Test of randomness with distributions of words

Hayato Takahashi

Random Data Lab. Inc.

Aug. 25, 2023
ICIAM 2023 Waseda Univ.

Pseudo random numbers test with distributions of words

Nonoverlapping Template Matching, Overlapping Template Matching, Longest Run of Ones in a Block, SP800-22 NIST[1]:

Chi square test with mean and variance of distributions of words.

Our approch:

Kolmogorov Smirnov Test (KS test) with exact distributions of words.

We test seven pseudo random number generators (PRNG) that are available from C++ <random> library.
Range: $[0, 2^{size} - 1]$.

Table: Type of PRNG

PRNG	minstd_rand0	minstd_rand	mt19937	mt19937_64
size	31	31	32	32
type	LCG	LCG	MT	MT

Table: Type of PRNG

PRNG	ranlux24	ranlux48	knuth_b
size	24	32	31
type	Ranlux	Ranlux	?

Exact distribution, Takahashi [2]

$G_{n,m}(x)$: the number of 0^m of size greater than or equal to m in x.
$L_n(x)$: the size of the longest run of 0s in x.

$$P(G_{n,m}(X_1^n) = t) = \sum_{t \le k \le \lfloor \frac{n+1}{m+1} \rfloor} (-1)^{k-t} \binom{n+1-mk}{t, k-t} 2^{-k(m+1)+1} - \sum_{t \le k \le \lfloor \frac{n}{m+1} \rfloor} (-1)^{k-t} \binom{n-mk}{t, k-t} 2^{-k(m+1)}.$$

$$P(L_n < m) = \sum_{0 \le k \le \lfloor \frac{n+1}{m+1} \rfloor} (-1)^k \binom{n+1-mk}{k} 2^{-k(m+1)+1} - \sum_{0 \le k \le \lfloor \frac{n}{m+1} \rfloor} (-1)^k \binom{n-mk}{k} 2^{-k(m+1)}.$$

Null hypothesis P: fair coin flipping.
run length: 6.
n : sample size.
iterate : 10000.
emp(x):= $|\{1 \leq i \leq iterate \mid G_{n,6}(X(i)) \leq x\}|/iterate,\ |X(i)| = n.$
dist: true distribution.
sup diff: $\sup_x |emp(x) - dist(x)|$.
arg max : $\mathrm{argmax}_x |emp(x) - dist(x)|$.
p-value: $P(\sup_x |emp(x) - dist(x)| >$ sup diff)

Table: KS test with $G_{n,m}$

PRNG	minstd_rand0	minstd_rand	mt19937	mt19937_64
n	3100	3100	3200	3200
sup diff	0.00654306	0.00987082	0.0106322	0.00690215
arg max	0	0	0	0
p-value	0.785314	0.284105	0.208279	0.727457

Table: KS test with $G_{n,m}$

PRNG	ranlux24	ranlux48	knuth_b
n	2400	3200	3100
sup diff	0.00739239	0.00610361	0.00787321
arg max	0	0	0
p-value	0.645304	0.850273	0.564903

Remark: $dist(0) = P(L_6 < n)$.

Table: KS test with Longest run

PRNG	minstd_rand0	minstd_rand	mt19937	mt19937_64
n	3100	3100	3200	3200
sup diff	0.00108607	0.00431393	0.0106533	0.00485326
arg max	11	11	9	9
p-value	1	0.992323	0.20643	0.972563

Remark: empirical distribution of longest run generated by minstd_rand0 is too fit to the true distribution. The p-value of the other side of KS test is almost 0, i.e. $P(\sup_x |emp(x) - dist(x)| \leq \text{ sup diff}) \approx 0$.

Table: KS test with Longest run

PRNG	ranlux24	ranlux48	knuth_b
n	2400	3200	3100
sup diff	0.00482436	0.00634618	0.00868296
arg max	11	13	10
p-value	0.974082	0.815406	0.437958

Reference I

[1] A. Rukhin, J. Soto, J. Nechvtal, M. Smid, E. Barker, S. Leigh, M. Levenson, M. Vangel, D. Banks, A. Heckert, J. Dray, and S. Vo.
A statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications, NIST Special Publication 800-22 Revised 1a.
US, 2010.

[2] H. Takahashi.
The explicit formulae for the distributions of words, 2023.
ICIAM 2023, Waseda Univ.