain the same number of numbers as is in the draw, namely r. A variation on this is to have tickets that can consist of $t < r$ numbers. Then such a ticket will be said to share the Jackpot if all of its t numbers are among the r numbers of the draw. Notice that now the Jackpot is shared because, for $t < r$ there can be more than one ticket with t numbers occurring in the draw. Denote the possible number of such tickets in a r/n lottery by N(t, r, n). Then, with an argument similar to that preceding (3), no matter which particular r numbers constitute the draw, the t numbers on a ticket will share the Jackpot if the other $r - t$ numbers of the draw are chosen from the $n - t$ numbers that are not in the ticket. And the number of such possible choices is ${}^{n-t}C_{r-t}$. Hence

$$N(t, r, n) = {}^{n-t}C_{r-t}\ . \tag{9}$$

Although (9) is derived on the basis of $t < r$, it also applies to $t = r$, for then $N(r, r, n) = 1$. And so for $t \leq r$

$$\begin{aligned} P_J(t, r, n) &= \text{probability that a ticket of } t(\leq r) \text{ numbers} \\ &\qquad \text{will share the Jackpot} \\ &= {}^{n-t}C_{r-t} \big/ {}^{n}C_r\ ; \end{aligned} \tag{10}$$

and for t = r it is clear that $P_J(r, r, n) = 1 \big/ {}^nC_r = P_J(r, n)$ of (2).

An alternative derivation of (10) is that for the r numbers of the draw there are ${}^rC_t$ possible tickets for sharing the Jackpot. And since the total number of possible tickets is ${}^nC_t$

$$P_J(t, r, n) = {}^rC_t/{}^nC_t . \tag{11}$$

The equivalence of (10) to (11) is that

$$\begin{aligned} {}^{n-t}C_{r-t} \big/ {}^nC_r &= \frac{(n-t)!}{(r-t)!(n-r)!} \, \frac{r!(n-r)!}{n!} \\ &= \frac{r!}{(r-t)!} \, \frac{(n-t)!}{n!} \\ &= \frac{r!}{(r-t)!t!} \, \frac{(n-t)!t!}{n!} \\ &= {}^rC_t \big/ {}^nC_t . \end{aligned} \tag{12}$$

And from (12)

$$P_J(t, r, n) = \frac{r(r-1) \cdots (r-t+1)}{n(n-1) \cdots (n-t+1)} \tag{13}$$

as shown in Bellhouse (1991), where it is attributed to Euler.

Comparable to the recurrence relationship (6) for $N_w(r, n)$ we have, for N(t, r, n) of (9)

$$N(t-1, r, n) = \frac{n-t+1}{r-t+1} N(t, r, n) . \tag{14}$$

Hence, with N(r, r, n) = 1 and for r = 6, we get from (14)

$$N(6, 6, n) = 1 \qquad N(3, 6, n) = \tfrac{1}{3}(n-3)[N(4, 6, n)]$$

$$N(5, 6, n) = n - 5 \qquad N(2, 6, n) = \tfrac{1}{4}(n-2)[N(3, 6, n)]$$

$$N(4, 6, n) = \tfrac{1}{2}(n-4)[N(5, 6, n)] \qquad N(1, 6, n) = \tfrac{1}{5}(n-1)[N(2, 6, n)]$$

$$N(0, 6, n) = \tfrac{1}{6} n \, [N(1, 6, n)] = {}^nC_6 .$$

Comparable to (5) we can also note that

$$\sum_{t=1}^{r} N(t, r, n) = {}^nC_{r-1} \tag{15}$$

and

$$\sum_{t=0}^{r} N(t, r, n) = {}^nC_r + {}^nC_{r-1} = {}^{n+1}C_r . \tag{16}$$

Proof of (15) and (16) is given in the Appendix.

THE KENO VARIATION

A game known as Keno, and available in several states, is that of having tickets with $t \leq r$ numbers on them with winning tickets being those having $w \leq t$ numbers occurring in the draw. Denote by $N_w(t, r, n)$ the number of possible tickets that have w of their t numbers occurring among the r numbers of the draw which, as usual, has been chosen from integers 1, 2, $\cdots$, n. In again using an argument similar to that of deriving (3) we now have, for given w, ${}^tC_w$ winning tickets for tickets of size t (i.e., t numbers). And for each to be a winning ticket the draw must have $r - w$ numbers chosen from the $n - t$ numbers that are not on the ticket. Therefore

$$N_w(t, r, n) = {}^tC_w \, {}^{n-t}C_{r-w} \, . \tag{17}$$

A recurrence formula for this, analogous to (6), is

$$N_{w-1}(t, r, n) = \frac{w(n-t-r+w)}{(t-w+1)(r-w+1)} N_w(t, r, n) \tag{18}$$

which does, of course, reduce to (6) when $t = r$. Special cases of (17) are shown in Table 1.

TABLE 1

Special cases of $N_w(t, r, n)$ of (17)

Special case:	$t = r$	$w = t$	$w = t = r$
$N_w(t, r, n)$ of (17):	$N_w(r, n)$ of (3)	$N(t, r, n)$ of (9)	1

And, for each value of t

$$\sum_{w=0}^{t} N_w(t, r, n) = \sum_{w=0}^{t} {}^tC_w \, {}^{n-t}C_{r-w} = {}^nC_r \tag{19}$$

as may be proven by exactly the same argument as used following (5).

From (17) and (19) we then have, comparable to (4),

$$\begin{aligned} P_w(t, r, n) &= \text{probability that a ticket of } t(\leq r) \text{ numbers will} \\ &\qquad \text{have } w(\leq t) \text{ numbers occurring in the draw} \\ &= N_w(t, r, n) \big/ {}^nC_r \\ &= {}^tC_w \, {}^{n-t}C_{r-w} \big/ {}^nC_r \, . \end{aligned} \tag{20}$$

For t = r this is $P_w(r, n)$ of (4), as one would expect; for w = t it is $P_J(t, r, n)$ of (10) and for w = t = r it is $P_J(r, n)$ of (2). And for calculating (19) we have from (20)

$$P_{w-1}(t, r, n) = \frac{w(n-t-r+w)}{(t-w+1)(r-w+1)} P_w(t, r, n) \ . \tag{21}$$

An alternative derivation of (20) similar to deriving (11), is that a ticket can win with exactly w of its numbers in ${}^rC_w$ ways; and for each, the ticket must also have t − w numbers selected from n − r that are not in the draw. This can be done in ${}^{n-r}C_{t-w}$ ways. Therefore, for given w, the number of winning tickets is ${}^rC_w \, {}^{n-r}C_{t-w}$. But the possible number, for all w = 0,⋯,t, is ${}^nC_t$. Hence

$$P_w(t, r, n) = {}^rC_w \, {}^{n-r}C_{t-w} \Big/ {}^nC_t \ .$$

That this equals (20) is so because

$$\begin{aligned} {}^rC_w \, {}^{n-r}C_{t-w} \Big/ {}^nC_t &= \frac{r!}{w!(r-w)!} \ \frac{(n-r)!}{(t-w)!(n-r-t+w)!} \ \frac{t!(n-t)!}{n!} \\ &= \frac{t!}{w!(t-w)!} \ \frac{(n-t)!}{(r-w)!(n-r-t+w)!} \ \frac{r!(n-r)!}{n} \\ &= {}^tC_w \, {}^{n-t}C_{r-w} \Big/ {}^nC_r \ , \end{aligned}$$

which is (20).

OTHER CONSIDERATIONS

Another whole set of combinatoric considerations is that of establishing the number of tickets in a r/n lottery that can contain different patterns of consecutive numbers; 6 consecutive numbers, or exactly 5 consecutive numbers and one other number, and so on. Morton (1987) deals with this in some detail.

REFERENCES

Bellhouse, D. R. (1991). The Genoese lottery. *Statistical Science* 6, 141-148.

Morton, R. H. (1990). Lotto, then and now. *The New Zealand Statistician* 25, 15-21.

Morton, R. H. (1989). Statistics and the Gaming and Lotteries Act. *New Zealand Law Journal*, October 1989, 365-367.

Morton, R. H. (1987). Comments on the game of lotto. *The New Zealand Statistician* 22, 39-45.

APPENDIX

Proof of (15) is as follows.

$$
\begin{aligned}
\sum_{t=1}^{r} N(t, r, n) &= \sum_{t=1}^{r} {}^{n-t}C_{r-t} \\
&= \sum_{t=1}^{r} \text{coefficient of } x^{n-r} \text{ in } (1+x)^{n-t} \\
&= \text{coefficient of } x^{n-r} \text{ in } \left[(1+x)^{n-1} + (1+x)^{n-2} + \cdots + (1+x)^{n-r}\right] \\
&= \text{coefficient of } x^{n-r} \text{ in } (1+x)^{n-r}\left[(1+x)^{r-1} + (1+x)^{r-2} + \cdots + 1\right] \\
&= \text{coefficient of } x^{n-r} \text{ in } (1+x)^{n-r}\left[(1+x)^{r} - 1\right] \Big/ x \\
&= \text{coefficient of } x^{n-r} \text{ in } \left[(1+x)^{n} \big/ x - (1+x)^{n-r} \big/ x\right] \\
&= \text{coefficient of } x^{n-r} \text{ in } \left[(1+x)^{n} \big/ x\right] \\
&= {}^{n}C_{n-r+1} = {}^{n}C_{r-1}\,.
\end{aligned}
$$

Proof of (16) is easy:

$$
\begin{aligned}
{}^{n}C_{r} + {}^{n}C_{r-1} &= \frac{n!}{r!(n-r)!} + \frac{n!}{(r-1)!(n-r+1)!} \\
&= \frac{n!}{r!(n-r+1)!}(n - r + 1 + r) \\
&= {}^{n+1}C_{r}\,.
\end{aligned}
$$