Entropy of Mersenne-Twisters

Fabien Le Floc’h

* Correspondence: F.L.Y.LeFloch@tudelft.nl

Abstract: The Mersenne-Twister is one of the most popular generators of uniform pseudo-random numbers. It is used in many numerical libraries and software. In this paper, we look at the Komolgorov entropy of the original Mersenne-Twister, as well as of more modern variations such as the 64-bit Mersenne-Twisters, the Well generators, and the Melg generators.

1. Introduction

The Mersenne-Twister [Matsumoto and Nishimura 1998] is one of the most popular generators of uniform pseudo-random numbers. It is used in many numerical libraries and software.

Several variant have been developed over the years, to accommodate some limitations of the original algorithm.

It has been enhanced to work directly with 64-bit numbers in [Nishimura 2000]. The same paper also describe a variant which increases the number of non-zero terms in the characteristic polynomial. Although rarely adopted in software, the latter has the tremendous advantage to escape much faster from zeroland. Indeed, the original Mersenne-Twister may require more than 700000 numbers to escape from a state with all bits set to zero, except for one. Although a good initialization scheme helps to minimize the issue, the issue may still unexpectedly appear. We give an example of innocuous initialization array, with many ones and zeros.

Panneton et al. [2006] propose a more general algorithm, based on the same underlying principles, but using more internal block matrices, while keeping a similar performance profile. This allows for maximal equidistribution properties, a feature not possible in the original design. It also permit the use of non Mersenne prime periods.

More recently, the algorithm has been slightly modified to obtain a maximally equidistributed variant over 64-bit numbers, by using a lung, but otherwise staying close to the (second) 64-bit Mersenne-Twister variant [Harase and Kimoto 2018].

While introducing his Mixmax random number generator, Savvidy [2015] discusses two properties of the original MT: a low entropy and eigenvalues falling on the unit circle, both of which are detrimental to randomness.

In this paper, we start by verifying the claims of Savvidy. We in fact find a larger entropy than stated in [Savvidy 2015]. We then compute the eigenvalues and the associated entropy of the different variants of the Mersenne-Twister and discuss their associated properties.

2. Measuring entropy

In [Savvidy 2015], the Kolmogorov entropy is defined as

$$h = - \sum_{k: |\lambda_k| < 1} \ln |\lambda_k|,$$

where $(\lambda_k)$ are the eigenvalues of the state transition matrix.

For a k-mixing system, the auto-correlation decay time $\tau_0$ is related to the entropy as $\tau_0 \leq 1/h$. It is thus desirable that most of the eigenvalues of the matrix should lie as far as possible away from it. Without involving
the theory around Komolgorov entropy, it is intuitive that randomness will increase as the eigenvalues of the linear system spread out.

An $\mathbb{F}_2$-linear generator may be defined through its state transition matrix by the relation [Panneton et al. 2006]

$$x_i = B x_{i-1} \mod 2,$$

$$y_i = C x_i \mod 2.$$  

where $x_i = (x_{i,k-1},...,x_{i,0})^T \in \mathbb{F}_2^k$ and $y_i = (y_{i,w-1},...,y_{i,0})^T \in \mathbb{F}_2^w$ are respectively the $k$-bit state and the $w$-bit output vector at step $i$, $B$ and $C$ are a $w \times k$ transition matrix and a $w \times k$ output transformation matrix with elements in $\mathbb{F}_2$, $k$ and $w$ are positive integers.

In the Mersenne-twister, the output transformation takes the first word $(x_{i,k-1},...,x_{i,k-1-w})$ of the state, and eventually tempers it to improve equidistribution properties.

For the original Mersenne-twister MT19937 [Matsumoto and Nishimura 1998], the matrix $B$ of size $(nw - r \times nw - r)$ is expressed in terms of word vectors $x_{i,j}^w$ (instead of bits) as follows

$$x_i = (x_{i,n-1}^w, x_{i,n-2}^w, ..., \overline{x}_{i,0}^w) = (x_{i-1,n-1}^w, x_{i-1,n-2}^w, ..., \overline{x}_{i-1,0}^w) B^T$$

where $\overline{x}_{i,j}^w$ represents the $w-r$ most significant $x_{i,j}^w$. The matrix $B$ reads,

$$B^T = \begin{pmatrix} I_w & I_w & \cdots & I_w \\ I_w & \cdots & \cdots & I_w \\ \cdots & \cdots & \cdots & \cdots \\ S \end{pmatrix},$$

with $n = 624, m = 397, w = 32, r = 31$, and

$$S = \begin{pmatrix} 0 & I_r \\ I_{w-r} & 0 \end{pmatrix} A.$$

The matrix $A$ is defined as

$$A = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \cdots & \cdots & \cdots & \cdots \\ a_{w-1} & a_{w-2} & \cdots & a_0 \end{pmatrix},$$

with $a = (a_{w-1}, a_{w-2}, ..., a_0) = 0x990B0DF$.

Thus, in the MT19937 generator, we have

$$x_{i,j}^w = x_{i-1,j+1}^w \quad \text{for} \quad j = n-2, ..., 0.$$  

This property is however not true in general for the other variants.

The transition matrix defines the iteration, and is thus the relevant one for the study of entropy. For $\mathbb{F}_2$-linear generators, the eigenvalues of the transition matrix correspond to the roots of its characteristic polynomial in $\mathbb{Z}$. In particular, those are different from the eigenvalues of the characteristic polynomial commonly used for skipping ahead, or verifying primitivity, which is stated in $\mathbb{F}_2$. For the original Mersenne-Twister, we give the polynomial in $\mathbb{Z}$ in Appendix B.
In practice however, we will not use the characteristic polynomial but work directly from the state transition matrix. For an $\mathbb{F}_2$-linear generator, the matrix can be recovered by seeding the standard basis successively, and fetching the state of the generator after one iteration as described in Algorithm 1.

### Algorithm 1 Print the state transition matrix for Mersenne-Twister generators

```plaintext
1: $i \leftarrow n \times w - 1$
2: while $i \geq 0$ do
3:   $indicator$ is an array of $n$ integers, each of $w$-bit size, initialized to zero
4:   $k \leftarrow i \times w$
5:   $indicator[k] \leftarrow 1 << (i \mod w)$ \{set the $i$-th bit to one\}
6:   initialize the random number generator with the state $indicator$
7:   compute the next random number
8: for $k \leftarrow 0$ to $n - 1$ do
9:   $indicator[k] \leftarrow state[k + 1 \mod n]$ \{copy the new state to the array $indicator$ with the correct indexing\}
10: end for
11: for $k \leftarrow n - 1$ to 0, $j \leftarrow w - 1$ to 0 do
12:   if $indicator[k] \& (1 << j) \neq 0$ then
13:     print 1
14:   else
15:     print 0
16:   end if
17: end for
18: print new line
19: end while
```

For the Melg generators of Harase and Kimoto [2018], the lung is always the last element of the state and does not move around. For the Well generators, the convention of the state matrix described in [Panneton et al. 2006] is to use the most significant bits of each word first. The algorithm is reversed, from $I = 0$ to $N \times w - 1$ (lines 1 and 2), $K$ and $J$ are reversed in similar fashion. This leads to Algorithm 2 in Appendix C.

Working directly with the state transition matrix avoids the numerical issues of the characteristic polynomial approach, which arise because of the large degrees involved.

### 3. Results

The eigenvalues of the original version of Mersenne-Twister, with a period of $2^{19937} - 1$, all fall close to unit circle (Figure 1). The modulus of the smallest eigenvalue is 0.99806, and the modulus of the largest eigenvalue is 1.00246. While Savvidy [2015] finds an entropy of 4.8, we measure an entropy of 10.49. We double checked our results by solving the roots of the characteristic polynomial with a robust polynomial roots solver [Aurentz et al. 2018] and quadruple precision arithmetic.

If we look at the eigenvalues of other $\mathbb{F}_2$-linear generators, we notice that for a very large period of $2^{19937} - 1$, and thus a characteristic polynomial degree of 19937, the eigenvalues always tend to be very close to the circle (see Figure A1 in Appendix E), even though the entropy increases substantially for some and may be close to the entropy of the generators with a smaller period. For example, the ID3 version of the 64-bit Mersenne-Twister (named MT19937x64ID3 thereafter) has twice the entropy of the ID1 version of the 64-bit Mersenne-Twister. With a large characteristic polynomial degree, the eigenvalues are likely to be very close to the unit circle: for a value strongly greater than one, the polynomial value explodes because of the dominant term of degree 19937.

In reality, only 32 bits (or 64 bits for the 64-bit generators) are produced by one application of the state transition matrix $B$. In order to be comparable to the Mixmax generator of Savvidy [2015], one would need
to consider \( n \) applications of the matrix, where \( n = 624 \) for the 32-bit Mersenne-Twister as the state of the Mersenne-Twister is composed of 624 words. And then, we would see much larger eigenvalues, since the eigenvalues of \( B^n \) are simply the \( n \)-th power of the eigenvalues of \( B \) (Figure 2). For the 32-bit and 64-bit Mersenne-Twisters, the \( n \)-th power is not merely the matrix to skip to the \( n \)-th point in the sequence, because the \( n - 1 \) previous numbers \( x_{n-1,1}, \ldots, x_{1,1} \) are also in the output vector, as consequence of Equation 3. This is however not true any more for the Melg and Well generators, we believe however that the eigenvalues for the \( n \)-th power are still representative of the entropy of the \( n \)-dimensional output.

For the same reason, \( \mathbb{F}_2 \)-linear generators working on 64-bit words have twice the entropy of the ones working on 32-bit words. If we want to compare the two kinds of generator, a more relevant measure is the entropy per bit of random number generated (of size one word).

From Table 1, we notice that the entropy per bit is the highest for the Well generators. In particular, Well607b has the largest entropy per bit. There seems to be a direct relationship with the blocks in the first block column of the transition matrix and the entropy. The MT19937 has two non-zero blocks of dimension \( w \times w \). The 64-bit version MT19937x64ID1 has the same configuration, but where \( w \) is twice the length, and this produces approximately twice the entropy. The version MT19937x64ID3 has four non-zero blocks (so another factor two), and leads to four times this entropy. Similarly, the Melg generators have four non-zero blocks, and the
Table 1. Entropy of the different random number generators. The performance is normalized towards the 32-bit Mersenne-Twister MT19937. MT19937 took 3.8ms to output one million double floating point numbers on an AMD Ryzen 1700 processor running OpenJDK 15 under Linux.

| Generator              | word size $w$ | Entropy | Entropy per bit | $N_1$ | Performance |
|------------------------|---------------|---------|-----------------|-------|-------------|
| MT19937                | 32            | 10.49   | 0.32            | 135   | 1.00        |
| MT19937-64ID1          | 64            | 20.75   | 0.32            | 285   | 0.87        |
| MT19937-64ID3          | 64            | 40.05   | 0.63            | 5795  | 0.49        |
| Well19937a             | 32            | 23.58   | 0.74            | 8585  | 0.68        |
| Well1024a              | 32            | 25.79   | 0.81            | 407   | 1.06        |
| **Well607b**           | 32            | 26.66   | **0.83**        | 313   | 0.83        |
| Melg607                | 64            | 42.83   | 0.67            | 313   | 0.85        |
| Melg19937              | 64            | 41.63   | 0.65            | 9603  | 0.85        |

Well generators have four to five non-zero blocks. We notice a linear relationship between the entropy and the number of non-zero blocks in the first block-column (also taking there size into account).

On Figure 3, we look at the state matrix, raised to the power 312, for the 64-bit Mersenne-Twisters MT19937x64ID1 and MT19937x64ID3. It corresponds to a full update of the state vector. As expected, the larger entropy translates to a pronounced increase of the spread between eigenvalues. The axis scale is multiplied by a factor larger than two hundred between the ID1 and the ID3 versions.

Regarding performance, the Well1024a is the fastest at generating one million random numbers in Java, closely followed by the 32-bit MT19937. The MT19937-64ID3 is the slowest, and the Melg generators are better performing alternative 64-bit generators, with good entropy properties.

4. Reaching and escaping zeroland

Following the methodology of Panneton et al. [2006], we consider the initial states filled with zero bits, except for one bit, and measure the number iterations it takes to reach a state with a balanced number of zero and one bits. There are $k$ such states, and we we let $y^{(j)}_i$ be the output vector $y_i$ at iteration $i$ for a given initial state $x_0 = e_j$ where $e_j$ is the $j$-th unit vector, for $j = 1, \ldots, k$. Let $H(x)$ be the Hamming weight of a bit vector $x$, i.e., the number of bits set to 1, the moving average $\overline{\gamma}_{n,p}$ is defined by

$$\overline{\gamma}_{n,p} = \frac{1}{p w} \sum_{i=n}^{n+p-1} \sum_{j=1}^{k} H\left(y^{(j)}_i\right).$$

(4)
Under the null hypothesis, $\gamma_{n,p}$ should be approximately normally distributed with mean $1/2$ and variance $1/(4pkw)$.

On Figure 4(a), we plot $\gamma_{n,100}$ as a function of $n$. In order to compare 32-bit with 64-bit generators, we normalized the number of iterations $n$ according to the size of one word: one iteration of a 64-bit generator counts for two iterations of a 32-bit generator. Similarly $p$ is halved for 64-bit generators. Well19937a requires around 730 iterations to reach a balanced state, while Melg19937 requires $2 \times 2040$ iterations, even though its characteristic polynomial has a larger number of ones ($N_1$ in Table 1). Even without normalization, the Melg generator requires more iterations than the Well generator. The measure of entropy per bit is however consistent with this result: it it higher for Well19937a than for Melg19937.

![Figure 4](image)

**Figure 4.** Hamming weight after $n$ iterations.

Of course, in practice, the random number generators will not be used with such a state. A good seed initialization procedure, such as the one used in the Melg generators, provides a simple remedy. The reference implementations of Well generators do not include such an initialization procedure, and this sometimes leads to artificially bad results in papers, which study the behaviour of random number generators with close seeds such as [Bhattacharjee et al. 2018]. A real world implementation will however implement a good initialization procedure.

With a good initialization, even if the probability is then very low, it is never fully guaranteed that, accidentally, some seed will lead to an unfavorable state. We give an example of an random looking seed, which could well be the produce of a good initialization procedure as an illustration in Appendix D. In order to find this special seed, we simply jump-ahead close to the end of the period, starting with a seed composed of one non-zero bit. Figure 4(b) shows an example of special seed, where $\gamma_{n,624}$ nearly reaches zero after 3500 iterations. There is less difference between the Well and the Melg generators on this particular example, but this could be because we only look at a specific seed, instead of an average over many seeds.

The equidistribution properties of the generators guarantee that those awkward states, with a very large percentage of zeros, are present in a period. In order to find a good initial seed, a good alternative may be to start with a seed composed of one non-zero bit, and jump-ahead sufficiently far away, but not too far. Then we may obtain a good starting point as many of the bad states are in the relatively recent past.

The issue is nearly invisible for generators with a smaller state such as Well607b: they escape zeroland after less than 50 iterations: too quickly to matter statistically.

5. Conclusion

The entropy reveals some properties of the $\mathbb{F}_2$ linear generators. In a similar fashion as the number of non-zero bits in the characteristic polynomial, a low entropy is related to a low mixing of bits. Entropy however
reveals more about the underlying generator as it seems linearly linked to the number of non-zero blocks in the first block row of the transition matrix.

The Well generators have the largest entropy per bit among the studied generators.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

**Acknowledgments:** Fruitful conversations with K. Savvidy.

Aurentz, Jared L, Thomas Mach, Leonardo Robol, Raf Vandebril, and David S Watkins. 2018. Fast and backward stable computation of roots of polynomials, part ii: backward error analysis; companion matrix and companion pencil. *SIAM Journal on Matrix Analysis and Applications* 39(3), 1245–1269.

Bhattacharjee, Kamalika, Krishnendu Maity, and Sukanta Das. 2018. A search for good pseudo-random number generators: Survey and empirical studies. *arXiv preprint arXiv:1811.04035*.

Harase, Shin and Takamitsu Kimoto. 2018. Implementing 64-bit maximally equidistributed f2-linear generators with mersenne prime period. *ACM Transactions on Mathematical Software (TOMS)* 44(3), 1–11.

Matsumoto, Makoto and Yoshiharu Kurita. 1992. Twisted gfsr generators. *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 2(3), 179–194.

Matsumoto, Makoto and Takuji Nishimura. 1998. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 32(1), 4–20.

Nishimura, Takuji. 2000. Tables of 64-bit mersenne twisters. *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 32(1), 4–20.

Panneton, François, Pierre L’ecuyer, and Makoto Matsumoto. 2006. Improved long-period generators based on linear recurrences modulo 2. *ACM Transactions on Mathematical Software (TOMS)* 32(1), 1–16.

Savvidy, Konstantin G. 2015. The mixmax random number generator. *Computer Physics Communications* 196, 161–165.

**Appendix A Twisted GFSR characteristic polynomial**

We consider the \((nw \times nw)\)-matrix \(B\) defined as

\[
B = \begin{pmatrix}
I_w & I_w & \cdots & I_w \\
I_w & I_w & \cdots & I_w \\
& I_w & \ddots & I_w \\
& & & I_w \\
S & & & \end{pmatrix},
\]

where, \(I_w\) is the identity matrix of size \(w\), in the first block column, the first row of \(I_w\) is at the row position \(((n-1-m)w+1)\) in \(B\).

In [Matsumoto and Kurita 1992, Appendix A.1], it is proven that, for the case \(r = 0\), the characteristic polynomial \(\phi_B(t) = \det(t I_{nw} - B)\) of \(B\) is

\[
\phi_B(t) = \phi_S(t^n + t^m),
\]

where \(\phi_S\) is the characteristic polynomial of the \((w, w)\)-matrix \(S\).

This is true in \(F_2\) but not in the general case. The correct formula is

\[
\phi_B(t) = \phi_S(t^n - t^m),
\]

(A1)
Proof. We have

$$\phi_B(t) = \det(tI_{n} - B) = \det \begin{pmatrix} tI_w & -I_w & \cdots & 0 \\ -I_w & tI_w & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -S & tI_w & \cdots & tI_{w-r} \end{pmatrix}.$$ 

The determinant stays unchanged by linear combinations of columns. We add second block column times $t$ to first column to obtain

$$\phi_B(t) = \det \begin{pmatrix} 0 & -I_w & \cdots & 0 \\ t^2 I_w & tI_w & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -I_w & tI_w & \cdots & tI_{w-r} \end{pmatrix}.$$ 

Then then we continue with $t^k$ times $(k+1)$-th block column, until line $(n-1-m)w+1$, we obtain

$$\phi_B(t) = \det \begin{pmatrix} 0 & -I_w & \cdots & 0 \\ 0 & tI_w & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ t^{n-m} I_w - I_w & tI_w & \cdots & tI_{w-r} \end{pmatrix}.$$ 

We now subtract $t^k(t^{n-m} - 1)$ times the $(n-m) + k$-th block column, until line $(n-1)w+1$:

$$\phi_B(t) = \det \begin{pmatrix} 0 & -I_w & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & tI_{w-r} \end{pmatrix}.$$ 

Now we reduce the $j$-th block column, by adding $t^k$ times the $k+j$ block, for $j = 2, \ldots, n-1$ and $k = 1, \ldots, n-j$. This leads to

$$\phi_B(t) = \det \begin{pmatrix} 0 & -I_w & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & t^2 I_w \end{pmatrix}.$$
We use \((n-1)\) permutations to move the first block column to the last position, we end up with a block-triangular matrix, the determinant is the product of the diagonal determinants.

\[
\phi_B(t) = \det(t^m(t^{n-m} - 1)I_w - S) = \phi_S(t^n - t^m).
\]

\[\square\]

**Appendix B Mersenne-Twister characteristic polynomial**

In the Mersenne-twister random number generator MT19937 [Matsumoto and Nishimura 1998], the matrix \(B\) corresponds to the state transition matrix, and, compared to the GFSR of Appendix A, is slightly modified to accommodate for a size \((nw - r \times nw - r)\). It is defined as follows:

\[
B = \begin{pmatrix} I_w & & & \\ & I_w & & \\ & & \ddots & \\ & & & I_{w-r} \\ S \\ \end{pmatrix},
\]

with \(n = 624, m = 397, w = 32, r = 31\), and the matrix \(S\) reads

\[
S = \begin{pmatrix} 0 & I_r \\ I_{w-r} & 0 \end{pmatrix} A.
\]

The matrix \(A\) is defined as

\[
A = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ a_{w-1} & a_{w-2} & \ldots & a_0 \end{pmatrix},
\]

with \(a = (a_{w-1}, a_{w-2}, \ldots, a_0) = 0x9908B0DF\). In particular, \(\phi_A(t) = t^w - \sum_{i=0}^{w-1} a_i t^{w-i-1}\).

For \(w - r = 1\), the \(S\) matrix reads

\[
S = \begin{pmatrix} 0 & 0 & 1 & & \\ & & & \ddots & \\ & & \ddots & & 1 \\ & a_{w-1} & a_{w-2} & \ldots & a_0 \\ 0 & 1 & 0 & \ldots & 0 \end{pmatrix},
\]

With similar column transformations as for the twisted GFSR generator (see Appendix A), the characteristic polynomial for the Mersenne-Twister reads

\[
\phi_B(t) = (t^n - t^m)^{w-r} (t^{n-1} - t^{m-1})^r - (t^n - t^m)^{w-r} \sum_{i=0}^{r-2} a_i (t^{n-i} - t^{m-i})^{r-1-i} - \sum_{i=r-1}^{w-r} a_i (t^n - t^m)^{w-i-1}. \quad (A2)
\]
For the MT19937, we have in particular,

\[
\phi(I) = (t^n - t^m)(t^{n-1} - t^{m-1})^3 - (t^n - t^m)(t^{n-1} - t^{m-1})^{30} - (t^n - t^m)(t^{n-1} - t^{m-1})^{29} - (t^n - t^m)(t^{n-1} - t^{m-1})^{28} \\
- (t^n - t^m)(t^{n-1} - t^{m-1})^{27} - (t^n - t^m)(t^{n-1} - t^{m-1})^{26} - (t^n - t^m)(t^{n-1} - t^{m-1})^{24} - (t^n - t^m)(t^{n-1} - t^{m-1})^{23} \\
- (t^n - t^m)(t^{n-1} - t^{m-1})^{18} - (t^n - t^m)(t^{n-1} - t^{m-1})^{17} - (t^n - t^m)(t^{n-1} - t^{m-1})^{15} - (t^n - t^m)(t^{n-1} - t^{m-1})^{11} \\
- (t^n - t^m)(t^{n-1} - t^{m-1})^{6} - (t^n - t^m)(t^{n-1} - t^{m-1})^{3} - (t^n - t^m)(t^{n-1} - t^{m-1})^2 - 1.
\]

Appendix C Algorithm for state transition matrix

Algorithm 2 Print the state transition matrix for Well generators

1: \( I \leftarrow 0 \)
2: while \( I \geq N \ast w - 1 \) do
3: \( K \leftarrow I / w \)
4: \( \text{indicator}[K] \leftarrow 1 \ll (w - i \mod w) \)
5: initialize the random number generator with the state \( \text{indicator} \)
6: compute the next random number
7: compute the next random number
8: for \( K \leftarrow 0 \) to \( N - 1 \) do
9: \( \text{indicator}[K] \leftarrow \text{state}[K + N - 1 \mod N] \) copy the new state to the array \( \text{indicator} \)
10: end for
11: for \( K \leftarrow 0 \) to \( N - 1 \), \( J \leftarrow 0 \) to \( W - 1 \) do
12: if \( \text{indicator}[K]\&(1 \ll w - j - 1) \neq 0 \) then
13: print 1
14: else
15: print 0
16: end if
17: end for
18: print new line
19: end while

Appendix D Bad seeds

Appendix D.1 Mielg19937
Appendix D.2 Bad seed for the Well1937a
Appendix E  Eigenvalues of some Mersenne-twister variants
Figure A1. Eigenvalues of the transition matrix of generators with a period of $2^{19937} - 1$: histogram (left) and complex plane representation (right).
Figure A2. Eigenvalues of the transition matrix of generators with a smaller period: histogram (left) and complex plane representation (right).