Random number generator: testing and whitening

Andrei Romashchenko & Alexander Shen ESCAPE / LIRMM

invitation to an internship (2020)

What is a random sequence of digits?



- 0

3/12

Can you guess which sequence here is truly random?

- 110010010000111111011010100001000001
 (btw, the last sequence is a binary expansion of *π*)

Can you guess which sequence here is truly random?

Can you guess which sequence here is truly random? (Yes, to prepare one of these sequences we tossed a real coin 40 times.)

3/12

Can you guess which sequence here is truly random? (Yes, to prepare one of these sequences we tossed a real coin 40 times.)

In some sense, all 40-bit sequences are equally random (they all have the same probability 2^{-40} in a fair coin model)

Can you guess which sequence here is truly random? (Yes, to prepare one of these sequences we tossed a real coin 40 times.)

In some sense, all 40-bit sequences are equally random (they all have the same probability 2^{-40} in a fair coin model), but some look more random than others.

What is the difference between truly random and non-random sequences?

4/12

What is the difference between truly random and non-random sequences?

Possible answer:

random = no regularities non-random = apparent or hidden regularities

What is the difference between truly random and non-random sequences?

Possible answer:

random = no regularities

non-random = apparent or hidden regularities = a short description

What is the difference between truly random and non-random sequences?

Possible answer:

random = no regularities non-random = apparent or hidden regularities = a short description

In practice:

various statistical tests (sequences with no regularities pass these tests)

What is the difference between truly random and non-random sequences?

Possible answer:

random = no regularities non-random = apparent or hidden regularities = a short description

In practice:

various statistical tests (sequences with no regularities pass these tests)

In theory: the notion of Kolmogorov complexity,

 $C(x_1...x_n) =$ length of the shortest program that prints $x_1...x_n$

What is the difference between truly random and non-random sequences?

Possible answer:

random = no regularities non-random = apparent or hidden regularities = a short description

In practice:

various statistical tests (sequences with no regularities pass these tests)

In theory: the notion of Kolmogorov complexity,

 $C(x_1...x_n) =$ length of the shortest program that prints $x_1...x_n$

$$x_1...x_n$$
 is random means $C(x_1...x_n) \approx n$
 $x_1...x_n$ is non-random means $C(x_1...x_n) \ll n$

cryptography (random secret keys etc.)

- cryptography (random secret keys etc.)
- randomized algorithms (Monte-Carlo methods etc.)

- cryptography (random secret keys etc.)
- randomized algorithms (Monte-Carlo methods etc.)
- random combinatorial structures (e.g., random error correcting codes)

- cryptography (random secret keys etc.)
- randomized algorithms (Monte-Carlo methods etc.)
- random combinatorial structures (e.g., random error correcting codes)

Big practical question: where do we get truly random bits?

| | | ANDER DIGITS | TABLE OF RANDOM BIERTS | |
|---|--|---|--|---|
| | 146.0 0 - | 94285 T6433 55216 66668 68548 Ato. | 00007 17000 00007 02137 10037 70511 4001 40 0772 03537 72597 99230 40580 13546 90367 10 0772 03547 02111 10744 0000 13464 90367 10 0114 01264 91555 0000 1446 913744 90367 10 0114 01264 91555 0010 1446 913744 91375 10 0114 0116 0005 0010 1446 913744 913744 91375 10 0114 0116 0005 0010 1446 913744 91375 10 0114 0116 0005 0010 1446 91374 91375 10 0114 0116 0005 0010 1446 91374 91375 10 0114 0116 0005 0010 1446 9138 10 0114 0116 0005 0010 1446 9138 10 0114 0116 0005 0010 1446 9158 10 0114 0116 0005 0010 1446 9158 10 0114 0116 0005 0010 1446 9158 10 0005 0005 0005 0005 0005 0005 0005 00 | 118 00 (981) 98044 11 1000 47048 11 1000 4700 |
| | a court 12344 \$2855 41080 | 1000 45T28 92990 58330 8880 1283 | 0 0011 56000 07432 89000 68636 1622 97391 30 0 5613 56000 07432 89000 68636 1622 97391 30 0 0017 8604 16175 86556 56070 07800 96068 11 0 0178 86046 8551 86556 56070 07800 56076 31 0 0178 86076 8501 85556 56070 07800 56076 31 0 0178 86076 8501 86556 56070 07800 56076 31 0 0178 86076 8501 85556 56070 07800 56076 30 0 0178 86076 85076 85076 30 0 0178 86076 85076 85076 85076 30 0 0178 86076 85076 85076 30 0 0178 86076 85076 85076 85076 30 0 0178 86076 85076 85076 85076 85076 85076 85076 85076 85076 85076 80 0 0178 86076 85 | 025 02180 98286 226 09480 70772 869 33656 06111 962 28616 06111 |
| 2044 2044 2044 | 8 24219 11000 2700 2700 20080 0 11000 82567 20019 28458 120141 88219 04327 47128 120141 88219 04612 42217 120141 88219 04612 42217 120441 84219 16412 42217 120441 84218 16418 1 | 10000 07716 17600 72899 02110 1040 | 04417 43097 44540 50019 31847 74878 3701 43 0451 31647 6332 70133 5496 5877 45878 0551 31647 6332 70135 5496 5877 4587 0478 0540 33970 5454 5478 1877 3122 3461 53 1987 1997 1997 1997 1997 1997 1997 3178 45897 1983 3710 24854 2778 34697 38789 4301 70 | 100 12746 12274 266 62655 20020 2 20027 201712 485 42052 98165 50141 20152 98166 |
| 1000 1000 1000 1000 | 08618 50635 94800 56467 08618 50635 50807 60711 00882 87848 50807 60711 87834 21530 5060 96885 87834 21530 70827 21531 | ALCO 198446 19972 31943 32926 4361 35545 07936 19472 31943 32269 4423 20177 05413 375658 2522 32269 4423 04828 19983 711237 42159 34138 3298 | 41405 41497 55835 52179 59834 53137 13624 20 97720 41497 55835 52179 59834 53137 13624 20 20135 91897 65377 47852 58566 6883 14183 80 1015 91897 65377 47852 58566 6883 14183 80 | 177 01801 89404 046 99532 84921 046 99532 84921 078 2565 43235 1724 12659 09722 05 41244 |
| 2004 1840 1840 2967 | 14545 21422 28254 48964 0 222682 49511 55285 55557 0 222682 49511 55285 55528 | 0444 5.547 13758 00436 45928 6534 4100 57534 34992 89946 98946 98046 4100 57534 34992 89946 98066 9803 4100 57534 34992 89946 98066 9803 6638 62718 78302 68833 69865 5493 6031 54014 52034 94428 54639 | 11456 02927 27646 21462 34243 22228 19728 4 11284 71872 08428 66073 72558 26915 93355 28 24129 62080 54642 47018 68338 66087 94434 5 4127 6917 26719 81188 (6097) 12388 94434 5 | ALL |
| 19879 35670 19877 | 6(218 36379 29797 72285 4 22938 44738 46792 90943 4 | 2058 83453 39418 34699 64009 8788 3007 643453 37749 33685 62343 4038 0685 12922 65442 28204 12356 3662 0688 20175 75059 38087 78628 2008 | 3 25423 18400 00247 88045 06084 44835 75018 8 54647 54776 95187 70425 38094 87250 18664 8 5155 0000 38294 47203 95329 80332 68378 0 5453 12449 51778 12818 45537 80382 68378 0 | NY78 98327 43856 |
| 13879 00680 16883 18883 18883 | 14387 T0068 54190 15450 6 | 8418 94310 28340 99862 80355 9609 5592 91281 93563 92355 27822 4187 5379 25995 64533 37083 59900 9266 | 98136 34362 95166 78256 20831 89983 45178 5 41000 86645 22822 91826 20592 81766 20535 8 66207 89235 34716 58103 96596 81923 54548 5 | 7284 99729 30256 2048 22165 41442 2027 5285 41442 |
| 55684 56883 20885 39885 | 00010 K2263 55.45 60624 G 20047 13353 35.455 60624 G 20028 229779 61300 48700 3 56644 38856 54846 31723 5 64643 20279 42341 28484 3 18005 78874 82366 37001 85 | 0050 17334 29553 25995 49379 0004 1007 14371 16727 10473 12774 0054 4400 00300 00338 58532 2656 8440 4400 00348 13559 05355 66928 0040 | 11847 87666 29213 11044 94390 69771 08846 0 00879 91116 80460 30312 20138 92975 14847 1 19958 60810 69776 05476 27892 83350 96897 | 8297 00413 34189 7808 24779 27983 7810 23734 99635 4195 22328 48846 4834 60554 92396 |
| 56489 28950 29952 20082 | 47790 60620 48517 81364 41 | 437 13462 73383 86023 50848 2138 436 60366 34988 26522 58803 4000 733 36899 55862 02633 2020 0019 749 6853 2000 10632 29355 8000 | 36644 28314 97164 12902 22228 17457 \$9449 - | 14444 STRIN 20100 |
| 30877 | 60655 61535 36774 84280 56 72484 8507 8280 35790 18 | NO 42073 73221 20834 75343 2084 578 70633 030706 64899 42346 6071 223 690333 036266 93352 81396 1386 4003 03070 64899 42346 60713 1386 231 690333 036266 93352 81396 1386 | | 65044 96510 95984 01253 20029 66982 02255 20029 95946 |
| L | | | 1000 MALE 0/000 12975 12349 50541 | 04527 34164 56472 |
| A Company | and the second second | THE STATE | T | |
| 2 | | TABLE OF RA | DOM DIGITS | |
| 00050 091 | 88 20097 | 32825 39527 0 | 220 86304 83389 | 87374 64278 58044 |

Rand Corporation, A Million Random Digits with 100,000 Normal Deviates (1955)

26269

20971 87749

66281 31003

[random digits kindly generated for us in 1955]

62290 64464

90429 12272

00682 27398

27124 67018

95375 05871

20714 53295

82760

41361

93823 43178

07706 17813

ESCAPE / LIRMM

00052

00053

00054

73189 50207

75768 76490

54016 44056



[random digits from a noise in an electric circuit]



[random bits from quantum phenomena]

Do these gadgets give perfectly random bits?

Do these gadgets give perfectly random bits? Not really.

Do these gadgets give perfectly random bits? Not really. Example:



100 kHz



default rate 2.5 MHz

| | | ana ana COURT ANA SURFACE | |
|---|--|--|--|
| 100 000 100 00000000 | | Anter Anter Sille Stati Anter Anter Anter Anter Anter Anter Anter Anter | |
| 2.03 명전 유지 유 | | and the second s | |
| | 100 B 1000 B 100 B | | 10 10 10 10 10 10 10 10 10 10 10 10 10 1 |
| State State State State State State State | | 1000 1000 1000 1000 1000 1000 1000 100 | 200 80 808 80 |
| 2년 년년 1 | | ····································· | |
| | 823 882 CT 8 | | |
| ····································· | | 144 144 144 144 144 144 144 144 144 144 | |

5 MHz

Do these gadgets give perfectly random bits? Not really. Example:



100 kHz



default rate 2.5 MHz



| | | | - 10 HE 10 10 10 10 10 10 10 10 10 10 10 10 10 |
|--|--------------------------------|---------------------------------------|--|
| | 202 BBA 122 | | (1) (1) (1) (1) (1) (1) (1) (1) (1) (1) |
| | | | |
| | 100 000 000 | | |
| 100 miles 100 mi | 100 200 100 | 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 | |
| 2.53 (115) (015) (01 010) 10 10 10 10 10 | 122 200 - 542 104 atr - 542 | 231 (326 659 #35 (568 669 | |
| 100000 | | | |
| | | | |
| | 認識 約 | | |
| | | 10.000 | 100 |
| | | · 法数 的复数 | 100 00 00 00 00 00 00 00 00 00 00 00 00 |
| 10 10 10 10 10 10 10 10 10 10 10 10 10 1 | 101 105 125 | 1012 2020 000 | |
| 1 10 10 10 10 10 10 10 10 10 10 10 10 10 | 848 882 102 828 882 115 | | |
| | | | 200 003 002 003 |
| 5 (1) (5 E) (8 E) (6 E) | | | 20 20 20 20 20 20 20 20 20 20 20 20 20 2 |

5 MHz

successful whitening: XOR of 3 (apparently biased) data flows looks pretty random

ESCAPE / LIRMM

Random number generator: testing and whitening

Randomly produced bits are usually not truly random

Randomly produced bits are usually not truly random, they fail standard randomness tests.

Randomly produced bits are usually not truly random, they fail standard randomness tests.

Second practical problem:

You cannot trust blindly the standard implementations of randomness tests

Randomly produced bits are usually not truly random, they fail standard randomness tests.

Second practical problem:

You cannot trust blindly the standard implementations of randomness tests (mathematically unsound tests, errors in the code).

The challenges:



cleanup/enhance existing randomness tests



- cleanup/enhance existing randomness tests
- experiments with physical random generators

- cleanup/enhance existing randomness tests
- experiments with physical random generators
- try new approaches to whitening

- cleanup/enhance existing randomness tests
- experiments with physical random generators
- try new approaches to whitening
- apply (certainly non-perfect) pseudo-random generators to produce useful random objects (e.g., error correcting codes)

Internship proposal:

Generation of random bits is a classical problem known in the context of pseudo-random generators and also in connection with of truly random physical processes (there exist electronic devices that produce random bits using an unpredictable physical noise or intrinsically nondeterministic quantum phenomena). However, the quality of physical generators of random bits remains badly founded and poorly tested. The first objective of this project is an experimental study of the validity and quality of several physical random numbers generators.

When we talk about the quality of random or pseudo-random generators, we have to use randomness tests. The second objective of the project is an inventory and revision of statistical tests for random and pseudo-random generators. We suggest to improve the quality of statistical tests and develop new techniques of "whitening" that improves the quality of non-ideal sources of random bits. Another axis of the project is a conversion of various probabilistic proofs into unconventional randomness tests.

Prerequisites: Basic knowledge of probability and statistics, and solid programming skills. The main tools in the project are pretty standard: C / gcc / Linux. The project requires not only writing your own code but also reading and maintaining the code that already exists.