数理解析研究所講究録 1240 巻 2001 年 161-171

Deviation of pseudorandom numbers and weight test*

松本 眞 (Makoto Matsumoto)[1] and 西村 拓士 (Takuji Nishimura)[2]

[1] 京大・総合人間学部
Faculty of IHS, Kyoto University, Kyoto 606-8501, JAPAN
matumoto@math.h.kyoto-u.ac.jp
[2] 山形大・理学部
Faculty of Science, Yamagata University, Yamagata 990-8560, JAPAN
nisimura@sci.kj.yamagata-u.ac.jp

Abstract. We introduce a theoretical test, named *weight discrepancy test*, on pseudorandom number generators. This test measures the χ^2-discrepancy between the distribution of the number of ones in some specified bits in the generated sequence and the binomial distribution, under the assumption that the initial value is randomly selected.

This test can be performed for most generators based on a linear recursion over the two-element field $\mathbb{F}_2$, and predicts with high precision for which sample size the generator will be rejected by a classical statistical test called the weight distribution test.

This test may be considered as a theoretical version of a one-dimensional random walk test. Differently from the empirical tests which can reject only very bad generators, this test assigns a ranking to generators. Thus it is useful to select good generators, similarly to the spectral tests and the k-distribution tests. This test rejects practically all generators linear over $\mathbb{F}_2$ that are known to fail in some physical tests although they pass k-distribution tests.

1 Necessity of a theoretical test on weight

Among the numerous pseudorandom number generators available, some are known to be defective, and some seem to be good.

For more than thirty years, GFSRs[11] based on three-term relations are known to suffer from statistical nonsymmetricity between 0 and 1, and to be rejected by χ^2-test on the goodness-of-fit to the binomial distribution [12][6][3][18][16].

However, these warnings were not loud enough to reach the users. These three-term GFSRs were introduced to the computational physics community by [8] suggesting the recursion $x_j = x_{j-103} \oplus x_{j-250}$, and became fairly popular. In middle 80's, physicists began to find the failure of these generators in simulations of physical models, such as Ising models [5][2][1] and random walks [4]. These physical models are simplified and proposed as tests of randomness in [23], which we call physical tests here.

In these works some physicists proposed two ways of improving GFSR: one is to increase the degree of the recurrence (e.g.[2]), and the other is to use five or more-term relations (e.g. [24]). These follow from an intuitive observation that few-term relations in a short range should lead to a deviation, and that increasing the number of terms or the range of the correlation will improve the deviation. These improvements are shown to be effective in the physical tests.

However, it is not clear which degree will be sufficient, or how many terms are enough for the required randomness. Five-term relations of degree 89 behave well for 10^6 samples, but are rejected for 10^8 samples by a random walk test [23]. A five-term relation of degree 1279 passed the test even for 10^9 samples. Is this enough? The computational power of the machines is increasing rapidly. Will some defect of such a generator be revealed in future? Or is it impossible for any future machine?

These physical tests are interesting in that they clearly exhibit the defects of random number generators in practical computational physics. However, they are not powerful enough to select good generators.

* The content of this manuscript is submitted to the proceedings of MCQMC 2000.

Actually, these tests reject only (1) three-term GFSRs, (2) five-term GFSRs with small degree, and (3) linear congruential generators with poor spectral properties. All these generators are known to fail in some simple statistical tests.

We shall introduce a theoretical test on the distribution of 1's and 0's in the bits of the sequence, named *weight discrepancy test*. This is not an empirical test but a figure of merit defined on the full period of the generator, like the spectral test[7] or the k-distribution test [9]. It predicts with high precision the sample size for which the generator is rejected by the weight distribution test, which is a classical empirical test equivalent to a random-walk test. For example, a generator (MT521) is shown to be quite safe since it would require 10^{156} samples to reject its output, whereas another generator called $R(11, 39, 95, 218)$ which passed all the physical tests in [24] is shown to be rejected if we take the sample size $> 600,000,000$ (see Table 7).

The weight discrepancy test gives an index δ which is a real number indicating the extent of the discrepancy between the distribution of the weight of the generated sequence and the ideal binomial distribution. Thus, smaller δ means better fit to the theoretical distribution, so we can choose the best one from a set of generators, even if they pass the physical tests. In this regard, our test is similar to the spectral test and the k-distribution test.

In the context of this discussion, we would like to point to a highly reliable random number generator. As mentioned above, physicists proposed the following improvements: (1) increase the number of terms [24], (2) increase the degree of recursion [2], and (3) use only a part of the sequence (decimation in [24] or discarding in [14]). Although these are effective, there is a generator adopting more advanced improvements, named Mersenne Twister (MT) [19]. This has period $2^{19937} - 1$, good k-distribution property, is based on more than 100-term relations, consumes only 624 words memory, passes practically all reasonable tests, and is very fast. Implementations in C, Fortran, and other languages are available from the following URL.

`http://www.math.keio.ac.jp/matumoto/emt.html`

2 χ^2-discrepancy

We begin with recalling the well-known χ^2-test for goodness-of-fit. Let Z_i $(i = 1, 2, \ldots, N)$ be independent identically distributed random variables conforming to the same discrete distribution such that the value k $(k = 0, 1, 2, \ldots, \nu)$ is taken with probability p_k (thus $p_k \geq 0$ and $p_0 + p_1 + \cdots + p_\nu = 1$).

Let $b_1, b_2, \ldots, b_N \in \{0, 1, \ldots, \nu\}$ be a sequence, computed from a pseudo random sequence, to mimic a sample sequence conforming to the random variables $Z_1, Z_2, \ldots, Z_N$. Our question is whether the null hypothesis H_0 that this sample comes from the random variables is justified or not.

We count the number of k among $b_1, \ldots, b_N$, and let it be Y_k for $0 \leq k \leq \nu$:

$$Y_k := \text{the number of } i \ (i = 1, 2, \ldots, N) \text{ with } b_i = k. \tag{1}$$

These are random variables which conform to binomial distribution $B(N, p_k)$ under the null hypothesis H_0. We compute the χ^2-value $\mathcal{X}$ by

$$\mathcal{X} := \sum_{k=0}^{\nu} (Y_k - Np_k)^2 / Np_k, \tag{2}$$

which measures a kind of discrepancy between the observed numbers Y_k and the expected value Np_k. Under the null hypothesis, it is known that this $\mathcal{X}$ is a random variable which approximately conforms to the χ^2-distribution with ν degrees of freedom, regardless of p_k.

Let $\mathcal{X}_b$ be the realization of $\mathcal{X}$ for an observed sample $b := (b_1, \ldots, b_N)$, and let χ^2_ν denote the random variable which conforms to the above χ^2-distribution. We compute the probability value

$$\text{Prob}(\chi^2_\nu < \mathcal{X}_b). \tag{3}$$

If this value is, say, $> .99$, then such a large value of $\mathcal{X}_b$ appears with probability $< .01$, thus the null hypothesis on the distribution of $b_1, \ldots, b_N$ is suspicious: the observed distribution is too far from the hypothetical distribution p_k. If this value is, say, .75, then such $\mathcal{X}_b$ appears with a moderate probability, so we do not reject the hypothesis.

Assume that our method to mimic Z_i has some deviation, and the probability to observe k in a trial is q_k, not p_k (independence is still assumed). We call this assumption the *nonnull assumption*. Then, $(b_1, b_2, \ldots, b_N)$ is a sample conforming to the distribution q_k, so Y_k in (1) is a random variable conforming to the binomial distribution $B(N, q_k)$ with expectation $E(Y_k) = Nq_k$ and variance $E(Y_k^2) - E(Y_k)^2 = Nq_k(1 - q_k)$.

Under this nonnull assumption, $\mathcal{X}$ in (2) approximately conforms to a *noncentral χ^2-distribution*. Recall its definition (see [21]). If $U_1, U_2, \ldots, U_\nu$ are ν independently normally distributed random variables, each having zero mean and unit standard deviation, and if $a_1, a_2, \ldots, a_\nu$ are ν constants, then

$$\chi'^2 := \sum_{i=1}^{\nu} (U_i + a_i)^2$$

is called a *noncentral χ^2-variate* having ν degrees of freedom with *noncentrality parameter* $\lambda := \sum_{i=1}^{\nu} a_i^2$. It is easy to check $E(\chi'^2) = \nu + \lambda$.

It is known (c.f. [21, P.279]) that the above $\mathcal{X}$ approximately conforms to the noncentral χ^2-distribution having ν degrees of freedom with noncentrality parameter $\lambda = N\delta$, where δ is the χ^2-discrepancy defined below.

Definition 1. We define the χ^2-discrepancy δ between the true distribution q_k and the expected distribution p_k by

$$\delta := \sum_{k=0}^{\nu} (q_k - p_k)^2 / p_k.$$

(The term χ^2-discrepancy appears in model selection theory, e.g. in [13].)

The expectation of $\mathcal{X}$ in (2) under the nonnull hypothesis is approximated by δ as follows.

Proposition 1. *$E(\mathcal{X}) \sim \nu + N\delta$. Here $\sim$ means that the absolute value of the difference is*

$$|E(\mathcal{X}) - (\nu + N\delta)| \leq \nu \max_{k=0,\ldots,\nu} \left|1 - \frac{q_k}{p_k}\right|,$$

and hence the error is negligible if $|1 - \frac{q_k}{p_k}| << 1$ for every k.

The first formula is implied by $E(\chi'^2) = \nu + \lambda$ and $\lambda = N\delta$ as stated above, but we show it by the following direct computation:

$$\begin{aligned} E(\mathcal{X}) &= \sum_{k=0}^{\nu} \frac{E((Y_k - Np_k)^2)}{Np_k} = \sum_{k=0}^{\nu} \frac{E(Y_k^2) - 2E(Y_k)Np_k + (Np_k)^2}{Np_k} \\ &= \sum_{k=0}^{\nu} \frac{E(Y_k^2) - E(Y_k)^2 + (E(Y_k) - Np_k)^2}{Np_k} \\ &= \sum_{k=0}^{\nu} \frac{q_k(1 - q_k)}{p_k} + N \sum_{k=0}^{\nu} \frac{(q_k - p_k)^2}{p_k} \simeq \sum_{k=0}^{\nu} (1 - q_k) + N\delta = \nu + N\delta. \end{aligned}$$

This supports the obvious fact that χ^2-test reveals the deviation if both χ^2-discrepancy δ and the sample size N are large.

The point is that a more quantitative analysis is possible. For $0 < p < 1$, $\mathcal{X}_p$ satisfying $\text{Prob}(\chi^2_\nu < \mathcal{X}_p) = p$ is approximated for large ν by the formula[1]

$$\mathcal{X}_p = \nu + \sqrt{2\nu}x_p + \frac{2}{3}(x_p^2 - 1) + o(\nu^{-\frac{1}{2}}), \tag{4}$$

where $x_p = 2.33$ for $p = .99$ and $x_p = 0.674$ for $p = .75$, (see e.g.[7]). Comparison of this with Proposition 1 yields the following

Theorem 1. *Let ν be moderately large, say $\nu \geq 5$. (For $\nu < 5$, we need to consult a table of χ^2-distribution.)*

1. *(Accepting sample size.) If the sample size N is small so that*

$$N \leq \frac{\sqrt{2\nu}x_p + \frac{2}{3}(x_p^2 - 1)}{\delta} \quad \text{for } x_p = 0.674,$$

then approximately $E(\mathcal{X})$ falls in the area with probability $p \leq .75$, and the χ^2-test will not reject the sequence.

2. *(Rejecting sample size.) If the sample size N is large so that*

$$N \geq \frac{\sqrt{2\nu}x_p + \frac{2}{3}(x_p^2 - 1)}{\delta} \quad \text{for } x_p = 2.33,$$

then approximately $E(\mathcal{X})$ falls in the area with probability $p > .99$, and the χ^2-test will reject the sequence.

Thus, the χ^2-discrepancy δ provides us with a guess at the sample size for which χ^2-test reveals the defect of the generator, as well as the size for which it does not.

Definition 2. A χ^2-discrepancy test means to obtain the χ^2-discrepancy δ for the simulation of the random variables Z_i by a pseudorandom number generator.

This test is similar to the spectral test(e.g. [7]) or to the k-distribution test(e.g. [9]), in the sense that it deals with the full-period behavior of the pseudorandom number generator, that it is not empirical, and that it gives a numerical estimate of the quality of the generator, differently from the statistical tests. The latter yield only probability values, which are sometimes confusing if they are on the border of .95 or 0.05, and differ every time we choose a new initial value.

3 Weight discrepancy test

3.1 Weight discrepancy test for $\mathbb{F}_2$-generators

We shall introduce a χ^2-discrepancy test on the distribution of 1's in the bits of the generated sequences, named a weight discrepancy test.

Let $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots$ be a pseudorandom sequence of w-bit integers generated by an $\mathbb{F}_2$-linear generator. Here, by an $\mathbb{F}_2$-linear generator we mean a machine (automaton) that has the p-bit state space $\mathbb{F}_2^p$, the linear state transition map $f : \mathbb{F}_2^p \to \mathbb{F}_2^p$, and the linear output function $b : \mathbb{F}_2^p \to \mathbb{F}_2^w$. Thus, we choose an initial seed $X_0 \in \mathbb{F}_2^p$, then generates a sequence of state vectors $X_0, X_1, X_2, \ldots$ by the recursion

[1] Knuth [7] presents this approximation for $\nu > 30$. An explicit computation shows that the maximal error ratio of this approximation for $p = 99\%$ and $\nu \geq 5$ is attained when $\nu = 5$, with true value 15.32 and the approximation 15.09. The error for $p = 75\%$ is even smaller. Thus, to guess the safe/dangerous sample sizes as in Theorem 1, $\nu \geq 5$ would suffice.

$X_{j+1} := f(X_j)$, and generate an output sequence of w-bit integers by $\mathbf{x}_j := b(X_j)$. This class of generators covers Tausworthe generators, M-sequences, GFSR, Combined Tausworthe, Twisted GFSR, Mersenne Twisters, and several other types.

We focus only on some bits of $\mathbf{x}$. For simplicity, we assume them to be the most significant s bits of $\mathbf{x}$, denoted by $\mathbf{x}^{(s)}$, although the following discussion makes sense for any choice of bits. We also fix a positive integer μ, and consider the distribution of the consecutive μ words of the sequence.

We shall count the number W (W for weight) of 1's appearing in the $m := s \times \mu$ bits in $\mathbf{x}_0^{(s)}, \mathbf{x}_1^{(s)}, \ldots, \mathbf{x}_{\mu-1}^{(s)}$. We assume that the initial seed is randomly uniformly chosen from the state space, consider W as a random variable, and look at the discrepancy between the distribution of W and the ideal binomial distribution. The term *weight* comes from coding theory where the (Hamming) weight $wt(\mathbf{x})$ of a vector $\mathbf{x} \in \mathbb{F}_2^m$ is defined as the number of 1's in the m components of $\mathbf{x}$.

We want to know the true distribution q_k of the weight W obtained from a pseudorandom number generator by the random selection of the initial seed. The exhaustive check of the seeds is intractable for generators with large state space, but for $\mathbb{F}_2$-generators one can compute q_k under some condition.

Assume that the generator is $\mathbb{F}_2$-linear. The mapping Φ from the state space to the $m := s \times \mu$ bits in the output is $\mathbb{F}_2$-linear, so the image $C \subset \mathbb{F}_2^m$ is a linear subspace. In coding theory we call C a linear code, and have the following terminology.

Definition 3. Let A_ℓ be the number of the vectors in C with weight ℓ $(0 \le \ell \le m)$, which is called the ℓ-th weight enumeration of C. We define the weight enumerator polynomial of C in indeterminates x, y by

$$W_C(x, y) := A_0 x^m + A_1 x^{m-1} y + \cdots + A_i x^{m-i} y^i + \cdots + A_m y^m.$$

Let r be the dimension of C. Since the mapping $\Phi : \mathbb{F}_2^p \to \mathbb{F}_2^m$ is linear, every vector in $C :=$ Image of Φ equally likely occurs under a random selection of the initial seed. Thus, the probability Q_ℓ that we have weight ℓ in the m-bits is given by

$$Q_\ell = A_\ell / 2^r, \tag{5}$$

where the desired probability P_ℓ of binomial distribution is

$$P_\ell = \binom{m}{\ell} / 2^m. \tag{6}$$

The problem is that the weight enumerations of C are in general intractable since it is an NP-complete problem [22]. So we use a fundamental theorem[15] in coding theory. Let us define an inner product on $\mathbb{F}_2^m$ by

$$\mathbf{y} \cdot \mathbf{x} = (y_1, \ldots, y_m) \cdot (x_1, \ldots, x_m) = y_1 x_1 + \cdots + y_m x_m \in \mathbb{F}_2, \tag{7}$$

and let $C^\perp \subset \mathbb{F}_2^m$ be the orthogonal dual to C, i.e. the subspace consisting of the vectors whose inner product with any vector in C is zero.

Theorem 2. (MacWilliams identity) *Let r be the dimension of $C \subset \mathbb{F}_2^m$. Then we have*

$$W_C(x, y) = \frac{1}{2^{m-r}} W_{C^\perp}(x + y, x - y).$$

Thus, if we have the weight enumerations of $C^\perp$, we know those of C. If we choose μ so that $m = \mu \times s$ is only slightly larger than r, then the dimension $m - r$ of $C^\perp$ is small enough that an exhaustive check for $C^\perp$ is possible. In particular, if $r = m$ or equivalently Φ is surjective, the weight enumerations of C coincide with the binomial coefficients, hence the discrepancy $\delta = 0$.

For approximations used in χ^2-test, the expectation Np_k for each k should not be too small, say at least five. For this, we need to group some low weights together, as well as high ones. We choose the

following categorization:

$$\begin{aligned} S_0 &= \{0, 1, \ldots, s_0\}, \\ S_k &= \{s_0 + k\} \quad (1 \leq k \leq m - 2s_0 - 1), \\ S_\nu &= \{m - s_0, m - s_0 + 1, \ldots, m\} \end{aligned} \tag{8}$$

for suitably chosen s_0, where $\nu := m - 2s_0$.

Let W be the random variable from the pseudorandom number generator as above. We define a random variable Z that takes the value k if W falls in S_k $(0 \leq k \leq \nu)$, and compute the χ^2-discrepancy for Z as in §2.

From (5) and (6), we have

$$q_k := \text{Prob}(W \in S_k) = \sum_{\ell \in S_k} A_\ell / 2^r, \tag{9}$$

$$p_k := \text{Prob}(W' \in S_k) = \sum_{\ell \in S_k} \binom{m}{\ell} / 2^m, \tag{10}$$

where W' is the random variable when the sequence is truly random. We choose s_0 so that $Np_0 = Np_\nu$ is not less than 1000.

3.2 Description of weight discrepancy test

Now we shall summarize the design of the weight discrepancy test.

1. Fix an $\mathbb{F}_2$-generator to test.
2. Determine which bits to test in each output word, say, s most significant bits.
3. Determine μ, for which we test the distribution of the s bits in consecutive μ words. Put $m := s \times \mu$.
4. Take a linear basis $\{\sigma_1, \ldots, \sigma_p\}$ of the initial seed space. For each σ_i, initialize the generator with seed σ_i and generate μ words of the corresponding output. Let Σ_i be the m-dimensional vectors consisting of m-bits in the output sequence. Let $C \subset \mathbb{F}_2^m$ be the span by Σ_i $(1 \leq i \leq p)$, and r be its dimension.
5. Compute a basis of the dual space $C^\perp$ of C. Obtain the weight enumerations $B_0, B_1, \ldots, B_m$ of $C^\perp$ by exhaustive enumeration. If the dimension $m - r$ of $C^\perp$ is too large to do an exhaustive check, then make μ smaller. If it is too small, then the power of the test is weaker, so make μ larger. Note that often $r = p$, so we can make a guess that the dimension of $C^\perp$ is $\mu s - p$.
6. By the MacWilliams identity, obtain the weight enumeration A_ℓ of C. Compute q_k, p_k by (9), (10), then the χ^2-discrepancy δ as in Definition 1. We obtain the safe and dangerous sample sizes by Theorem 1.

   An explicit formula for δ is given as:

$$\begin{aligned} \delta = &\frac{[\sum_{j=1}^m (\sum_{\ell=0}^{s_0} M_{\ell j}) B_j]^2}{2^m \sum_{\ell=0}^{s_0} \binom{m}{\ell}} + \sum_{\ell=s_0+1}^{m-s_0-1} \frac{[\sum_{j=1}^m M_{\ell j} B_j]^2}{2^m \binom{m}{\ell}} \\ &+ \frac{[\sum_{j=1}^m (\sum_{\ell=m-s_0}^{m} M_{\ell j}) B_j]^2}{2^m \sum_{\ell=m-s_0}^{m} \binom{m}{\ell}}, \end{aligned} \tag{11}$$

   where M_{ij} is defined by

$$(x + y)^{m-j} (x - y)^j = \sum_{i=0}^{m} M_{ij} x^{m-i} y^i. \tag{12}$$

Here we mention that in the case of $s = 1$, the use of the weight of the dual space is introduced by [6]. Our method generalizes this concept and combines it with the χ^2-discrepancy to obtain a statistical test.

3.3 Weight distribution test and random walk

As explained in §2, χ^2-discrepancy test is designed to make a prediction on the result of the empirical χ^2-test. In the case of the weight discrepancy test, the corresponding empirical test is a classical test sometimes called the weight distribution test (c.f. [16]), which we shall briefly recall.

Fix s, μ, N, s_0 as in §3.1. Choose an initial seed, and generate μ words of pseudorandom number sequences. Look the s bits in each words, and let W_1 be the number of ones in $m = \mu \times s$ bits. Then again generate μ words, count the number of ones and let W_2 be this number. Iterate this N times to obtain $W_1, W_2, \ldots, W_N$. If the sequence is truly random, this should conform to the binomial distribution. We apply the χ^2-test to these N samples, using the categories (8). We obtain one value of the χ^2-statistics, and the final result is the corresponding probability value.

A slight difference on the assumption from the weight discrepancy test is that in weight distribution test we initialize the generator only once, not on every μ-th generations. This seems not significant, because the state of a usual pseudorandom number generator transits as if the next state is uniformly randomly selected. This expectation is confirmed by experiments in the next section. We will see that the results of the weight distribution tests are in close accordance with the forecasts obtained by the weight discrepancy tests.

Note that this test is nothing but a one-dimensional random walk test for $s = 1$, where a moving point starts at the origin of the real line, moves to the right or left by one according to the most significant bit of the generated number is 1 or 0, respectively. After $m = \mu$ steps, the final position is $W - 2m$ where W is the weight of the collection of the most significant bits in the m consecutive words. This type of simple random-walk test is an essence of all physical tests including Ising models, as explained in [24].

4 The result of tests

4.1 GFSRs

The first example is a 3-term GFSR of degree 89, based on the recursion $x_{j+89} := x_{j+38} + x_j$ over $\mathbb{F}_2$, whose period attains the maximal $2^{89} - 1$. We look only at the most significant bit, i.e., put $s = 1$, and look at the $m = 94 = 89 + 5$ consecutive words. For $s = 1$, $r = m$ holds if $m \leq p$ and $r = p$ holds if $m > p$. Thus the dimension of $C^\perp$ is $m - r = 5$, and the exhaustive check of $C^\perp$ is easy. The result of the weight discrepancy test is shown in Table 1, where $\nu = 30$ denotes the degree of freedom, from which the categorizing parameter s_0 in (8) can be computed by $\nu = m - 2s_0$ (i.e. $s_0 = 32$). The column δ shows the χ^2-discrepancy, the column "safe," "risky" respectively shows the safe, risky sample size implied by Theorem 1. Thus if the sample size is less than 25,000 then the sequence will not be rejected in average, but if it is more than 120,000 then the sequence will be rejected with significance level 0.99 in average.

Table 1. Weight discrepancy test on the generator $x_{j+89} := x_{j+38} + x_j$

m	ν	δ	safe	risky
94	30	1.80×10^{-4}	2.69×10^4	1.16×10^5

We also empirically test the same generator by the weight distribution test with the same parameters, and show the result in Table 2. We choose five different initial values randomly, and tested the generator for 3 different sample sizes N, namely, 25,000, 120,000, and 500,000. The weight discrepancy test predicts that $N = 25,000$ will pass, but $N = 120,000$ will be rejected with probability value .99 in average. Since Proposition 1 shows that $E(\mathcal{X})$ will increase linearly in N, $N = 500,000$ will be definitely rejected. The empirical results of five tests are in good accordance.

Table 2. Weight distribution test on the same GFSR with Table 1

N	1st	2nd	3rd	4th	5th
2.5×10^4	30.2%	61.4%	62.2%	83.9%	26.3%
1.2×10^5	99.3%	88.4%	99.8%	85.3%	99.9991%
5.0×10^5	100%	100%	100%	100%	100%

Table 3. Weight discrepancy test on a 5-term GFSR of degree 89

m	ν	δ	safe	risky
94	30	3.01×10^{-7}	1.62×10^7	6.99×10^7

Table 3 shows the same result on a generator based on a five-term relation $x_{j+89} := x_{j+57} + x_{j+23} + x_{j+15} + x_j$ with the same period. By comparing Tables 1 and 3, we see the effect of increasing the number of terms as the decrease of δ by a factor of roughly $1/600$, and consequently as the increase of the safe and risky sample sizes by the factor of 600 in this example. Table 4 shows the corresponding empirical weight distribution tests for safe and risky sample sizes, namely $N = 16,000,000$ and $70,000,000$, which again show a good accordance with the weight discrepancy test. Next example is same type of generator

Table 4. Weight distribution test on the same GFSR as in Table 3

N	1st	2nd	3rd	4th	5th
1.6×10^7	51.4%	94.5%	56.6%	77.4%	14.0%
7.0×10^7	97.7%	65.1%	99.8%	99.1%	99.3%

with degree 521 and period $2^{521} - 1 \simeq 6.86 \times 10^{156}$. To see the effect of the number of terms, we searched

Table 5. Weight discrepancy tests on GFSR of degree 521, with eight different numbers of nonzero terms

# of terms	safe N	risky N	min. weight
3	7.54×10^6	3.05×10^7	3
5	1.97×10^{11}	7.98×10^{11}	5
15	4.92×10^{28}	1.99×10^{29}	15
25	6.11×10^{42}	2.47×10^{43}	25
51	6.96×10^{71}	2.82×10^{72}	51
99	3.94×10^{109}	1.59×10^{110}	99
157	2.41×10^{138}	9.74×10^{138}	157
259	3.46×10^{156}	1.40×10^{157}	246

Table 6. Weight distribution test on the first generator in Table 5

N	1st	2nd	3rd	4th	5th
7.5×10^6	99.0%	20.0%	83.9%	92.0%	60.2%
3.1×10^7	100%	99.9%	100%	99.4%	94.9%

for eight primitive polynomials with $3, 5, 15, 25, 51, 99, 157, 259$ terms, respectively. We apply the weight discrepancy test on the most significant $s = 1$ bit for the consecutive $m = 526 = 521 + 5$ words. Table 5

shows the number of terms, the safe sample sizes, the risky sample size, and the minimum weight of the dual space for these eight generators. Table 6 shows the result of the weight distribution tests for the 3-term generator written in the first row in Table 5, confirming the accordance.

This example illustrates the power of weight discrepancy test. The 3-term generator at the first low will be rejected only if the sample size is more than 10^7, but it would take time and effort to notice this by experiments. Some researches reported that 5-term relations with degree 521 seem defectless, but our result shows that for sample sizes larger than 8×10^{11}, it will be rejected. This size seems large enough for present computers, but may be not in future. On the other hand, it seems very difficult to reject the 15-term generator in near future, since it will require the sample size at least 5×10^{28}. To reject 259-term generators, it requires the sample size N roughly the same order as the period. It is impossible to deduce this kind of result from empirical tests. Also, it is noteworthy that the ratio between safe and risky sample sizes is only about four, which seems rather tight.

The above results suggest that the increase of the number of terms implies the exponential decrease of discrepancy δ. An intuitive account for this is as follows. According to explicit computations, it seems that $M_{\ell,j}$ in (12) satisfy the convexity

$$|M_{\ell,1}| >> |M_{\ell,2}| >> |M_{\ell,3}| >> \cdots << |M_{\ell,m-2}| << |M_{\ell,m-1}| << |M_{\ell,m}|$$

for near at the both ends (like $j \leq 5$ and $j \geq m-5$), for most of ℓ. For example, if $m = 94$ and $\ell = 20$, $M_{20,j} = M_{20,94-j}$ is $7.76 \times 10^{19}, 4.36 \times 10^{19}, 2.39 \times 10^{19}, 1.28 \times 10^{19}, 6.59 \times 10^{18}$ for $j = 1, 2, 3, 4, 5$, respectively.

This and (11) imply that the main terms in δ would come from the first nonzero weight enumeration B_d, where d is the minimum weight of $C^\perp - \{0\}$, or the last nonzero $B_{d'}$. If $C^\perp$ is an "average" subspace, then d is moderately large and d' is not near to m, as shown in the proof of Shannon's theorem on the existence of good codes. Now the definition of the dual and the inner product (7) implies that $C^\perp$ contains the coefficient vector of the defining relation. That is, if the pseudorandom bit sequence is generated by the recursion

$$x_{j+n} = \sum_{i=0}^{n-1} a_i x_{j+i},$$

then an m-dimensional vector $(-1, a_{n-1}, a_{n-2}, \ldots, a_1, a_0, 0, \ldots, 0)$ obtained from the coefficient vector by supplementing 0's at the right (we assume $m > n$) lies in $C^\perp$ (also its right-shifts as well). Thus, k-term relations imply the existence of weight k vector in $C^\perp$. For small k, it would be often the case that k is the minimum weight of $C^\perp$, and often no very-high weight vector exists in $C^\perp$. These would imply that the number of the terms will mostly determine δ, which agrees with the results of tests. A quantitative analysis on this observation is a possible future work.

Next we see the effect of increasing the dimension of $C^\perp$. Table 7 shows the result on the five-term GFSR $x_j := x_{j-11} + x_{j-39} + x_{j-95} + x_{j-218}$ proposed as $R(11, 39, 95, 218)$ in [24], which is equivalent to decimation of every 7th output of $x_j := x_{j-11} + x_{j-218}$. We choose $m = 228$ and 238, for which the dimension of the dual space is 10, 20, respectively. The result says that the latter is more powerful than the former, and that the risky sample size is 600,000,000 for the latter. Similarly to the above, we confirmed that the weight distribution test for this risky sample size rejects the generator, although the result is omitted. This result can be compared to the experiments in [24], where the generator passes his

Table 7. Weight discrepancy test on a 5-term GFSR of degree 218 with $m = 228, 238$

m	ν	δ	safe	risky
228	46	1.29×10^{-8}	4.72×10^8	1.96×10^9
238	48	4.37×10^{-8}	1.43×10^8	5.90×10^8

random walk test up to 2×10^6 samples, but is reported to show an error for 10^8 samples. Note that his random walk is two-dimensional, and consumes much more random numbers in one trial than 238 in our test. We also tested five-term relations of degree 250 and 1279 in [24] which passed all the tests there. For example, the result of weight discrepancy test for degree 1279 with dual dimension 20 shows that the risky sample size is 4.38×10^{12}, which is larger than those used in [24], explaining the success of this generator in the tests.

Table 8 shows an example where $s = 4$. The first row shows the result on a twisted GFSR generator named T800 [16]. This generator is known to have a 3-term linear relation on the most significant three bits for 26 consecutive words, although the most significant bit behaves very well [17]. Its period is $2^{800} - 1 \simeq 6.67 \times 10^{240}$. We choose $s = 4$, $\mu = 30$ so that $m = 120$. It turns out that $C^\perp$ is 15-dimensional. We choose $s_0 = 43$ and $\nu = 34$. The first row of Table 8 and Table 9 show the results of the weight discrepancy test and the weight distribution test, respectively. This defect was successfully removed in TT800 by tempering method in [17] (see also [20]). The second row in Table 8 shows the result of the weight discrepancy test on TT800, where $s = 4$ and $\mu = 204$. This $\mu = 204$ is far larger than the previous 30, but is necessary to have nontrivial $C^\perp$, which is 16-dimensional in this case. The order of 10^{49} would be large enough for any future machines, but is not the order of the period which seems best possible in other examples. The third row of Table 8 shows the result of weight discrepancy test on a small

Table 8. Weight discrepancy test on T800, TT800, and MT521

generator	m	ν	δ	safe	risky	min. weight
T800	120	34	7.77×10^{-4}	6.69×10^3	2.85×10^4	3
TT800	816	74	3.23×10^{-49}	2.43×10^{49}	9.70×10^{49}	26
MT521	536	62	3.55×10^{-156}	2.01×10^{156}	8.13×10^{156}	210
TAUS88	104	32	2.63×10^{-26}	1.91×10^{26}	8.22×10^{26}	31

Table 9. Weight distribution test on T800

N	1st	2nd	3rd	4th	5th
6.6×10^3	71.6%	56.8%	99.6%	49.0%	99.4%
2.9×10^4	93.6%	99.7%	94.0%	99.3%	100%

Mersenne Twister[19] MT521 with period $2^{521} - 1 \simeq 6.86 \times 10^{156}$. We choose $s = 4$, $\mu = 134$, and $C^\perp$ turns to be 15-dimensional. We put $s_0 = 237$. It shows that MT521 has much better δ than TT800. We do not know whether this phenomenon is by chance or not. The fourth row shows the result on a combined Tausworthe generator TAUS88[10] of period $\simeq 2^{88} \simeq 3.09 \times 10^{26}$ for $s = 4$ and $\mu = 26$ with $m - r = 16$, which seems fairly good. We did not test the standard Mersenne Twister MT19937 because its size $p = 19937$ exceeded the ability of Mathematica, but expect to have a good quality similarly to MT521.

5 Future works

We introduced the weight discrepancy test on specified m-bits of the generated sequence, which is closely related to physical empirical tests, but is more powerful and easier to handle in selecting a good generator.

Some shortcomings of our method are that we do not know which choice of the m-bits leads to a rejection, that m can not be chosen freely, and that the relation to the number of terms in the recursion is not very clear. It is desirable to obtain an approximation formula on δ depending only on the numbers of low weight vectors in $C^\perp$, not on the medium weight vectors.

Acknowledgment. The authors are indebted to Kunio Shimizu and Masakazu Jimbo for some knowledge on statistics, to Eiichi Bannai for coding theory, and to Hikoe Enomoto for constant help and encouragements. They are thankful to the anonymous referee for pointing out the relevance of noncentral chi-square distributions and for many other valuable comments. The first author is supported by the Kakenhi grant of the Ministry of Education, No.13440005.

References

1. Coddington, P. D. (1994) Analysis of random number generators using Monte Carlo simulation. Int. J. Mod. Phys. C **5**, 547–560.
2. Ferrenberg, A. M., Landau, D. P., and Wong, Y. J. (1992) Monte Carlo simulations: hidden errors from 'good' random number generators. Phys. Rev. Lett. **69** 3382–3384.
3. Fredricsson, S. A. (1975) Pseudo-randomness properties of binary shift register sequences. IEEE Trans. Inform. Theory, **IT-21**, 115–120.
4. Grassberger, P. (1993) On correlations in 'good' random number generators. Phys. Lett. A **181** 43–46.
5. Hoogland, A., Spaa, J., Selman, B., and Compagner, A. (1983) A special-purpose processor for the Monte Carlo simulation of Ising spin systems. J. Comput. Phys. **51**, 250–260.
6. Jordan, H. F. and Wood, D. C. M. (1973) On the distribution of sums of successive bits of shift-register sequences. IEEE Trans. Computers **C-22**, 400–408.
7. Knuth, D. E. (1997) The Art of Computer Programming. Vol. 2. Seminumerical Algorithms 3rd Ed. Addison-Wesley, Reading, Mass.
8. Kirkpatrick, S. and Stoll, E. P. (1981) A very fast shift-register sequence random number generator. J. of Computat. Phys. **40**, 517–526.
9. L'Ecuyer, P. (1994) Uniform random number generation. Ann. Oper. Res. **53**, 77–120.
10. L'Ecuyer, P. (1996) Maximally equidistributed combined Tausworthe generators. Math. Comput. **65**, 203–213.
11. Lewis, T. G. and Payne W. H. (1973) Generalized feedback shift register pseudorandom number algorithms. J. ACM **20**, 456–468.
12. Lindholm, J. H. (1968) An analysis of the pseudo-randomness properties of subsequences of long m-sequences. IEEE Trans. Inform. Theory **IT-14**, 569–576.
13. Linhart, H. and Zucchini, W. (1986) Model Selection. John Wiley & Sons, New York.
14. Lüscher, M. (1994) A portable high-quality random number generator for lattice field theory simulations. Computer Physics Communications **79**, 100–110.
15. MacWilliams, F. J. and Sloane, N. J. A. (1977) The Theory of Error-Correcting Codes. North-Holland.
16. Matsumoto, M. and Kurita, Y. (1992) Twisted GFSR generators. ACM Trans. on Modeling and Computer Simulation, **2**, 179–194.
17. Matsumoto, M. and Kurita, Y. (1994) Twisted GFSR generators II. ACM Trans. on Modeling and Computer Simulation, **4**, 254–266.
18. Matsumoto, M. and Kurita, Y. (1996) Strong deviations from randomness in m-sequences based on trinomials. ACM Trans. on Modeling and Computer Simulation **6**, 99–106.
19. Matsumoto, M. and Nishimura, T. (1998) Mersenne Twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Trans. on Modeling and Computer Simulation **8**, 3–30.
20. Matsumoto, M. and Wigenkittl, S. Getting rid of correlations among pseudorandom numbers: discarding versus tempering. ACM Trans. on Modeling and Computer Simulation, *to appear.*
21. Tiku, M. (1981) Noncentral chi-square distribution. In S. Kotz and N.L. Johnson, editors, Encyclopedia of Statistical Sciences, vol. 6, 276–280. John Wiley.
22. Vardy, A. (1997) The intractability of computing the minimum distance of a code. IEEE Trans. Inform. Theory **43**, no. 6, 1757–1766.
23. Vattulainen, I., Ala-Nissila, T., and Kankaala, K. (1994) Physical tests for random numbers in simulations. Phys. Rev. Lett. **73**, 2513–2516.
24. Ziff, R. M. (1998) Four-tap shift-register-sequence random-number generators. Computers in Physics **12**, 385–392.