Can One Fake Randomness?

Chad Franzen, Rishabh Marya, Robert Weber, A.J. Hildebrand (Faculty Mentor)

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Project Goals

- *Create a set of interactive online experiments and tools to test a user's ability to simulate true randomness and to test user-supplied data for randomness.*
- *Develop a suite of statistical tests that can serve as effective "fake randomness detectors", optimized for each experiment.*
- *Potential applications include educational uses, testing real-world data such as win/loss sequences in sports for randomness, and the use as a fraud detection tool.*

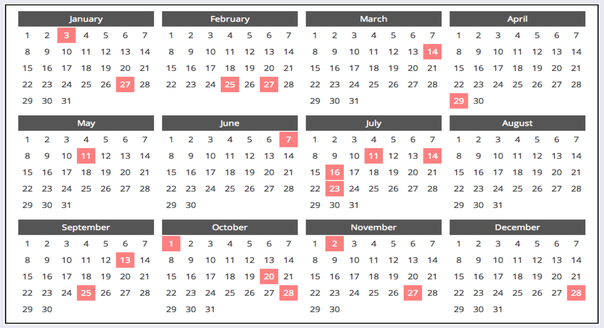
Quantifying Randomness

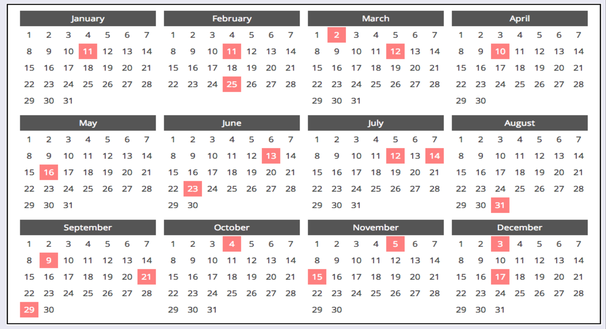
To quantify randomness and identify "fakeness" in a binary sequence, we are using a variety the statistical tests originally developed for testing random number generators (Knuth, 2005; Marsaglia, 1996). Our current implementation employs the following tests:

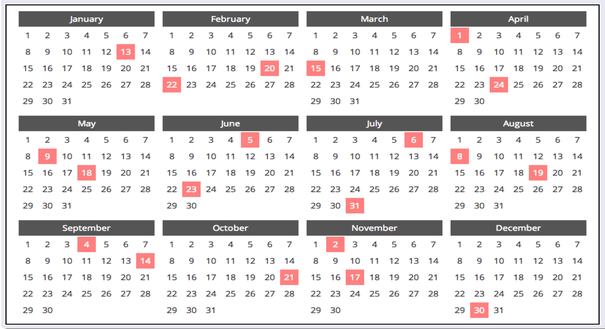
- **Frequencies of blocks.** The string is broken up into non-overlapping blocks of length 3 and 4, the frequencies for each of the 8 blocks of length 3 and 16 blocks of length 4 are computed, then a chi-squared test is run to determine a p-value for each block size.
- **Coupon Collector Test.** The coupon collector test runs along the string, reading substrings of length 4, attempting to "collect" all 16 different strings.
- **Frequencies of gaps between 1's.** Frequencies of gaps of sizes $0, 1, 2$ and ≥ 3 are computed and the chi-squared statistic for these frequencies is calculated.
- **Alternation test.** Humans are more likely to generate blocks "1010" and "0101" than computers. We count the number of these blocks in the string.
- **Calculating the Fakeness Score.** Each of the above tests produces a p-value, the probability that a truly random string would produce a result as extreme as the one found. We combine these values into a single value $\hat{p}$, defined as the harmonic mean of the p-values. The harmonic mean has the effect of attaching more weight to the extreme p values, and less weight to non-extreme ones. We define our fakeness score as $f = 1 - \hat{p}$.
- **Interpreting the Fakeness Score.** The fakeness score represents the likelihood that the sequence is not random: On average, only about 13% of computer-generated random sequences have a fakeness score of 0.9 or greater and only 6% of strings have a fakeness score of 0.95 or greater.

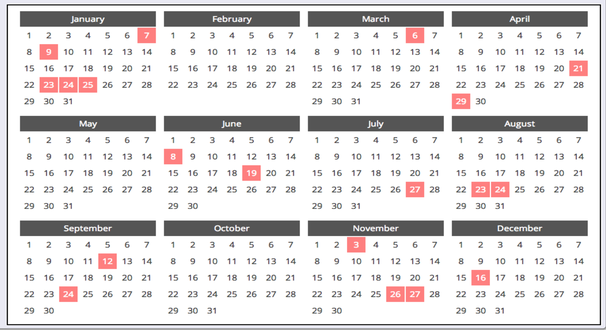
Can You Spot the Fakes?

Below are 4 calendars with 20 dates selected. One was randomly generated by a computer, one is the birthdays of our Math 199 CHP class, and two were chosen by humans trying to imitate randomness. Can you tell which is which?





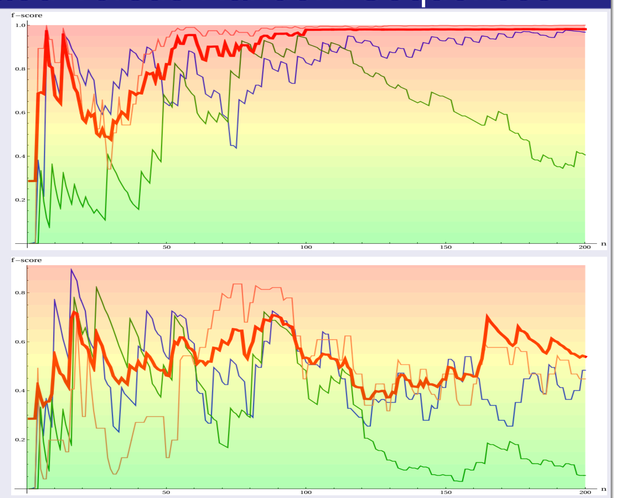




Fakeness Data: Human-Generated Versus Random Sequences

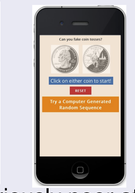
Source	Fakeness Score	3-Blocks Test	4-Blocks Test	Gaps Test	Alternation Test	Coupon Test
Computer	0.146	0.097	0.262	0.008	0.161	0.160
Computer	0.164	0.004	0.306	0.271	0.045	0.106
Computer	0.169	0.035	0.012	0.341	0.191	0.179
Computer	0.232	0.124	0.549	0.022	0.045	0.075
Computer	0.280	0.059	0.306	0.104	0.161	0.531
Computer	0.365	0.021	0.459	0.223	0.469	0.454
Human	0.394	0.124	0.549	0.104	0.419	0.513
Computer	0.408	0.124	0.359	0.341	0.555	0.494
Computer	0.460	0.007	0.459	0.638	0.555	0.284
Computer	0.461	0.004	0.103	0.081	0.756	0.494
Computer	0.474	0.124	0.505	0.428	0.469	0.632
Computer	0.477	0.007	0.054	0.295	0.756	0.494
Computer	0.537	0.081	0.598	0.022	0.631	0.714
Computer	0.562	0.124	0.405	0.842	0.191	0.011
Computer	0.589	0.046	0.638	0.175	0.775	0.632
Computer	0.621	0.059	0.598	0.104	0.469	0.850
Computer	0.637	0.035	0.035	0.318	0.883	0.413
Human	0.769	0.870	0.306	0.175	0.631	0.884
Human	0.953	0.279	1.000	0.223	0.523	0.090
Human	0.966	0.925	0.220	0.968	0.992	0.413
Human	0.984	1.000	0.944	1.000	1.000	0.769
Human	0.985	0.999	0.924	1.000	1.000	0.950
Human	0.986	0.978	0.989	1.000	1.000	0.924
Human	0.988	1.000	1.000	1.000	1.000	0.371
Human	0.988	0.998	0.994	1.000	1.000	0.513
Human	0.988	1.000	1.000	1.000	1.000	0.601
Human	0.988	1.000	1.000	1.000	1.000	0.924
Human	0.989	1.000	1.000	1.000	1.000	0.976
Human	0.989	0.999	0.984	1.000	1.000	0.992
Human	0.990	1.000	1.000	1.000	1.000	1.000
Human	0.990	1.000	1.000	1.000	1.000	1.000
Human	0.990	1.000	1.000	1.000	1.000	0.993

The table shows the numerical fakeness scores for each of our tests for 16 strings of 200 heads and tails generated by Math 199 CHP students instructed to try to mimick true randomness, along with 16 computer generated strings of the same length.



The graph on top shows the overall fakeness score (red graph) and the fakeness scores of some of the individual tests for a human generated string of 200 heads and tails by a Math 199 student; the bottom graph shows the corresponding scores for a computer generated string.

Try our Fake Coin Toss Detector



Scan the QR code to try our fake coin toss app. Can you make it up to 200 coin tosses without triggering our "fakeness" detector?

Humans are notoriously poor at mimicking randomness (Bakan, 1960), and our data confirms that (see table on left). Very few are able to generate 200 truly fake coin tosses.

How it works

- User inputs a sequence of heads (H) and tails (T) by clicking on one of the coins shown.
- Program applies a series of randomness tests to this sequence and calculates an overall fakeness score, as described on the left.
- Program provides continuous feedback to the user in the form of a fakeness probability, with associated color codes, until the sequence reaches a length of 200.

Fakeness Scores of Famous Sequences

Real World Data	Fakeness Score
Binary Expansion of Golden Ratio	0.796
Binary Expansion of √2	0.401
Binary Expansion of √3	0.644
Copeland Erdös Constant	0.599
2013 Arizona Diamondbacks	0.275
2011-2012 Ottawa Senators	0.709
2012-2013 Boston Celtics	0.306
Super Bowl Coin Toss	0.389

Fakeness scores of sequences of binary digits of mathematical constants, win/loss sequences of sports teams, and the sequences of coin toss results of Super Bowls. All of these scores were consistent with scores produced by computer-generated random sequences.

References

- Bakan, P. Response-tendencies in attempts to generate random binary series. *American Journal of Psychology*, 1960, **73**, 127-131.
- Knuth, Donald Ervin. *The art of computer programming.* Pearson Education, 2005.
- Marsaglia, George. "DIEHARD: a battery of tests of randomness." (1996).

Answers to Calendar Experiment

Upper left: computer generated
Lower left: human generated

Upper right: human generated
Lower right: class birthdays

Poster produced as part of class project in Math 199 CHP (Spring 2014)