Statistical evaluation of Google's Random Number Generator

Majda KEMMOU
AppStat Bonus Homework 2

Introduction

Random number generators (RNGs) are crucial for a variety of tasks, including simulations, cryptographic applications, and sampling. An RNG, in theory, will produce numbers with equal probability within the range requested. This study will put Google's web-based RNG through strict testing for uniformity when requested to produce integers from 1 to 10.

This report records an experiment in which 300 numbers were drawn manually from Google's RNG (to prevent bot detection) and were subjected to certain statistical tests, such as a chi-squared test, bootstrap resampling, and confidence interval estimation, to verify if the generator behaves uniformly.

Data collection

The dataset consists of 300 values ranging from 1 to 10 that were drawn manually using Google's RNG interface. At first, automation was attempted but Google flagged the bot activity, necessitating manual entry. Thus, the values were recorded and saved in a CSV file for analysis in Python.

Under a truly uniform RNG, each integer from 1 to 10 should appear with approximately equal frequency, i.e., 30 times in a sample of 300. Any deviation from this expectation could suggest a bias in the RNG.

Methodology

1. Frequency analysis

The first step was to visualize the empirical distribution of the 300 draws using a histogram. This offered a quick visual check for any irregularities in the frequencies.

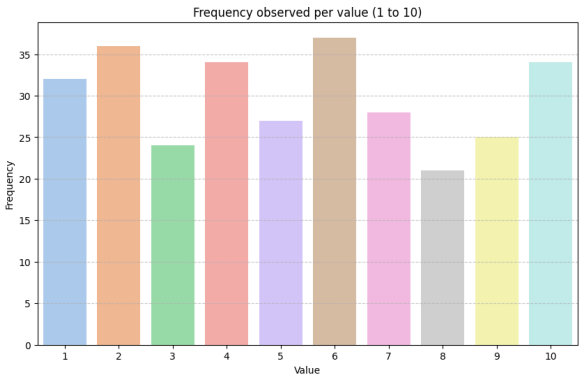

Figure 1: Observed Frequencies of Google's RNG (Sample size: 300)

2. Chi-Squared goodness-of-fit test

We applied a chi-squared goodness-of-fit test to statistically evaluate the uniformity. The null hypothesis (H_0) is that the values are uniformly distributed across $[1, 10]$. The alternative hypothesis (H_A) is that the distribution is not uniform.

$$H_0 : \text{Google RNG is uniform}$$
$$H_A : \text{Google RNG is not uniform}$$

The test statistic is:

$$\chi^2 = \sum_{i=1}^{10} \frac{(O_i - E_i)^2}{E_i},$$

where O_i is the observed frequency and $E_i = 30$ is the expected frequency.

Using SciPy's `chisquare` function, we obtained a p-value of approximately 0.65. Since this value is far above the typical significance threshold of 0.05, we do not reject H_0. This suggests that the deviations observed are consistent with what one would expect from a uniform RNG under random sampling.

3. Bootstrap resampling and confidence intervals

To reinforce the analysis, we used bootstrap resampling to construct empirical confidence intervals for the frequency of each number. Bootstrap is a non-parametric method that involves repeatedly sampling with replacement from the observed data and calculating the statistic of interest.

In this case, we resampled the dataset 10,000 times and computed the 95% confidence interval for the frequency of each number. The goal was to see whether the expected frequency (30) falls within the CI for each value.

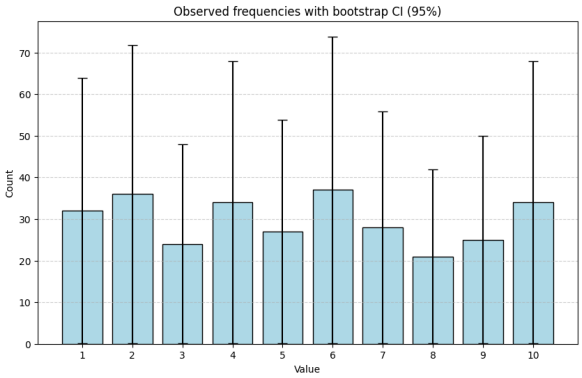

Figure 2: Observed Frequencies of Google's RNG with bootstrap CI

All the bootstrap intervals comfortably contained 30, which supports the hypothesis that the deviations are due to random fluctuations rather than systemic bias. It should be noted, however, that the wid confidence intervals reflect greater uncertainty which is a common feature of non-parametric methods when sample counts are low.

Discussion

The combination of visual analysis, classical hypothesis testing, and bootstrap confidence intervals all indicate that Google's RNG is statistically consistent with a uniform distribution over $[1, 10]$.

The chi-squared test gave the formal rejection criterion, and no reason to reject H_0. The p-value of 0.65 is especially good, as it suggests that the observed frequencies could quite easily have arisen by random variation.

This is confirmed by the bootstrap analysis, which gives non-parametric confidence intervals. Not only is 30 included in these intervals, but no number had an outlier frequency indicating systematic bias.

Conclusion

From what we have tested, we have no indication in the form of statistics that Google's random number generator is not uniform over the interval $[1, 10]$. The experiment, though hampered by the hand collection of data, was extensive in terms of using multiple statistical tools.

Our results highlight the need to supplement visual examination with formal statistical procedures. Both the chi-squared test and bootstrap CI analysis lead to the same conclusion: the RNG is operating correctly.