1.10 Statistical properties of LFSRs

The study of the statistical properties of LFSR sequences of maximum period $2^l - 1$, where l is the length of the LFSR, goes back to Golomb [2].

Here are some results:

1. Each full period contains exactly 2^{l-1} ones and $2^{l-1} - 1$ zeroes.

   Proof Each of the 2^l state vectors $\in \mathbb{F}_2^l$ (except 0) occurs exactly once, corresponding to the integers in the interval $[1 \dots 2^l - 1]$. Of these integers 2^{l-1} are odd, the remaining ones are even, and their parities yield the exact output sequence of the LFSR.

2. A **run** in a sequence is a constant subsequence of maximum length.

   Example: $\dots 0111110 \dots$ is a run of ones of length 5.

   Noting that the pieces of length l of the LFSR sequence are exactly the different state vectors $\neq 0$ we immediately see that a full period contains:

   - no run of length $> l$,
   - exactly one run of 1's and no run of 0's of length l—otherwise the zero state vector would occur, or the all-1 state would occur more often than once,
   - exactly one run of 1's and exactly one run of 0's of length $l - 1$,
   - more generally exactly 2^{k-1} runs of 1's or 0's each of length $l - k$ for $1 \leq k \leq l - 1$,
   - in particular exactly 2^{l-1} runs of length 1, exactly half of them consisting of 0's or 1's.

3. For a periodic sequence $x = (x_n)_{n \in \mathbb{N}}$ in $\mathbb{F}_2$ of period s the **autocorrelation** w. r. t. the shift t is defined as

$$\begin{aligned} \kappa_x(t) &= \frac{1}{s} \cdot [\#\{n \mid x_{n+t} = x_n\} - \#\{n \mid x_{n+t} \neq x_n\}] \\ &= \frac{1}{s} \cdot \sum_{n=0}^{s-1} (-1)^{x_{n+t} + x_n} \end{aligned}$$

   (as in Part II for Boolean functions). Consider a sequence x generated by an LFSR of length l,

$$x_n = a_1 x_{n-1} + \dots + a_l x_{n-l} \quad \text{for } n \geq l,$$

   and the sequence $y_n = x_{n+t} - x_n$ of its differences. This sequence is obviously generated by the same LFSR. If the start values $y_0, \dots, y_{l-1}$ are all 0, then the y sequence is constant $= 0$, the t-th state

vector $x_{(t)} = x_{(0)}$, hence t is a multiple of the period, and $\kappa_x(t) = 1$. Otherwise—and if x has the maximum possible period $s = 2^l - 1$—a full period of y consists of exactly 2^{l-1} ones and $2^{l-1} - 1$ zeroes by Remark 1. Thus

$$\kappa_x(t) = \begin{cases} 1, & \text{if } s|t, \\ -\frac{1}{s}, & \text{else.} \end{cases}$$

Hence the auto-correlation is uniformly small, except for shifts by a multiple of the period.

Golomb called these statements the three randomness postulates. They tell us that the sequence is very uniformly distributed. Therefore electrical engineers are fond of LFSR sequences of maximum period, and call them PN sequences (= pseudo-noise sequences).

Executing the Sage code sample 1.4 with the parameter 1024 instead of 20 yields the output (without parentheses and commas):

```
11001000110101100011001111000000 00111011100011100000100011101111
01001001111001010111110010111001 00010010110001100111001111010111
11000100011000001110011000010111 01101010101110110001010111011000
11110000010000100010111100011110 10100111000001111000100001011000
01010101000101111110110011011101 11001001110111110001011000100010
11100100101111110011011001010011 00001100100001100110100011100100
11101000100101110110011011001010 11011100100110111001011100000011
00100010111101111000110000010001 01110100001110011111101000100101
00111010001111000100000000110110 10000101110101110001100000010001
11011011011110111001000110101001 10001111110110101010011111100001
11101110111101011001010110001010 00000100001001100110001110100110
00010100101110100000010101100100 10010110101011111110111111011101
11001010010100010010110111111110 10100101001111110110100100010001
10111100011001111001011111010110 01110111010100100010100101101111
01100111011000000111011111010000 11011101111111110000010001000100
10010111111110101011101110111111 01110010110000010001111001100111
```

The visualization in Figure 1.10 shows that the output looks quite random, at least at first sight.

By the way the LFSR of this example generates a sequence of maximum period $2^{16} - 1 = 65535$ since its characteristic polynomial

$$T^{16} + T^{14} + T^{13} + T^{11} + 1 \in \mathbb{F}_2$$

is primitive.

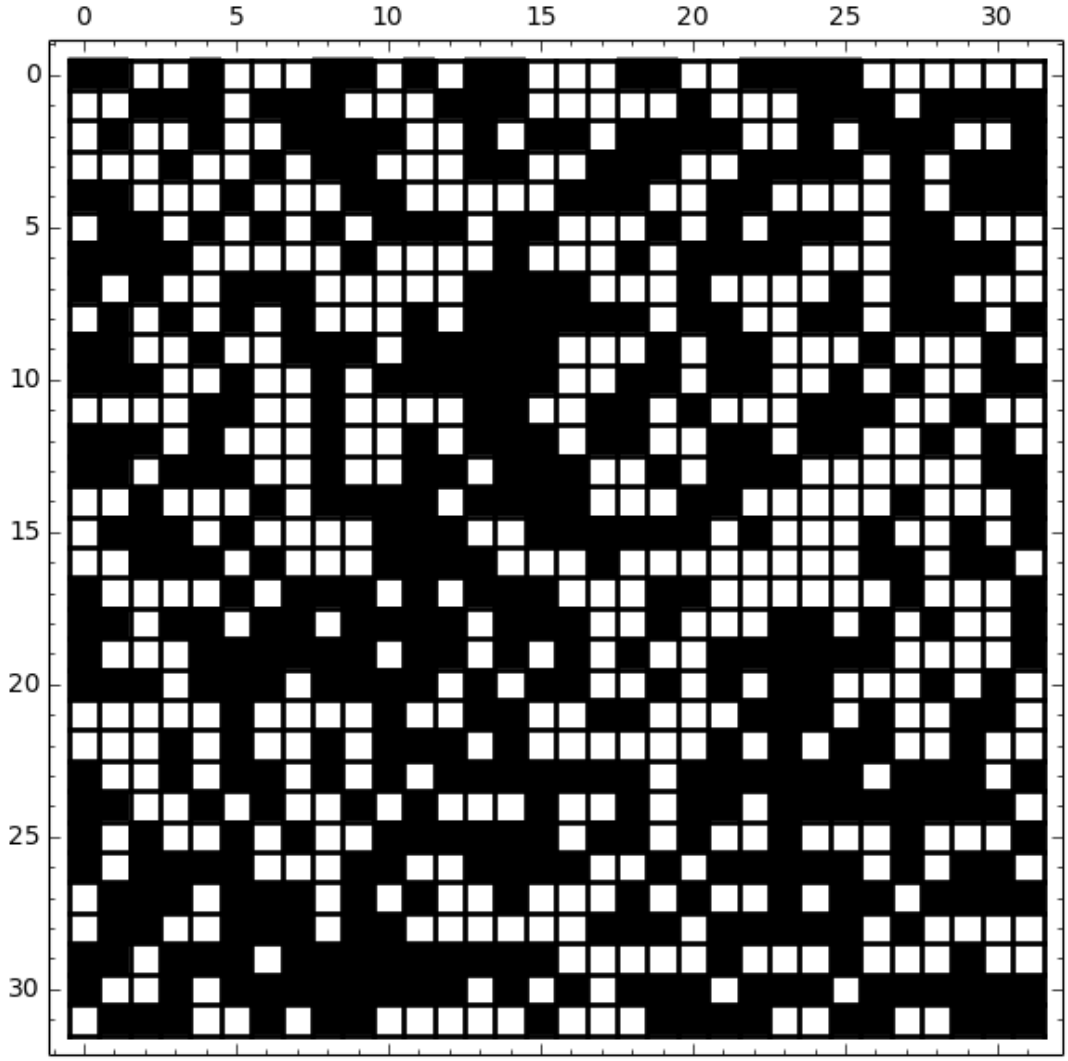

Figure 1.10: An LFSR sequence