Modelling and Simulation of Systems
Exercise 3: Pseudorandom number generators

Agata Radys, Paweł Cejrowski, Łukasz Myśliński

November 17, 2016

1 Generator 1

GFSR (generalized feedback shift register) with parameters: $j = 1$, $k = 2$, $m = 2^{32}$.

1.1 Equation

Numbers are generated according to equation 1.

$$x_n = x_{n-1} \oplus x_{n-2} \quad (mod\ 2^{32}) \tag{1}$$

where x_0 and x_1 are given.

1.2 Questions

- **What is the minimum(maximum) possible value of the period? Give an example of initial values for which the period is small(large).**
  Minimum value of period is equal 1, when both x_0 and x_1 are equal 0. Based on lecture, we know that maximum value of period is less than $(2^k - 1) \cdot k = (2^2 - 1) \cdot 2 = 6$, but this value of period was not found during tests. Found 3 when x_0 or x_1 was different from 0(e.g. $x_0 = 0$ and $x_1 = 16$).

- **What is the minimum(maximum) possible mean value? Give an example of initial values for which the average value is small(large).**
  Minimum possible mean value is equal 0, when both x_0 and x_1 are equal 0. Maximum found mean value within period is equal $\frac{2}{3} \cdot (2^{32} - 1) = 2863311530$, when $x_0 = x_1 = 2^{32} - 1$

- **What is the minimum(maximum) possible variance? Give an example of initial values for which the variance is small(large).**
  Minimum possible variance value is equal 0, when both x_0 and x_1 are equal 0. Maximum found variance value within period is equal $4.0992764589155 \cdot 10^{18}$, when $x_0 = x_1 = 2^{32} - 1$

- **Does the generator meet the requirements that good generators should satisfy? If not, which of the requirements are not satisfied and why?**
  The generator does **not** meet all requirements for good generators.
  Characteristics of good generators:

  1. **generated numbers distributions are as close as possible to the desired one**
     Not satisfied: 3 or even 6 point distribution is not enough to be called uniform distribution within any interval.
  2. **subsequences of the produced sequence are mutually independent**
     Not satisfied: each subsequence of length 3 contains the same 3 numbers.
  3. **long period, with length at least $\sqrt{l}$, where l is the length of the used subsequence**
     Not satisfied: short period with maximum length of 6.
  4. **the ability to make jumps, i.e. to compute x_j from x_i for every $j > i$**
     Satisfied: x_j can be computed from single number x_i due to short period, and modulo operation on index.
  5. **repeatable, portable and efficient**
     Satisfied: generator is deterministic and for the same seeds returnes exactly the same values. It is portable, because it can be implemented in almost every environment and is efficient, because it does not require any time-consuming operations.

- **Is the generator suitable for use in the cryptography? If not, why?**
  The generator is **not** suitable for use in the cryptography, because it does not satisfy any of the desired conditions.
  Characteristics of generators suitable for use in the cryptography:

  1. **it must be impossible to predict its seed and internal state even if we have a large sample of the numbers it produced**
     Not satisfied: it is very easy to predict the number when the period is at most 3 numbers long.
  2. **it must have a long period for every possible value of its seed**
     Not satisfied: do not have long period for any seed.
  3. **it should be unpredictable to the public, i.e. the probability of predicting the subsequent numbers should be low even if have a large sample of the numbers it produced**
     Not satisfied: having 2 consecutive numbers one can predict any other easily.

2 Generator 2

LCG(linear congruential generator) with parameters $a = 3$, $c = -1$, $m = 2^{32}$.

2.1 Equation

Numbers are generated according to equation 2.

$$x_n = 3 \cdot x_{n-1} - 1 \ \ (mod\ 2^{32}) \tag{2}$$

where x_0 is given.

Equation 2 might be modified:
$x_1 = 3 \cdot x_0 - 1$
$x_2 = 3 \cdot (3 \cdot x_0 - 1) - 1$
$x_3 = 3 \cdot (3 \cdot (3 \cdot x_0 - 1) - 1) - 1$
$\ldots$
$x_n = 3^n \cdot x_0 - 3^{n-1} - 3^{n-2} - \ldots - 3^0$
$x_n = 3^n \cdot x_0 - \sum_{i=0}^{n-1} 3^i = 3^n \cdot x_0 - 1 \cdot \frac{1-3^n}{1-3} = 3^n \cdot x_0 + \frac{1}{2}(1 - 3^n)$
to finally achieve to following formula:

$$x_n = 3^n \cdot (x_0 - \frac{1}{2}) + \frac{1}{2} \tag{3}$$

which might be generalized to:

$$x_j = 3^{j-i} \cdot (x_i - \frac{1}{2}) + \frac{1}{2} \tag{4}$$

, where $j > i$.

2.2 Questions

Below corner cases were found by 10000 generator executions with randomized n, m and *seed* by shell `$RANDOM`, which generates pseudorandom number from range $[0, 32767]$.

- **What is the minimum(maximum) possible value of the period? Give an example of initial values for which the period is small(large).**
  Minimum value of period was not found, because in tests it was always of the length of generated sequence (up to 10^8). Based on lecture, we know that maximum value of period is 2^{30} and is reached when x_0 is odd.
- **What is the minimum(maximum) possible mean value? Give an example of initial values for which the average value is small(large).**
  Minimum mean value for period cannot be found, because minimum period is not set. Maximum found mean value is 2373236302.0435 for seed 28208, but still it is only empirical.
- **What is the minimum(maximum) possible variance? Give an example of initial values for which the variance is small(large).**
  Minimum and maximum variance cannot be correctly set, because it is too big. For all tests and not full periods it was over 10^{18}.

- **Does the generator meet the requirements that good generators should satisfy? If not, which of the requirements are not satisfied and why?**
The generator meets all requirements for good generators.
Characteristics of good generators:

  1. **generated numbers distributions are as close as possible to the desired one**
  Satisfied: K^+ and K^- satisfied in 99.9% cases with the $\alpha = 0.05$. All satisfied with $\alpha = 0.15$. It is caused by exponential jumps and many 'iterations' over the interval $[0, 2^{32}]$.
  2. **subsequences of the produced sequence are mutually independent**
  Not satisfied: despite the fact that chi-square test was always satisfied in tests on the $\alpha = 0.05$, the consecutive numbers are autocorrelated, thus they are not independent.
  3. **long period, with length at least $\sqrt{l}$, where l is the length of the used subsequence**
  Satisfied: periods are long, because function 3 is monotonous and increasing.
  4. **the ability to make jumps, i.e. to compute x_j from x_i for every $j > i$**
  Satisfied. jumps can be made based on equation 4.
  5. **repeatable, portable and efficient**
  Satisfied: generator is deterministic and for the same seeds returnes exactly the same values. It is portable, because it can be implemented in almost every environment and is efficient, because it does not require any time-consuming operations.

- **Is the generator suitable for use in the cryptography? If not, why?**
The generator is **not** suitable for use in the cryptography, because it does not satisfy two of the desired conditions.
Let's consider characteristics of generator for seed equal 0 suitable for use in the cryptography:

  1. **it must be impossible to predict its seed and internal state even if we have a large sample of the numbers it produced**
  For this internal state can be calculated based on formula 3.
  2. **it must have a long period for every possible value of its seed**
  Satisfied: as described above.
  3. **it should be unpredictable to the public, i.e. the probability of predicting the subsequent numbers should be low even if have a large sample of the numbers it produced**
  Not satisfied: the formula is easy to guess based on consecutive numbers.

3 Generator 3

3.1 Equation

Numbers are generated according to equation 5.

$$\begin{cases} x'_n = x'_{n-1} \oplus x'_{n-2} \;\; (mod\ 2^{32}) \\ x''_n = 3 \cdot x''_{n-1} - 1 \;\; (mod\ 2^{32}) \\ x_n = x'_n \cdot x''_n \;\; (mod\ 2^{32}) \end{cases} \tag{5}$$

where x'_0, x'_1 and x''_0 are given.

3.2 Questions

Below corner cases were found by 10000 generator executions with randomized range $[n, m]$ and seeds x'_0, x'_1, x''_0 by shell `$RANDOM`, which generates pseudorandom number from range 0-32767.

- **What is the minimum(maximum) possible value of the period? Give an example of initial values for which the period is small(large).**
Minimum value of period is equal 1, when all seeds are equal 0. Maximum period was not found, because it was greater than 10^9 for not full evaluation. It is shorter than 2^32.

- **What is the minimum(maximum) possible mean value? Give an example of initial values for which the average value is small(large).**
Minimum mean value is equal 0 for all seeds equal 0 and maximum is undefined, because it is bigger than 10^9.

- **What is the minimum(maximum) possible variance? Give an example of initial values for which the variance is small(large).**
  Minimum variance value is equal 0 for all seeds equal 0 and maximum is undefined, because it is bigger than 10^9.

- **Does the generator meet the requirements that good generators should satisfy? If not, which of the requirements are not satisfied and why?**
  The generator does **not** meet all requirements for good generators.
  Characteristics of good generators:

  1. **generated numbers distributions are as close as possible to the desired one**
     Satisfied: K^+ and K^- satisfied in 99.9% cases with the $\alpha = 0.011$. All satisfied with $\alpha = 0.10$, with similar arguments as above.
  2. **subsequences of the produced sequence are mutually independent**
     Satisfied: chi-square always satisfied on the $\alpha = 0.05$. Cannot find any contrarguments because autocorrelation is removed by `xor` operation.
  3. **long period, with length at least $\sqrt{l}$, where l is the length of the used subsequence**
     Not satisfied: there is a combination of seeds that gives period of length 1 (seeds: 0, 0, 0).
  4. **the ability to make jumps, i.e. to compute x_j from x_i for every $j > i$**
     Satisfied: it is possible to make jumps using short period of first equation and formula 4 for second generator.
  5. **repeatable, portable and efficient**
     Satisfied: generator is deterministic and for the same seeds returnes exactly the same values. It is portable, because it can be implemented in almost every environment and is efficient, because it does not require any time-consuming operations.

- **Is the generator suitable for use in the cryptography? If not, why?**
  The generator is **not** suitable for use in the cryptography.
  Characteristics of generators suitable for use in the cryptography:

  1. **it must be impossible to predict its seed and internal state even if we have a large sample of the numbers it produced**
     Not satisfied for seeds 0, 0, 0.
  2. **it must have a long period for every possible value of its seed**
     Not satisfied for seeds 0, 0, 0.
  3. **it should be unpredictable to the public, i.e. the probability of predicting the subsequent numbers should be low even if have a large sample of the numbers it produced**
     Not satisfied for seeds 0, 0, 0.