PROBABILITY AND THE ODDS OF REPEATING NUMBERS

By Helena Pereira, Marketing Director, Szrek2Solutions

SHOULD WE QUESTION THE VALIDITY OF A LOTTERY DRAW IF THE SAME WINNING NUMBERS ARE REPEATED WITHIN A COUPLE OF DAYS OR WEEKS?

ooner or later you may need to explain the outcomes from your draws and deal with customer and stakeholder disbelief. Showing the proof of integrity is crucial for digital draws. Given enough opportunities, the seemingly impossible becomes plausible. The law of large numbers guarantees stable long-term results (e.g. the average of one billion tosses of a coin will be close to 500 million heads), but it also allows for unlikely short-term occurrences (e.g. a string of 100 consecutive heads). There are so many lotteries around the world drawing numbers each day that at some point we will come up with a draw in which the random numbers will repeat themselves on a proximate or consecutive day. Consider also how our attention works and how our media functions: When something unusual happens, much more attention is brought to it. We do not write about lottery draws having usual outcomes!

Consider, then, situations when the same identical string of numbers replayed itself for a lottery draw. On September 6, 2009, the Bulgarian lottery's winning numbers were 4, 15, 23, 24, 35 and 42. On September 10, the Bulgarian lottery randomly selected the same winning numbers 4, 15, 23, 24, 35 and 42, although the numbers appeared in a different order. "This is happening for the first time in the 52-year history of the lottery. We are absolutely stunned to see such a freak coincidence, but it did happen," a spokeswoman was quoted as saying in a September 18 Reuters article that year. This "freak coincidence" is actually the Improbability Principle at work. Lottery games are taking place regularly all over the world, hence creating numerous opportunities for numbers to repeat. Frequent lottery draws over time allow for repetition of outcomes.

In Israel's Mifal HaPayis state lottery, the numbers drawn on October 16, 2010—13, 14, 26, 32, 33, and 36— were exactly the same as those drawn a few weeks earlier, on September 21. The North Carolina Cash 5 lottery draw produced the same winning numbers on July 9 and 11, 2007.

In 1986, when our founder, Walter Szrek, worked at GTECH (now IGT) there was a similar situation that occurred at the Rhode Island Lottery: The same draw outcomes for a 4-digit game repeated themselves (1-9-1-0). The software director at GTECH asked what the probability was of the same 4-digit number coming up twice in a row; one university statistician responded that it was a 1 in 100 million chance, because the probability of this specific 4-digit number occurring twice was 1/10,000 x 1/10,000. Walter reasoned differently, asking

TABLE 1:Probabilities of Recurring Numbers

Q1	Probability of 1-9-1-0 occurring on a given day	1/10,000
Q2	Probability of the same number occurring on subsequent days	1/10,000
Q3	Probability of 1-9-1-0 repeating itself on subsequent draws	1/10,000 x 1/10,000=1/100,000,000
Q4	Probability of ANY NUMBER occurring twice in subsequent draws over 8 years	Using Wolfram's Calculation*, the end result is .2534

* For calculations of probabilities over time (Q4 in the chart), we use Wolfram's calculation of the probability of exceedance at www.worframalpha.com. Three fields need to be filled, and these can be interpreted as follows: (i) Number of times exceeded is the number of times the outcome occurs; (ii) Observed time period is the number of possible occurrences; and (iii) The prediction time period is the number of draws or time period over which a probability is estimated. To calculate the probability of the same numbers occurring on subsequent draws over the lottery's eightyear lifetime, we estimate: i) "1" occurrence of a certain 4-digit combination entails ii) "10,000" different possible combinations (1/10 * 1/10 * 1/10, which iii) can be estimated over a period 2922 days (eight years).

what was the chance that any same number combination – not the specific combination – repeated itself in a short time period. Leonard Stefanski made the same point in his analysis of the North Carolina Cash 5 coincidence in the The American Statistician (May 2008). His analysis would be of interest to readers with more mathematical inclination.

For a given day, the probability of having a specific combination is 1 in 10,000. The probability of any 4-digit combination occurring on subsequent draws is also 1 in 10,000 (the probability of the first

draw combination occurring again). But what is the probability that this same number combination repeats on subsequent draws at least once during the lifetime of the lottery? In this case, the probability of the same number combination repeating itself was estimated at over 25 percent for the lifetime of the game (which was eight years old at the time).

The theoretical chances of repeated draw outcomes discussed above assume the ideal world, where the draw devices used to generate random numbers do not interfere with randomness and all possible outcomes are equally probable. Now, it is possible that mechanical machines can have some differentially weighted balls and so all outcomes are not equally probable. An infamous example involved the Pennsylvania lottery scam in which insiders managed to substitute some of the balls for heavier balls. Today, most lotteries adopt procedures in which they verify the weight of mechanical balls prior to and after a draw. In the rest of the article we focus on repeat numbers generated from electronic draw machines which are generally less transparent than mechanical draw machines.

Electronic draws using RNGs

The industry has seen cases in which the draw numbers were repeated because of hardware or software problems, including Arizona, Delaware and Kansas. Hence, the coincidence of repeat lottery outcomes needs to be viewed with more care when it comes to electronic drawing machines.

The problems in electronic draw machines are not confined to repeating lottery outcomes. Even if the numbers are not repeated, hardware or software problems could result in the draw results becoming nonrandom, such that not all draw results are possible or have the same chance to occur. This is important because for a game to be fair all draw outcomes

LIKEWISE, WE SHOULD DEMAND PROOF OF INTEGRITY FOR EVERY DRAW, EVEN WHEN THE NUMBERS DO NOT REPEAT AND "SEEM RANDOM."

must be possible and must occur with the same probability.

The need to have fair outcomes has led to the practice of certifying electronic draw machines – testing random number generating systems and the data they generate to make sure that the outcomes display random properties. However, certification only guarantees that electronic draw machines generate random numbers when they are functioning properly and when there is no interference, say from an intruder. Pre-and posttesting were introduced to check that the hardware works properly during the draw. However, it is possible that both the pre-and post-testing produce results that "look random" without actually being random. One example is that the results could come from a limited range of generated numbers, which could be a result of a system configuration error (e.g. exclude 0s or not allow for repeat digits). It is also possible that non-random results are caused by draw system hardware or software errors or occur because of an undetected draw system seeding device reset.

Since we cannot tell if the draw outcomes are random by observing them, we must always be aware that there could be a problem that is not visible. We should always verify each draw to make sure the results were randomly generated, legitimate and did not experience any draw system problems or fraud.

As we look at repeat numbers that occurred by chance, and other sets of draw outcomes that were a result of a problem or manipulation, we realize that proof of integrity is the most important baseline the lottery industry can offer to the public. When we see repeat numbers in draw results, we question the legitimacy of the draw and demand proof of integrity. Likewise, we should demand proof of integrity for every draw, even when the numbers do not repeat and "seem random."

Besides testing and certification, a reliable verification method needs to be used to guarantee the integrity of the electronic draw. More specifically, to prove the integrity of an electronic draw, a lottery needs to be able to guarantee (i) that there were no hardware or software errors and (ii) that there was no interference with the draw. To satisfy these requirements, the draw system must first fulfill some important design criteria, such as "only use a valid RNG seed and do not allow for the generation of outcomes from a corrupted seed or defective hardware." Furthermore, the lottery must also be able to detect any intrusion to or manipulation of the draw system.

Szrek's Trusted Draw is one example of a system that guarantees proof of integrity and provides full transparency into the draw. If the same numbers were authentically drawn twice, these types of systems can prove their validity independently, and there is no doubt of the correctness of each individual outcome. On the other hand, if the outcomes were invalid due to a hardware problem, software problem or insider fraud, the system will immediately detect the anomaly. Lotteries using these types of independent systems will not experience draw problems due to hardware/software glitches and will not have any undetected fraud. They will be able to show all stakeholders that their RNG draws are correct and valid.