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 test the null hypothesis that, for the German Lotto, the distribution of the minimum distance between lottery numbers, d, is as it would be in a uniformly sampled lottery. The χ^2_6 goodness-of-fit test resulted in a p-value of 0.4989. 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 5. 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 6 by a brief discussion of the limitations of our work.

2 Dataset

Section 3 investigates 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 believ 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.

Figure 1: Origin of numbers making up our dataset.

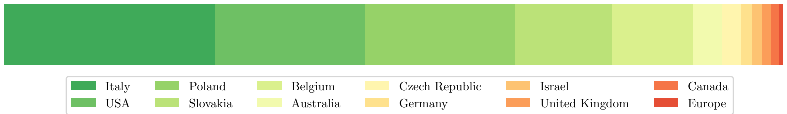

Figure 2: National composition of the "other" numbers.

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

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. Their author recommends using 10 to 12 MB of random bits, which is substantially more than 28.4 KiB of Lotto numbers. Therefore we add winning numbers from other lotteries to our dataset. By data volume, our data is dominated by the New York Quick Draw lottery, because it has been drawn every 4 minutes for the last decade, adding up to to 25,663,100 numbers at our collection of the dataset. Washington DC's Keno is quite similar, reaching 2,299,563 numbers over 3 years of existence. Our dataset is completed by the "other" category, consisting of 1,917,890 winning numbers combined from lotteries drawn less often worldwide.

The "other" dataset comes from European or english-speaking countries. The reason is twofold - we have been looking for them with English search queries and these nations are more likely to subscribe to ideas like open data and thus offer such dataset as csv files. The European nationality stands for European transnational lotteries Eurojackpot and Euromillions.

In total, we have gathered winning numbers from 37 different lotteries.

3 Methods

3.1 Testing the distribution of d

We conducted the following hypothesis test.

H_0: the distribution of d is what it we expect when the lottery numbers are sampled uniformly.

H_A: the distribution of d is different.

We used Pearson's χ^2 test, a commonly recommended test for the probabilities of observing categorical data. This is a popular test for performing exactly the type of hypothesis test we intend to perform. Furthermore, it is the same type of test performed by Drakakis et al., which allows us to check our (and their) results. We can also satisfy the assumptions of the χ^2 test, and follow the common rules-of-thumb for using the test. **Motivate the choice of this particular test.**

3.1.1 Description of minimum-distance statistic.

To create the frequency tables for the $\chi^2 test$, 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 $1, ..., n$, 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:

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

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.

3.1.2 Description of the test statistic.

The χ^2 goodness-of-fit test tests how well an expected discrete distribution fits an observed discrete distribution.

We computed the expected distribution using the following formula proved by Drakakis (2007).

Let the lottery game be a sample of m integers drawn from the integers $1, ..., n$. (In the German Lotto, $n = 49$ and $m = 6$.) Let $r_1, ..., r_m$ be the numbers drawn in the sample of size m.

The minimum distance $d = min_{1 \leq i < j \leq m} |r_j - r_i|$ has the following distribution. For $k = 1, ..., \lfloor \frac{n-1}{m-1} \rfloor$,

$P(d < k) = 1 - \frac{\binom{n-(k-1)(m-1)}{m}}{\binom{n}{m}}$

We define $\alpha = 0.05$ as our significance level because it is popular and Krakakis used the same significance level.

To prepare the German Lotto 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 type of data 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.

This prepration yielded the dataset in Table 1.

We estimated no parameters from the data, so we use $p - 1 = 7 - 1 = 6$ degrees of freedom for the χ^2 statistic.

3.2 Diehard tests

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

4 Results

4.1 Results of hypothesis test

We computed $\chi^2_6 = 5.3574$, with a p-value of 0.4989. This is not significant at $\alpha = 0.05$, and is not significant at any commonly used significance level.

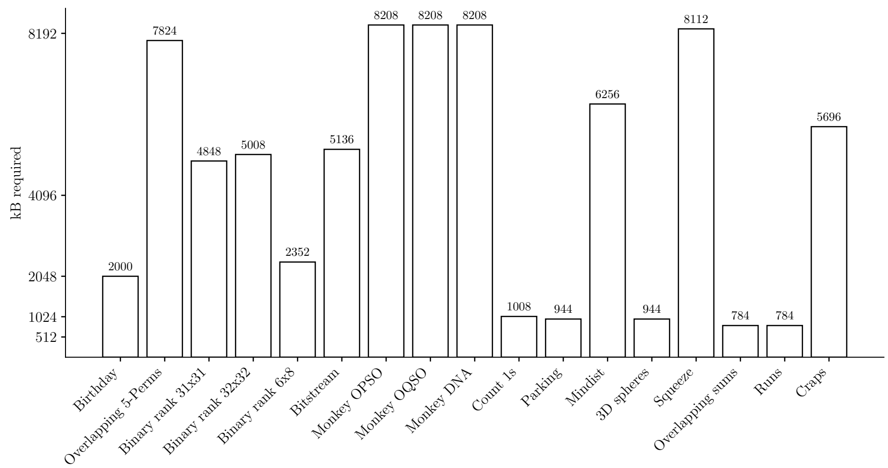

Figure 3: Amount of data required to run tests at default settings.

This demonstrates that the German Lotto's distribution of d looks very much like a distribution we would expect under a properly conducted lottery. There is no reason to believe, from this result, that any tampering has occurred or that the numbers are sampled with unequal probabilities.

5 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 superceded today by other suites, including its derivatives such as Dieharder or TestU01. In comparison with the alternatives, the Diehard tests consume a lot less data, making its use feasible to limited data cases such as our project.

Nonetheless even less data greedy test still requires quite a lot of data. We are thus limited in which tests we can run, see 1 for our dataset size.

5.1 Data augmentation

Diehard battery expects a random stream of zero and one bits as input. To satisfy this, we transform our lottery numbers distributed in 1-n range by first subtracting 1 and taking 5 lowest bits from every number for $n \in \langle 32, 64)$ or 6 for $n \in \langle 64, 128)$.

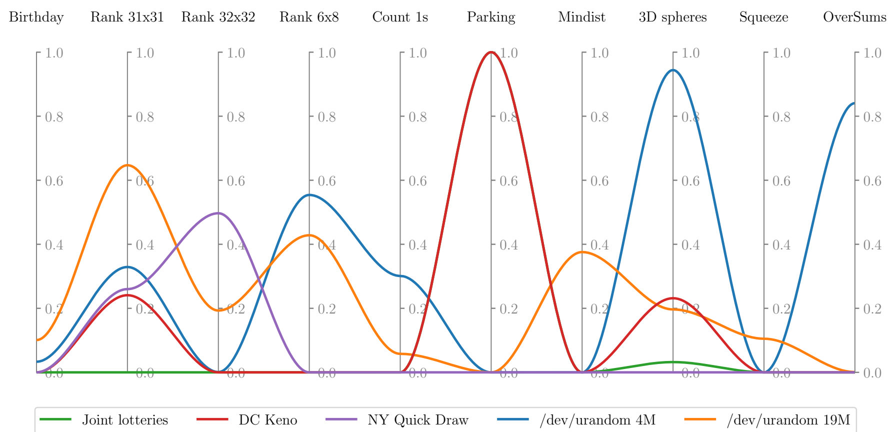

Figure 4: Score at various Diehard tests.

This transformation doesn't however solve all problems. The winning numbers are sometimes provided in sorted, ascending order in a single draw. We discuss this in further sections. Furthermore, the numbers are drawn without replacement. We argue that this is an acceptable deviation. **A bit of combinatorics, combinations with replacement as approximation for combinations without replacement**.

5.2 Results

1. About Diehard.
2. The following part is divided into several section. 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 asymetric 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.

6 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)

Some other choices of hypothesis tests were possible. We discuss why we did not choose some alternatives.

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 common goodness-of-fit tests when each day contributes a single event. However, another

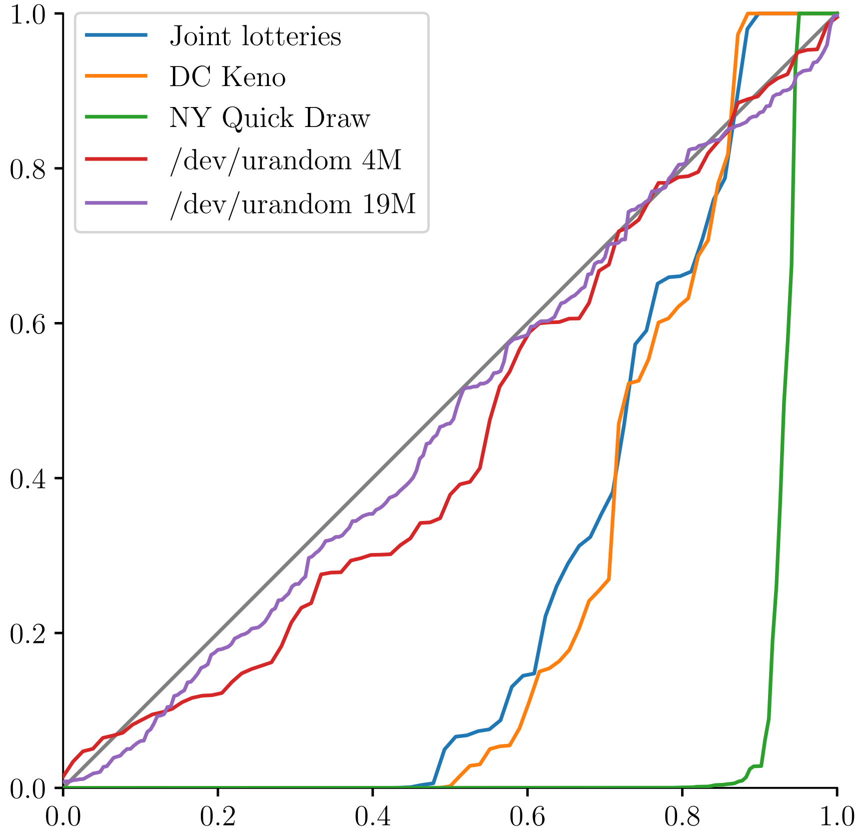

Figure 5: Randomness quality as measured by Diehard. Gray line is perfect randomness.

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. We provide an implementation of this apporach in R.

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.