Sam C. Saunders, and **Michael E. Moody**
Department of Pure and Applied Mathematics
Washington State University
Pullman, Washington 99164-2930

Great Expectations: or Playing the Odds in the State Lotteries

Abstract

Within the past quarter century there has been a proliferation of state sponsored lotteries in the United States of America. Legislatures regard lotteries as a successful means of imposing a voluntary tax on a citizenry that has become resistant to increased taxation on property, income, sales, or service. Since its introduction in 1982, the Washington State Lottery has returned $300 million to the state general fund. Because by law 40 percent of the total monies wagered must be awarded to the bettors and the same amount to the general fund, an easy calculation gives the total wager by persons in the hope of becoming a millionaire. To evaluate the actual chances of winning using some of the currently accepted betting systems a mathematical model is developed which describes the stochastic behavior of the common state lottery sytems. The model is compared with the classical occupancy problem, from which asymptotic results are derived for the distribution of the number of jackpot winners. Results for the expected number of repeated numbers in successive lotteries are computed as a means of checking the touted betting strategy of playing "hot" numbers. The true expected value of the lottery is determined, and the results are compared with data from the Washington State Lottery. It is found that:

1. As the jackpot increases, the expected return does not significantly change because increased participation raises the likelihood of multiple winners with the payoff split among them.
2. The pattern and cluster systems do not offer a basis for increased expectation of success because the winning combination is itself randomly generated (and carefully monitored) and thus the sequence of winning numbers will obey the unintuitive laws of probability and not those of some "system."
3. The observed persistence of winning numbers is an expected artifact of true random number generation; but which numbers will be repeated cannot be known beforehand so does not offer any basis for making predictions.
4. Our results on expectation may not be persuasive to those individuals whose personal utility function is convex, instead of linear, that is a person for whom a few dollars spent on lottery tickets means virtually nothing.
5. Our results should not be regarded as moralistic fervor against play, since the lottery is fairly run, its participation is voluntary, and it has a higher expected return than many other gambling games with which it competes.

Introduction

The first modern state-sponsored lottery was initiated in New Hampshire in 1964. It proved to be so successful that, as of 1986, twenty-two states, and the District of Columbia, have instituted lotteries which now generate about $16 billion in business annually. Herein we shall concentrate principally upon the Washington state game, *Lotto,* although our results also apply to similar lotteries, such as in New York, Oregon, Maryland, or Delaware where the prizes are split among the winners.

The *Lotto*-style games are played as follows: For a fee (usually $1) the player selects two six-number combinations without repetitions from the first μ whole numbers. The value of μ is adjusted to the population of the state; in New York $\mu = 48$, in Washington $\mu = 44$, in Oregon 42, in Maryland 40, while in Delaware $\mu = 30$. Thus in any lottery of this type there are

$$m = \binom{\mu}{6} = \frac{\mu!}{6!(\mu-6)!}$$

different combinations possible. In Washington $m = 7{,}059{,}052$ and in New York $m = 12{,}271{,}512$. Each week the State randomly selects the winning combination. In Washington's *Lotto* game the largest jackpot is shared equally by all persons selecting winning combinations; a smaller jackpot is shared by all who have selected 5 out of the 6 and another is shared by those selecting 4 of 6. In all of these games the salient feature is that the jackpots accumulate if not won, and thus there is, intermittently, an extremely large jackpot which invariably draws increased participation from among the betting citizens. For example, in Washington on March 1, 1986 there was an $8.5 million jackpot that was shared by two winners. The previous jackpot, which started growing on February 8, 1985, was not won at $1 million; there was no winner the following week so the prize rose to $2 million; there was no winner the next week (new jackpot $4 million) and the jackpot then increased to $8.5 million which was then claimed by two winners. Since the introduction of the current version of *Lotto* there have been, as of this writing, 37 weeks without winners out of 62 weeks of play, and 8 of the 25 times the jackpot has been won it was shared by at least two winners.

As with many stochastic games, it is quite easy to arrive at erroneous or naîve expectations for one's prospects. This is not surprising since frequently probabilistic behavior will test the intuition of even the trained investigator. For example, the fact that many of the larger state lottery jackpots have been claimed by multiple winners has been observed by many players of the lottery. A person of our acqaintance, who recently won about $1000, gives the following advice:

> Pick an unusual number combination, one that someone else will not likely select. Thus, if you win, you will not have to divide the prize.

The reasoning is that the winning number will be chosen at random and you are equally likely to win whether you have chosen an unusual combination or not. Thus if you follow this advice your expected winnings will be higher. However, we are left with the difficulty of identifying "unusual" numbers.

Although the "systems" employed by individuals are undoubtedly numerous and to lesser or greater degrees inventive, perhaps the best known advocate of a systematic approach to better the odds is Gail Howard of New York. She is a former stock broker and commodities trader, who at one time was among the nation's top commodities forecasters. Her system (marketed in several publications[1]) consists of monitoring "hot" numbers, and by tracking "randomness which forms patterns that are predictable to a certain extent." In many respects her approach to the lottery is much the same as that of an economic forecaster. She claims to "ferret-out" winning numbers by charting the past numbers which have won (*Washington Times.* August 12, 1986, Section B, Page 1):

> About half of all winning numbers have hit within the previous three games; about two-thirds of all winning numbers have hit within the previous five games; and about 87 percent of all winning numbers have hit within the last ten games. So if you play recent winning numbers you increase your chances of winning.

In a recent year, it is reported (*loc. cit.*) that she won 72 second and third place prizes in the various lotteries using her strategy. However, it was not reported how the prizes were distributed between the two places, nor how much was invested to have acquired such a record.

One of the purposes of this note is to analyze some of these popular strategies from a probabilistic framework. We also point out that the frequent accumulation of large jackpots shared by multiple winners, although perhaps not intuitive, is in fact to be expected. Further, we calculate the expected fraction of time there will be no winner or multiple winners in each drawing as a function of the level of participation. The expected fraction of winning numbers that will have been repeated in preceding drawings is also determined, and compared to Howard's empirical evidence. Finally, we utilize all this information and calculate the true expected value of the lottery to a player, and then formulate this into the best advice possible to increase one's expected return.

Expected Return

Whether to play a game of chance and risk losing one's bet should depend upon the expected payoff of the game. This is, of course, related to the odds of winning and the system of rewards. In the Washington *Lotto* two 6-number combinations are selected for each $1 ticket, either of which has the same chance of matching all 6 numbers; it is

$$p_6 - \binom{44}{6}^{-1} \doteq 1.417 \times 10^{-7}.$$

For comparison this is about the same chance as getting a straight-flush followed by three-of-a-kind in two successive hands at poker; or of obtaining heads on each of 23 consecutive tosses of a fair coin; or of getting 11 of 13 spades in a hand of bridge. None of these will occur frequently.

Prizes are also given for matching 5 out of 6, 4 out of 6 and two free Lotto tickets are given

[1]Mrs. Howard offers at least two publications giving advice on how to play the state lotteries. Among her offerings are: *Lottery Buster,* and for the Canadian lottery, *Lottery Advantage.* Her many articles (e.g., Howard, 1984; 1985a,b) are also illuminating. She is the most prominent lottery forecaster, and many tens of thousands of individuals subscribe to her methodology.

if one matches 3 out of 6. Using the hypergeometric probabilities we calculate

$$p_5 = \binom{6}{5}\binom{38}{1}/\binom{44}{6} \doteq \frac{1}{30961},$$

$$p_4 = \binom{6}{4}\binom{38}{2}/\binom{44}{6} \doteq \frac{1}{670},$$

$$p_3 = \binom{6}{3}\binom{38}{3}/\binom{44}{6} \doteq \frac{1}{42}.$$

However the amount to be won in each category varies from week to week, at the discretion of the Gaming Commission, depending upon the outcome of the preceding week's lottery. Let us assume some typical values, namely that the payoff is \$1 million, \$750, and \$30, respectively for 6, 5, and 4 matches. Since in Washington's *Lotto* each dollar purchase yields two 6-number combinations, the value (or expected payoff) ν of the game must satisfy the equation

$$2(10^6 p_6 + 750 p_5 + 30 p_4 - \nu p_3) - 1 = \nu, \tag{1}$$

assuming, for simplicity, that the value in the next game will be the same as in the present one. Performing the arithmetic yields $\nu = -0.655$ in dollars. Thus for each dollar spent the expected loss is about 65 cents. Remember that in a fair game the expected loss is, by definition, zero. But if the grand prize increases, to \$3 million, the value becomes positive. If we were to scale each prize by 3 in eqn. (1) we obtain the value to be $\nu = \$1.34$.

If the jackpot value becomes large, say to \$8 million, not taking into account any other payoff, one sees the value would be

$$8 \times 10^6 (2.83 \times 10^{-7}) - 1 = \$1.26.$$

We might conclude from this calculation that, even though the chance of winning is slight, the game is nonetheless advantageous: the more one bets the more one expects to win. Thus, the odds would seem to favor the players whenever the pot becomes large!

This simple, and erroneous, reasoning seems to be at least partially responsible for the vastly increased participation in *Lotto* whenever the jackpot reaches about \$3 million. However, one's enthusiasm, regardless of how large the prize, should take into account:

- except in the Canadian lottery, the present value of the prize is only about half the nominal value since it is paid in an annuity over 20 years, and
- it is more likely, as participation increases, that the prize will be split among several winners thus reducing the value of the pot to each winner.

The exact mathematical calculation of this last contingency is difficult, and is necessary for a true appraisal of the expected gain to a *Lotto* player. We will return to this point after developing an accurate mathematical model for the game.

A Mathematical Model for *Lotto*

Consider all different number combinations possible in a state lottery as *cells* numbered 1, 2, ... , m. Assume that one of these cells is chosen at random with each purchase of a lottery ticket. These numbers are chosen with replacement, i.e. a given number combination has the same probability of being chosen at each successive drawing whether previously selected or not. This is certainly the case for the player who uses the pseudo-random numbers generated by machine through the method called *Fast-Pic* in Washington or analogous systems elsewhere. There are, to be sure, a number of players who bet "lucky" numbers, birthdays, hunches, etc. which, strictly speaking, may not be random, but this percentage is small and their selections are not likely to vitiate our assumption of randomness.

If this assumption is true then playing a lottery is exactly analogous to tossing a ball at random into one of m cells, each one of which has the same probability of being selected. To calculate the probability that a certain number of cells remain empty after a given number of balls have been so distributed is a difficult combinatorial exercise known as the *Classical Occupancy Problem* (Feller, 1968).

Let X_0 denote the random number of empty cells when n balls have been tossed into m cells. (X_0 corresponds to the number of combinations that have not been selected by any lottery player.) The exact distribution of X_0 is derived in Feller *loc. cit.* It is given for $k = 0, 1, \ldots, n$ by

$$\Pr[X_0 = k] = p(k : n, m) = \binom{m}{k} \sum_{j=0}^{m-k} \binom{m-k}{j} \left(\frac{m-k-j}{m}\right)^n. \tag{1}$$

Unfortunately this probability is neither simple nor computationally tractable except for very small n, m.

However an ingenious partial result on both the expected value and the asymptotic behavior of X_0 was given by R. von Mises (1939), who reasoned as follows: If n balls are tossed into m cells some of the cells will have no balls in them, some one, some several. We say a cell has i-

occupancy if it contains i balls, for $i = 0,1,\ldots,n$. Let X_i denote the random number of cells with i-occupancy. Now it is clear that

$$\sum_{i=0}^{n} X_i = m, \quad \text{the number of cells} \tag{2}$$

and

$$\sum_{i=0}^{n} iX_i = n, \quad \text{the number of balls.} \tag{3}$$

Consider the indicator variable

$$X_{ij} = \begin{cases} 1 & \text{if the } j^{th} \text{ cell has } i \text{ occupants} \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Now $X_i = \sum_{j=1}^{m} X_{ij}$. Consider the number of balls in the first cell. After n tosses have been made this number will be a binomial random variable since each toss has the same probability, *viz.* $1/m$, of going in the first cell. Thus from the definition

$$\Pr[X_{i1} = 1] = \binom{n}{i}\left(\frac{1}{m}\right)^{i}\left(1 - \frac{1}{m}\right)^{n-i}, \quad i = 0,1,\ldots,n. \tag{5}$$

By symmetry each cell has the same probabilistic behavior and so by (4) and (5)

$$E[X_i] = mE[X_{i1}] = m\Pr[X_{i1} = 1] = m\binom{n}{i}\frac{(m-1)^{n-i}}{m^n}. \tag{6}$$

If we define the *factorial polynomial* of degree r for any variate x as

$$(x)_r = \prod_{j=1}^{r}(x+1-j) \tag{7}$$

then it becomes possible, by a somewhat more involved argument given by von Mises *loc. cit.*, to obtain the general r^{th} factorial moment of X_k,

$$E[(X_k)_r] = \frac{(m)_r n!(m-r)^{n-rk}}{(n-rk)!(k!)^r m^n} \tag{8}$$

for n, k, r non-negative integers such that $n \geq rk$, $r \leq m$. The distribution of X_k for $k \geq 1$ is known; it is given in [5], but is quite complicated (like that of X_0), and suffers from the same computational difficulty.

Now, for example, we can easily obtain the mean and variance of X_0, namely,

$$E[X_0] = m\left(\frac{m-1}{m}\right)^n \tag{9}$$

and

$$\begin{aligned} \mathrm{Var}[X_0] &= E[(X_0)_2] + E[X_0] - (E[X_0])^2 \\ &= E[X_0]\left[1 + (m-1)\left(\frac{m-2}{m-1}\right)^n - m\left(\frac{m-1}{m}\right)^n\right]. \end{aligned} \tag{10}$$

The asymptotic distribution of X_0 was found by von Mises (1939) under certain conditions, Theorem 1 *If* $m, n \to \infty$ *so that* $me^{-n/m} \to \lambda$ *(finite) then the distribution of* X_0 *becomes, in the limit, a Poisson distribution with mean* λ, *i.e., for any* k = 0,1, ...

$$p(k : n, m) - \frac{e^{-\lambda}\lambda^k}{n!} \to 0. \tag{11}$$

Thus for large values of m, n for which $me^{-n/m}$ is moderate the distribution of X_0 can be adequately approximated by a Poisson distribution. Table 1 (after Feller, 1968) displays the approximation.

But if we are interested in the probability that $X_0 = 0$ for other values of m, n then (1) gives

$$p(0 : n, m) = \sum_{j=0}^{m}(-1)^j\binom{m}{j}\left(\frac{m-j}{m}\right)^n. \tag{12}$$

TABLE 1. Poisson Approximation to $p(k : n, 1000)$, the Probabilities of Finding Exactly k Empty Cells When n Balls are Randomly Distributed Into 1000 Cells

		$p(k : n, 1000)$							
		k							
n	λ	0	1	2	3	5	7	9	11
5000	6.74	0.0012	0.0080	0.0269	0.0604	0.1371	0.1482	0.0935	0.0386
5500	4.09	0.0167	0.0685	0.1400	0.1909	0.1596	0.0636	0.0148	0.0023
6000	2.48	0.0838	0.2077	0.2575	0.2128	0.0655	0.0096	0.0008	
6500	1.50	0.2231	0.3347	0.2510	0.1255	0.0141	0.0008		
7000	0.91	0.4027	0.3661	0.1666	0.0506	0.0021			
7500	0.55	0.5777	0.3163	0.0873	0.0162	0.0003			
8000	0.34	0.7126	0.2406	0.0414	0.0049				
8500	0.20	0.8187	0.1637	0.0164	0.0011				
9000	0.12	0.8869	0.1064	0.0064	0.0003				

This calculation is not feasible even by machine for n, m very large. Notice that if $n = m \ln(m/\lambda)$ as $m \to \infty$ then not only does

$$m\left(\frac{m-1}{m}\right)^n \to \frac{\lambda^k}{k!} \quad \text{but also} \quad p(0:n,m) \to e^{-\lambda}$$

since

$$\binom{m}{k}\left(\frac{m-k}{m}\right)^n \to \frac{\lambda^k}{k!} \quad \text{as } m \to \infty.$$

But in the circumstance that $m \to \infty$ with $\nu = n/m$ fixed then each term of (12), except the first, diverges. A simple method of calculating this probability is yet to be found in this case.

Unfortunately the von Mises criterion cannot always be realistically met for lotteries. Consider the case in the New York lottery where $m \doteq 1.2 \times 10^7$ and $n \doteq 2 \times 10^7$; then $\lambda \doteq 2.27 \times 10^5$, an impractical value for a Poisson approximation. Moreover in all state lotteries convergence is appropriately modelled by

$$n \sim m\nu \quad \text{for some fixed } \nu \text{ as } m \to \infty, \tag{13}$$

in which case the distribution of X_0 does not converge to a Poisson. However, this is not the impression conveyed by Feller, *loc. cit.*, (p. 105)

> In practice we may use the (Poisson as an) approximation whenever (our number) m is great; for moderate values of m an estimate of the error is required . . . but we shall not enter into it.

One might conclude this if one only examined eqn. (9), since for m as small as 10, $E[X_0] \doteq me^{-n/m}$ for any value of n; but examining eqn. (10) one sees otherwise. We state

Theorem 2 *If* $\nu = n/m$ is fixed for n,m large then

$$E[X_0] \simeq me^{-\nu} \tag{14}$$

and

$$\mathrm{Var}[X_0] \cong me^{-\nu}\left(1 - (1+\nu)e^{-\nu}\right). \tag{15}$$

Proof: The first claim is obvious since under these conditions

$$\lim_{m\to\infty}\left(\frac{m-1}{m}\right)^{m\nu} = e^{-\nu}.$$

From eqn. (10) the quantity in square brackets becomes

$$\begin{aligned}[\cdots] &= 1 - \left(\frac{m-1}{m}\right)^n \left[m - (m-1)\left(1 - \frac{1}{(m-1)^2}\right)\right] \\ &= 1 - \left(1 - \frac{1}{m}\right)^n \left\{1 + (m-1)\left[1 - \left(1 - \frac{1}{(m-1)^2}\right)^n\right]\right\}. \end{aligned} \tag{16}$$

By setting

$$m - 1 = \frac{1}{t}, \quad \kappa(t) = \nu\left(1 + \frac{1}{t}\right),$$

the term in curly braces in eqn. (16) becomes in the limit

$$\lim_{m\to\infty}\{\cdots\} = 1 + \lim_{t\downarrow 0}\left[\frac{1-(1-t^2)^{\kappa(t)}}{t}\right]. \tag{17}$$

Applying L'Hopital's rule in (17) and using the expansion $-\ln(1 - t^2) = t^2 + o(t^2)$ we infer

$$\lim_{t\downarrow 0}[\cdots] = \lim_{t\downarrow 0}\left[\frac{2t\kappa(t)}{1-t^2} - \kappa'(t)\ln(1-t^2)\right] = \nu,$$

which gives the result claimed.■

For the distribution of X_0 to converge to the Poisson it is necessary that the variance converge to the mean and we see in this circumstance that it will not. But the case $\nu = n/m$ fixed for n,m large is exactly the case we are interested in when we consider state lotteries. We note that virtually nothing more than the moments of X_k for $k \geq 1$, as given in eqn. (8), is known about the exact distribution, since computation is so difficult.

Let us normalize X_k by considering $U_k = X_k/m$, the fraction of cells having k-occupancy among the m. Let $\nu = n/m$ be the fraction of tickets sold per combination, called the *relative participation*. From eqn. (6) we then see, as $m \to \infty$, that

$$E[U_k] = \left(\frac{m-1}{m}\right)^n \frac{\nu^k}{k!}\prod_{j=1}^{k}\left(\frac{1+\frac{1-j}{m\nu}}{1-\frac{1}{m}}\right) \to \frac{e^{-\nu}\nu^k}{k!}. \tag{18}$$

These asymptotic values yield an obvious interpretation with eqn. (2) and (3).

Moreover, one may check that

$$\mathrm{Var}[U_k] \to 0 \quad \text{as } m \to \infty.$$

We display this convergence for $k = 2$, $\nu = 2$ in Table 2.

TABLE 2. The variance of U_2 for various m and $\nu = 2$

m					
10	10^2	10^3	10^4	10^5	10^6
3.373×10^{-2}	4.089×10^{-3}	4.164×10^{-4}	4.171×10^{-5}	4.172×10^{-6}	4.172×10^{-7}

We assert that for realistic values of n and m in a state lottery the random U_k are virtually deterministic, and for all practical purposes the fractions

$$\rho_k(\nu) \equiv \mathrm{E}[U_k] \doteq \frac{e^{-\nu}\nu^k}{k!} \quad \text{with } \nu - \frac{n}{m} \tag{19}$$

give the probability there will be k winners sharing the grand prize in the lottery as a function of ν. This behavior is illustrated in Table 3 for a typical value of the relative participation $\theta = 3/2$:

TABLE 3. The expected proportion for various occupancy numbers, $\nu = 1.5$.

Occupancy number (k)	0	1	2	3	≥4
Expected proportion (E[U_k])	0.223	0.335	0.254	0.126	0.065

Thus if the *Lotto* game sells n = 10.5 million tickets for which $m = \binom{44}{6} \doteq 7 \times 10^6$ combinations are available, *there is nearly 1 chance in 4 that no person will obtain a 6-match and win the jackpot.* This seems surprisingly high to most people.

Usually fewer tickets are sold; suppose only 3.5 million. Then we have $\nu = 1/2$ and Table 4 reveals that there is now a 60 percent chance of no jackpot winner.

TABLE 4. Expected proportion when $\nu = 0.5$.

Occupancy number (k)	0	1	≥2
Expected proportion (E[U_k])	0.607	0.303	0.09

To see the behavior of the probability of various multiplicities as a function of the betting participation ν we plot in Figure 1 ϱ_0, ϱ_1, and ϱ_+

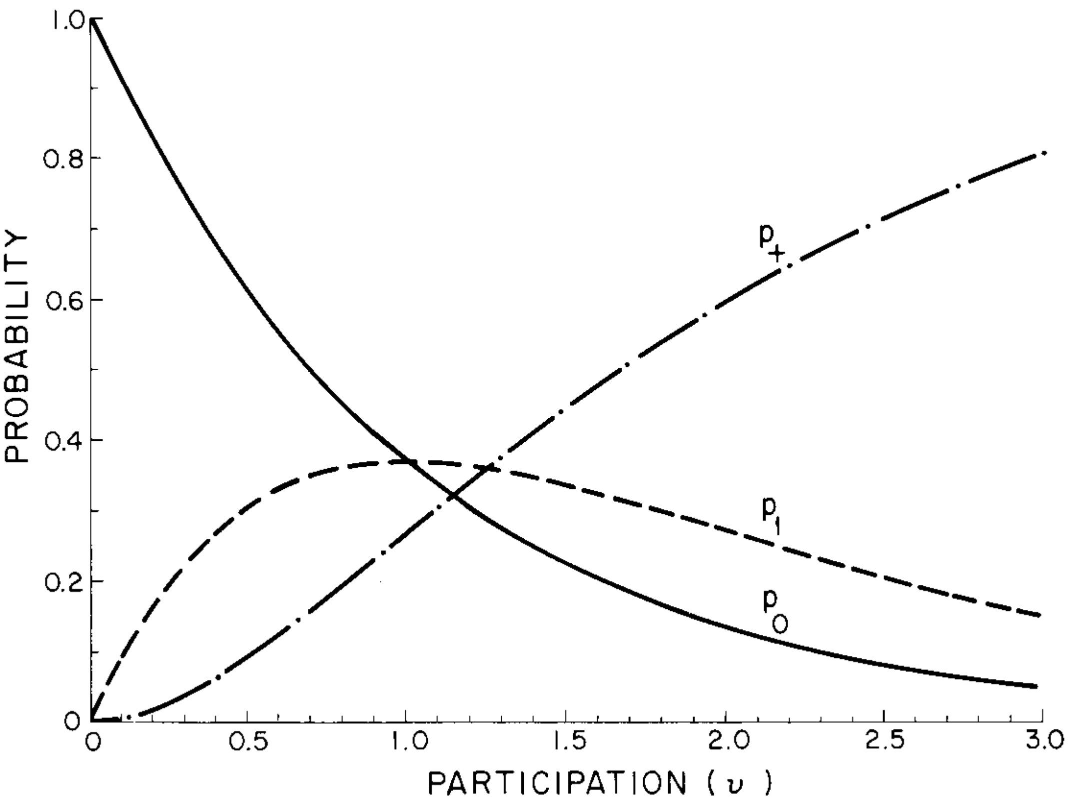

Figure 1. The probability of no winner (ϱ_0), exactly one winner (ϱ_1) or multiple winners ($\varrho_+ = \varrho_2 + \varrho_3 + \cdots$), versus the relative participation θ. The probability that there is no winner, so that the jackpot increases, is surprisingly high even with large relative participation.

$= \varrho_2 + \varrho_3 + \cdots$ (the chance of multiple winners of the jackpot).

The Persistence of Winning Numbers

As we have noted, Howard's system is based upon the empirical claim that certain numbers among those that have been chosen as winning lottery combinations remain "hot": these numbers reoccur with extraordinarily high frequency, far beyond what could be explained by a random model. Some of her startling empirical frequency records have been previously quoted.

We will now show, surprisingly enough, that the fractions she gave are in virtual accord with the predictions of a probabilistic model for the percentage of reoccurrences *of one particular state lottery.* (Although the state upon which her data is based is not specified, such data clearly could not be true for all lotteries independent of μ.)

The winning combination is a subset of size $\zeta = 6$ drawn at random without replacement (by State authorities) from the set of the first μ natural numbers. Recall that the number of such combinations is $m = \binom{\mu}{6}$. Assume that winning combinations are drawn repeatedly for $\omega \geq 1$ times. Label these sets of winning numbers $W_1, \ldots, W_\omega$, and record the *span* $K_\omega = \cup_{i=1}^{\omega} W_i$, the set of all different winning numbers obtained in the first ω weeks. A final winning combination, say W_0, is drawn. The repeated set is

$$R = W_0 \cap K_\omega,$$

i.e. the set of numbers in the final winning combination that have appeared in the preceding ω week's drawing. We ask for the expected value of the size of R, call it $R^\#$. (Here $R^\#$ means the cardinality of the set.) By the theorem on conditional expectations

$$E[R^\#] = E\left\{E[R^\# \mid K^\#_\omega]\right\}, \qquad (1)$$

and a little reasoning shows that

$$\Pr[R^\# = r \mid K^\#_\omega = k] = \binom{k}{r}\binom{\mu - k}{\zeta - r}\Big/\binom{\mu}{\zeta}.$$

This is a hypergeometric distribution which has known means $\zeta k/\mu$. Hence the conditional expectation in eqn. (1) is

$$E[R^\# \mid K^\#_\omega] = \frac{\zeta K^\#_\omega}{\mu}. \qquad (2)$$

Now consider the random set $K_\omega = K_{\omega-1} \cup W_\omega$, and thereby the conditional random variable defined by

$$[K^\#_\omega \mid K^\#_{\omega-1} = k_{\omega-1}] = k_{\omega-1} + k \quad \text{w.p.} \quad \binom{k_{\omega-1}}{\zeta - k}\binom{\mu - k_{\omega-1}}{k}\Big/\binom{\mu}{\zeta}.$$

By taking expectations we have

$$E[K^\#_\omega \mid k_{\omega-1}] = k_{\omega-1} + \frac{\zeta}{\mu}(\mu - k_{\omega-1}) - \zeta + k_{\omega-1}\left(1 - \frac{\zeta}{\mu}\right).$$

By analogy we have for $j = 0, \ldots, -2$

$$E[K^\#_{\omega-j} \mid k_{\omega-j-1}] = \zeta + k_{\omega-j-1}\left(1 - \frac{\zeta}{\mu}\right).$$

Thus by iteration

$$E[K^\#_\omega \mid k_{\omega-2}] = \zeta + \left(1 - \frac{\zeta}{\mu}\right)\left[\zeta + k_{\omega-2}\left(1 - \frac{\zeta}{\mu}\right)\right]$$

$$= \zeta + \zeta\left(1 - \frac{\zeta}{\mu}\right) + k_{\omega-2}\left(1 - \frac{\zeta}{\mu}\right)^2.$$

Continuing in like manner we have

$$E[K^\#_\omega \mid k_{\omega-(\omega-1)}] = \zeta\sum_{j=0}^{\omega-2}\left(1 - \frac{\zeta}{\mu}\right)^j + k_1\left(1 - \frac{\zeta}{\mu}\right)^{\omega-1}$$

and after simplification we find the conditional expectation to be

$$E[K^\#_\omega \mid K^\#_1] = \mu\left[1 - \left(1 - \frac{\zeta}{\mu}\right)^{\omega-1}\right] + K^\#_1\left(1 - \frac{\zeta}{\mu}\right)^{\omega-1}.$$

But since $E[K^\#_1] = E[W_1] = \zeta$, we obtain for the unconditional expectation

$$E[K^\#_\omega] = \mu\left[1 - \left(1 - \frac{\zeta}{\mu}\right)^{\omega-1}\right] + \zeta\left(1 - \frac{\zeta}{\mu}\right)^{\omega-1}.$$

Hence from (2), after simplification, we have the remarkably simple result which we state in

Theorem 3 *A sample of size ζ is drawn without replacement from the set of the first μ natural numbers. This is repeated $\omega + 1$ times. The set of repeated numbers is R, i.e. the set of numbers in the last sample which had occurred in the preceding ω samples. The number of elements in R is called $R^\#$ and it has the expectation*

$$E[R^\#] = \zeta\left[1 - \left(1 - \frac{\zeta}{\mu}\right)^{\omega}\right] \doteq \zeta\left[1 - e^{-(\zeta\omega/\mu)}\right]. \qquad (3)$$

This method of derivation actually yields the distribution of $R^\#$, more than is required for the expectation. But from the simplicity of $E[R^\#]$, as shown in (3), an easy explanation presents itself. The expectation is merely the sample size ζ times the probability of any fixed fraction ζ/μ not being covered in ω independent trials.[2] We have the immediate

Corollary 1 *If the samples of size ζ are drawn with replacement for $(\omega + 1)$ times, the expected number of repetitions in the last sample among the preceding ω is*

$$E[R^\#] = \zeta\left[1 - \left(1 - \frac{1}{\mu}\right)^{\zeta\omega}\right] \doteq \zeta\left[1 - e^{-(\zeta\omega/\mu)}\right].$$

In order to see the values of the fraction of repetitions we tabulate $E[R^\#/\zeta]$ for various ω and μ with $\zeta = 6$ in Table 5.

---

[2]The authors are indebted to a referee for this observation.

TABLE 5. The fraction of repetitions with $\zeta = 6$.

		ω						
	μ	3	4	5	6	10	11	15
NY	48	0.330	0.414	0.487	0.551	0.737	0.769	0.865
WA	44	0.356	0.444	0.519	0.585	0.769	0.801	0.889
OR	42	0.370	0.460	0.537	0.603	0.786	0.817	0.901
MD	40	0.386	0.478	0.556	0.623	0.803	0.833	0.913
VT	38	0.403	0.497	0.577	0.643	0.821	0.849	0.924
DE	30	0.488	0.590	0.672	0.738	0.893	0.914	0.965

Thus for any state with $\mu \leq 38$, as it was formerly in Oregon, the expected persistence of winning numbers agrees with sufficient accord with the empirical fractions reported by Howard *loc. cit.* But they do not agree so closely with states with $\mu > 38$. However, whatever the state, it is clear that the information gleaned from charting the behavior of past winning combinations cannot be used to better one's odds of winning. By analogy, sequentially and independently tossing a fair coin naturally leads to runs of heads and tails; but the observation of a run of heads will not lead us reasonably to expect that the chance of a head will be anything other than one-half on any future toss.

The True Expectation of the *Lotto* Player

We will now compute the expected gain of a player taking into account the increased chance of multiple winners with higher relative participation. Let us assume that if the jackpot for a particular week's lottery has a present value of P dollars the ensuing relative participation will be ν, *i.e.*, sales of $n = \nu m$ tickets. We opine that ν will generally increase with P. Indeed, this is observed. Here, for simplicity, we shall be concerned only with the jackpot, which in *Lotto* is usually about 10 times the amount shared among 5-match and 4-match winners.

The probability that a particular ticket will be a 6-match winner of the jackpot is $p_6 = 1/m$. Given that this ticket is a winner, its owner's share of the prize will depend upon the occupancy of that number combination by the other players. If their play results in k-occupancy, the fraction of possible combinations of which is given by eqn. III(18), then with probability $e^{-\nu}\nu^k/k!$ the share will be $P/(k + 1)$. Hence the player's expected share of the prize is

$$\frac{1}{m}\sum_{k=0}^{\infty}\frac{P}{k+1}\frac{e^{-\nu}\nu^k}{k!} = \frac{P}{m\nu}(1 - e^{-\nu}). \tag{4}$$

Thus we have

Theorem 4 *In Lotto the expected value of the jackpot when the prize has present value P is*

$$\frac{P}{m\nu}(1 - e^{-\nu}) \quad \textit{per ticket,} \tag{5}$$

where

$$m = \binom{\mu}{6}$$

and $n = m\nu$ is the number of tickets sold.

Proof: Another way of looking at eqn. (5) is instructive. Each ticket sold must have the same expected value; there are n tickets sold. The probability the prize will not be claimed is $e^{-\nu}$. Thus the expected gain g in dollars to the player for each ticket (tickets in *Lotto* are 2 per dollar) is

$$g = \frac{P\left(1 - e^{-n/m}\right)}{n} - \frac{1}{2} \tag{6}$$

which is merely a rewriting of eqn. (5). ■

In order to obtain an idea of possible fluctuations in the player's expected gain for different values of the jackpot J which determine in some stochastic manner the relative participation, we make a few calculations for the Washington lottery.

Case 1: A prize of J = \$3 million attracted buyers of $n = 6.6 \times 10^6$ numbers. Since $\nu = 0.93$, with $m \doteq 7 \times 10^6$ we find from eqn. (3) the expected gain $g = -0.224$.

Case 2: $J = 2 \times 10^6$, with $n = 4.6 \times 10^6$, so that $\nu = .65$ and thus $g = -0.292$.

Case 3: $J = 5 \times 10^6$ produced $n = 8.7 \times 10^6$, thus $g = -0.093$.

Case 4: $J = 6 \times 10^6$, with $n = 13.1 \times 10^6$, hence $g = -0.114$.

It is not surprising that the expected gain is always negative but the relatively small fluctuation in the gain regardless of the size of the jackpot is not intuitive to most people. But we see that because of bigger prizes the consequent publicity and increased participation *does not allow the player's gain to increase significantly, if at all.* This is contrary to the usual expected value calculations done in Section II, and which seem to govern betting patterns.

Suppose that some knowledge of the distribution of ticket-sales is obtained, determined empirically as a function of the size of the proffered jackpot. The distribution of participation could then be controlled through the proper choice of μ. (Recently Oregon has changed μ from 38 to 42). Thus, in a sense, the return to the state can be optimized by balancing the expected gain between the two alternatives of offering prizes that are large which return too much to the bettors, or offering prizes too small to elicit the volume of ticket sales necessary. Were the distribution of sales, as a function of jackpot size, known *a priori* this calculation of expected gain would be possible as well as the variance of the gain.

To illustrate this point we exhibit in Figure 2 a plot of the relationship between the nominal jackpot prize (usually paid in a 20-year annuity) and the consequent participation for 62 consecutive weeks in the Washington State Lottery. Note that if the present value of the prize were plotted, instead of the nominal value, the ordinate of each point would be reduced by about one-half, depending upon interest rates at the time the prize was awarded. This latter way is how the state regards the relationship between prize and participation.

The utility of the nominal prize may be different for each bettor, depending upon age or taxable income, and could change depending upon the size of the prize. In general the utility function seems to be convex for most people, who prefer to wager one dollar at astronomical odds in the hope of winning a million dollars rather than wager one dollar at the corresponding odds

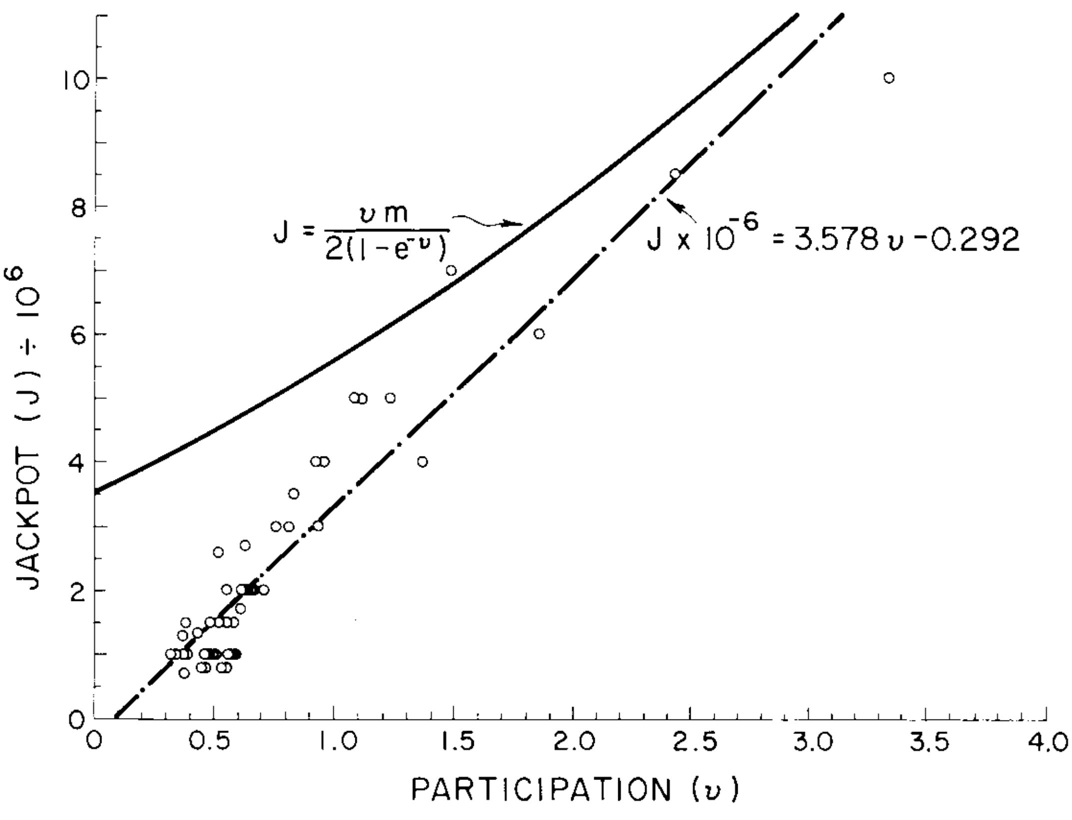

Figure 2. The scattered circles (○) plot the jackpot (in millions) versus relative participation for the first 61 weeks of the Washington *Lotto*. The solid curve is a plot of the function $2J = \nu m(1 - e^{-\nu})^{-1}$, the relationship between the jackpot and the relative participation which would give zero for the expected nominal gain ($g = 0$). For one week only, the expected gain to the player was positive (the single point above the curve); for this week it "made sense" to play the game. The line plots the linear regression of P on ν.

in order to win a thousand dollars. This was clearly the case in the New York state lottery, when a $41 million prize caused a "feeding frenzy" of participation by effecting sales of 80 million tickets in New York and surrounding states.

Discussion

As with many stochastic phenomena, the behavior of the various lottery games is often counter-intuitive. However, the observed "patterns" to these games are in fact consistent with the simplest null model of independent and equally-likely outcomes. For example, our analysis of the "pick-6" lottery games reveals that the weekly pattern of winning (*i.e.* no winner, one winner, multiple winners), though somewhat surprising, is easily explained. The empirical pattern of coincidences of lottery numbers in successive weeks, proposed by some observers as a useful "strategic" device in forming one's lottery selections, is also in accord with our (purely random) model of the lottery. The occurrences of "clusters" or "runs" is a salient feature of random processes, and their presence should not be misconstrued as offering a means to divine the future. This misconception seems to lie behind the most common schemes for advancing the prospects of lottery players. However, we state categorically that which we all know in our minds to be true: None of the various empirical (or metaphysical) systems for choosing lottery numbers will improve a player's chances of winning in a fair lottery. This sage advice is consonant with that presented by Packel (1981) and Adler (1985), which have not been "best sellers." This is in marked contrast to the evangelistic fervor of Howard (1984, 1985a,b) for the persistence of winning numbers, whose publications sell very well.

We conclude with the remark of a legislator of an earlier day that "lotteries are a tax on fools" and the sage advice "the only sure way to double one's money is to fold it once and put it back in one's pocket."

Literature Cited

Adler, W. 1985. *The Lottery Book.* Quill/William Morrow, Fairfield, New Jersey.

Feller, W. 1968. An *Introduction to Probability Theory and Its Applications.* Third Edition. John Wiley, New York.

Howard, G. 1984 State lotteries: how to get in it. *Gambling Times.* Issue 12, p. 78.

Howard, G. 1985a. How to handicap the Lotto numbers. *Gambling Times.* Issue 2, p. 78.

Howard, G. 1985b. Handicapping the Lotto numbers—Part II. *Gambling Times.* Issue 3, p. 78.

Mises, R. V. 1939. Über Aufteilungs- und Besetzungswahrscheinlichkeiten. Revue de la Faculté des Sciences de l'Université d'Istanbul, N.S., 4:313-334.

Packel, E. 1981. *The Mathematics of Games and Gambling.* The Mathematical Association of America. Mathematical Association of America, Washington, D.C.

Received 29 January 1987
Accepted for publication 26 May 1987