
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. First, we consider Drakakis, Taylor, and Rickard's minimal distance between the winning numbers as a test statistic, and apply it to the German Lotto data. **We obtain the following p-value.** Furthermore, to evaluate the randomness of the random number generators underlying different lotteries, we apply the Diehard battery of tests **with the following results.**

1 Introduction

Games such as lottery and dice are likely one of the first application of randomness in human culture. But for the same time organizers have cheated in lotteries, skewing the uniform distribution of drawn numbers expected by common sense for their own monetary gain and misleading customers.

We begin by describing the data-gathering process in Section 2. Given the enormous demands placed on data volume by the second part of this paper, a sizable portion of our work involved preparing the input data.

In Section 3, we investigate the distribution of answers for the German Lotto lottery drawn from 1955 to the present day.

We continue by reformulating lottery as a process generating random numbers and explore their quality via the Diehard battery of tests [1] in Section 4. Through their means we attempt to prove whether lottery holds properties such as mutual uncorrelatedness or an absence of a period of repetition.

This paper is concluded in Section 5 by a brief discussion of the limitations of our work.

2 Dataset

The main dataset that we investigate is the numbers drawn from the German lottery Lotto, from 1955 onwards. We have chosen this dataset because it is relevant to us and readers of this analysis as current residents of Germany, and it contains almost 70 years of data, which we believe is enough data to perform a meaningful analysis.

Furthermore, Drakakis, Taylor, and Rickard did not apply their tests to this data, so performing the test on the German Lotto data is a meaningful new result.

The dataset has been compiled by Johannes Friedrich, a software developer who has made the data publicly available via [his Github repository](#).

Perform verification of the correctness of the dataset by sampling from it and manually inspecting the data.

To prepare the dataset for the hypothesis test, we computed d , the minimum distance between winning numbers, for each lottery day. We counted the frequencies of each value of d , and combined the counts of the two largest possible values, 7 and 8. This is the same preparation that was performed in the paper, and we did it to satisfy the common rule of thumb for usage of the χ^2 goodness-of-fit test: the expected frequency in each bin must be ≥ 5 .

Is there a dataset for a rigged lottery? This looks sketchy as hell.

However even 70 years of Lotto numbers is not sufficient to produce enough data for the diehard tests. These require 10 to 12 MiB of random bits, which is substantially more than 28.4 KiB of Lotto numbers. Therefore we downloaded other lottery datasets and combined them together. In total, our dataset reached more than 750 000 numbers drawn in 18 different lotteries. These majority of these lotteries come from various english-speaking countries such as USA, Australia or UK because we have been looking for them with English search queries. Most of the complement is formed by other european nations like Italy, Czech Republic or Germany.

Merge datasets?

The numbers are drawn individually, but their order within a single lottery draw does not matter - but maybe it does for some of our tests?

Investigate this

3 Distribution testing

3.1 Testing uniformity

QUESTION FOR MICHAEL: which test should we use, between: MINIMUM DISTANCE, BINNED SUMS, SIMPLE NUMBER FREQUENCIES. Is it okay if the analysis is essentially: replicating the methods in a paper, but on a new dataset?

We conducted the following hypothesis test.

H_0 : each lottery number is drawn with equal probability.

H_A : the lottery numbers are drawn with unequal probabilities: some numbers are more likely to appear than others.

We used Pearson's χ^2 test, a commonly recommended test for the probabilities of observing categorical data. **Motivate the choice of this particular test.**

3.1.1 Description of minimum-distance statistic.

To create the frequency tables for the χ^2 test, we used the minimum distance statistic d used by Drakakis et al.

write the definition from Drakakis et al.

This statistic is useful for detecting human tampering because it is known that humans usually do a poor job of imitating the random choices that occur under the null hypothesis. Boland and Pawitan showed that when humans are asked to sample m integers from the set $1, \dots, n$, the value of this statistic is greater on average than it is under uniform sampling. In particular, they observed that the true probability of $d = 1$ is greater than 0.5, a result that is highly unintuitive to humans: humans tend to underestimate how likely it is that two consecutive numbers are picked. In particular, Boland and Pawitan computed a χ^2 goodness-of-fit statistic of the human-produced d against the uniformly produced d , which if they had conducted a hypothesis test, would have yielded a p-value of < 0.0000001 .

Therefore, a low p-value could be consistent with human tampering, especially if we observe an unusually small number of lottery drawings where $d = 1$, as Drakakis et al. did in the French lottery.

Table 1: Frequencies of d statistic for German Lotto

d	Expected frequency	Actual frequency
1	2310.595965	2383
2	1266.759926	1247
3	639.888057	615
4	290.255446	273
5	113.589766	105
6	35.857667	35
7 and 8	9.025145	8

3.1.2 Description of the test statistic.

Explain the χ^2 goodness-of-fit test. Include a table of the expected vs. observed frequencies. Report the results.

The χ^2 statistic is a function of how much our actual observed frequencies deviate from these expected frequencies. Larger deviations result in higher values of the χ^2 statistic and therefore lower p-values.

Justify the choice of degrees of freedom.

Decide whether we should spell out the computation of the p-value, as if we computed it manually.

Report the value of the χ^2 statistic and the p-value.

3.2 Diehard tests

For more information about the Diehard battery of tests we refer the reader to the original paper [2].

4 Diehard tests

Fairness entails more than the question whether are lottery numbers from the expected distribution. For instance, the Kolmogorov-Smirnov test used in the first part of this paper does not concern itself with the order the numbers are drawn. However if we saw a lottery whose numbers were always drawn in a descending sequence, for example, we would become suspicious.

Thus a more comprehensive test is clearly required to establish a more detailed answer to our question. We approach this problem by reformulating lottery as a process producing a stream of (supposedly) random numbers, which themselves are simply bit sequences. Under this formulation, we can deploy standard statistical tests developed for testing random number generators: we have a file of one and zero bits and wish to investigate if its bits are correlated, repeating with a period or other quantities undesirable for randomness.

A number of these test suites has been developed over time. Donald Knuth presented an initial set of empirical tests in the second volume of his computer science bible *The Art of Computer Programming* in 1969. Many general cryptography textbooks such as *Handbook of Applied Cryptography* or *Foundations of Cryptography* contain multiple tests of their own. The American National Institute of Standards & Technology has published a *guideline* discussing this matter too.

We decided to use the Diehard battery of tests, which was developed by the American statistician George Marsaglia in the nineties. While this package used to be quite popular in its day, it has been superseded today by other suites, including its derivatives such as Dieharder or TestU01. **Why did we pick this one?**

1. About Diehard.
2. The following part is divided into several sections. In section 1, we discuss data gathering, processing and general creation of input files for Diehard. Special attention is given to the problem of attaining input file of sufficient size and the asymmetric requirements of individual Diehard tests.

3. Part 2 highlights result on Diehard suite, interprets them and compares them against commonly used PRNGs.
4. Section 3 clarifies the shortcomings of above approach.

5 Conclusion

1. Lottery draws without replacement
2. Some datasets are sorted in draw order
3. Reminder: too few numbers for comfort
4. The Diehard tests are flawed ([Linear Feedback Shift Registers](#))

We initially wanted to perform a hypothesis test of the distribution of the numbers. In particular, we considered performing a test of the goodness-of-fit of the discrete uniform distribution from 1 to 49 for the observed frequencies of each Lotto number. However, we discarded this approach because the common goodness-of-fit tests (χ^2 , Kolmogorov-Smirnov) require that the events are i.i.d., and the Lotto numbers are not i.i.d.: in each Lotto drawing, the numbers are sampled without replacement, so each number drawn on a particular day depends on the numbers that were drawn before on that day.

Each lottery day is i.i.d. from a multivariate hypergeometric distribution, so we can apply one of these goodness-of-fit tests when each day contributes a single event. However, another rule-of-thumb for the χ^2 goodness-of-fit test is that each category should have an expected count ≥ 5 . Directly counting the frequencies of each unique combination would not satisfy this rule-of-thumb. To handle this, we could compute a function of lottery draws and use that to group lottery draws together: for example, we could compute the sum of the 6 Lotto numbers from a single drawing, and count the frequencies of the sums being in specific bins. Then, we could safely apply the χ^2 goodness-of-fit test. However, it is unknown how well this approach may detect human tampering. Prior research has shown that humans do a poor job of reproducing the distribution of d , so we felt that it was more appropriate for tampering detection. However, this approach may still be useful for a player of the lottery: a deviation from the expected distribution of sums may guide a player to choose a combination of numbers that is more likely than would be expected under a "fair" null hypothesis.

References

- [1] 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.
- [2] George Marsaglia. A current view of random number generators. In Elsevier Science Publishers, editor, *Computer Science and Statistics, Sixteenth Symposium on the Interface*, 1985.