Lab1 TNM098 – Coin Toss

Jens Jakobsson and Jonathan Bosson

Introduction

Given a file with 60 datapoints that all contained 200 tossed coins. The task was to write a program to distinguish between the real people and the computer. Ones will represent head and zeros tail.

Method

To distinguish the group implemented two different kinds of analysis. The first checks the deviation between the total number of heads and tails in a specific datapoint. A deviation beyond the set threshold will be determined as a human. The group argued that a good RNG would not deviate from a 50/50 result by a large amount. Large in this case is a deviation of 18 or more.

This analysis was used as a first cull to take away the extreme datapoints that differed greatly from the rest. The datapoints that did not clear the first analysis were sent to the second analysis.

The second analysis checked the number of repetitive values in a sequence in a datapoint. If the same value occurred a certain number of times it was assumed that datapoint was generated by an RNG. This since the group argued it likely that a human would vary more between ones and zeros, and repetitive sequence of seven or more of the same value was deemed unlikely.

Result

The first analysis found 8 datapoints that were humans while the second analysis found 20 more. The table below shows the datapoints' line number in the data file that were found to be human, making a total of 28 human generated datapoints from the initial 60.

5	**6**	**9**	**10**	**16**	**17**	**18**
20	22	26	28	29	30	31
32	34	35	36	38	39	42
44	45	46	49	50	54	57

Discussion

The chosen thresholds in both of the analysis are no more than educated guesses which could prove to be poorly chosen. Further analysis regarding the used RNG's random distribution as well as human behavior could help to achieve improved chosen thresholds. Checking specific patterns and number of occurrences of a certain pattern in a datapoint could perhaps result in a more detailed analysis.