# Is lottery fair?

**Michal Horáček**
Matrikelnummer 6373382
michal.horacek@student.uni-tuebingen.de

**Carson Zhang**
Matrikelnummer 6384481
carson.zhang@student.uni-tuebingen.de

## Abstract

We investigate the uniformity of various lotteries. To detect potential tampering for both the German Lotto and New York Quick Draw lottery, we test the null hypothesis that the distribution of the minimum distance between lottery numbers, $d$, is as it would be in a uniformly sampled lottery. Furthermore, we evaluate lottery as a random number generator using the Diehard battery of tests.

## 1 Introduction

Lottery is a prominent source of chance outcomes in everyday life. However, we believe it is not a given that lotteries are conducted fairly. This assumption has been tested in prior research [1], and we believe it is worth testing with new data, especially with the German Lotto data, which is relevant to the authors and readers of this analysis as residents of Germany. Furthermore, even if a lottery has not been tampered with, assessing the randomness of different lotteries is useful to determine if they are truly fair, and to demonstrate that patterns do exist: players might investigate further to find and exploit these patterns for monetary gain.

This report consists of four sections. Section 2 describes the dataset collection. In section 3, we choose two lotteries to apply the minimum distance test from [1] to. Subsequently, we explore the randomness of lottery through the Diehard battery of tests in section 4. Finally we briefly discuss the limitations of our work in section 5. Our code and data are available here.

## 2 Dataset

We have collected data from 37 different lotteries organized in 30 unique countries, which in total provide over 32 million winning numbers.

The German lottery Lotto, which is tracked at [2] from 1955 onward, and on the New York Quick Draw lottery, which makes up the vast majority of our collected data: almost 26 million numbers.

The Diehard tests are applied to three datasets: the NY Quick Draw [3], Washington, DC's Keno (5,174,779 numbers) [4], and 1,917,890 other winning numbers combined from worldwide lotteries, which make up the Joint Lotteries dataset [5, 6, 7, 8]. The overwhelming majority of our data comes from official sources. As a sanity check, [2] does indeed match the actual results for 24.12.2022 [9].

Given the variety of sources, there are notable differences in their format. Some lotteries draw from the interval 1 - 80 instead of 1 - 49, and their draws are of various length. Most crucially, a portion of lottery CSV files record numbers in individual draws in draw order, while others sort them in an ascending order. We demonstrate that this has enormous impact on the results of Diehard tests.

Table 1: Frequencies of $d$ statistic for German Lotto

| d | 1 | 2 | 3 | 4 | 5 | 6 | 7 and 8 |
|---|---|---|---|---|---|---|---|
| Expected | 2310.596 | 1266.76 | 639.888 | 290.255 | 113.59 | 35.858 | 9.025 |
| Observed | 2383 | 1247 | 615 | 273 | 105 | 35 | 8 |

## 3 Distribution testing

For the German Lotto and NY Quick Draw, we conducted a hypothesis test where $H_0$ is true when lottery numbers are sampled with equal probability, and $H_A$ is true otherwise. We used Pearson's $\chi^2$ goodness-of-fit test because we can satisfy the assumptions and rules-of-thumb its usage, and it is the same test used in previous work, so our results can add on to theirs [1]. We used $\alpha = 0.05$ because it is popular and Drakakis et al. used the same $\alpha$.

### 3.1 Minimum-distance statistic

To create the frequency tables for the $\chi^2$ test (example in Table 1), we used the minimum distance statistic $d$ [1] . It is useful for detecting human tampering because humans do a poor job of imitating its distribution under the null hypothesis. Boland and Pawitan [10] asked humans to simulate a lottery where 6 numbers are sampled from 42, and computed a $\chi^2$ goodness-of-fit statistic of 36.45 when comparing the human-produced $d$ against the uniformly produced $d$ (assuming $p - 1$ degrees of freedom, in a hypothesis test this would have a p-value of $< 0.0000001$).Therefore, a low p-value could be consistent with human tampering.

We computed the expected distribution using the following formula [11].

Let the lottery game be a sample $r_1, ..., r_m$ of $m$ integers drawn from the integers $1, ..., n$. (In the German Lotto, $n = 49$ and $m = 6$.) Then, $d$ has the following definition and distribution.

$$d = \min_{1 \leq i < j \leq m} |r_j - r_i| \tag{1}$$

$$P(d < k) = 1 - \frac{\binom{n-(k-1)(m-1)}{m}}{\binom{n}{m}}, \quad k = 1, ..., \left\lfloor \frac{n-1}{m-1} \right\rfloor \tag{2}$$

### 3.2 The test statistic

The $\chi^2$ goodness-of-fit test tests how well an expected discrete distribution fits an observed discrete distribution. Under $H_0$, each lottery drawing is independent, which is a requirement for this test [12]. We combined the counts of adjacent categories when necessary to satisfy this rule of thumb for usage of the $\chi^2$ goodness-of-fit test: the expected frequency in each bin must be $\geq 5$ [1].

We define the $\chi^2$ statistic as follows: let $O_i$ be the observed frequency of category $i$, and $E_i$ be the expected frequency. Then, for our data,

$$\chi^2 = \sum_{i=1}^{p} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2(p-1) \tag{3}$$

We use $p - 1$ because we estimated no parameters from our data [1].

### 3.3 Results

For the German Lotto, we computed $\chi^2_6 = 5.357$, with a p-value of $0.498$. For the NY Quick Draw, we computed $\chi^2_1 = 1.407$, with a p-value of $0.235$. These results are not significant at $\alpha = 0.05$ or any other common significance level. For both lotteries, the distribution of $d$ looks like a typical distribution for a fair lottery. We do not suspect human tampering from these results.

### 3.4 Discussion

We considered testing the goodness-of-fit of the frequency of the drawn numbers to a discrete uniform distribution, but the lottery numbers are not independent: the numbers are sampled without

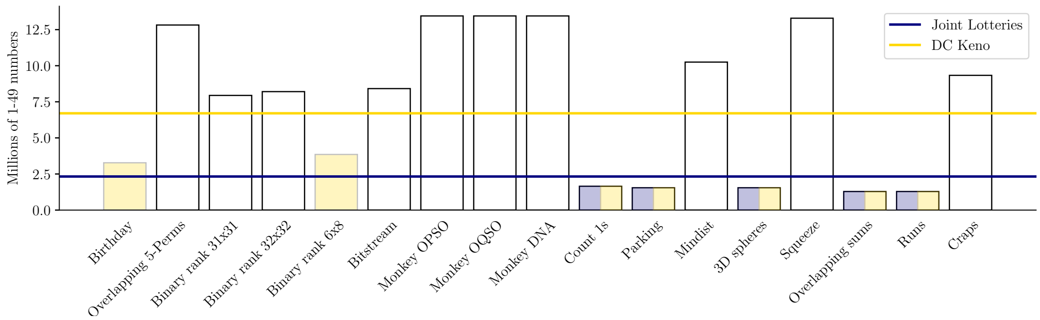



Figure 1: Amount of data required to run Diehard tests at default settings and dataset size. NY Quick Draw dataset is not present because at over 32 million numbers it falls out of scope.

replacement, so they are multivariate hypergeometric under $H_0$. Thus, their frequencies have non-zero covariance [13]. This also means sequences of lottery numbers will not be truly random.

Testing the distributions of other statistics, such as the sum of a lottery drawing, could provide additional insight into the fairness of the lottery, and into which numbers a player might choose. We provide an implementation of this test.

This analysis is appropriate for detecting human-generated lottery numbers, but it is unknown how well it will more sophisticated forms of unfairness, i.e. non-uniform sampling.

# 4 Diehard tests

Fairness entails more than the question of whether lottery numbers appear with the same frequency, so more comprehensive tests are required to detect deeper patterns. We use RNG testing suites to further test the fairness of lotteries.

We used Marsaglia's Diehard battery of tests [14] because of its popularity [15], and because it requires the least amount of data to run [16]. Even so, given the time constraints of this project and the fact that there does not exist so much publicly available lottery data, we were unable to collect enough data to run the entire battery of tests.

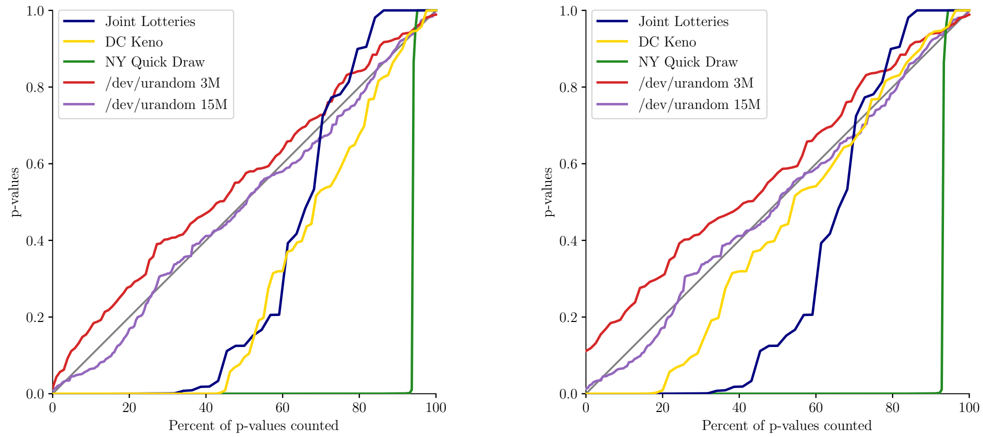
## 4.1 Data augmentation

The Diehard tests accept streams of bits as input, so we use the binary representations of the lottery numbers. However, we discard numbers $> 32$ for 1-49 lotteries and $> 64$ for 1-80 lotteries. Otherwise, the higher bits would not have equal probabilities of being active and inactive (in the 1-49 case, we would expect $\frac{17}{49} \neq \frac{1}{2}$ of the numbers to have $1$ in the highest bit). We discarded about 20 % of our data for this reason, which exacerbated our problem of having too little data.

## 4.2 Results

There are 15 Diehard tests in the battery. All these tests operate with the null hypothesis that the data is truly random. Each of them outputs one or more p-values, which are uniformly distributed under the null hypothesis [17]. Because each test measures randomness in a different way, it is common for a single dataset to pass some tests and not others (i.e. to observe wildly different p-values for the same dataset), especially because certain tests are easier to pass than the others. An example is the Overlapping Permutations Test, which divides the sequence into subsequences of 5 and tests whether each possible subsequence occurs with equal probability. For more information on the Diehard tests, consult either the project source code or one of Marsaglia's papers [18, 19].

As a baseline, we compare our lottery datasets against randomness collected from /dev/urandom, the standard source of randomness in Linux systems [20, 21].




(a) All tests possible for each dataset.

(b) Removed p-values from Binary Rank 6x8 test.

Figure 2: Distribution of p-values for datasets. Uniform distribution of perfect randomness in gray.

We immediately notice that the NY Quick Draw catastrophically failed to pass the Diehard tests. We believe this is because the numbers are presented in sorted order, which is obviously non-random. This hypothesis is supported by the superior performance of drawn-order DC Keno dataset. In tests composed of multiple runs, the drawn-order part of Joint Lotteries dataset also results in much higher p-values than its ascending-order remainder.

The DC Keno dataset displays a remarkable even slope after it starts rising. Moreover, the 25 of the 32 zero p-values for DC Keno come from the Binary Rank 6x8 test. If we disregard this test, the resulting p-values are fairly uniformly distributed, although the dataset still struggles at some of them, corresponding to 20 % of zero p-values. This is reasonable considering some enduring faults of the dataset (lottery drawn without replacement).

### 4.3 Discussion

There are many pitfalls to take into account. Some of them have already been addressed, such as data size, drawn/ascending order or lottery being drawn without replacement and thus not providing a truly uniform distribution.

The Diehard tests were designed to be used together as a battery. In the case of our medium-sized dataset DC Keno, we base our observations on 81 out of 213 p-values produced by the entire Diehard battery due to aforementioned size constraints. This inevitably leads to inaccuracy. We do not know how certain are our results and cannot rule out a stroke of bad luck.

Furthermore, even the Diehard battery of tests possesses flaws of its own. A critique can be found at [15, 22]. For instance, Linear Feedback Shift Registers pass the Diehard battery of tests [23] despite being predictable.

## 5 Conclusion

We analyzed the uniformity of lottery results in two major ways. We performed a hypothesis test of the distribution of the minimum distance $d$, a statistic that is known to be poorly imitated by humans. Based on the results for the German Lotto and NY Quick Draw, there is no strong evidence for human tampering, but it is unknown how well this analysis may detect other forms of unfairness, and it is unknown how well this would inform a player's choices.

The Diehard analysis was greatly limited by having too little data, the without-replacement nature of lottery sampling, and some obviously non-random (sorted) data. Despite these concerns, we believe that the performance of DC Keno on the Diehard tests suggests the fairness of this lottery.

## References


[1] Konstantinos Drakakis, Ken Taylor, and Scott Rickard. A statistical test to detect tampering with lottery results. 01 2009.

[2] Johannes Friedrich. Lotto number archive. https://github.com/JohannesFriedrich/LottoNumberArchive, 2023.

[3] State of New York. Lottery quick draw winning numbers: Beginning 2013. https://catalog.data.gov/dataset/?q=lottery&sort=views_recent+desc&ext_location=&ext_bbox=&ext_prev_extent=-150.46875%2C-80.17871349622823%2C151.875%2C80.17871349622823&tags=new-york-lottery, 2023.

[4] DC Lottery. Dc keno past winning numbers. https://dclottery.com/winning-numbers/past-draw-numbers?game=1, 2023.

[5] Tipos. Archive of drawn numbers. https://www.tipos.sk/loterie/ciselne-loterie/informacie-k-loteriam/archiv-vyzrebovanych-cisel, 2023.

[6] National Lottery. Results. https://www.national-lottery.co.uk/results, 2023.

[7] Lotterywest. Lottery results. https://www.lotterywest.wa.gov.au/results/, 2023.

[8] Lottoland. Offizielle eurojackpot zahlen & quoten. https://www.lottoland.com/eurojackpot/zahlen-quoten, 2023.

[9] Lottoland. Offizielle lotto zahlen & quoten. https://www.lottoland.com/en/german-lottery/results-winning-numbers/24-12-2022, 2023.

[10] Philip J. Boland and Yudi Pawitan. Trying to be random in selecting numbers for lotto. *Journal of Statistics Education*, 7(3):null, 1999.

[11] Konstantinos Drakakis. A note on the appearance of consecutive numbers amongst the set of winning numbers in lottery. *Math. Inform*, 22:1–10, 01 2007.

[12] David Diez, Mine Çetinkaya Rundel, and Christopher D Barr. *OpenIntro Statistics*. OpenIntro, 2019.

[13] Kyle Siegrist. *Probability, Mathematical Statistics, Stochastic Processes*. Random Services, Hunstville, Alabama, 2022.

[14] George Marsaglia. The marsaglia random number cdrom including the diehard battery of tests of randomness. https://web.archive.org/web/20160125103112/http://stat.fsu.edu/pub/diehard/, 1995.

[15] Robert G. Brown. Dieharder: A random number test suite. https://webhome.phy.duke.edu/~rgb/General/dieharder.php, 2004.

[16] Crypto Stackexchange. How to compute the dataset size required by dieharder tests? https://crypto.stackexchange.com/questions/90076/how-to-compute-the-dataset-size-required-by-dieharder-tests, 2021.

[17] Lawrence Bassham, Andrew Rukhin, Juan Soto, James Nechvatal, Miles Smid, Stefan Leigh, M Levenson, M Vangel, Nathanael Heckert, and D Banks. A statistical test suite for random and pseudorandom number generators for cryptographic applications, 2010-09-16 2010.

[18] G. Marsaglia and Arif Zaman. Monkey tests for random number generators. *Computers & Mathematics with Applications*, 26:1–10, 11 1993.

[19] George Marsaglia, B. Narasimhan, and Arif Zaman. The distance between random points in rectangles. *Communications in Statistics - Theory and Methods*, 19:4199–4212, 01 1990.

[20] Linux manual. https://man7.org/linux/man-pages/man4/random.4.html.

[21] Thomas' Digital Garden. Myths about /dev/urandom. https://www.2uo.de/myths-about-urandom/.

[22] Dirk Eddelbuettel. Sts critique. https://github.com/eddelbuettel/dieharder/blob/master/NOTES, 2019.

[23] Nidhi Gupta and G. P. Biswas. Wep implementation using linear feedback shift register (lfsr) and dynamic key. In *2011 2nd International Conference on Computer and Communication Technology (ICCCT-2011)*, pages 422–427, 2011.