Bayesian Approach to Hypothesis-Based Randomness Estimation

Alexander Rozenkevich[1]
Adam Street, Building 4, Apartment 3, Jerusalem, Israel

Abstract

A Bayesian method for dynamic hypothesis-based randomness estimation of a sequence of experimental data is proposed. Examples of pseudorandom number generator testing are given.
Keywords: Bayesian method.

The prior probability of getting heads or tails is 0.5. Therefore, with repeated tossing, the number of heads or tails should be approximately the same. If heads came up five times in a five-fold toss, tails are intuitively expected in the next toss. Tossing a coin is a physical process, and it is impossible to take into account all the influencing factors. Previously, it was believed that the ideal random number generator is tossing a “fair” coin, but researchers [1] in 2023 (!) as a result of 350,757 tests proved that, as a rule, coins fall on the same side they started with.

This paper examines the possibility of using the Bayesian method, based on experimental data and observations, to easily and reliably identify and evaluate discrepancies with theory, which in turn can guide the search for physical causes of such discrepancies.

Observations of coin tosses (event B) can be taken into account when calculating the posterior probability, confirming or not confirming the hypothesis of one side or the other of the coin in the next N+1 tosses. In the Bayesian method, the posterior probability is:

$$P(R/B)=\frac{p(R)\cdot p(B/R)}{p(R)\cdot p(B/R)+p(O)\cdot p(B/O)} \qquad (1)$$

$$P(O/B)=\frac{p(O)\cdot p(B/O)}{p(R)\cdot p(B/R)+p(O)\cdot p(B/O)} \qquad (2)$$

Here: $P(R/B)$ and $P(O/B)$ are the conditional (a posteriori) probabilities of confirming the hypotheses, namely: the next N+1 coin tosses will result in tails or heads, respectively, provided that the results of N observations (event B) are true;

– $p(R)$ and $p(O)$ are the a priori probabilities of tails and heads, respectively;

– $p(B/R)$ and $p(B/O)$ are the conditional probabilities that events R and O, respectively, will occur.

In the general case, for a polyhedron with S sides, the posterior probability of the hypothesis of the appearance of the face Sk in N+1 observations is:

$$P(S_k/B_k) = \frac{p(S_k)\cdot p(B_k/S_k)}{\sum_{1}^{S} p(S_k)\cdot p(B_k/S_k)} \qquad (3)$$

Here: $p(S_k)$ is the a priori probability of the face S_k, for a regular polyhedron - 1/ S ;

$p(B_k/S_k)$ - the influence of the observation results on the confirmation of the hypothesis:

$$p(B_k/S_k) = \sum_{i=1}^{N} \frac{1}{(S_{ki}+1)} \qquad (4)$$

where N is the number of observations of the face S_k in N trials.

S_{ki} = 1 when the face is observed and S_{ki} = 0 when it is not.

The inverse frequency in formula (4) suggests that the given facet should not appear in the next observation, and the unit in the denominator confirms the hypothesis through recursion.

Figure 1 shows graphs of the results of testing popular pseudo-random number generators (simulation of tossing a "fair" coin) in C++: Xorosshiro128+, Linear_Congruential, Squares_RNG and Mersenne_Twister. The X-axis shows the number of observations, 100 in total, and the Y-axis shows the ratio of the

natural logarithms of the probability of hypotheses for two sides - Ln(P1/P2), calculated using formulas (3) and (4). The prior probability for both sides is taken to be 0.5.

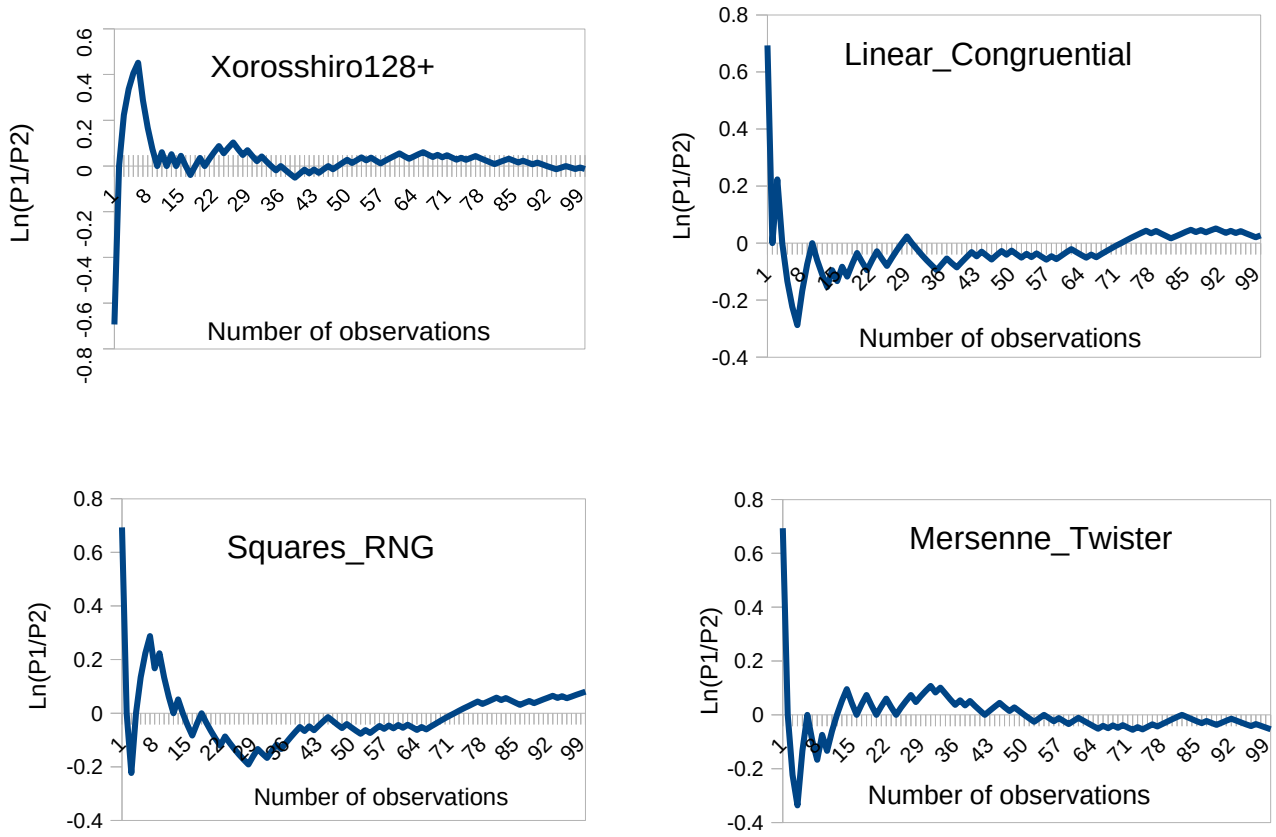

Fig. 1 Dependence of the logarithm of the ratio of probabilities of hypotheses on the number of observations

Figure 2 shows correlograms, graphs of the function of dependence of the autocorrelation coefficients (ACF) of the logarithms of the probability of hypotheses on the lag.

The table summarizes the results of testing hypotheses on the ACF, the fulfillment of the law of the iterated logarithm (LIL law) [5], the frequency of sign changes and the relative pseudo-period of randomness (local period of ergodicity). The law of the iterated logarithm was tested using the formula:

$$\frac{\left|\sum \ln(P1/P2)\right|}{\sigma\sqrt{2n\ln\ln(n)}}$$

here: σ^2 – is the variance, n is the number of observations, in our case 100.

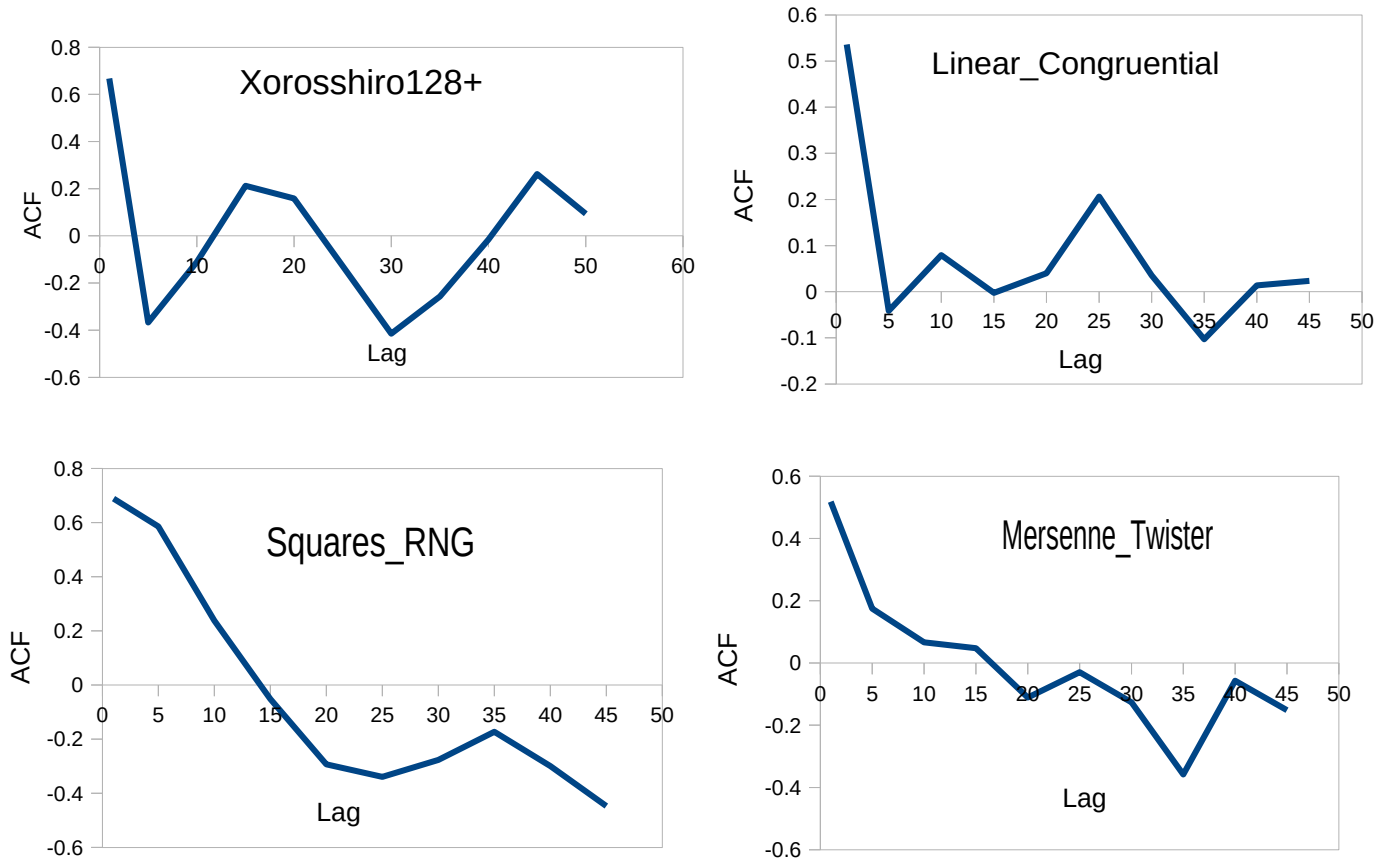

Fig.2 Correlograms, autocorrelations of the ratios of the logarithms of hypotheses (ACF).

Generators	*ACF*	σ^2	*Law LIL*	*Frequency of sign change*	Period of ergodicity
Xoroshiro128+	≠ 0	*0.012*	1.6363	*26*	*1*
Linear Congruential	≠ 0	*0.093*	1.3050	*12*	*0.46*
Mersenne_Twister	≠ 0	*0.089*	1.6507	*20*	*0.77*
Squares _RNG	≠ 0	*0.013*	0.5728	*24*	*0.92*

According to the table, the autocorrelation check (ACF) shows a lack of randomness in the sequence of hypothesis logarithms. Furthermore, a deviation from the law of the iterated logarithm (LIL) indicates a connection between the internal parameters of two mutually exclusive hypotheses.

In our view, the regularity and absence of random parameters in the hypotheses enable an evaluation of the efficiency and quality of generators by examining the stability of hypothesis logarithm oscillations (Fig. 2)

Here are the signs indicating the presence of structures in a random number sequence:

The lower the dispersion of wandering logarithms (as per the LIL law), meaning smaller deviations from 1, the higher the predictability of the numbers;

A low frequency of sign changes at large amplitudes suggests a hidden structure, whereas a high frequency of sign changes points to a high pseudo-random period, or a longer local ergodicity period;

Prolonged oscillations above or below zero indicate one hypothesis 'dominates' the other, which implies an imbalance in the generator;

In a truly random sequence, the oscillations of hypothesis logarithms should cross zero frequently and avoid staying within any single region for extended periods. Conversely, prolonged waves of oscillations may indicate trends or patterns, pointing to hidden or explicit structures.

Based on these criteria, we conclude that all tested generators meet the requirements of pseudorandom number generators, albeit with varying performance quality. The best results were achieved by Xoroshiro128+ and Squares_RNG.

The proposed method for the dynamic analysis of hypothesis logarithm oscillations, grounded in experimental and observational data, is not limited to testing generators.

References

[1]. Fair coins tend to land on the same side they started: Evidence from 350,757 flips arXiv:2310.04153v3

[2]. "Bayes Theorem - Formula, Statement, Proof | Bayes Rule". *Cuemath*. Retrieved 2023- 10-20

[3]. *"Bayes' Theorem: Introduction"*. *Trinity University*. Archived from the original on 21 August 2004. Retrieved 5 August 2014.

[4]. ttp://csrc.nist.gov/groups/ST/toolkit/rng/stats_tests.html

[5]. https://en.wikipedia.org/wiki/Law_of_the_iterated_logarithm

[6]. Вентцель Е.С. Теория вероятностей. М.:Высш.шк. 2001

[1] Email: alexroz2008@gmail.com