Coin Flip Series Randomness Tester

Ivan Silajev and Conor Bateman
Contacts:
https://www.linkedin.com/in/ivan-silajev-04957a18b/
https://www.linkedin.com/in/conor-bateman-b742131b7/

September 18, 2021

1   Introduction

This document serves as supplementary material for the "Random Coin Flips" WDSS Summer Research Project. The project involves creating a program (an R Shiny app) for measuring a user's ability to simulate a believable sequence of independent Bernoulli events, akin to a series of coin flips. The supplementary material provides a review of the theory used in the project and the program design.

1.1   Project Motivation

Random events happen without an underlying pattern, meaning no consistent algorithm can produce a series of truly random events. After numerous psychological tests, scientists found that people often approach the concept of randomness with underlying assumptions about its nature, usually ones that contradict its definition. Humans have an inherent tendency to look for predictability in anything they encounter to improve their understanding of it. Unfortunately, this instinct often leads to a paradox where one attempts to make sense of randomness as something predictable, worsening their understanding of it instead.

With the development of statistical mathematics, humans can now "measure" the randomness of events with appropriate hypothesis testing methods. The motivation behind this project is to create a program that, with the help of hypothesis testing theory, grades a user's ability to create plausible random sequences. This project is interested only in measuring randomness in human-generated binary sequences.

1.2   Statistical Tests

The statistical tests used in this project include:

- The Bernoulli distribution test

- The generalised runs test

- The longest run length test

- The Hadamard-Walsh test

- The head count test

- The multinomial test

- The last equalisation test

There exist time-series methods for measuring the predictability of a sequence, of which none we used in the project due to time constraints and limited intellectual resources. The rest of this document will explain the theory behind each test and the program's design.

2 Preliminaries

There is not yet an official way to measure randomness, but there are statistical tests that, when combined, accomplish such a function to a reasonable degree. The project utilises four* tests deemed to be useful enough for detecting implausible random sequences. We have to justify the usage of the four tests before explaining how to carry them out.

2.1 Binary Sequences

A binary sequence is an ordered string consisting of two different symbols only. Commonly, computers express binary sequences using the symbols **1** and **0**. An example of such a binary sequence is **100011** and has a length of six because it consists of six symbols. Numeric sequences, in general, possess a wide range of characteristics expressible as values; binary sequences are no exception.

Some numerical properties of binary sequences include:

- The length of the sequence

- The number of **0**s and **1**s

- The proportion of **0**s or **1**s to the length of the sequence

- The longest run of one symbol (the second-longest, third, etc.)

- The number of runs

- The number of runs of a specified length

- The frequency of runs

Some multivalued properties directly address the relationships between each term in the sequence:

- The Hadamard-Walsh transform

- Autocorrelation properties

By no means is the above list of properties exhaustive.

2.2 Hypothesis Test Review

A hypothesis is a statement with an unspecified truth value that depends on what people argue it to be. People can't fully validate a hypothesis about the properties of a variate (e.g. the mean, the variance) without analysing all members of the population related to the variate, which is often impossible. Therefore, statisticians use samples, a limited number of members from the population, to yield partial evidence for the properties of a variate. A hypothesis test is a tool for producing partial evidence for a hypothesis from samples, the performance of which depends on the amount of information available for testing and the nature of the test.

A hypothesis test often conveniently gives evidence as a p-value, a value that supports the hypothesis the closer it is to one and opposes it, the closer it is to zero. A p-value is commonly defined as the probability that a variate abiding by its hypothesised properties produced a given sample. The smaller the probability that the hypothesised variate could have produced the sample, the smaller the p-value. By this logic, a p-value of a test can serve as a measure of how closely a series of Bernoulli events can resemble a given binary sequence.

2.3 Bernoulli Distribution Test

The standard Bernoulli distribution test tests the plausibility that a series of identical, independent Bernoulli events can produce a given binary sequence. A Bernoulli event is essentially a coin toss with a set bias towards a specified side of the coin. Based on this interpretation of a Bernoulli event, the Bernoulli distribution test is highly appropriate for the project's purpose. The test uses the proportion of 1s to the sequence length as its only input, which carries no information about the arrangement of the sequence.

Thus, a sample with a **1**s to length ratio that strongly deviates from the hypothesised proportion would result in a small p-value upon applying the Bernoulli distribution test. It is necessary to utilise a different set of tests altogether to consider the arrangement of entries in a given sequence.

2.4   Generalised Runs Test

The generalised runs test is a hypothesis test inspired by the non-parametric Wald-Wolfowitz runs test. The Wald-Wolfowitz runs test tests the serial independence of entries in a binary sequence of a given length, using the number of runs in the series as an input. Unlike the Wald-Wolfowitz test, the generalised runs test does not assume that samples come from a finite population and does not assume that the two outcomes are equally likely.

Unfortunately, unlike the Bernoulli distribution test, the generalised runs test has an underdeveloped theory. Statisticians do not yet know the limiting distribution of runs in a Bernoulli trial sequence. Regardless, the test considers the arrangement of **1**s and **0**s in the sequence, which the Bernoulli distribution test does not.

A sequence with one run is a repetition of one symbol. In contrast, a sequence with maximal runs is a periodic sequence of isolated **1**s and **0**s. The two possibilities mentioned strongly express patterned behaviour, meaning they are weak contenders for being random sequences. Therefore, the number of runs on a sequence would be between one and the sequence length to plausibly represent a series of independent Bernoulli events.

2.5   Longest Runs Length Test

Similar to the generalised runs test, the longest run test is underdeveloped but less so. The test uses the length of the longest run in a binary series as its input. This test also considers the arrangement of 1s and 0s in the sequence, though differently this time.

A sequence whose maximum run length is one is a periodic sequence of isolated 1s and 0s. Alternatively, a sequence with a maximum run as long as the series consists of only one symbol. Again, both possibilities are weak contenders for being random sequences for being too patterned. So, the longest run in a sequence should fall between the two extreme cases to plausibly represent a series of independent Bernoulli events.

2.6   Hadamard-Walsh Test

Similar to how one can decompose functions, with continuous domains, into a sum of sine and cosine waves with a Fourier transform, one can decompose binary sequences into sums of Walsh functions with a Hadamard-Walsh transform. A Walsh function is a periodic square wave function that takes one and negative one as its values. One can decompose any binary sequence into Walsh functions because they constitute a complete set of orthogonal discrete functions.

The transform yields a vector that numerically indicates how strongly a given binary sequence resembles each Walsh function. The disadvantage of the Hadamard-Walsh transform is it works only with vectors with lengths equal to powers of two, so the project will work only with binary sequences of lengths 32 and 64 to account for the limitation. Theoretically, a random sequence is more likely to resemble every Walsh function weakly. The more a sequence resembles a single Walsh function, the more patterned it is.

2.7   Head Count Test

For $p = \frac{1}{2}$, the most probable number of heads is $\frac{n}{2}$ for even n. For values further away from $\frac{n}{2}$, the probability decreases, for example a sequence of all heads becomes increasingly improbable as n increases. This pattern is similar for $p \neq \frac{1}{2}$. We can analyse this behaviour to test the hypothesis that our coin throws are independent and come from a Bernoulli distribution with probability p. for $p \neq \frac{1}{2}$,

2.8   Multinomial Test

The natural extension to the head count test is looking at pairs of tosses. For $p = \frac{1}{2}$, we expect there to be an equal frequency ($\frac{n}{4}$), of Head Head, Head Tail, Tail Head and Tail Tail pairs. Generally,

the more spread these frequencies have, the more improbable it is that the tosses were independent and came from Bernoulli($\frac{1}{2}$) distribution. For example having all Head Tail pairs suggests patterned behaviour, and so are weak contenders for being random sequences. This somewhat compliments the Head Count test, since the sequence **101010101010** will be given a p-value 1 by the head count test, but a very low p-value for the multinomial test for $p = \frac{1}{2}$. This test can be easily extended to $p \neq \frac{1}{2}$, as has been done in the program.

2.9 Last Equalisation Test

By modelling the coin tosses as a Markov Chain, it is possible to calculate the probability distribution for the last time in which the number of heads will equal the number of tails. We can then use our binary input to find the last equalisation in our observed sequence.

We define an equalisation in our binary sequence as the time step in which the number of **1**s and**0**s are equal. With this distribution and observed statistic, we can calculate the associated p-value of the observed value.

3 Bernoulli Distribution Test Theory

The Bernoulli distribution test is derived as follows.

We define a sample $\mathbf{X}$ as a vector of n random variables, each representing an entry X_i in a possible binary sequence:

$$X = (X_i)_{i=1}^n$$

We also define the following statistical model for the sample:

$$(S, \{f(x|\theta) : \theta \in \Theta\}) = (\{0,1\}^n, \{p^{n\bar{x}}(1-p)^{n(1-\bar{x})} : p \in [0,1]\})$$

where $\bar{x}$ is the mean of the realised sample of $\mathbf{x}$.

The statistical model explicitly states that the possible outcomes of the sample are binary sequences of zeros and ones. The distribution fitted to each sample entry is a Bernoulli distribution, each with the same parameter p taking values from zero to one.

By the theorems of sufficient statistics, the sample mean $\bar{\mathbf{X}}$ is enough to test whether the hypothesised Bernoulli statistical model is a good fit for a given realised sample. By the theorems of point estimation, the sample mean is the best-unbiased estimator of the parameter p. We can say the more the sample mean deviates from a hypothesised value of p, the less likely a Bernoulli variate with parameter p could have produced the sample.

We can formally state the hypothesis we want to test in terms of p and its hypothesised value p bar.

$$\mathbb{H}_0 : p = \bar{p}$$

$$\mathbb{H}_1 : p \neq \bar{p}$$

With the help of the central limit theorem, for sufficiently large values of n, the p-value for the hypothesis that a Bernoulli variate with parameter p produced a given sample is given by:

$$P_{value} = 2\left(1 - \Phi\left(\frac{|\bar{x} - \bar{p}|}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}}\right)\right)$$

Since the project uses sample sizes of 32 and 64, the usage of the central limit theorem is justified.

4 Generalised Runs Test Theory

The generalised runs test is derived as follows.

Instead of fitting the distribution to the sequence, we fit the run distribution of a Bernoulli sample to the number of runs in a given realised sample.

The sample $\mathbf{X}$ of coin flips and the statistical model tied to it is the same as before. The statistic used to calculate the number of runs in a sample is as follows:

$$r(X) = 1 + \sum_{i=1}^{n-1} |X_i - X_{i+1}|$$

The distribution of this statistic does not yet possess an elementary expression. Fortunately, the pmf is calculable through combinatorial deduction. The pmf of r($\mathbf{X}$) is given by:

$$\mathbb{P}(r(X) = r) = \sum_{h=0}^{n} (k(n, h, r) + k(n, n - h, r))p^h(1 - p)^{n-h}$$

The k(n,h,r) coefficient enumerates the number of binary sequences, starting with a **1**, of length n, with h **1**s that have r runs. The k coefficients follow the following recursive rule:

$$k(n, h, r) = \begin{cases} k(n - 1, h - 1, r - 1) + k(n - 1, h - 1, r), & \text{if r is odd} \\ k(n - 1, h, r - 1) + k(n - 1, h, r), & \text{if r is even} \end{cases}$$

Sequences of length one can only have one run; therefore, k(1, 0, 1) is zero and k(1, 1, 1) is one. With this basic knowledge, we can deduce all other k coefficients. In the special case that p and q are equal (both a half), the pmf simplifies to:

$$\mathbb{P}(r(X) = r) = \binom{n - 1}{r - 1} \frac{1}{2^{n-1}}$$

where the column vector is the binomial coefficient. Since there is no actual runs parameter for the Bernoulli distribution, the best we can do is hypothesise that the sample consists of identically distributed independent Bernoulli variates.

$$\mathbb{H}_0 : X_i | X_j = X_i \sim Bernoulli(\bar{p}) \forall i, j \in [1, n]_{\mathbb{Z}}$$

$$\mathbb{H}_1 : X_i \not\sim Bernoulli(\bar{p}) \forall i, j \in [1, n]_{\mathbb{Z}}$$

The closer the number of runs is to the extreme cases, the more likely the alternative hypothesis becomes. Therefore, we calculate the p-value using the classical two-sided approach, given that p is

$$P_{value} = \begin{cases} \mathbb{P}(r(X) \leq \bar{r}) + \mathbb{P}(r(X) \geq \underline{r}) : \mathbb{P}(r(X) \leq \underline{r}) \leq 0.5 \\ \mathbb{P}(r(X) \geq \bar{r}) + \mathbb{P}(r(X) \leq \underline{r}) : \mathbb{P}(r(X) \geq \underline{r}) \leq 0.5 \end{cases}$$

where $\bar{r}$ is equal to r(x), the number of runs in a realised sample, and $\underline{r}$ defined as follows for the two possible cases for $\bar{r}$:

$$\max \mathbb{P}(r(X) \geq r) = \mathbb{P}(r(X) \geq \underline{r}) : \mathbb{P}(r(X) \geq \underline{r}) \leq \mathbb{P}(r(X) \leq \bar{r})$$

$$\max \mathbb{P}(r(X) \leq r) = \mathbb{P}(r(X) \leq \underline{r}) : \mathbb{P}(r(X) \leq \underline{r}) \leq \mathbb{P}(r(X) \geq \bar{r})$$

5 Longest Run Test Theory

The longest run test is derived as follows.

Here, we fit the distribution for the longest run of a Bernoulli sample to the longest run length in a given realised sample. The statistic used to calculate the run of maximal length in sample X is given by:

$$m(X) = \max\{k : \{X_{i+j}\}_{j=0}^{k-1} \in \{\{1\}, \{0\}\} \wedge i, k \in [1, n]_{\mathbb{Z}}\}$$

The formula for the statistic translates to "the maximum number of consecutive entries in sample X that are all equal to either zero or one", where sample X still follows the statistical model mentioned before. The pmf of m(X) is combinatorially more complex than r(X).

$$\mathbb{P}(M(X) = m) = \sum_{h=0}^{n}(t(n, h, m) - t(n, h, m-1))p^h(1-p)^{n-h}$$

The t(n, h, m) coefficient enumerates the number of binary sequences of length n, with h 1s with a maximum run length of m. The t coefficients follow the following recursive rule:

$$t(n, h, m) = \sum_{i=0}^{m} t(n-1-i, h-1, m) - \sum_{i=1}^{m} t(n-1-m-i, h-1-m, m) + e(n, h, m)$$

Where the e coefficients are given by:

$$e(n, h, m) = \begin{cases} 1, & \text{if } h = 0 n - h \in [0, m]_{\mathbb{Z}} \\ -1, & \text{if } h = m + 1 n - h \in [0, m]_{\mathbb{Z}} \\ 0, & \text{otherwise} \end{cases}$$

Again, there is no maximum run length parameter for the Bernoulli distribution, so we hypothesise that the sample consists of identically distributed independent Bernoulli variates as we did for the generalised runs test. The closer the maximum run length is to the extreme cases, the more likely the alternative hypothesis becomes. The p-value is calculated using the two-sided approach, with p given as $\bar{p}$, as last time:

$$P_{value} = \begin{cases} \mathbb{P}(m(X) \leq \bar{m}) + \mathbb{P}(m(X) \geq \underline{m}) : \mathbb{P}(m(X) \leq \underline{m}) \leq 0.5 \\ \mathbb{P}(m(X) \geq \bar{m}) + \mathbb{P}(m(X) \leq \underline{m}) : \mathbb{P}(m(X) \geq \underline{m}) \leq 0.5 \end{cases}$$

where $\bar{m}$ is equal to m(x), the number of runs in a realised sample, and $\underline{m}$ defined as follows for the two possible cases for $\bar{m}$:

$$\max \mathbb{P}(m(X) \geq m) = \mathbb{P}(m(X) \geq \underline{m}) : \mathbb{P}(m(X) \geq \underline{m}) \leq \mathbb{P}(m(X) \leq \bar{m})$$

$$\max \mathbb{P}(m(X) \leq m) = \mathbb{P}(m(X) \leq \underline{m}) : \mathbb{P}(m(X) \leq \underline{m}) \leq \mathbb{P}(m(X) \geq \bar{m})$$

6 Walsh-Hadamard Test Theory

The Walsh-Hadamard test is derived as follows.

The recursive formula for the Hadamard matrix is:

$$H_{2n} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \otimes H_n : H_1 = (1)$$

where the cross operator denotes the Kronecker product.

Let **X** be the sample as before. The statistic for the Walsh-Hadamard test is given by:

$$\omega(X) = H_n X$$

As before, we hypothesise that sample X consists of identically distributed independent Bernoulli variates. We calculate the p-vector (vector of p-values) for this test as follows:

$$p_v = (p_i)_{i=1}^{n} : p_i = \begin{cases} 2\left(1 - \Phi\left(\frac{|\omega(X)_i - n\bar{p}|}{\sqrt{n\bar{p}(1-\bar{p})}} \right) : i = 1 \\ 2\left(1 - \Phi\left(\frac{|\omega(X)_i|}{\sqrt{n\bar{p}(1-\bar{p})}} \right) : i \in [2, n]_{\bar{Z}} \end{cases}$$

The first term of the p-vector is just the p-value for the Bernoulli distribution test. We mentioned that the Walsh-Hadamard test is valid only for orders that are powers of two, so we chose the project to utilise sample sizes of only 32 and 64, enough to justify using the central limit theorem for the p-vector formula.

We can turn the Walsh-Hadamard test p-vector into a single metric by checking how closely a sample of values from a standard uniform distribution resembles the p-vector. The following statistic gives the measure of uniformity of values in the p-vector:

$$u(p_v) = \left| 1 - \frac{p'_v \cdot v}{p'_v \cdot p'_v} \right| : v = \left(\frac{i}{n} \right)^n_{i=1}$$

where the dashed p-vector is the p-vector with the order of entries arranged from smallest to largest.

The closer the u(p-vector) statistic is to zero, the better the standard uniform distribution fits the p-vector. Thus, it is more probable that the realised sample consists of identically distributed independent Bernoulli variates.

We derived the u(p-vector) statistic by applying linear regression techniques to a Kolmogorov-Smirnov plot with the p-vector. The more the plot resembles the identity function, the better the standard uniform distribution fits the p-vector (the p-vector being the predictor variable).

7 Head Count Test Theory

The statistic used in the head count test, the number of heads is given by

$$h(X) = \sum_{i=1}^{n} X_i$$

It is known that the sum of independent and identically distributed Bernoulli random variables is binomially distributed with the following rule

$$(Y_i)_{i=0}^n \sim \text{i.i.d } (\text{Bernoulli}(q)) \Rightarrow \sum_{i=1}^{n} Y_i \sim \text{Binomial(n,q)}$$

Using this implication, we can simply use the binomial distribution as our distribution for h($\mathbf{X}$). First, we formulate our hypotheses:

$$\mathbb{H}_0 : \text{for } i \in \{1, 2, ..., n\}, X_i \sim i.i.d(Bernoulli(p))$$

$$\mathbb{H}_1 : \text{for } i \in \{1, 2, ..., n\}, X_i \not\sim i.i.d(Bernoulli(p))$$

For $p = \frac{1}{2}$, we can use the symmetry of the binomial distribution to generate our p-values. However, for $p \neq \frac{1}{2}$, our distribution will be skewed, and so the following formula is used.

$$P_{value} = \sum_{k=0, \mathbb{P}(H(X)=k) \leq \mathbb{P}(H(X)=h(x))}^{n} \mathbb{P}(H(X) = k)$$

8 Multinomial Test Theory

The multinomial test is derived as follows. There are four possibilities when tossing two coins, they are **11, 10, 01,** and **00**, where 1 denotes a head, and 0 a tails. To maintain independence of the frequencies of these variables, the sum of these frequencies will be $\frac{n}{2}$, by considering the 1st and 2nd, the 3rd and 4th toss and so on. For example the sequence **1011** has 1 **10** and 1 **11** for the purpose of this test

Under individual coin tosses having identical Bernoulli distributions as well as being independent, these 4 possibilities are distributed according to the multinomial distribution. Allowing p to represent the probability of head in an individual coin toss and n the total number of tosses, we have

$$\mathbb{P}(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3, Y_4 = y_4) = \frac{(\frac{n}{2})!}{y_1! y_2! y_3! y_4!} (p^2)^{y_1} (p(1-p))^{y_2} ((1-p)p)^{y_3} ((1-p)^2)^{y_4}$$

where Y_1, Y_2, Y_3, Y_4 are the random variables representing the number of **11s, 10s, 01s and 00s** respectively.

We then formulate our hypotheses:

$$\mathbb{H}_0 : \text{for } i \in \{1, 2, ..., n\}, X_i \sim i.i.d(Bernoulli(p))$$

$$\mathbb{H}_1 : \text{for } i \in \{1, 2, ..., n\}, X_i \not\sim i.i.d(Bernoulli(p))$$

We can then generate p-values using the multinomial test. This test is applicable for this distribution since more 'extreme' values are those with lower probability. Define

$$\text{Z} = \{\frac{(\frac{n}{2})!}{z_1!z_2!z_3!z_4!}(p^2)^{z_1}(p(1-p))^{z_2}((1-p)p)^{z_3}((1-p)^2)^{z_4} : \sum_{i=1}^{4} z_i = \frac{n}{2} \wedge z_1, z_2, z_3, z_4 \in \mathbb{N}_0\}$$

Then, denoting $y = (y_1, y_2, y_3, y_4)$ as our observed vector, and pr(y) as the probability of having y as our observed value under the null hypothesis define

$$p_{value} = \sum_{z \in Z, z \leq pr(y)} z$$

Note that for $p=0.5$, there is a more simple representation of the p-value, but this definition is used to encompass all possible values for p

9   Last Equalisation Test Theory

The last equalisation test is derived as follows. Another statistic we can look at is the last time in which the number of head and tail tosses are the same, or in the context of our binary sequence, the last term in which the number of **0**s and **1**s up to and including that term are equal. For example the sequence **100111** has its last equalisation on term 4.

Another perspective to take on the last equalisation is considering the $(X_i)_{i=1}^{n}$ as defined previously as a Markov Process. However, in this context it is more useful to assign a value 1 to a head toss and value -1 to a tails toss. Then, we have a 1 dimensional random walk starting at the origin.

We can define

$$p_n(k) := \mathbb{P}(\text{A binary sequence of length n having last equalisation at term k})$$

For the purpose of the test we will need to calculate the distribution of $p_32(k)$ and $p_64(k)$ but the general derivation for even n will be given (with the distribution for n and n+1 being equivalent for even n, so the odd cases can be easily extracted). We note that for odd k $p_n(k) = 0$ since on odd terms either the number of heads will be odd and the number of heads even, or vice versa. In either case the number of heads and tails cannot be equal.

Next we have that

$$p_n(n) = \binom{n}{\frac{n}{2}} p^{\frac{n}{2}}(1-p)^{\frac{n}{2}}$$

. This is because for the random walk to end back at the origin after all n steps it is necessary and sufficient to have the same number of head and tail tosses.

We know, since there must be a last equalisation that for k>=1

$$\sum_{k=0}^{n/2} p_n(2k) = 1$$

and so

$$p_n(0) = 1 - \sum_{k=1}^{n/2} p_n(2k)$$

We also have that $p_n(2k) = p_{2k}(2k)p_{n-2k}(0)$ by the Markov Property. Using these two identities, we can get the distribution for the last equalisation. Our hypotheses remain the same in that the null hypothesis is that our sample consists of independently distributed Bernoulli variates. For $p=\frac{1}{1}$ the distribution for the last equalisation is symmetric, so the p-values for the 2-sided test can be computed

quite simply. However, for $p \neq \frac{1}{2}$, we must utilise a different method to calculate our p-values. We calculate them as follows

$$P_{value} = \sum_{k=0, p_n(2k) \leq p_n(y)}^{n/2} p_n(2k)$$

where y is our observed value of the last equalisation.

10   App Design

All tests this document covers are incorporated into a single R shiny app that automatically calculates the required p-values from an input binary string.
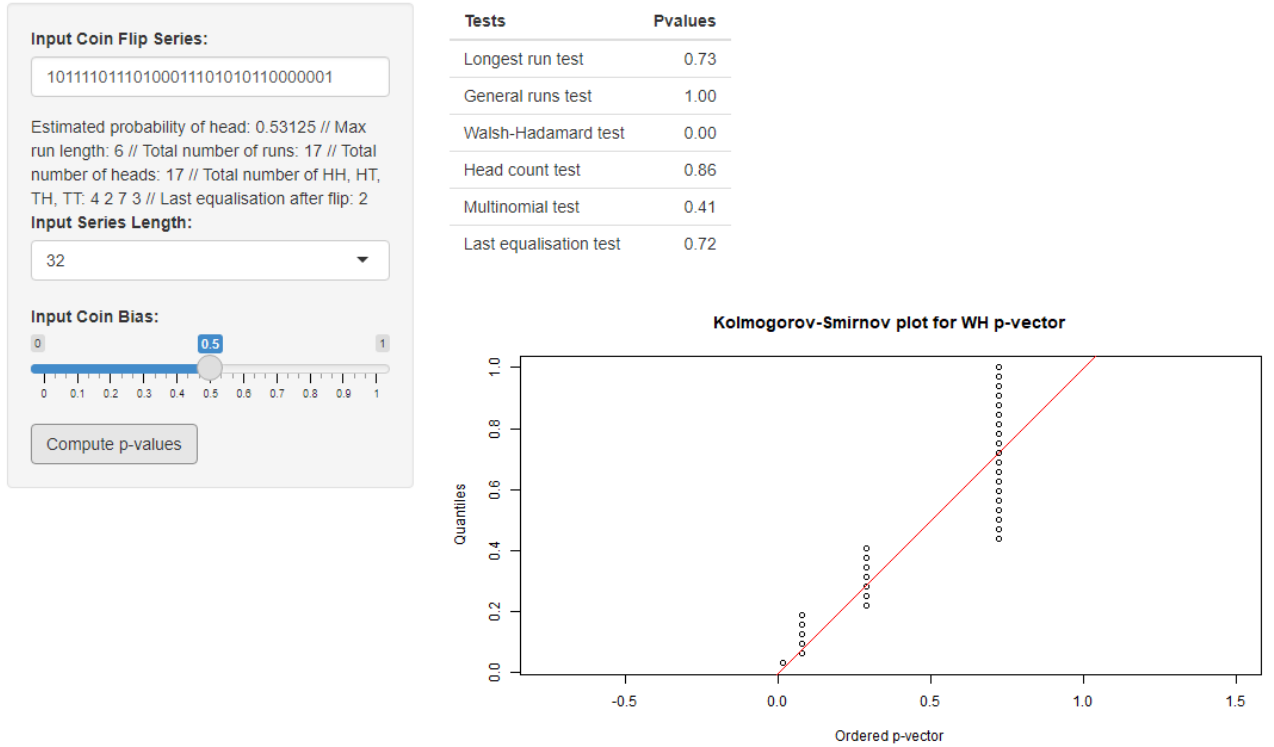

Figure 1: Screenshot of the app interface

   The app user introduces the inputs on the left side of the app (grey box), while the relevant outputs appear on the blank space on the right.
   The "Input Coin Flip Series" box accepts binary strings consisting of ones and zeros. The program treats any characters that are neither one nor zero as zeros. The "Input Series Length" box grants the user a choice between an input sample of length 32 or 64. If the input binary string has the incorrect length, the program will inform the user about the current length of the input string, so the user will know how many terms to add or subtract. The "Input Coin Bias" slider allows the user to set the bias of the Bernoulli distribution fitted to the input sample. The program will not compute the p-values if the set bias is an integer; the program will warn the user if it is an integer. The "Compute p-values" button will generate a table of p-values and a KS plot of the p-vector generated by the Walsh-Hadamard test on the output space.
   The p-value table shows the p-values from the longest run length test and the generalised runs test, and the measure of uniformity of the p-vector from the Walsh-Hadamard test. Overall, the closer the first two values are to one, the better fitted the Bernoulli distribution is to the input sequence. The

closer the red line on the plot is to the black line, the smaller the Walsh-Hadamard test metric is, and the more random the input sequence is.

11   Evaluation   Conclusion

The final project app functions perfectly; however, it takes significantly longer to calculate p-values for samples of length 64 than 32. The excess processing stems from the calculation of the CDFs of both the max run test and the general runs test, requiring a large number of computations to complete. One can reduce the program computation time by deriving the limiting distributions of the discrete CDFs, allowing for significantly faster calculations that way. It would be ideal also to add time series methods to test the independence between sample entries, though time restrictions prevented the realisation of such implementation. All in all, the project was successful.

12   Bibliography

M. Bar-Hillel (1991), 'The perception of randomness', Advances in Applied Mathematics, Vol 12, Issue 4, 428-454

J. V. Bradley (1960), Distribution-Free Statistical Tests, Ohio, United States Air Force, 195-215

I. Fazekas, Z. Karacsony, Z. Libor (2010), 'Longest runs in coin tossing. Comparison of recursive formulae, asymptotic theorems, computer simulations', Acta Universitatis Sapientiae, Mathematica, Vol 2, No. 2, 215-228

H. Liero, S. Zwanzig (2011), Introduction to the Theory of Statistical Inference, CRC Press, Taylor Francis Group

A. M. Mood (1940), 'The Distribution Theory of Runs', Annals of Mathematical Statistics, Vol 11, No. 4, 367-392

A. G. Oprina, A. Popescu, E. Simion, G. Simion (2009), 'Walsh-Hadamard Randomness Test and New Methods of Test Results Integration', Bulletin of the Transilvania University of Brasov, Series III, Vol 2, No. 51, 93-106

T. Read, N. Cressie (1988), Goodness-of-Fit Statistics for Discrete Multivariate Data, Springer-Verlag
M. F. Schilling (1990), 'Longest Run of Heads', The College Mathematics Journal, Vol 21, Issue 3, 196-207