Optimal play for multiple lottery wins

Richard Stong and Skip Garibaldi

IDA Center for Communications Research-La Jolla
4320 Westerra Ct
San Diego, CA 92121, U.S.A.

`rstong@ccrwest.org`, `skip@garibaldibros.com`

Abstract

We determine the optimal strategy for a family of lottery games involving repeated drawings with the same conditions, which includes Brazil's jogo do bicho and games typically offered by state lottery organizations in the United States under names such as Daily 4 or Cash 3. The proof that the strategy is optimal and the resulting formula for the probability of success both rely on a solution to a recursion that generalizes the usual Pascal recursion for binomial coefficients, which itself relies on a count of lattice paths. We illustrate how our result can be applied to focus on-the-ground investigations of suspicious patterns of lottery wins.

Mathematics Subject Classifications: 60G40, 60C05

1   Introduction

Motivation and context

Most US states operate a lottery and most of them will provide, if asked in the right way, a list of the prizes (of at least some minimum size, often $600) they have given out. The list will include additional information about each prize awarded, such as the game played and the name and city of residence of the person receiving the prize. It is rather rare for someone to win a prize that would appear on such a list. For a typical scratcher game, a gambler might expect to spend $30,000 on lottery tickets to claim one such prize. Yet the lists of prizes awarded in some states include persons who have claimed more than 100 prizes. It is natural to wonder if such a person is truly a legitimate gambler (the null hypothesis) or is up to some kind of possibly-illegal shenanigans.

With this question in mind, the paper [1] developed a method to provide a lower bound on how much the person would have had to spend on lottery tickets under the null, in which case they win prizes in lottery games with a probability that is deducible from the rules of the games. For some specific people, it was found that they had to have spent

more than \$1000 per day on lottery tickets over a period of years [7]. Spending that much on lottery tickets is time consuming, raising doubts about the null hypothesis. Moreover, it is easy to then check the null hypothesis with "shoe leather": under the null, the person was spending a very large amount of money over a long period of time on lottery tickets, so store clerks and others will remember them and have stories to share.

The mathematical analysis saves labor by focusing on-the-ground investigations. Following the first application of the techniques from [1] in [7], they have been used to analyze frequent prize winners in at least 15 other states, including a governmental analysis in South Carolina [9]. Investigations have led to jail time for some frequent winners, see for example [3]. It has also buttressed court cases concerning various shenanigans, in the state of Georgia and elsewhere.

This paper's contribution

We address here a missing piece of the analysis, concerning prizes won in games with repeated drawings such as California's Daily 4 or Florida's Play 4 or Brazil's jogo do bicho ("animal game"). These games are very similar to the "numbers" and "policy" games that predate the current generation of state-run lotteries in the US, see e.g. [2, Ch. VI] or [11]. The gambler bets on one or more 4-digit numbers. (Of course, versions with 3 digits, etc., also exist.) Later, at a pre-scheduled time, the winning number is revealed. If the gambler had chosen to bet on the winning number, then he or she would win a prize. Because each instance of the game, each drawing, has the same odds of winning and the same payoff, we call these games *repeated draw games*. These are in contrast with instant-win or scratcher games, where the gambler immediately learns whether they have won, and whose outcomes are well-approximated by the binomial distribution.

Our two main contributions are to answer the following about repeated draw games:

- Given a total amount M to wager and a number k, *what betting strategy maximizes the probability of achieving k wins?* See Theorem 5.1.

- *What is the probability of achieving at least k wins with this strategy?* See Proposition 4.1.

In §2, we give a precise statement of an abstracted version of the problem and define the strategy that will be proved to be optimal. Section 3 can be read independently of the rest of the paper; it gives a general result about a weighted version of the Pascal recursion for binomial coefficients, whose proof relies on counts of lattice paths. The solution to the recursion is then used in the proofs of each of the main results (§4 and §5).

The optimal strategy we describe here is *adaptive*, in the sense that the outcome of the next drawing will influence what the strategy says to do for the following drawing. In contrast to this, a naive strategy is to wager 1 unit per drawing for M drawings, in which case the probability of success is a straightforward application of the binomial distribution as in §6. We observe that the optimal strategy has at most twice the probability of success as the naive strategy. The final section of the paper, §7, illustrates how this can be applied to analyzing frequent prize claimants.

2 A strategy and its probability of success

The precise problem statement

Consider a game where in each round you draw one card (with replacement) from a deck of N equally likely cards. On any round of the game, you can bet on any subset of the cards, at a cost of 1 coin per card. You win if the drawn card is any of the cards in your chosen set.

You will wager up to M coins (your "bankroll") in total across all rounds of the game, and you want to get at least k wins. There is no benefit to you to getting more than k wins, and anything less than k wins counts as a failure. We ignore any monetary payout for a win.

Note that after each round you are back at another instance of the same game, with your bankroll lowered by your previous bets and your goal number of wins lowered by the number of wins you have already achieved. Therefore to describe the optimal strategy it suffices to describe your first move for any pair (M, k); these are the parameters that are specific to the gambler.

An adaptive strategy

Here is a strategy described in terms of three cases. (We will prove below in Theorem 5.1 that this strategy achieves the optimal probability of success.)

I. If $M \geqslant kN$, then you bet on all N possibilities in the next draw. (Note that this will lower M by N and also guarantee a win, so k will decrease by 1. Thus you will be back in this same case on the next round, and you will be guaranteed success.)

II. If $(k-1)N < M < kN$, then you bet on $M - (k-1)N$ possibilities in the next draw. (Note that if you win, you will be left with $(k-1)N$ coins needing $k-1$ more wins. Hence you will be in case (I) and guaranteed success.)

III. If $M \leqslant (k-1)N$, then bet on one card in the next round.

Here is another way of describing this strategy. Say that you are ridiculously rich if your bankroll exceeds N times the number of desired wins. You are ridiculously rich if you could buy out some full number of draws, guaranteeing you win, and still have money left over at the end. The basic strategy is to always bet on one card, with one exception. If you look ahead and see that a win would make you ridiculously rich, then you increase the number of cards you bet on (to $M - (k-1)N$) to maximize the chance that this happens in the current round.

Probability of success

Let $p_k(M)$ be the probability that you achieve success with the strategy described above. Then the description of the cases shows that $p_0(M) = 1$ for all M and

$$p_k(M) = \begin{cases} 1 & \text{if } M \geqslant kN, \\ \frac{M-(k-1)N}{N} + \frac{kN-M}{N}p_k((k-1)N) & \text{if } (k-1)N < M < kN, \\ \frac{1}{N}p_{k-1}(M-1) + \left(1 - \frac{1}{N}\right)p_k(M-1) & \text{if } k \leqslant M \leqslant (k-1)N \\ 0 & \text{if } 0 \leqslant M < k. \end{cases} \tag{2.1}$$

The first three cases represent (I)–(III) and the last says that you cannot succeed if your bankroll is less than the number of wins you want to achieve.

Example 2.2. When $k = 1$, if we are in case (II), then we bet on M cards in the first round. Note that this is optimal since regardless of how we bet the expected total number of wins is $\frac{M}{N}$ and this is therefore an upper bound for the probability of getting at least one win. We have:

$$p_1(M) = \begin{cases} \frac{M}{N} = 1 - \frac{N-M}{N} & \text{if } 0 \leqslant M \leqslant N \\ 1 & \text{if } N \leqslant M. \end{cases}$$

3 Interlude: a weighted Pascal recursion

This section may be read independently of the rest of the paper. It provides a key relation that we will use in a few different places in the rest of the paper.

Suppose we have a family $B_1, B_2, \ldots$ of polynomials that satisfy the recursion

$$B_k(M) = \frac{1}{N}B_{k-1}(M-1) + B_k(M-1) \tag{3.1}$$

for some constant N, with the convention that $B_k = 0$ for k less than some bound. We recognize this as a weighted version of the usual Pascal Rule for the binomial coefficients, which is the case $N = 1$ and $B_k(M) = \binom{M}{k}$.

Example 3.2. From the similarity to Pascal's Rule, we see that a family of polynomials of the form

$$B_k(M) = N^{b-k}\binom{M+a}{k-b}$$

for integer a, b is a solution. Since the recursion (3.1) is linear we can assemble these into more complicated solutions. In particular, the solution with given initial values $B_k(0)$ is given by

$$B_k(M) = \sum_{j=0}^{k} B_{k-j}(0)\binom{M}{j}N^{-j}.$$

The recursion (3.1) can be rewritten as saying that the finite difference $B_k(M) - B_k(M-1) = \frac{1}{N} B_{k-1}(M-1)$ of B_k is proportional to (a shift of) B_{k-1}. In particular, if B_j is the first nonzero polynomial on the list, then B_j is constant and B_k has degree $k-j$ for $k \geqslant j$.

Iterating the recursion in a clever way provides the following extremely useful identity.

Proposition 3.3. *With the notation above,*

$$B_k(M) = \sum_{j=1}^{\infty} c_j B_{k+1-j}(M - jN) \tag{3.4}$$

where the constant c_j is defined to be

$$c_j := \frac{(jN-1)(jN-2)\cdots(jN-j+2)}{(j-1)!\, N^{j-2}}. \tag{3.5}$$

We note that $c_1 = c_2 = 1$ and $c_3 = \frac{3N-1}{2N}$.

Proof. Applying the recursion (3.1) a second time to every term on the right side gives

$$B_k(M) = \frac{1}{N^2} B_{k-2}(M-2) + \frac{2}{N} B_{k-1}(M-2) + B_k(M-2).$$

If we keep iterating the recursion we see the binomial coefficients appear

$$B_k(M) = \sum_{j=0}^{r} N^{-j} \binom{r}{j} B_{k-j}(M-r).$$

Think of the binomial coefficient $\binom{r}{j}$ as the number of lattice paths from $(0,0)$ to $(r-j, j)$. To get a more interesting identity instead of iterating the recursion on every term on the right, we choose a subset of the terms and agree to stop iterating when we hit one of them. The specific case we want is to stop when we encounter a term of the form $B_{k+1-j}(M-jN)$ for some j, which corresponds to a lattice path ending at $(jN - j + 1, j - 1)$. The result will be a sum of the required form, where the terms c_j have not yet been determined.

Note that there is no issue with paths escaping to infinity since if j gets too large all the terms are zero anyway. As a result the sum (3.4) is not really infinite, as for each fixed k only finitely many terms are nonzero. The constants c_j (depending on N only) are N^{1-j} times a combinatorial term counting lattice paths. In this case a direct translation of the iteration above describes this combinatorial term as the number of the lattice paths that start at $(0,0)$ and have absorbing points at $(jN - j + 1, j - 1)$ that are absorbed at the j-th such point. This can be rephrased as a more standard lattice path count. Note that if we add an extra absorbing point at $(1, -1)$ (corresponding to $j = 0$), then we can think of our lattice path as starting at $(0, -1)$. The paths that leave $(0, -1)$ horizontally are immediately absorbed so they contribute nothing new. Shifting everything up so that we again start at $(0,0)$ we see that the absorbing points shift to $(j(N-1) + 1, j)$ for

$j \geqslant 0$. Next we note that a path absorbed at $(j(N-1)+1, j)$ must have ended with a horizontal move from $(j(N-1), j)$. Thus rather than absorbing points, we can just count paths that stay on or above the line $y = x/(N-1)$. Putting this together we conclude that $N^{j-1}c_j$ is the number of lattice paths starting at $(0,0)$, ending at $(j(N-1), j)$, and always on or above the line $y = x/(N-1)$. The explicit formula for c_j, then, is either a result of Spivey [10] or of Niederhausen [8], depending on taste. $\qquad\square$

4 A precise formula for $p_k(M)$

In this section we give a formula for $p_k(M)$ that is less recursive than (2.1). The probability $p_k(M)$ is most naturally thought of as breaking up into ranges, which to a first approximation are indexed by $r = \lceil M/N \rceil$. We have already done this for the case $k = 1$ (Example 2.2) and the formula is trivial for $r > k$ (case (I) with $p_k(M) = 1$), so we are left to address the cases $k \geqslant r \geqslant 1$.

Polynomials Q_k and $B_{k,r}$

First we define polynomials $Q_k(M) \in \mathbb{Q}[M]$ for $k \geqslant 2$ of degree $k-2$ via

$$Q_k(M) := \sum_{j=0}^{k-2} (k-j-1)\binom{M-k+j}{j} N^{-j},$$

i.e., $Q_k(M)$ is the Laurent series for the function

$$\left(k - 1 - \frac{M}{N}\right)\left(1 - \frac{1}{N}\right)^{k-2-M}$$

taken up to the N^{2-k} term. For example,

$$Q_2(M) = 1 \quad \text{and} \quad Q_3(M) = \frac{M-2}{N} + 2.$$

We set $Q_k := 0$ for $k < 2$.

We use the Q_k to build polynomials $B_{k,r}(M) \in \mathbb{Q}[M]$ for $k > r \geqslant 1$ of degree $k-r-1$ via

$$B_{k,r}(M) := \sum_{i=r}^{\infty} c_i Q_{k+1-i}(M - iN)$$

for c_i as in (3.5). Note that the sum has only finitely many nonzero terms and that $B_{k,r}$ is a polynomial of degree $k-r-1$. For example,

$$B_{r+1,r}(M) = c_r \quad \text{for } r \geqslant 1$$

and (as in the previous section)

$$B_{k,1}(M) = Q_k(M).$$

We define $B_{k,r}(M) = 0$ for $k \leqslant r$.

Proposition 4.1. *For* $(r-1)N \leqslant M \leqslant rN$ *and* $k \geqslant r \geqslant 1$, *we have*

$$p_k(M) = 1 - \frac{kN - M}{N} A_r + B_{k,r}(M) \left(1 - \frac{1}{N}\right)^{M+2-k} \tag{4.2}$$

where A_r *is given by*

$$A_r := 1 - \sum_{i=1}^{r-1} c_i \left(1 - \frac{1}{N}\right)^{iN-i+1}$$

with c_i *as defined in* (3.5).

Note that A_r depends only on r and N. The first few values are:

$$A_1 = 1, \quad A_2 = 1 - \left(1 - \frac{1}{N}\right)^N, \quad A_3 = 1 - \left(1 - \frac{1}{N}\right)^N - \left(1 - \frac{1}{N}\right)^{2N-1}.$$

Therefore, in case $k = 2$, the proposition asserts that

$$p_2(M) = \begin{cases} 1 - \frac{2N-M}{N} + \left(1 - \frac{1}{N}\right)^M & \text{if } 0 \leqslant M \leqslant N+1 \\ 1 - \frac{2N-M}{N}\left(1 - \left(1 - \frac{1}{N}\right)^N\right) & \text{if } N \leqslant M \leqslant 2N \\ 1 & \text{if } 2N \leqslant M. \end{cases}$$

For $k = 3$, the proposition asserts that

$$p_3(M) = \begin{cases} 1 - \frac{3N-M}{N} + \left(\frac{M-2}{N} + 2\right)\left(1 - \frac{1}{N}\right)^{M-1} & \text{if } 0 \leqslant M \leqslant N+2 \\ 1 - \frac{3N-M}{N}\left(1 - \left(1 - \frac{1}{N}\right)^N\right) + \left(1 - \frac{1}{N}\right)^{M-1} & \text{if } N \leqslant M \leqslant 2N+1 \\ 1 - \frac{3N-M}{N}\left(1 - \left(1 - \frac{1}{N}\right)^N - \left(1 - \frac{1}{N}\right)^{2N-1}\right) & \text{if } 2N \leqslant M \leqslant 3N \\ 1 & \text{if } 3N \leqslant M. \end{cases}$$

Note that we have allowed the ranges to overlap slightly here and in the statement of the proposition, as long as the formulas give the same numerical value. Indeed we will see that adjacent ranges give formulas that always agree on at least one value. We will refer to each such formula as a layer of the full formula. So for example, the formula for $p_3(M)$ consists of four layers. In general the formula for $p_k(M)$ will consist of $k+1$ layers, the last of which will just be $p_k(M) = 1$ for $M \geqslant kN$. Excluding this trivial case, the r-th layer will hold for $(r-1)N \leqslant M \leqslant rN$ and will have the functional form displayed in the statement of the proposition. One consequence of the slight overlap of the layers is that we may always assume $M-1$ and M are included in the same layer. (We have also included, for example, that the $(k-1)$-st layer and the k-th layer overlap at $M = (k-1)N + 1$, because in that situation (II) says to bet one coin on the next draw.)

Proof of Proposition 4.1. Example 2.2 shows that the claim holds when $k = 1$. For $k > r$, define $B'_{k,r}(M)$ via

$$p_k(M) = 1 - \frac{kN - M}{N} A_r + B'_{k,r}(M) \left(1 - \frac{1}{N}\right)^{M+2-k}$$

and set $B'_{r,r} = 0$. We aim to prove that $B'_{k,r} = B_{k,r}$ for $k > r \geqslant 1$.

Suppose now that $r < k$, i.e., case (III), so $p_k(M)$ satisfies the recursion

$$p_k(M) = \frac{1}{N} p_{k-1}(M-1) + \left(1 - \frac{1}{N}\right) p_k(M-1). \qquad (4.3)$$

This recursion is linear in p_k and is satisfied by the constant function 1 and by the function $kN - M$. Therefore the $B'_{k,r}$ satisfy the weighted Pascal recursion (3.1), separately for each r. The recursion (3.1) has a number of interesting features. In particular, the polynomials $B'_{k,r}(M)$ are determined inductively (in k) by the recursion, with a fairly explicit formula, if we have one value of $B'_{k,r}$ for each k.

Continue to assume $r < k$. Since $p_k(0) = 0$ for $k \geqslant 2$, we see that

$$B'_{k,1}(0) = (k-1) \left(1 - \frac{1}{N}\right)^{k-2}$$

and hence Proposition 3.3 gives

$$B'_{k,1}(M) = \sum_{j=0}^{k-2} (k - j - 1) \binom{M}{j} N^{-j} \left(1 - \frac{1}{N}\right)^{k-j-2}.$$

Note that we have omitted the last two terms in the sum since they vanish. If instead we use the initial values $p_k(k-1) = 0$ we get $B'_{k,1}(k-1) = k-1$ and hence a different (but equivalent) formula $B'_{k,1}(M) = Q_k(M)$, verifying the claim for $1 = r < k$.

For the inductive step, suppose we have proved the result for layer $r-1$. Thus we only care about $p_k(M)$ for $k \geqslant r$. For $k = r$ we are in case (II) of the strategy and therefore we have the recursion

$$p_r(M) = \frac{M - (r-1)N}{N} + \frac{rN - M}{N} p_r((r-1)N).$$

This recursion can be restated as saying that for M in the interval $[(r-1)N, rN]$ the probabilities $p_r(M)$ interpolate linearly between $p_r((r-1)N)$ at $M = (r-1)N$ and $p_r(rN) = 1$ at $M = rN$. We can also compute $p_r((r-1)M)$ from the $(r-1)$-st layer formula giving

$$
\begin{aligned}
p_r((r-1)N) &= 1 - A_{r-1} + B_{r,r-1}((r-1)N) \left(1 - \tfrac{1}{N}\right)^{(r-1)N+2-r} \\
&= 1 - A_{r-1} + c_{r-1} \left(1 - \tfrac{1}{N}\right)^{(r-1)N-r+2} \\
&= 1 - A_r.
\end{aligned}
$$

We conclude that

$$p_r(M) = 1 - A_r \frac{rN - M}{N}$$

for $(r-1)N \leqslant M \leqslant rN$ since p_r linearly interpolates between $1 - A_r$ and 1, verifying the claim for $r = k$.

For the general case, since $B'_{k,r}$ satisfies the recursion (3.1) it suffices to show that the stated formula is correct for one value of M and for this value we take $M = (r-1)N$. The equality of the $(r-1)$-st and r-th layer formulas for $p_k((r-1)N)$ says (after cancelling out equal terms and using the formula we just found for $A_r - A_{r-1}$):

$$B'_{k,r}((r-1)N)\left(1 - \frac{1}{N}\right)^{(r-1)N+2-k} = B_{k,r-1}((r-1)N)\left(1 - \frac{1}{N}\right)^{(r-1)N+2-k}$$
$$+ (k-r+1)c_{r-1}\left(1 - \frac{1}{N}\right)^{(r-1)N-r+2}$$

or

$$B'_{k,r}((r-1)N) = B_{k,r-1}((r-1)N) + (k-r+1)c_{r-1}\left(1 - \frac{1}{N}\right)^{k-r}.$$

Since we recognize that $Q_{k-r+2}(0) = (k-r+1)\left(1 - \frac{1}{N}\right)^{k-r}$, we conclude that

$$B'_{k,r}(M) = B_{k-1,r}(M) + c_{r-1}Q_{k-r+2}(M - (r-1)N) = B_{k,r}(M)$$

and hence the formula above for $B_{k,r}$ is correct. $\qquad\square$

5 Proof of optimality

Theorem 5.1. *The function $p_k(M)$ defined in (2.1) is the probability of obtaining at least k wins while betting a total of M coins (across any number of drawings), allocating bets in an adaptive manner to maximize said probability.*

In fact we will prove something slightly stronger.

Lemma 5.2. *In case* (III), *the expected return on betting on b cards for the next round (with $1 \leqslant b \leqslant N$),*
$$\frac{b}{N}p_{k-1}(M - b) + \frac{N - b}{N}p_k(M - b),$$
is a non-increasing function of b.

Proof. In formulas, the claim says that

$$\frac{b}{N}p_{k-1}(M - b) + \frac{N - b}{N}p_k(M - b) \geqslant \frac{b+1}{N}p_{k-1}(M - b - 1) + \frac{N - b - 1}{N}p_k(M - b - 1).$$

Setting $M' = M - b$ and writing this as

$$\frac{b}{N}\left(p_{k-1}(M') + p_k(M'-1) - p_k(M') - p_{k-1}(M'-1)\right)$$
$$+ \left(p_k(M') - \frac{1}{N}p_{k-1}(M'-1) - \left(1 - \frac{1}{N}\right)p_k(M'-1)\right) \geqslant 0$$

and noting that the last term is the recursion satisfied by p_k since the pair (M', k) is also in case (III), we see that proving monotonicity is equivalent to showing that

$$p_{k-1}(M') + p_k(M'-1) \geqslant p_k(M') + p_{k-1}(M'-1).$$

Dropping the primes, this amounts to showing that

$$p_{k-1}(M) + p_k(M-1) \geqslant p_k(M) + p_{k-1}(M-1)$$

as long as M is in case (III) of the strategy. Since $M - 1$ and M always share a layer, we may restrict to just a single layer of the formula for p_k, say the r-th layer. The contributions of the constant 1 and the linear terms $A_r \frac{kN-M}{N}$ to both sides of this inequality cancel. Hence we see that it suffices to show that

$$B_{k-1,r}(M)\left(1 - \frac{1}{N}\right)^{M+3-k} + B_{k,r}(M-1)\left(1 - \frac{1}{N}\right)^{M+1-k}$$
$$\geqslant B_{k,r}(M)\left(1 - \frac{1}{N}\right)^{M+2-k} + B_{k-1,r}(M-1)\left(1 - \frac{1}{N}\right)^{M+2-k}$$

or

$$(N-1)^2 B_{k-1,r}(M) + N^2 B_{k,r}(M-1) \geqslant N(N-1)(B_{k,r}(M) + B_{k-1,r}(M-1)).$$

In the particular case $r = 1$, for which $B_{k,1}(M) = Q_k(M)$, we find that the left hand side of this inequality minus the right hand side is

$$R_k(M) = \frac{(M-1)(M-2)\cdots(M-k+2)}{(k-2)!\, N^{k-3}}. \tag{5.3}$$

This can be simply ground out from the formulas for Q_k. An alternative (slightly less computational) proof is to note that

$$R_k(M) = (N-1)^2 Q_{k-1}(M) + N^2 Q_k(M-1) - N(N-1)(Q_k(M) + Q_{k-1}(M-1))$$

satisfies the same linear recursion as Q_k and we compute that $R_2(1) = N$ and for $k > 2$

$$R_k(1) = (N-1)^2\left(k - 2 - \frac{1}{N}\right)\left(1 - \frac{1}{N}\right)^{k-4} + N^2(k-1)\left(1 - \frac{1}{N}\right)^{k-2}$$
$$- N(N-1)\left(\left(k - 1 - \frac{1}{N}\right)\left(1 - \frac{1}{N}\right)^{k-3} + (k-2)\left(1 - \frac{1}{N}\right)^{k-3}\right)$$
$$= 0.$$

Hence we get (5.3). Since $R_k(M) \geqslant 0$ trivially holds, we get the $r = 1$ case of the monotonicity inequality. The inequality above is linear and invariant under shifts of M. Hence it follows from the formula for $B_{k,r}$ that the general r inequality reduces to

$$R_k(M) - c_1 R_k(M - N) - c_2 R_{k-1}(M - 2N) - \cdots - c_{r-1} R_{k-r+2}(M - (r-1)N) \geqslant 0.$$

But this is obvious since R_k (as a binomial coefficient) also satisfies the recursion (3.1) and hence the identity

$$R_k(M) = \sum_{j=0}^{\infty} c_j R_{k+1-j}(M - jN)$$

and in particular subtracting any truncation of the sum from $R_k(M)$ leaves a nonnegative remainder (since every term in the sum is nonnegative). $\qquad\square$

Proof of Theorem 5.1. Lemma 5.2 proves the optimality of $p_k(M)$ as long as we are in case (III). The optimality in case (I) when $p_k(M) = 1$ is trivial. For case (II), we note that by the monotonicity we only need to consider bets of at most $M - (k-1)N$. But in this range the quantity we wish to maximize

$$\frac{b}{N} p_{k-1}(M - b) + \frac{N - b}{N} p_k(M - b)$$

is a quadratic function of b (since $p_{k-1}(M - b) = 1$ and $p_k(M - b)$ is linear). Further the coefficient of b^2 is positive (since $p_k(M - b)$ has negative slope as a function of b). Thus the maximum is attained at the endpoints, that is at either $b = 0$ or $b = M - (k-1)N$. Note that the values at these points are equal since $b = 0$ is equivalent to a "no-op" which has the same probability of success as doing the intended move immediately. $\qquad\square$

6 An upper bound

It is natural to wonder how the probability $p_k(M)$ of success under the optimal strategy compares to the probability $b_k(M)$ of success of the naive strategy of betting one coin in each of M drawings. That is, $b_k(M) = \mathbb{P}(X_M \geqslant k)$ where X_M is a Binomial$(M, 1/N)$ random variable, so

$$b_k(M) = I_{1/N}(k, M - k + 1) \quad \text{for } M \geqslant k,$$

where I denotes the regularized Beta function

$$I_x(a, b) := \frac{\int_0^x t^{a-1}(1 - t)^{b-1}\, dt}{\int_0^1 t^{a-1}(1 - t)^{b-1}\, dt}.$$

We have:

Proposition 6.1. $b_k(M) \leqslant p_k(M) \leqslant 2b_k(M)$.

Proof. The first inequality is trivial from Theorem 5.1. For the second, we apply induction on k and M.

The base case is $k = 0$, where $p_k(M) = b_k(M) = 1$. If $M < k$, then clearly $p_k(M) = b_k(M) = 0$.

Now we assume that we have verified $p_k(M) \leqslant 2b_k(M)$ in case at least one of k or M is diminished. When $k \leqslant M \leqslant (k-1)N$, we are done by induction using (2.1). If $(k-1)N < M < kN$, we have, for $\theta = k - M/N$:

$$p_k(M) = (1 - \theta) + \theta p_k((k-1)N).$$

Now $b_k(kN) \geqslant 1/2$ as in [5] or [6]. Therefore,

$$p_k(M) \leqslant (1 - \theta)2b_k(kN) + \theta 2b_k((k-1)N).$$

As the density function of the binomial distribution (probability of k successes on x trials) is increasing for $x \leqslant kN$ (elementary), the survival function $b_k(x)$ is concave down, whence $p_k(M) \leqslant 2b_k(M)$.

For $kN \leqslant M$, we have $p_k(M) = 1$ and $2b_k(M) \geqslant 2b_k(kN) \geqslant 1$. $\qquad\square$

In one extreme case, if $M = 1$, then $b_k(1) = p_k(1)$ for all $N \geqslant 1$, so the first inequality in the proposition is sharp.

The other inequality also cannot be improved, in the sense that

$$\lim_{k \to \infty} p_k(kN)/b_k(kN) = 2.$$

To see this, note that X_{kN} is a sum of kN i.i.d. Bernoulli random variables, so $Y_k := (X_{kN} - k)/\sqrt{kN}$ converges to a mean zero Gaussian, whence $\mathbb{P}(X_{kN} \geqslant k) = \mathbb{P}(Y_k \geqslant 0)$ converges to $1/2$.

7 An application to lottery security

Let's apply the main results to lottery security, as suggested in the introduction. Suppose a person claimed k prizes in a repeated draw game where each drawing has N possible outcomes, where k and N are known. For any hypothetical amount M of money spent on tickets, one may calculate using Theorem 5.1 and Proposition 4.1 an upper bound $p_k(M)$ on the probability that the gambler won so many times. This presents three challenges.

First, it may be tiresome to compute $p_k(M)$ for large k, even in the form provided in Proposition 4.1. We address this using Proposition 6.1.

Second, this probability is only immediately relevant for detecting illicit behavior if one began by selecting that person for inspection and only later discovered the property that they were a frequent prize claimant. However, the typical situation is that the focus is on the frequent claimant because of the fact they are a frequent claimant, in which case one must adjust the probability to account for that. (See, for example, [4] for a discussion of this issue in the specific context of the lottery.)

Third, examining $p_k(M)$ has the disadvantage of involving two numbers that require the user to pick them or interpret them, namely the spending level M and a threshold for the probability.

The paper [1] addressed the second and third challenges. It argued that winning so many times has a negligible probability if, when we imagine the entire gambling population spending M per capita on tickets, the probability of at least one of them winning so many times is at most 10^{-6}. In this way, one can deduce a (rather conservative) lower bound $\underline{M}$ for the frequent winner's spending, assuming they are obtaining tickets in the usual way of buying them.

Example 7.1. Suppose someone claims 100 prizes in Play 4, a repeated draw game offered by the Florida lottery game where players bet \$1 to wager on a a 4-digit number ($N = 10^4$). The population of Florida is about 21 million. Taking $k = 100$ and solving the equation

$$2b_k(\underline{M}) = 10^{-6}/(21 \text{ million})$$

provides a lower bound $\underline{M} \approx 420{,}000$ on M. That is, if the prize claimant is obtaining tickets by buying them in the usual legal way (as opposed to engaging in shenanigans), then there is a negligible probability that they spent less than \$420,000 on tickets. In particular, if these wins were claimed over a period of one year, we would infer that a legitimate gambler who claimed so many prizes would have been spending at least \$1150 per day on lottery tickets for that year.

The above example is somewhat unusual, in that a typical frequent prize claimant will have claimed prizes in a variety of games. The paper [1] explains how to compute an overall lower bound on spending, assuming one can calculate a lower bound on spending for each individual game. Here we have provided such a bound for games consisting of repeated drawings.

References

[1] R. Arratia, S. Garibaldi, L. Mower, and P.B. Stark, *Some people have all the luck*, Math. Mag. **88** (2015), 196–211.

[2] H. Asbury, *Sucker's progress: an informal history of gambling in America*, Dodd, Mead and Co., 1938.

[3] District of Massachussetts US Attorney's Office, *Lynn man sentenced for multi-million dollar lottery ticket scam*, Press release, September 25 2019.

[4] J.A. Hanley, *Jumping to coincidences: defying odds in the realm of the preposterous*, The American Statistician **46** (1992), no. 3, 197–202.

[5] R. Kaas and J.M. Buhrman, *Mean, median and mode in binomial distributions*, Statistica Neerlandica **34** (1980), no. 1, 13–18.

[6] N. Lord, *Binomial averages when the mean is an integer*, Math. Gazette **94** (2010), no. 530, 331–332.

[7] L. Mower, *Gaming the lottery*, The Palm Beach Post, March 30 2014.

[8] H. Niederhausen, *How many paths cross at least l given lattice points?*, Proceedings of the thirteenth Southeastern conference on combinatorics, graph theory and computing (Boca Raton, Fla., 1982), vol. 36, 1982, pp. 161–173.

[9] South Carolina General Assembly Legislative Audit Council, *A review of the South Carolina education lottery and its oversight of retailers and players*, June 2019.

[10] M.Z. Spivey, *Enumerating lattice paths touching or crossing the diagonal at a given number of lattice points*, Electronic Journal of Combinatorics **19** (2012), no. 3, #P24.

[11] S. White, S. Garton, S. Robertson, and G. Whiteis, *Playing the numbers: gambling in Harlem between the wars*, Harvard University Press, 2010.