# SOME PEOPLE HAVE ALL THE LUCK

## RICHARD ARRATIA, SKIP GARIBALDI, LAWRENCE MOWER, AND PHILIP B. STARK

## Contents

1.	Warm-up: one kind of lottery ticket, estimation only	2
2.	Why Florida?	3
3.	Sanity check: getting on the list	3
4.	One kind of lottery ticket, forward bounds	4
5.	What this has to do with Joe DiMaggio	5
6.	One kind of lottery ticket, reverse bounds	7
7.	Multiple kinds of tickets, reverse bounds	8
8.	More about Louis Johnson	9
9.	The man from Hollywood	10
10.	What some people get up to	12
11.	Questions for further investigation	14
12.	Appendix: No dependent wins is almost as good as independent bets	15
13.	Appendix: Remarks on the optimization problem $(7.5)$	16
14.	Appendix: How long can a gambler gamble?	17
15.	Summary	18
Ref	erences	19

Lawrence Mower, an investigative reporter, wondered if someone knew how to "beat" the lottery. Maybe someone clever figured out how to tell whether a scratcher ticket was a winner, without scratching it, perhaps by cryptanalyzing the exposed serial number, as Mohan Srivastava is alleged to have done [Leh11]. Such a person might be easy to identify: she should show up in lottery records as having won lots of prizes in scratcher games. So Mower requested a list of all lottery prizes given out in Florida—he lives in West Palm Beach. He found something unexpected: at least 10 people had each collected more than 80 prizes worth \$600 or more. Most of them had won prizes in many different scratcher games or in a game called "Play 4."

When Mower asked Florida Lottery Secretary Cynthia O'Connell about these gamblers, she replied that they could be lucky: "That's what the lottery is all about. You can buy one ticket and you become a millionaire" [Mow14]. While we agree that it's *possible* for someone to win 80 times or more, it requires a suspicious degree of luck.

When someone wins the jackpot in a high-value lottery like Mega Millions or Powerball, we don't think anything of it: after all, usually somebody wins. But if someone were to win such a large jackpot *twice*, we would take notice.<sup>1</sup> She would

<sup>&</sup>lt;sup>1</sup>To the best of our knowledge, no one has won a Mega Millions or Powerball jackpot twice.

### 2 RICHARD ARRATIA, SKIP GARIBALDI, LAWRENCE MOWER, AND PHILIP B. STARK

certainly seem to be very lucky! And if someone were to win such a large jackpot three times, we might strongly suspect foul play.<sup>2</sup> These conclusions are intuitive and don't seem to require careful justification—the odds of winning three times are just too minuscule.

The gamblers Mower identified represent cases that are less clear cut, because they won games with more favorable odds than Mega Millions or Powerball. To assess how suspicious their winnings are requires some mathematics. We will look at specific, real-life examples of Floridian gamblers and what we know of their wins from public lottery records. Some of these gamblers may have been just lucky. For others, no plausible amount of luck can account for the winnings: there is strong statistical evidence that some gamblers are chiseling. We describe some methods of cheating that have been uncovered elsewhere, but we do not accuse any particular gambler in our study of any particular scheme.

The conclusions reached in this article were used to support the investigative report by Mower that appeared in the Palm Beach Post [Mow14]. One of our aims is to record the mathematics behind that exposé, since, as you'll see, it's too much for a newspaper article.

Acknowledgements. SG's research was partially supported by NSF grant DMS-1201542.

### 1. WARM-UP: ONE KIND OF LOTTERY TICKET, ESTIMATION ONLY

It is easiest to explain the mathematics when the gambler only buys one kind of lottery ticket, so we pretend for a moment that this is the case. Put c for the cost of a ticket and suppose each ticket has probability p of winning a prize large enough to appear in our records. From here on, a "win" means a win large enough to be recorded; for Florida, the threshold is \$600.

If a gambler buys n tickets, that costs cn and, on average, will win np times. This is intuitively obvious, and follows formally by modeling a lottery ticket as a binomial trial with probability of success p, so in n trials we expect np successes.

We don't know n, and the gambler is unlikely to tell us. But based on the calculation in the preceding paragraph, we might estimate that a gambler who won W times bought W/p tickets. Indeed, an unbiased estimate for W is

$$(1.1) \qquad \qquad \hat{n} := W/p$$

corresponding to the gambler spending  $c\hat{n}$  on tickets.<sup>3</sup> Since p is very small, like  $10^{-4}$ , the number  $\hat{n}$  is big—and so is the estimated amount spent,  $c\hat{n}$ .

**Example 1.2** (Louis Johnson 1). One Florida gambler, Louis Johnson, claimed 57 \$5,000 prizes from "straight" Play 4 tickets. A straight Play 4 ticket costs \$1 (so c = \$1) and wins \$5,000 if the player guesses the winning 4-digit number; i.e., it wins \$5000 with probability  $p = 10^{-4}$  and wins nothing otherwise. Because W = 57, we estimate that Louis played  $\hat{n} = W/p = 570,000$  tickets at a cost of \$570,000.<sup>4</sup>

 $<sup>^{2}</sup>$ As quipped by the Bond villain Goldfinger: "Once is happenstance. Twice is coincidence. The third time it's enemy action." [Art64]

<sup>&</sup>lt;sup>3</sup>This is the dollar amount handed to a clerk to buy tickets. If the gambler rolls all winnings back in to buy more tickets, she can get by with a smaller initial bankroll.

<sup>&</sup>lt;sup>4</sup>Because this bet has expected rate of return of 50% (ignoring taxes, for which see [BC03]), a gambler who rolls all winnings back into buying more tickets can hope to achieve this result with an initial bankroll of half that, just \$285,000; see Proposition 14.1 below.

Of course, a gambler could be lucky and win by buying far fewer than 570,000 tickets. After all, when someone wins the Mega Millions jackpot, which happens for any given ticket with probability about 1-in-259 million, we do not conclude that she purchased 259 million tickets.

In undergraduate statistics, a standard problem involving the Binomial distribution asks how well you can estimate p if you know W and n. Here, we want to draw conclusions about n from W and p. This question is slippery, as indicated by the suggestive titles of, e.g., [Cas86], [Kah87], and [Hal94]. Here we will use only elementary probability and not statistics.

# 2. Why Florida?

We focus on gamblers in Florida because we have great data for that state. Every prize worth at least \$600 from any Florida lottery game (from Powerball to a \$1 scratcher ticket) has to be claimed at a Florida Lottery office, and each prize claimed is public record. The record gives the date the prize was claimed, the store where the ticket was purchased, the gambler's name, the gambler's city of residence, the game, and the value of the prize. A comprehensive list is available as an electronic spreadsheet, making it easy to look for anomalies, such as individual gamblers who have claimed many prizes. Our calculations are based on a list containing all prizes claimed between January 4, 2000 and August 7, 2013.

Our lead example is Florida's Play 4, which requires guessing a 4-digit number. This popular game format, offered in most states that have a lottery, is based on the numbers or policy game formerly offered by organized crime, described in [KM70] and [Sel63]. But there are variations by state that would be tiresome to elucidate. For instance, in Florida all prizes are of fixed amounts, but in California (where it is called Daily 4) all prizes are pari-mutuel. This is another reason to focus on a single state.

## 3. SANITY CHECK: GETTING ON THE LIST

We have a list of every prize of \$600 or more claimed in the Florida lottery; we have identified some gamblers who won many such prizes (i.e., appear in the list many times); and we wonder whether they cheated. To assess the evidence, we need to compare their performance to a control: someone who does not cheat. That is, we want to test the null hypothesis that they won so often honestly. If it is extremely unlikely that their winnings resulted from fair play—under the most optimistic but fair strategy—we reject the null hypothesis and conclude that they chiseled.

So let us imagine a hypothetical gambler, Mr. Badger, who is honest but has a serious gambling problem. Badger wants to appear in the list as many times as possible. (If the lottery were a video game, we might imagine that he is hoping to win the "most \$600+ prizes" badge.)

Suppose Badger plays a single game; that his initial bankroll is worth n tickets of that game; that with probability p that game gives a unique payout over \$600 worth the cost of exactly j tickets; and that he uses any winnings to buy more tickets, continuing to play until he busts. By Proposition 14.1, the expected number of tickets Badger gets to buy before his eventual ruin is n/(1-pj): his initial bankroll of n tickets, divided by the expected loss of tickets per wager, 1-pj. The expected number of wins before he busts is np/(1-pj).

#### 4 RICHARD ARRATIA, SKIP GARIBALDI, LAWRENCE MOWER, AND PHILIP B. STARK

How much will Badger start with? Surely some gamblers have lost houses, so let us say he starts with a bankroll worth \$175,000, an amount between the median list price and the median sale price of a house in Florida [Zil14].

One game we will focus on is straight Play 4, which for \$1 pays \$5,000 with probability  $p = 10^{-4}$ . If Badger plays that game, he can expect to collect about

$$np/(1-pj) = 17.5/(1-0.5) = 35$$
 prizes.

Badger could do better by betting instead on the Play 4 6-way box: In that game, you pick two different digits. You win if the winning number contains both of those digits twice (e.g., guessing  $\{1,2\}$  wins if the winning number is any of the 6 permutations 1122, 1212, etc.).<sup>5</sup> A \$1 bet on Play 4 6-way pays \$800 with probability  $p = 6 \times 10^{-4}$ , so with this bet Badger can expect to win about

$$np/(1-pj) = 6 \times 17.5/(1-0.48) \approx 202$$
 times.

This is a big number, and it would already put him amongst the top handful of winners in the history of the Florida lottery. The expected total amount spent on lottery tickets—how much Badger actually hands over the counter to the clerk, including his initial bankroll of \$175,000 and all the prize money won—is about \$336,500. Because we are relatively confident there are gamblers who have lost this much on the lottery, we conclude the number of wins alone does not give evidence that a gambler cheated. We must take into account the particulars of the winning bets.

## 4. One kind of lottery ticket, forward bounds

If you confronted the gambler with the calculation of  $\hat{n}$  from (1.1), the gambler might quite reasonably object that she is just very lucky, and that the true number of tickets she bought, n, is much smaller than your estimate  $\hat{n}$ . Because we paid attention in our undergraduate probability class, we can calculate the exact probability D(n; w, p) that the gambler will win at least w times from buying n tickets using the binomial distribution. That is, we want to know the chance that in nindependent Bernoulli trials, each of which has probability p of success, there will be at least w successes:

(4.1) 
$$D(n; w, p) := \begin{bmatrix} \text{Probability of at least } w \text{ wins} \\ \text{with } n \text{ tickets} \end{bmatrix} = \sum_{k=w}^{n} \binom{n}{k} p^{k} (1-p)^{n-k}.$$

Alternatively, we could view this as a negative binomial distribution:

$$D(n; w, p) = \begin{vmatrix} \text{Probability that it takes } n & \text{or} \\ \text{fewer tickets to win } w & \text{times} \end{vmatrix}$$

This function is built into Matlab as nbincdf(n-w, w, p), into Mathematica as CDF[NegativeBinomialDistribution[w, p], n-w], and into the Python SciPy library as nbinom.cdf(n-w, w, p). It also can be written in terms of the *regular-ized Beta function* 

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

<sup>&</sup>lt;sup>5</sup>This bet is relatively unpopular. In 2012, about \$254.6 million worth of Play 4 tickets of all kinds were sold, and only about \$1.5 million or 0.6% were for this bet. Total Florida lottery sales were around \$4.45 billion [Lot13], so this bet accounted for about 0.03% of lottery wagers.

(4.2) 
$$D(n; w, p) = I_p(w, n - w + 1).$$

Hence we can evaluate D(n; w, p) in Matlab as betainc(p, w, n-w+1), in Mathematica as BetaRegularized[p, w, n-w+1], and in the Python Scipy library as scipy.special.betainc(w, n-w+1, p)\*scipy.special.beta(w, n-w+1).

**Example 4.3** (Louis Johnson 2). The probability that a gambler wins 57 Play 4 "straight" games by buying 175,000 tickets or fewer is

$$D(175000; 57, 10^{-4}) \approx 6.3 \times 10^{-14}.$$

For context, imagine an omniscient being picks a star in one of 40 galaxies approximately the size of our Milky Way. The probability above is about the same as the chance of randomly guessing which star the being picked: it is utterly implausible that a gambler wins 57 times by buying 175,000 or fewer tickets. Or, to put it more flippantly, "Yeah, it could happen... yeah, and monkeys might fly out of my butt" [Pic92].

Table 1 gives the chance that a gambler who buys \$5 scratcher tickets will get a \$600+ prize several times in a year by betting various sums per week.

#### 5. What this has to do with Joe DiMaggio

The computation in the last section is far from a direct answer to the question of whether Louis Johnson is lucky or up to something shady. The most glaring problem is that we have calculated the probability that an innocent gambler who buys \$175,000 of Play 4 tickets would win so many times (" $\mathbb{P}(\text{win 57 times} \mid \text{innocence})$ "). But we want to know whether someone who wins so many times is an innocent gambler, i.e., something like (" $\mathbb{P}(\text{innocence} \mid \text{win 57 times})$ ").<sup>6</sup>

Among other things, we need to check whether so many people are playing Play 4 frequently that it's reasonably likely that at least one of them would win so many times. If so, Louis Johnson might be that person, just like Powerball: no particular person has a big chance of winning, but there's a big chance *someone* wins.

And we do have to be careful about this, since the news media has publicized some lottery coincidences as astronomically unlikely, yet these coincidences have turned out to be relatively unsurprising in view of the enormous number of people playing the lottery. The most famous example is probably the woman who won the New Jersey lottery twice in four months, discussed in [DM89, esp. p. 859], [Kol90], and [SM86]. Another example is described in [Ste08], and this general subject is the focus of a chapter in the book [Han14].

We take an approach that is similar to how baseball and probability enthusiasts attempt to answer the question, *Precisely how awesome was Joe DiMaggio?* As you likely know, Joe DiMaggio is famous for having the longest hitting streak in baseball: hits in 56 consecutive games. The closest modern player is Pete Rose, who had a 44-game streak in 1978. One probabilistic approach to this is to consider

 $<sup>^{6}</sup>$ Conflating the two is an example of *The Prosecutor's Fallacy*, see e.g. [TS87]. For the lottery, we have no "prior" probability that a gambler is innocent, so the latter conditional probability is only notional.

on a typical f winning a	day (column 1) that the odds o	iven sum per	ler who bets a g	$\dots$ ) that a gamb	gainst (1-in	nate odds a or 52 weeks	Approxii	TABLE 1.
$10,\!523,\!700$	$1,185,\!430$	146,583	20,084	3,083	538	109	26	200/week
	>100,000,000	$43,\!072,\!300$	2,972,290	230,362	$20,\!354$	$2,\!090$	256	100/week
			>100,000,000	$22,\!549,\!500$	1,001,120	51,773	3,204	50/week
				> 100,000,000	56,378,800	$1,\!460,\!710$	$45,\!380$	25/week
11  wins	10  wins	9  wins	8 wins	7 wins	6  wins	5  wins	4  wins	

so scratcher game for 52 weeks would win at least 4, 5, ..., 11 times. Here we assume that the odds of winning a \$600+ prize with a given ticket are 1-in-8300 — the actual odds will depend on the precise game.

the probability that a randomly-selected player gets a hit in a game,<sup>7</sup> find the probability that there is at least one hitting streak at least 56 games long in the entire history of baseball. If a DiMaggio-sized streak is likely, then the answer to the question is "not so awesome," DiMaggio just happened to be the person who had the unsurprisingly long streak. If it is very unlikely that there would be such a long streak, then the answer is: DiMaggio was truly awesome.<sup>8</sup>

Let's use this reasoning to consider the case of Louis Johnson from Example 4.3. Suppose that N gamblers bought Play 4 tickets during the relevant time period, each of whom spent at most \$175,000. Then an upper bound on the probability that at least one such gambler would win at least 57 times is the chance of at least one success in N Bernoulli trials, each of which has probability no larger than  $p \approx 6.3 \times 10^{-14}$  of success. (Louis Johnson represents a success.) The trials might not be independent, however, because different gamblers might pick the same numbers for the same game. Nonetheless, the chance that at least one of the N players wins at least 57 times is at most Np by the inclusion-exclusion principle, which implies that for events  $A_1, \ldots, A_N$  we have  $\mathbb{P}(\cup_{i=1}^N A_i) \leq \sum_{i=1}^N \mathbb{P}(A_i)$ . What is N? Let us say it's the current population of Florida, approximately

What is N? Let us say it's the current population of Florida, approximately 19 million. Then the chance at least one person would win at least 57 times is no larger than  $19 \times 10^6 \times 6.3 \times 10^{-14} = 0.0000012$ , just over one in a million.

This estimate is crude because the estimated number of gamblers is very rough and of course the estimate is not at all sharp (it gives a lot away in a safe direction) because most people spend nowhere near \$175,00 on the lottery. We are, in fact, giving even more away because Louis Johnson won many other bets many other times (yet the total winnings are still dwarfed by the expect cash outlay). Considering all these factors, one might reasonably conclude that either Louis Johnson has a source of hidden of money—perhaps he is a wealthy heir with a gambling problem<sup>9</sup>—or he is up to something.

## 6. One kind of lottery ticket, reverse bounds

There is an artificial quality to Example 4.3 in that we picked the \$175,000 spending level almost out of thin air, based on Florida house prices as in §3. Let's avoid this by doing the same calculation in reverse.

**Example 6.1** (Louis Johnson 3). We re-do Example 4.3, but instead of starting with the spending level of \$175,000 and deducing a probability, we will start with a probability  $\varepsilon = 5 \times 10^{-14}$  and infer a spending level. If Louis buys *n* tickets, then he wins at least 57 times with probability D(n; 57, p) by equation (4.2), so we want to find the smallest *n* such that  $D(n; w, p) \ge \varepsilon$ . We can do this by solving numerically the equation

$$D(n_0; 57, 10^{-4}) = \varepsilon.$$

 $<sup>^{7}</sup>$ This is not as well connected to the facts as the lottery calculation is, because for the lottery, the randomness is part of the game, while for baseball, we have to imagine a hypothetical situation where players are selected at random.

<sup>&</sup>lt;sup>8</sup>The conclusions in DiMaggio's case have been equivocal, see the discussion in [GPS11, pp. 30–38]. There have also been allegations that DiMaggio's streak was rigged in game 30, see e.g. [Rob07].

<sup>&</sup>lt;sup>9</sup>There are certainly rich people with gambling problems, like the professional golfer John Daly who says that he lost \$50–60 million in 12 years of heavy gambling [Dal07].

This gives  $n_0 = 174,000$ .<sup>10</sup> Repeating the Joe-DiMaggio-style calculation from the previous section, we find that the probability that anyone in Florida could win Play 4 so many times whilst spending at most \$174,000 is less than one in a million.

#### 7. Multiple kinds of tickets, reverse bounds

Real lottery gamblers tend to wager on many different games. Suppose they play b different bets. (It might feel more natural to say "games", but on Play 4— one game—you might bet on straight, straight/box, etc. We write "bets" because it is more accurate.) Number the bets  $1, 2, \ldots, b$ . Bet i costs  $c_i$  dollars and has probability  $p_i$  of winning at least the threshold amount. The gambler won more than the threshold on bet  $i w_i$  times. We don't know  $n_i$ , the number of times the gambler wagered on bet i. So far this is the same as the single bet version, but with subscripts.

As in (1.1), we estimate that the gambler bet  $\hat{n}_i = w_i/p_i$  times on bet *i*, for an estimate cash spent on lottery tickets of

$$\hat{c} = \sum_{i=1}^{b} c_i \hat{n}_i$$

Since all sums here are going to be for i = 1, ..., b, we will write  $\sum$  instead of  $\sum_{i=1}^{b}$ . If  $\sum w_i$  is big—i.e., if the gambler won many times— $\hat{c}$  will be very big, and the gambler will appear to be crooked.

If we knew the exact number of tickets bought of each kind—i.e., if we knew the vector  $\vec{n} = (n_1, n_2, \dots, n_b)$ —then we could calculate the probability:

(7.1) 
$$P := \begin{pmatrix} \text{Probability of winning at least } w_i \text{ times on bet } i \\ \text{with } n_i \text{ tickets, for all } i \end{pmatrix}.$$

If the different bets are on independent events (say, each bet is a different kind of scratcher ticket), then

(7.2) 
$$P = \prod_{i=1}^{b} \left( \begin{array}{c} \text{Probability of winning at least } w_i \text{ times on bet } i \\ \text{with } n_i \text{ tickets} \end{array} \right),$$

where the probabilities on the right given by (4.2). But gamblers might make dependent bets. For example, bet 1 might be a Play 4 straight ticket on the number 1234 and bet 2 a Play 4 24-way box on the number 1234, for the same drawing. Then (7.2) does not hold. Apart from Mr. Hollywood (examined in §9), the actual gamblers we investigate, however, did not turn in dependent winning tickets. That is, when we examine their claimed wins, we do not see such overlapping wagers. As we explain in the appendix, *if there are no dependent wins, it is almost as if the wagers were independent*. To make this precise, we replace the definition (7.1) of P with

$$P := \begin{pmatrix} \text{Probability of winning at least } w_i \text{ times on bet } i & \text{no wins on de-} \\ \text{with } n_i \text{ tickets, for all } i & \text{pendent bets} \end{pmatrix};$$

<sup>&</sup>lt;sup>10</sup>Mathematica could not solve this equation numerically; we had to resort to zooming in on the graph of  $D(n; w, p) - \varepsilon$  as a function of n. Matlab could solve it using **fzero**.

and we replace the equality (7.2) with the inequality

(7.3) 
$$P \leq \prod_{i=1}^{b} \left( \begin{array}{c} \text{Probability of winning at least } w_i \text{ times on bet } i \\ \text{with } n_i \text{ tickets} \end{array} \right).$$

Combining this with (4.2), we find:

(7.4) 
$$P \le \prod_{i=1}^{b} D(n_i; w_i, p_i).$$

**Turning it into an optimization problem.** Let's tie this to the discussion of Joe DiMaggio from the previous section. We want to find a level of play (number of tickets)  $\vec{n}$  such that the probability P from (7.4) is at least  $\varepsilon$ . Amongst all possible such choices for  $\vec{n}$ , we are generous to the player and pick the one that has the least actual cost. (Remember that a Mega Millions ticket costs \$1 but scratchers can cost \$1, \$5, \$20, \$25, etc.) We will also be generous to the player and allow her to pick non-integer values for the entries in  $\vec{n}$ . The cheapest way to have chance at least  $\varepsilon$  of winning bet i at least  $w_i$  times, for all i, is to wager on bet  $i \ \vec{n}_i^*$  times, where

(7.5) 
$$\vec{n}^* = \operatorname{argmin}_{\vec{n}} \vec{c} \cdot \vec{n}$$
 s.t.  $n_i \ge w_i$  and  $\prod_{i=1}^b D(n_i; w_i, p_i) \ge \varepsilon$ 

We can find a local minimizer  $\vec{n}^*$  of (7.5) using the Matlab function fmincon or the Python Scipy function scipy.optimize.minimize. We now show that a local minimum is in fact a *global* minimum, which allows us to conclude that it is utterly implausible that the gambler spent less than  $\vec{c} \cdot \vec{n}^*$ .

**Proposition 7.6.** A local minimizer  $\vec{n}^*$  for the optimization problem (7.5) is a global minimizer.

*Proof.* Because  $I_{p_i}(\alpha_i, \beta_i)$  with  $0 < \alpha_i < \beta_i$  is a log-concave function of  $\beta_i$ , by [FR97, Cor. 4.6(iii)], the product  $\prod_i D(n_i; w_i, p_i)$  is a log-concave function of  $\vec{n}$ , and so is quasi-concave. Therefore, in the optimization problem (7.5), the feasible set is a convex subset of  $\mathbb{R}^b$ . Since the function  $\vec{n} \mapsto \vec{c} \cdot \vec{n}$  is concave on  $\mathbb{R}^b$ , the claim follows.

The key point in the proof that  $\beta \mapsto I_x(\alpha, \beta)$  is log-concave. In the cases studied in this paper, the second derivative of  $I_x(\alpha, \beta)$  with respect to  $\beta$  changes sign, so  $I_x(\alpha, \beta)$  itself is neither concave nor convex.

# 8. More about Louis Johnson

Let's illustrate the method described in the previous section by discussing in detail the gambler Louis Johnson from Pompano Beach, Florida, whom we have already mentioned in Examples 1.2, 4.3, and 6.1. The list of winners provided by the lottery operators says he claimed 252 prizes worth \$600 or more. We selected some subset of those games as illustrated in Table 2 and solved the optimization problem (7.5) for these games, giving an estimated minimum spending level for him of about \$2.2 million.

In the table, the column labeled "minimum spending" should be interpreted as a group, not singly. These are the entries in the solution vector  $\vec{n}^*$  to the optimization problem (7.5), that is, they represent the amount a gambler would spend on that

particular game in order to have chance at least  $\varepsilon$  of winning as much as Louis Johnson while minimizing total spend on all games. It does not mean that Louis Johnson must have spent this amount on each of these particular games—however, if he spent less on one of the games, he would have to spend more on others to make up for it.

Applying the recipe from the previous section, with a cutoff of  $\varepsilon = 5 \times 10^{-14}$ , we would say that in order to have a larger probability for Johnson to win this many times on these bets—in order to pass the Wayne's World test for winning so many times—we would claim he would need to spend at least about \$2.2 million on lottery tickets, broken down as in the last column of Table 2.

When asked about his winnings, Johnson said "Lottery? I've never won the lottery." He suggested that someone might be cashing in tickets in his name. He also said that the IRS had sent him a letter saying he owes about \$50,000 in unpaid taxes on his lottery winnings [Mow14].

#### 9. The man from Hollywood

So far, for the sake of doing a detailed example, we have analyzed Louis Johnson, who claimed an astounding 252 prizes worth \$600 or more in the period 10/1/2007 to 7/30/2013. But this record is beaten by a man from Hollywood, Florida, whom we refer to as "H." During that same period, H claimed 570 prizes, more than twice as many as Johnson.

Mower's news report stimulated a law enforcement action against Johnson but not against H. Why? As hinted in the Badger example from §3, the sheer number of prizes claimed is not sufficient to distinguish between gamblers who are suspicious and those who are rich/lucky/addicted. When we look at the games won by these two gamblers, we find that Louis Johnson and our other top winners claimed many prizes in scratcher games and in Play 4, whereas all but one of H's prizes are from Play 4. The game Play 4 is really different from scratcher tickets, in that if you buy \$100 worth of scratcher tickets for a single \$1 game, this amounts to 100 independent binomial trials that are each equivalent to playing a single \$1 scratcher ticket. But with Play 4, you can bet any multiple of \$0.50 on a number to win a given drawing; if you win (which happens with probability  $p = 10^{-4}$ ), then you win 5000 times whatever you bet. So if you bet \$100 on a single Play 4 draw, your odds of winning remain  $10^{-4}$ , but your possible jackpot becomes \$500,000, and if you win, the lottery records this in the list of claimed prizes as 100 separate prizes.

So, to infer how much H had to spend on the lottery for his wins to be unsurprising, we have to first estimate how much he bets on each drawing. Unfortunately, we cannot deduce this from the list of claimed prizes, because they only record the date the prize was claimed and not the specific drawing the ticket was for. Louis Johnson's Play 4 prizes were all claimed on distinct dates, so it is reasonable to assume they correspond to different draws. (A few other frequent winners claimed more than Play 4 prize on the same day, in which case we ignored all but one of those Play 4 prizes, which is conservative.) The Palm Beach Post paid the Florida lottery to dig out of their archives a sample of H's winning tickets, which led them to believe that H's winning plays were as in Table 3.

In order to find a lower bound on H's spending using the optimization technique in §7, we imagine that he played several different Play 4 games, distinguished by their bet size. For the sake of argument, let us pretend that a player can bet \$1,

"min. spending"	\$29,300	\$30,600	\$85,400	\$1,015,700	\$121,100	\$88,700	\$50,500	\$67,100	\$61,100	\$41,900	\$65,800	\$30,700	\$47,600	\$22,800	\$43,500	\$60,700	\$46,800	\$29,500	\$51,600	\$57,000	\$62,600	\$45,100	\$33,000	\$2,188,000
est. spending $c_i w_i/p_i$	\$21,700	\$22,500	\$80,000	\$1,270,000	\$188,800	\$132,600	\$107,400	\$123,900	\$96,100	\$59,900	\$117,300	\$110,900	\$87,400	\$27,700	\$66,200	\$94,200	\$82,700	\$82,600	\$116,800	\$183,800	\$102,400	\$73,500	\$240,000	\$3,488,400
$\operatorname{cost}c_i$	\$1	\$1	\$0.50	\$1	\$20	\$10	\$20	\$10	\$25	\$10	\$30	\$20	\$20	\$10	\$5	\$5	\$10	\$20	\$2	\$20	\$5	\$20	\$1	
wins $w_i$	13	6	16	127	7	5	2	33	с,	2	с,		2		2	33	2		2	2	с,	2		212
prob. $p_i$	$6  imes 10^{-4}$	$4 \times 10^{-4}$	$10^{-4}$	$10^{-4}$	0000741667	0.000377	0.00037233	0.000242167	0.000780624	0.000333824	0.000767379	0.000180399	0.0000457516	0.000360667	0.000150969	0.000159302	0.000241933	0.000242063	0.000034259	0.000217593	0.000146528	0.000543873	0.000004167	Total
Game ID	14	14	14	14	1007	1025	1048	1060	1167	655	730	750	784	786	1011	1053	1067	1088	1107	1119	1130	1142	1189	
$\operatorname{Bet}$	Play 4 6-way box	Play 4 4-way box	Play 4 straight	Play 4 straight	Billion dollar	2m cash	Lucky a yr for life	Cash bonanza	Millionaire	Mega Gold	\$600m high roller	Gold rush	Sapphire blue	\$200m spectacular	Ruby Red 7's	250k monopoly	Lucky \$100k	Millionaire club	Monopoly	Gold rush tripler	Winning resolution	\$3m jubilee	Bronze bucks	
i		2	ŝ	4	5	9	2	$\infty$	6	10	11	12	13	14	15	16	17	18	19	20	21	22	23	

TABLE 2. Data for some of the prizes claimed by Louis Johnson.

date	number played	amount wagered
12/6/2011	6251	\$52
??	????	\$1
11/11/2012	4077	\$101
12/31/2012	1195	\$2
2/4/2013	1951	\$212
3/4/2013	1951	\$200

12 RICHARD ARRATIA, SKIP GARIBALDI, LAWRENCE MOWER, AND PHILIP B. STARK

TABLE 3. H's Play 4 wins during 2011–2013

\$50, \$100, or \$200 and say that we have observed H winning these bets 2, 1, 1, and 2 times, respectively. Using these as the parameters in (7.5) and the same probability cutoff  $\varepsilon = 5 \times 10^{-14}$  gives a minimum spending of just \$96,354.

But we can find a number tied more closely to his circumstances. In 2011–2013, he claimed \$2.84 million in prizes. These are subject to income tax, and if we presume his tax rate is about 35% then he would have taken home about \$1.85 million. It might be more useful to ask: if he spent that sum on Play 4 tickets, what is the probability that he would have won so much? We can find this by solving the following optimization problem with  $p = 10^{-4}$ ,  $\vec{w} = (2, 1, 1, 2)$ , and  $\vec{c} = (1, 50, 100, 200)$ :

$$\max_{\vec{n}} \prod_{i=1}^{4} D(n_i; w_i, p) \quad \text{s.t.} \quad w_i \le n_i \le w_i / p_i \quad \text{and} \quad \vec{c} \cdot \vec{n} \le 1.85 \times 10^6.$$

Solving this numerically shows that his chance winning so many times could have been as high as 0.0016, or one-in-625: it is plausible that H just got lucky. This impression was also supported by face-to-face interviews conducted by Mower.

It seems that H was successful by making large, dependent bets, while we know from earlier examples in this paper that betting a similar total amount on smaller, independent bets is less successful. This is a concrete illustration of the well known principle of casino gambling from [DS65, p. 170] or [Mos87, #37]: bold play is better than cautious play. That is, if you plan to spend \$200 on an evening of roulette, and you only care about your finances at the end of the evening, you are better off wagering \$200 on one spin and then stopping, rather than placing 200 \$1 bets.

## 10. What some people get up to

Above, we argued that some gamblers have claimed too many lottery prizes to be playing legitimately. You might wonder: if they aren't playing fair, what might they be up to? In this (non-mathematical) section, we describe various schemes that people have used to win "more than they should."

Methods specific to scratchers. Scratcher games have been compromised by various methods specific to scratchers, such as insider information or cryptanalytic techniques [Leh11].

Store owners have additional opportunities [Mar07; Ros14]. They have been known to take an entire roll of scratch-offs, scratch them all looking for winners, and if there are none (or not enough to pay for the entire roll), they'll report them stolen. They've also been known to take a pin and lightly scratch the wax on a ticket, revealing just enough of the barcode underneath to be able to scan it, as described in [McD13] or [Mar07, paragraph 75]. If they scan it and it's not a winner, they'll sell it to a customer, who may not notice the very faint scratches on the card. Happily, this scheme has already become impossible in many places, because the lottery operators have replaced the primitive linear barcode with a 2-dimensional barcode such as a PDF417 barcode.

But these "cheats" do not apply to Play 4, and the frequent winners described above have numerous Play 4 wins. So what might a suspicious Play 4 winner be up to? We are not suggesting that any of the specific gamblers discussed in this article have done any of these things; these are simply schemes that some people have engaged in in the past.

We ignore schemes like doctoring a losing ticket to make it appear to be a winner [WH12], or defacing the bar code on a paid-out winning ticket and getting paid for it again [Smi12]. These schemes are easily detected by the lottery operator.

**Ticket theft.** If you talk to people who play, or just go to a store and watch them play, you'll notice that they go to the store with their winning ticket and ask the cashier to check and see if it's a winner. The clerk has a terminal behind the counter that can scan the barcode on a ticket to see if it's a winner. (Barcodes for scratch-offs are under the wax.) The clerk will see that it's a winner and then has a choice: be honest and pay the person or tell the customer that the ticket is a loser and pocket the ticket for themselves—we call this *ticket theft*. Clerks are honest the vast majority of the time, of course, but they've been known to be dishonest, see for example the *NBC Dateline* exposé [Han09], [Sim12], [Bra12], [Osu12], [Dun], [Lou], [Yee13], [Mar07, paragraphs 47, 48, 80, 146], or the November 2013 scandal in Long Island [Bon13].<sup>11</sup>

Money laundering for beginners. It can be hard for a criminal to launder proceeds from illegal business operations, meaning to convert illegally-obtained cash into money in a bank account. This is because all cash deposits over \$10,000 have to be reported to the government, and presumably many such large deposits from someone with a small legal income will attract the interest of the authorities. However, one can launder money through, for example, using the illegal cash to buy scratcher tickets. Some such have an expected value of 75% of face value, meaning that spending \$10,000 on them is expected to produce \$7,500 of laundered money. All prizes over \$600 will be claimed at a lottery office, which implicitly provides a receipt and therefore an alibi for where the money came from.

Note that the above scheme loses 25% to the laundering operation—this fee is usually called the "vig." Perhaps this seems high, but for comparison purposes, Herbalife paid 43% vig to (legally) convert Venezuelan Bolivars to US dollars according to [Her12].

**Ticket aggregation.** There are lots of reasons a gambler might not want to claim a lottery prize at the lottery office. In order to claim a prize, the gambler has to present government-issued photo identification and provide a social security number; illegal immigrants or those with outstanding police warrants might be reluctant to produce such information. Furthermore, 2013 Florida statute 24.115, paragraph 4, requires that the lottery operator first pay any outstanding child support or debt the gambler has with any state agency before paying the gambler. A

<sup>&</sup>lt;sup>11</sup>The gambler who is worried about this threat is encouraged to sign the back of her ticket and to use self-service ticket checking machines.

gambler who owes large debts or a gambler reluctant to present identification might happily sell her ticket for less than face value to someone without such worries. Note that both people involved in this transaction commit a first-degree misdemeanor according to Florida statute 24.101, paragraph 2.

In this scheme, the person who buys the ticket should set the price to earn a profit—besides the time and effort and the legal risk involved in committing a crime, there is income tax on the winnings. Discussions with experts suggest that a typical winner might expect \$600 for a \$1000 ticket. If the ticket buyer is ambitious, she might offer this service to many nervous gamblers, in which case we call her a *ticket aggregator*.<sup>12</sup> Aggregators stand out in the list of claimed prizes because their winning tickets are typically purchased at many different stores (as opposed to typical gamblers, who mostly buy tickets at only a few stores). When we examined the list of winners in Florida, we noticed that the ones who appeared to be ticket aggregators (because they had many prizes from many different stores) had an improbable absence of very large prizes given their numerous prizes in the \$1000to-\$5000 range. Nervous gamblers may be less likely to trust a ticket aggregator with truly large prizes.

Money laundering (advanced). Ambitious money launderers (as in the previous subsection) might combine money laundering with ticket aggregation. In this scheme, instead of using illegally-obtained cash to buy lottery tickets from retailers, they instead buy winning lottery tickets from nervous gamblers. It seems clear that this strategy should provide launderers with a lower vig rate. This intuition is backed up by the fact that this strategy has been observed in the wild—see [Ana09], [Bet], [Tor11], [Ric96], [Ove04]—and was even used by famous Boston mobster Whitey Bulger [Cul95] and allegedly by Spanish politician Carlos Fabra [Fab13]. As a concrete data point, one such operation in Puerto Rico paid about 17% vig [Cot].

## 11. QUESTIONS FOR FURTHER INVESTIGATION

In this note, we have looked at some specific gamblers who claimed many lottery prizes. We used probability and optimization to distinguish some of the gamblers as "implausibly lucky" (i.e., up to something) and others as "plausibly lucky." However, our tools are somewhat blunt in that they rely on the gambler having claimed many nontrivial prizes. In order to deal with gamblers who have claimed few but very large prizes, we need finer information. One test case might be Diane and Kerry Carmichael of Phoenix, Arizona, who won the \$2.5 million top prize in The Pick (the Arizona state lottery) in December 1995 and a \$1 million second prize in Powerball in March 2013. There is no reason to doubt that they are just lucky, but our methods are too coarse to say much about it quantitatively.

How much do real gamblers spend? It would be helpful to know, by state, how many people wager more than a given threshold annually. This spending distribution seems to be skewed to the right: "the top 10 percent of players ... account for 50 percent of the total amount wagered, while the top 20 percent wager about 65 percent of the total." [CC89, p. 92]. For more on this question, see [Gar+14].

<sup>&</sup>lt;sup>12</sup>This line of work is easy for store clerks to fall into, because a nervous gambler is probably already going to them to collect smaller prizes that do not require going to a lottery office.

**Other states.** Forty-four US states have lotteries (not Alabama, Alaska, Hawaii, Mississippi, Nevada, nor Utah), so it seems likely that examining the list of prizes in some of the other 43 states would reveal similar anomalies. Georgia provided us a list of prizes claimed valued at \$5000 or more, which revealed 19 players who had claimed at least 50 such prizes since 2003.

**Exercise.** Get a list of lottery prizes claimed in your state or a neighboring one. Determine the ten gamblers who have claimed the most prizes. Analyze each of them using the methods of this paper.

# 12. Appendix: No dependent wins is almost as good as independent bets

This appendix justifies equation (7.3), which says that *having observed* no wins on dependent bets, an overall upper bound is given by treating the bets as if they were independent. We maintain the notation of  $\S7$ .

Think of a game like Play 4, which consists of a series of drawings, and in each drawing one can make a variety of bets such as 1234 straight, 1344 6-way box, etc., whose outcomes are dependent; of course, two bets on different drawings are always independent. (In solving the optimization problem (7.5), we treated each bet as a different game.) Abstractly, we envision a finite series of d independent drawings. For each drawing  $j, j = 1, \ldots, d$ , the gambler may bet any amount on any of b different bets, whose outcomes—for drawing j—may be dependent, but whose outcomes on different draws are independent. We write  $p_j$  for the probability that a bet on j wins in any given drawing.

For i = 1, ..., b and j = 1, ..., d, write  $n_{ij} \in \{0, 1\}$  for the indicator that the gambler wagered on bet i in draw j, so that  $n_i := \sum_j n_{ij}$  is the total number of times that bet i has been made. We call the entire system of bets B, so that B is identified with the b-by-d zero-one matrix  $B = [n_{ij}]$ .

We assume that we observe event

I := (In every drawing, the gambler using B wins at most one bet).

This event corresponds to not winning any dependent bets. The following proposition justifies (7.3).

**Proposition 12.1.** Suppose that, for each i, a gambler wagers on bet i in  $n_i$  different drawings, as specified by B, given above. Then

(12.2) 
$$\mathbb{P}\begin{pmatrix} at \ least \ w_i \ wins \ on \ bet \ i \ for \ all \ i, \\ and \ at \ most \ 1 \ win \ in \ drawing \ j \\ for \ all \ j \end{pmatrix} \leq \prod_{i=1}^b D(n_i; w_i, p_i).$$

The proof relies on the BKR inequality, conjectured in [BK85] and proved in [Rei00] and [BF87] (or see [CPS99]), which requires some new notation. Given the bets *B* laid by the gambler, we consider the events

 $W_i :=$ (gambler wins at least  $w_i$  wins on bet i, with bets B).

The BKR inequality uses the notation  $W_i \square W_j$  to denote the event that both  $W_i$ and  $W_j$  are satisfied, and furthermore there are  $w_i + w_j$  draws such that bet *i* is won in  $w_i$  of them and bet *j* is won in  $w_j$  of them; sometimes people say "events  $W_i$  and  $W_j$  both happen, and they do so for disjoint reasons".

Lemma 12.3.  $\mathbb{P}(W_1 \square W_2 \square \cdots \square W_b) \leq \prod_{i=1}^b \mathbb{P}(W_i).$ 

*Proof.* For b = 1, there is nothing to prove, and for b = 2 it is the famous BKR inequality. More generally, define  $V_i$  to be the event  $W_1 \square W_2 \square \cdots \square W_i$ , so  $V_i = V_{i-1} \square W_i$ . Then applying repeatedly the BKR inequality gives

$$\mathbb{P}(V_b) \le \mathbb{P}(V_{b-1})\mathbb{P}(W_b) \le \mathbb{P}(V_{b-2})\mathbb{P}(W_{b-1})\mathbb{P}(W_b) \le \dots \le \prod_{i=1}^b \mathbb{P}(W_i). \qquad \Box$$

We can now prove the proposition.

Proof of Proposition 12.1. Evidently we have an inclusion

(12.4) 
$$I \bigcap \left( \cap_{i=1}^{b} W_i \right) \subseteq W_1 \square \cdots \square W_b,$$

hence the left side of (12.2) is at most  $\mathbb{P}(W_1 \Box \cdots \Box W_b)$ , which by the lemma is at most  $\prod \mathbb{P}(W_i)$ . As  $\mathbb{P}(W_i) = D(n_i; w_i, p_i)$ , the proof is complete.  $\Box$ 

Note that in (12.4) we needn't have equality. For example, with b = 2 and 6 draws, suppose that  $w_1 = w_2 = 3$  and the gambler lays both bets on all 6 draws. The outcome where both bets win on all 6 draws is not in the left side of (12.4) but is in  $W_1 \square W_2$ .

## 13. Appendix: Remarks on the optimization problem (7.5)

**Monotonicity.** Some of the gamblers we studied for the investigative report claimed prizes in more than 50 different lottery games. In such cases it is convenient to solve (7.5) using only a subset of these games to ease computation by reducing the number of variables (or to be lazy and avoid calculating the probability  $p_i$  for all the different games). It is intuitively clear that this still gives a lower bound for the gambler's minimum spending, but we can also verify it in the notation of (7.5).

Suppose we have solved (7.5), but then decide to include the prizes from one more game in the calculation. Put  $w_0$  for the (positive) number of claimed prizes in the new game,  $p_0$  for the probability of a random ticket winning a prize in that game, and  $c_0$  for the (positive) cost to play. Then, writing, for example,  $(n_0, \vec{n})$  for the vector with entries  $n_0, n_1, n_2, \ldots, n_b$ , the new problem is to solve

(13.1) 
$$\vec{u}^* = \operatorname{argmin}_{\vec{u}}(c_0, \vec{c}) \cdot \vec{u}$$
 s.t.  $n_i \ge w_i$  and  $\prod_{i=0}^b D(u_i; w_i, p_i) \ge \varepsilon$ .

As  $D(u_0; w_0, p_0)$  is positive, we find that

$$\prod_{i=1}^{b} D(u_i; w_i, p_i) \ge \varepsilon / D(u_0; w_0, p_0) > \varepsilon,$$

so the vector  $(u_1, u_2, \ldots, u_b)$  is in the feasible set of (7.5) and

$$(c_0, \vec{c}) \cdot \vec{u}^* > \vec{c} \cdot (u_1, u_2, \dots, u_b) > \vec{c} \cdot \vec{n}^*,$$

as claimed.

Verifying the solution. Given that Mathematica could not solve (7.5) numerically, you might be justified in distrusting another program's answer claimed value for  $\vec{n}^*$ . If you want to check it, you needn't verify the full Kuhn-Tucker conditions. Indeed, in the cases we examined, we found that the constraints  $n_i \geq w_i$  were unsurprisingly never binding, and we had equality  $\prod D(n_i^*; w_i, p_i) = \varepsilon$ . Therefore, one need only check that  $\vec{c}$  points in the same direction as the gradient of the function  $\vec{n} \mapsto \prod D(n_i; w_i, p_i)$  at  $\vec{n}^*$ , i.e., that the dot product of the vectors is (approximately) the product of their lengths.

**Poisson approximation.** Some readers may want to approximate the binomial distribution with a Poisson distribution to estimate (4.1). This approximates D(n; w, p) with

$$U(n; w, p) = \sum_{k=w}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!}$$
 with  $\lambda = np$ .

This can be calculated using the survival function of the Poisson distribution, but it also can be expressed in terms of the regularized Gamma function

$$U(n; w, p) = P(w, np) = \frac{\int_0^{np} t^{w-1} e^{-t} dt}{\Gamma(w)}$$

which can be evaluated in Mathematica as GammaRegularized[w, 0, np], in Matlab as gammainc(np, w), and in Python Scipy library as scipy.special.gammainc(np, w). This approximation does not simplify the calculations in this paper. However, there are a couple of interesting points:

- While it is challenging to verify that D(n; w, p) is a log-concave function of n (and we referred to [FR97] for it), it is simple to check it for U(n; w, p).
- In the typical case where  $np \leq w$ , not only is U(n; w, p) an approximation for D(n; w, p), it is even an upper bound:  $U(n; w, p) \geq D(n; w, p)$  by [AS65].

## 14. Appendix: How long can a gambler gamble?

Suppose a gambler starts with an initial bankroll of  $S_0$  and repeatedly buys some kind of lottery ticket over and over again. If he spends all of his initial bankroll on these tickets and all his winnings, how much would he expect to spend in total?

**Proposition 14.1.** Suppose a ticket costs c and pays j with probability p and nothing otherwise. If the expected value of the ticket, pj - 1 is negative and j is an integer multiple of c, then the expected value of the total amount spent on tickets is  $cS_0/(1-pj)$ .

In the Louis Johnson Example 1.2,  $p = 10^{-4}$ , c = \$1, and j = \$5000, so a gambler betting on this game would expect to spend  $2S_0$  on tickets.

While the proposition above is enough to handle most of the lottery games offered in Florida, it is just as easy to prove an analogous result for a more general kind of ticket, which we do now. Let the random variable X denote the value of a ticket, payoff minus cost. (In the Louis Johnson example, we had  $\mathbb{P}(X = \$4999) = 10^{-4}$ .) We assume that

$$(14.2) \mathbb{E}(X) < 0,$$

because that is the usual setting for a scratcher or Play 4 ticket.<sup>13</sup> This assumption and the Law of Large Numbers say that the gambler eventually will run out of money with probability 1. The question we want to answer is: *how fast?* 

Write c > 0 for the cost of the ticket, so

(14.3) 
$$\mathbb{P}(X \ge -c) = 1 \quad \text{and} \quad \mathbb{P}(X = -c) \neq 0.$$

The gambler's bankroll after t bets is

$$S_t := S_0 + X_1 + X_2 + \dots + X_t,$$

where  $X_1, \ldots, X_t$  are i.i.d. random variables distributed the same as X and  $X_i$  tells the value of the *i*-th ticket. The gambler can continue to buy tickets until time T, the smallest  $t \ge 0$  for which  $S_t < c$ .

Proposition 14.4. In the notation of the preceding paragraph,

$$\frac{S_0 - c}{\mathbb{E}(X)|} < \mathbb{E}(T) \le \frac{S_0}{|\mathbb{E}(X)|},$$

with equality on the right if  $S_0$  and all possible values of X are integer multiples of c.

In most situations,  $S_0$  will be much larger than c, and the two bounds will be almost identical.

## *Proof.* By the definition of T and (14.3),

(14.5) 
$$0 \le \mathbb{E}(S_T) < c$$

with equality on the left in case  $S_0$  and X are integer multiples of c. Now the crux is to relate  $\mathbb{E}(T)$  with  $\mathbb{E}(S_T)$ . If T were constant (instead of a random variable), then we could simply write out

(14.6) 
$$\mathbb{E}(S_T) = \mathbb{E}(S_0 + \sum_{i=1}^T X_i) = S_0 + \mathbb{E}(T)\mathbb{E}(X)$$

(where, of course,  $\mathbb{E}(T) = T$ ) and combining this with (14.5) would give the claim. The miracle is that equation (14.6) holds even though T is a random variable; this is Wald's Equation (see, e.g., [Dur13, §5.4]).

You might recognize in this discussion that we are considering a version of the gambler's ruin problem but with an unfair bet and where the house has infinite money; for bounds on gambler's ruin without these hypotheses, see, e.g., [EK02].

## 15. Summary

We studied how to determine whether a gambler who claimed many lottery prizes was plausibly lucky or suspiciously lucky. One cannot conclude either way based purely on the number of wins; more information is necessary. Of the two specific gamblers we studied, we concluded that one of them (Johnson) was almost certainly up to something and the other (Hollywood) could easily just be lucky, even though Hollywood had claimed more prizes and more prize money than Johnson did. Finally, we surveyed ways in which one might claim many lottery prizes without being an innocent gambler.

 $<sup>^{13}\</sup>mathrm{Although}$  see e.g. [GM06] or [AG10] for cases where lottery tickets can have positive expectation.

We investigated 11 frequent winners using these methods. Nine of these 11 appear to be something other than lucky heavy gamblers. Perhaps they are rich for some reason unrelated to the lottery, and that allows them to lose large sums buying tickets. But barring some explanation like that, calculations described in this paper give reason to suspect that they are engaged in an activity like those described in §10. We believe that the lottery operators are obligated to investigate frequent winners like these. If a frequent winner is a store clerk who sells lottery tickets, at least one court has ruled that the operator has a legal obligation to investigate [Mar07, paragraph 62]. And we also believe the government has a moral obligation to not allow a service it provides (the lottery) to be used for money laundering. Finally, state lotteries were created not just to raise money for the states and not just to provide a reliable lottery-based schemes (like those surveyed in [Asb38]) as a source of income to organized crime; we believe that prosecuting ticket aggregators is in line with this tradition.

**Specific recommendation.** Sources like [Mar07] or [Mow14] make specific recommendations of new procedures that lottery operators could pursue to decrease the risk of criminal activities of the kinds we have investigated here, such as installing more self-service ticket checking machines, giving close scrutiny to prize-winners who work at lottery retail outlets, and real-time tracking of prizes claimed by frequent winners. To this list, we add an additional recommendation.

One complication that arose in our analysis was how to determine whether claimed prizes were for independent outcomes or not (the main obstacle was Play 4, where the gambler may bet any multiple of \$0.50 on a single number for a single drawing). Consequently, we urge those lottery operators that are required to report information on prizes claimed in Play 4 and similar games to also report the drawing that each ticket belonged to. This information would have made our job easier (namely, in the construction of Table 3) and would not violate anyone's privacy nor help bad guys, who already know the name, hometown, and approximate winnings of all winners of large prizes.

**Further developments.** Once the newspaper story [Mow14] appeared, the Florida lottery quickly announced reforms to curb the activities that we highlighted; see [O'C14]. The police also raided 14 stores, including three stores that provided most of Louis Johnson's winning tickets. As of this writing (shortly after the raid) there have not yet been any arrests, but the lottery has barred those stores from selling lottery tickets and seized their lottery machines.

#### References

- [AG10] Aaron Abrams and Skip Garibaldi. "Finding good bets in the lottery, and why you shouldn't take them". In: Amer. Math. Monthly 117 (2010), pp. 3–26.
- [Ana09] George Anastasia. "Lawyer sentenced in mob probe". In: The (Philadelphia) Inquirer (2009). URL: http://articles.philly.com/2009-03-10/news/25278392\_1\_digiacomo-perjury-charge-gamblingrelated-charges.
- [Art64] United Artists. *Goldfinger*. movie. 1964.

- [AS65] T.W. Anderson and S.M. Samuels. "Some inequalities among binomial and Poisson probabilities". In: Proc. Fifth Berkeley Symp. Math. Statist. Probab. Vol. 1. 1965, pp. 1–12.
- [Asb38] Herbert Asbury. Sucker's progress: an informal history of gambling in America. Dodd, Mead and Co., 1938.
- [BC03] Charles W. Blau and Jason B. Coutant. "Federal tax treatment of gains and losses from gambling transactions". In: *Gaming Law Review* 7.5 (2003), pp. 319–322.
- [Bet] Leaders of the notorious Black Mafia gang will spend 30 year apiece behind bars. BET.com. 2008. URL: http://www.bet.com/news/news/ 2008/09/18/newsarticlecrimeblackmafiagang30years.html.
- [BF87] J. van den Bergh and U. Fiebig. "On a combinatorial conjecture concerning disjoint occurrences of events". In: Annals of Probability 15.1 (1987), pp. 354–374.
- [BK85] J. van den Bergh and H. Kesten. "Inequalities with applications to percolation and reliability". In: J. Appl. Prob. 22 (1985), pp. 556–569.
- [Bon13] Kathianne Boniello. "Deli owners give lotto winner \$1K for \$1M ticket". In: New York Post (2013).
- [Bra12] Max Brantley. Lawsuit filed over lottery ticket tampering. Arkansas Times blog. 2012. URL: http://www.arktimes.com/ArkansasBlog/archives/ 2012/04/27/lawsuit-filed-over-lottery-ticket-tampering.
- [Cas86] George Casella. "Stabilizing binomial *n* estimators". In: Journal of the American Statistical Association 81.393 (1986), pp. 172–175.
- [CC89] Charles T. Clotfelter and Philip J. Cook. *Selling hope: state lotteries in America*. Harvard University Press, 1989.
- [Cot] Danica Coto. Santero priest tarted in PRico drug operation. URL: http: //www.denverpost.com/ci\_20317421/santero-priest-targetedprico-drug-operation#.
- [CPS99] Jennifer Tour Chayes, Amber L. Puha, and Ted Sweet. "Independent and dependent percolation". In: Probability theory and applications (Princeton, NJ, 1996). Vol. 6. IAS/Park City Math. Ser. Amer. Math. Soc., Providence, RI, 1999, pp. 49–166.
- [Cul95] Kevin Cullen. "US orders lottery to hold Bulger's winnings". In: Boston Globe (1995), p. 1. URL: http://www.boston.com/news/packages/ whitey/globe\_stories/1995/0718\_us\_orders\_lottery\_to\_hold\_ bulger\_s\_winnings.htm.
- [Dal07] John Daly. My life in and out of the rough: the truth behind all that bull\*\*\*\* you think you know about me. Harper, 2007.
- [DM89] Persi Diaconis and Frederick Mosteller. "Methods for Studying Coincidences". In: J. Amer. Stat. Assoc. 84 (1989), pp. 853–861.
- [DS65] Lester E. Dubins and Leonard J. Savage. *How to gamble if you must: inequalities for stochastic processes.* McGraw-Hill, 1965.
- [Dun] Maryland liquor store owner charged in lottery scam. Lottery Post. 2011. URL: http://www.lotterypost.com/news/229041.
- [Dur13] Rick Durrett. Essentials of stochastic processes. 2nd. Springer, 2013.
- [EK02] S.N. Ethier and Davar Khoshnevisan. "Bounds on gambler's ruin probabilities in terms of moments". In: Methodology and Computing in Applied Probability 4 (2002), pp. 55–68.

- [Fab13] María Fabra. "The law finally catches up with former Castellón cacique Fabra". In: *El País in English* (2013). URL: http://elpais.com/ elpais/2013/10/02/inenglish/1380734613\_504452.html.
- [FR97] H. Finner and M. Roters. "Log-concavity and inequalities for chi-square, F and Beta distributions with applications in multiple comparisons". In: *Statistica Sinica* 7 (1997), pp. 771–787.
- [Gar+14] Skip Garibaldi et al. Lottery spending: a non-parametric analysis. preprint. 2014.
- [GM06] Kent R. Grote and Victor A. Matheson. "In search of a fair bet in the lottery". In: *Eastern Economic Journal* 32 (2006), pp. 673–684.
- [GPS11] Charles M. Grinstead, William P. Peterson, and J. Laurie Snell. Probability tales. Vol. 57. Student Mathematical Library. American Mathematical Society, 2011.
- [Hal94] Peter Hall. "On the erratic behavior of estimators of N in the binomial N, p distribution". In: Journal of the American Statistical Association 89.425 (1994), pp. 344–352.
- [Han09] Chris Hansen. How lucky can you get? Lotto clerks uncovered. Dateline NBC. 2009. URL: http://www.nbcnews.com/id/30550956/ns/ dateline\_nbc-the\_hansen\_files\_with\_chris\_hansen/t/howlucky-can-you-get/.
- [Han14] David J. Hand. The improbability principle: why coincidences, miracles, and rare events happen every day. Farrar, Straus, and Giroux, 2014. ISBN: 978-0374175344.
- [Her12] Herbalife. United States Securities and Exchange Commission Form 10-K. 2012. URL: http://www.sec.gov/Archives/edgar/data/1180262/ 000119312513065327/d452887d10k.htm.
- [Kah87] William D. Kahn. "A cautionary note for Bayesian estimation of the binomial parameter n". In: The American Statistician 41.1 (1987), pp. 38– 40.
- [KM70] Lawrence J. Kaplan and James M. Maher. "The economics of the numbers game". In: American Journal of Economics and Sociology 29.4 (1970), pp. 391–408.
- [Kol90] Gina Kolata. "1-in-a-trillion coincidence, you say? Not really, experts find". In: The New York Times (1990).
- [Leh11] Jonah Lehrer. "Cracking the scratch lottery code". In: *Wired* (2011). URL: http://www.wired.com/magazine/2011/01/ff\_lottery/.
- [Lot13] Florida Lottery. Brighter than ever: annual report 2011-2012. available at floridalottery.com. 2013.
- [Lou] Kentucky store clerks indicted for lottery fraud. Lottery Post. 2011. URL: http://www.lotterypost.com/news/233545.
- [Mar07] André Marin. A game of trust: investigation into the Ontario Lottery and Gaming Coproration's protection of the public from fraud and theft. Ombudsman Report. 2007. URL: http://www.ombudsman.on.ca/ Investigations/SORT-Investigations/Completed/Lotteries---em-A-Game-of-Trust--em-.aspx?lang=en-CA.
- [McD13] Kimberly McDougall. Clerk accused of tampering with Lotto tickets. 2013. URL: http://fox40.com/2013/04/30/clerk-accused-oftampering-with-lotto-tickets/.

- [Mos87] F. Mosteller. *Fifty challenging problems in probability with solutions*. reprint of the 1965 Addision-Wesley edition. Dover, 1987.
- [Mow14] Lawrence Mower. *Gaming the lottery*. The Palm Beach Post. 2014. URL: http://www.mypalmbeachpost.com/gaming-the-lottery/.
- [O'C14] Cynthia O'Connell. "Point of view: lottery working to improve security, fight fraud". In: Palm Beach Post (2014). Opinion page. URL: http: //www.palmbeachpost.com/news/news/opinion/point-of-viewlottery-working-to-improve-security-/nfQmq/.
- [Osu12] Steve Osunsami. Brothers charged in \$5M Lotto scam. ABC News blog. 2012. URL: http://news.yahoo.com/blogs/abc-blogs/ny-brothers-5m-lottery-scam-police-125531344--abc-news-topstories. html.
- [Ove04] Torsten Ove. "Cocaine suspect may lose property". In: Pittsburgh Post-Gazette (2004). URL: http://www.post-gazette.com/frontpage/ 2004/06/09/Cocaine-suspect-may-lose-property/stories/ 200406090162.
- [Pic92] Paramount Pictures. *Wayne's World*. movie. 1992.
- [Rei00] David Reimer. "Proof of the van den Berg-Kesten Conjecture". In: Combinatorics, Probability and Computing 9.1 (2000), pp. 27–32.
- [Ric96] Warren Richey. "Officials: drug cash laundered via lottery". In: Sun-Sentinel (1996). URL: http://articles.sun-sentinel.com/1996-07-12/news/9607110636\_1\_lottery-winners-florida-lotterylottery-ticket.
- [Rob07] David Robbeson. "56\*". In: *The Walrus* (2007). URL: http://walrusmagazine. com/article.php?ref=2007.10-joe-dimaggio-56-games.
- [Ros14] Jeffrey S. Rosenthal. "Statistics and the Ontario Lottery Retailer Scandal". In: *Chance* 27(1) (2014).
- [Sel63] Thorsten Sellin. "Organized crime: a business enterprise". In: Annals of the American Academy of Political and Social Science 347 (1963), pp. 12–19.
- [Sim12] Abby Simons. Hopkins clerk accused of scratch-ticket fraud. Minneapolis Star Tribune. 2012. URL: http://www.startribune.com/local/west/ 160284665.html.
- [SM86] Stephen M. Samuels and George P. McCabe, Jr. "More Lottery Repeaters Are on the Way". In: *The New York Times* (1986). letter to the editor.
- [Smi12] Jordan Smith. The great \$100 lottery scam: Texas Lottery pursues bearers of 'previously paid' but forgotten scratch-off tickets. The Austin Chronicle. 2012. URL: http://www.austinchronicle.com/news/2012-06-08/the-great-100-lottery-scam/all/.
- [Ste08] Leonard A. Stefanski. "The North Carolina lottery coincidence". In: The American Statistician 62.2 (2008), pp. 130–134. DOI: 10.1198/ 000313008X306844.
- [Tor11] Yeganeh June Torbati. "Prosecutors: alleged drug dealer laundered money via Md. lottery". In: The Baltimore Sun (2011). URL: http://articles. baltimoresun.com/2011-04-27/news/bs-md-blackwell-indictment-20110427\_1\_tahirah-carter-blackwell-case-alleged-drugdealer.

- [TS87] William C. Thompson and Edward L. Schumann. "Interpretation of statistical evidence in criminal trials: the prosecutor's fallacy and the defense attorney's fallacy". In: Law and Human Behavior 11.3 (1987), pp. 167–187.
- [WH12] Deanese Williams-Harris. "Elgin woman claims record lottery jackpot, gets criminal record instead". In: Chicago Tribune (2012). URL: http: //articles.chicagotribune.com/2012-05-01/news/chi-elginwoman-tries-to-game-the-system-with-phony-lottery-ticketbut-gets-felony-charges-instead-20120501\_1\_mega-millionslottery-officials-numbers.
- Vivian Yee. "Three held in theft of \$75,000 lottery ticket". In: New [Yee13] York Times (2013). URL: http://www.nytimes.com/2013/05/ 09/nyregion/3-suffolk-clerks-accused-of-stealing-75000lottery-ticket.html?\_r=0.
- [Zil14] Zillow.com. Florida home prices and home values. 2014. URL: WWW. zillow.com/local-info/FL-home-value/r\\_14/.

ARRATIA: DEPARTMENT OF MATHEMATICS, UNIVERSITY OF SOUTHERN CALIFORNIA, DEPART-MENT OF MATHEMATICS, 3620 S. VERMONT AVE., KAP 104, LOS ANGELES, CA 90089-2532  $E\text{-}mail \ address:$  rarratia at usc.edu

GARIBALDI: INSTITUTE FOR PURE AND APPLIED MATHEMATICS, UCLA, 460 PORTOLA PLAZA, Box 957121, Los Angeles, California 90095-7121 E-mail address: skip at member.ams.org

Mower: Palm Beach Post, 2751 S Dixie Highway, West Palm Beach, FL 33405

STARK: DEPARTMENT OF STATISTICS, 3860 UNIVERSITY OF CALIFORNIA, BERKELEY, CA 94720-3860