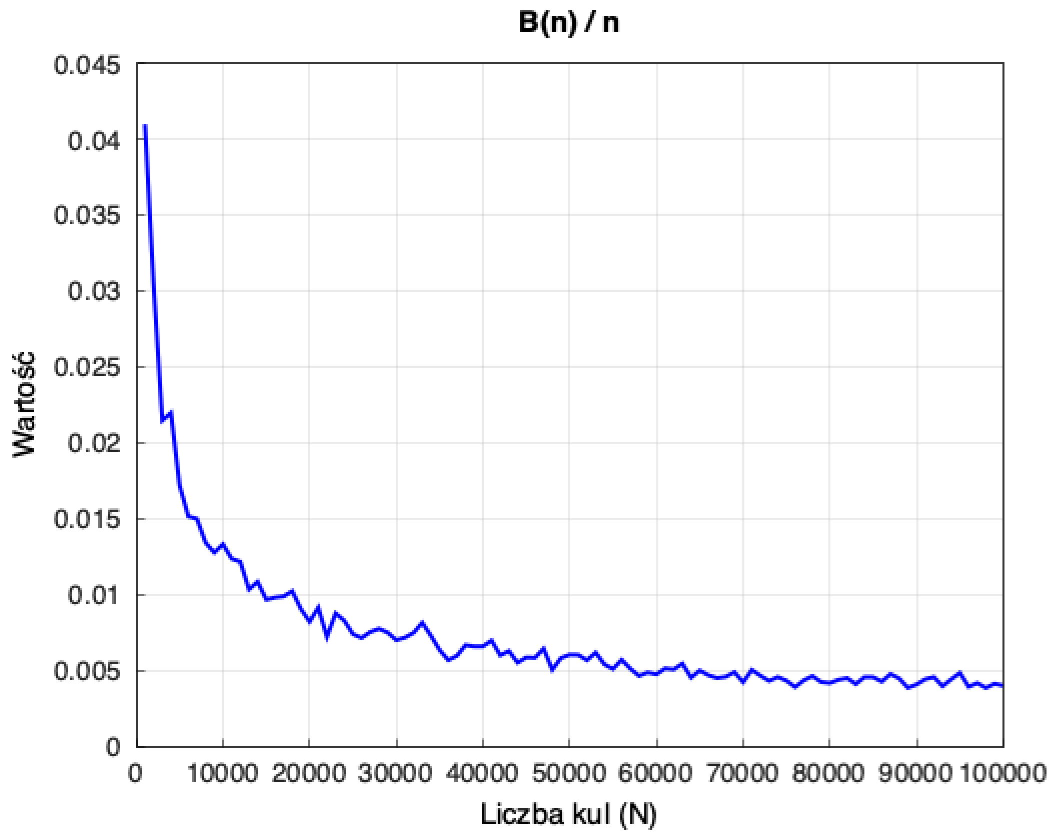
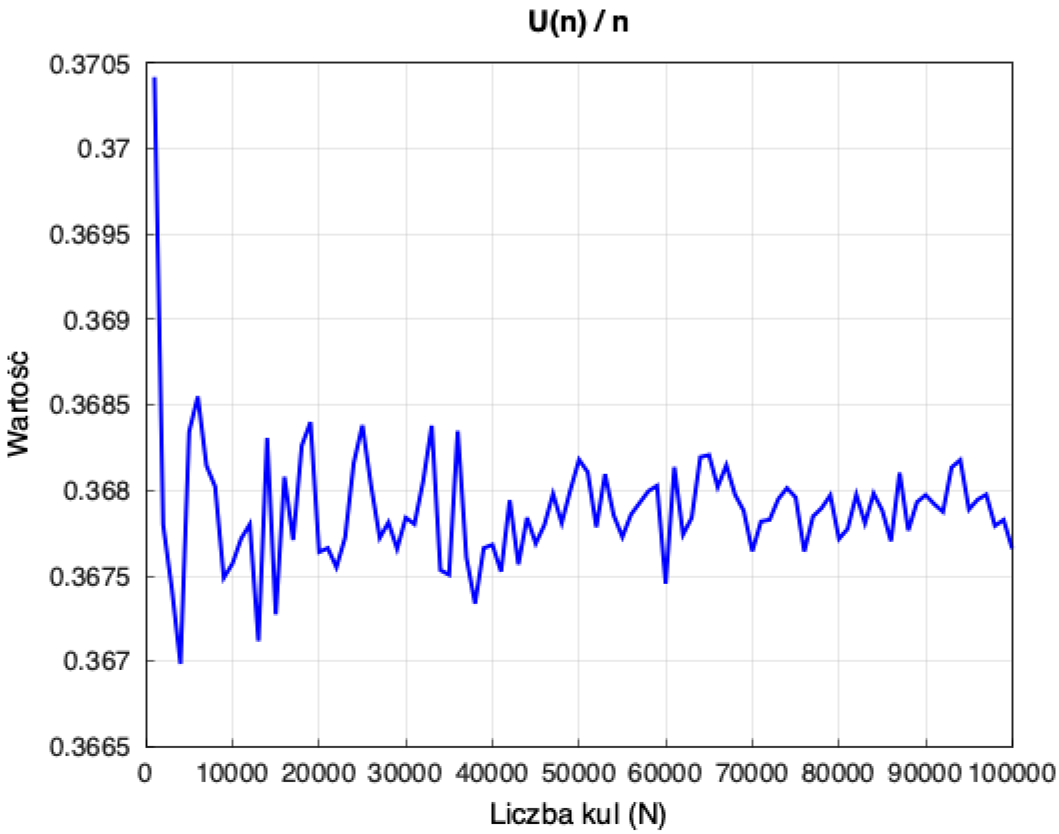
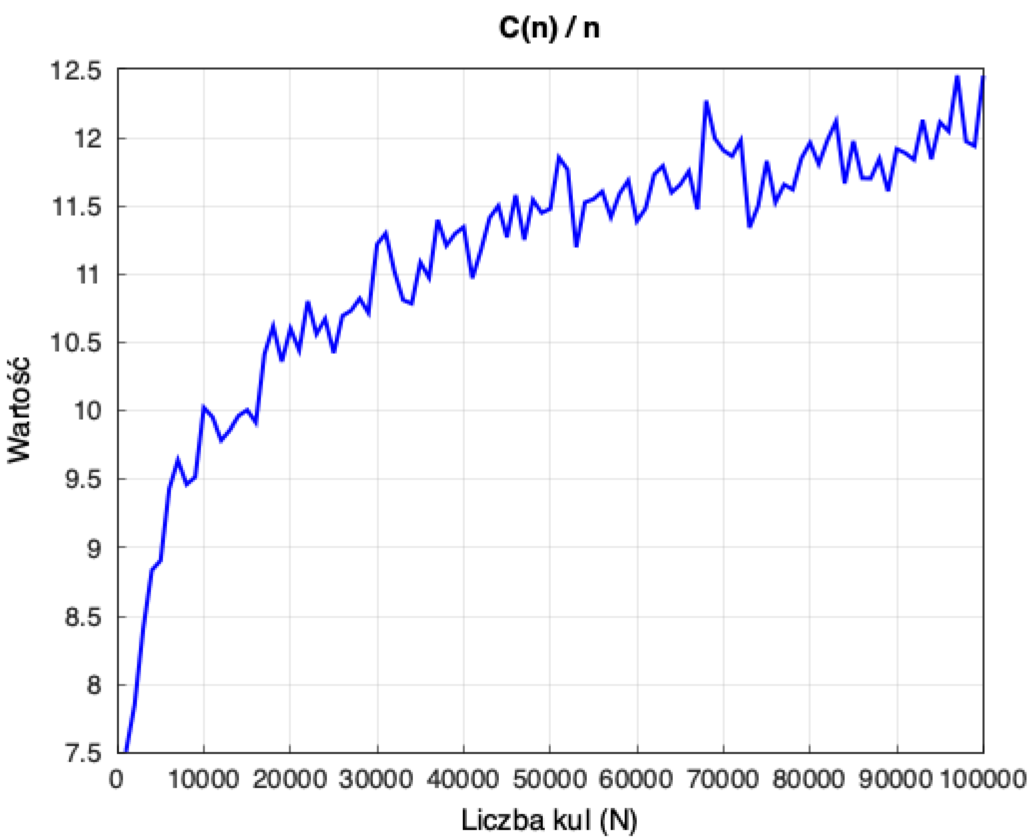
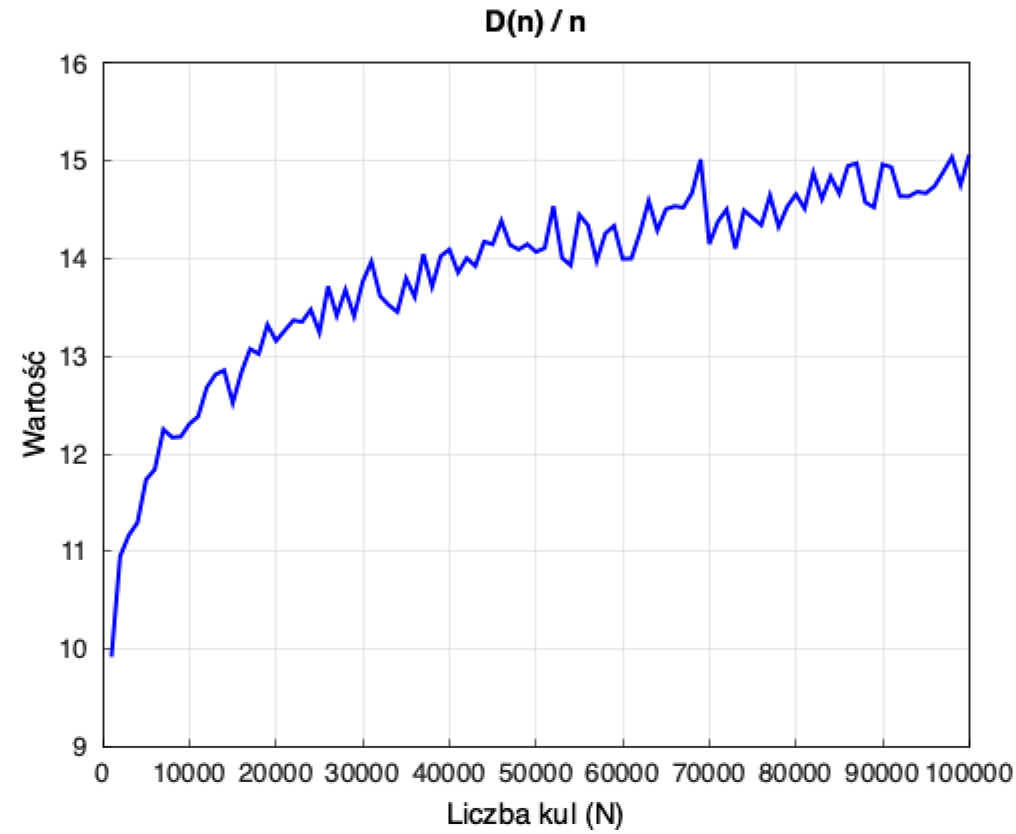
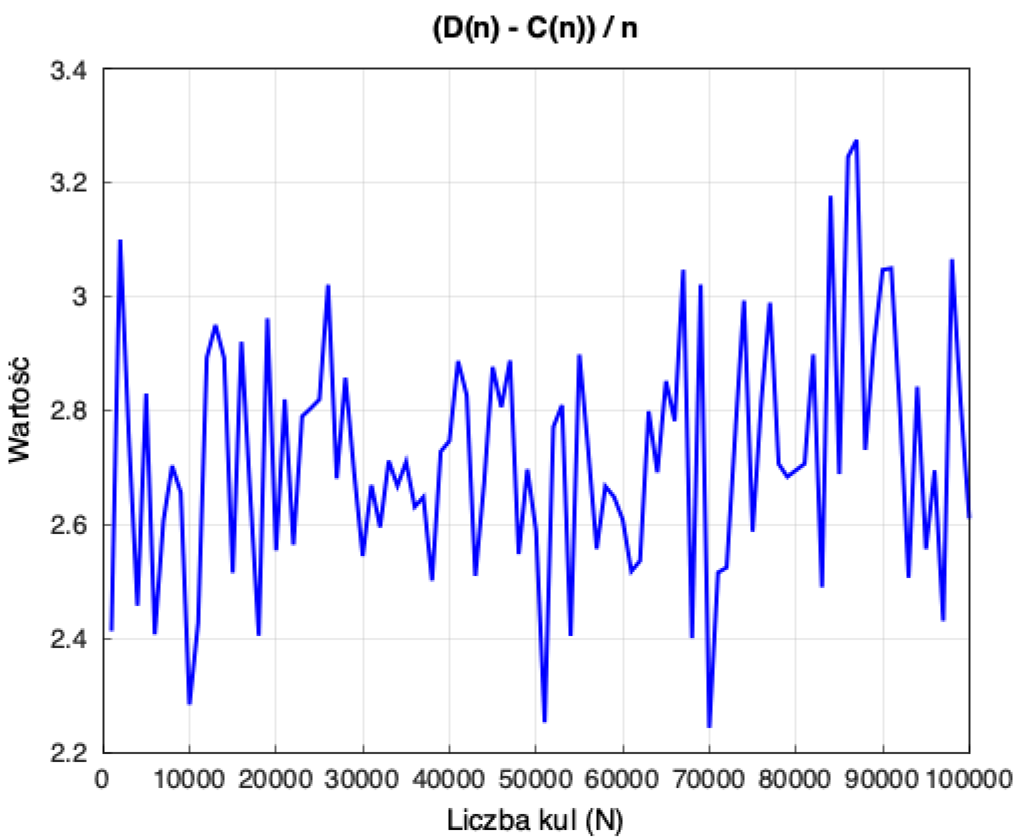
Homework2
Anna Grelewska
Sprawozdanie

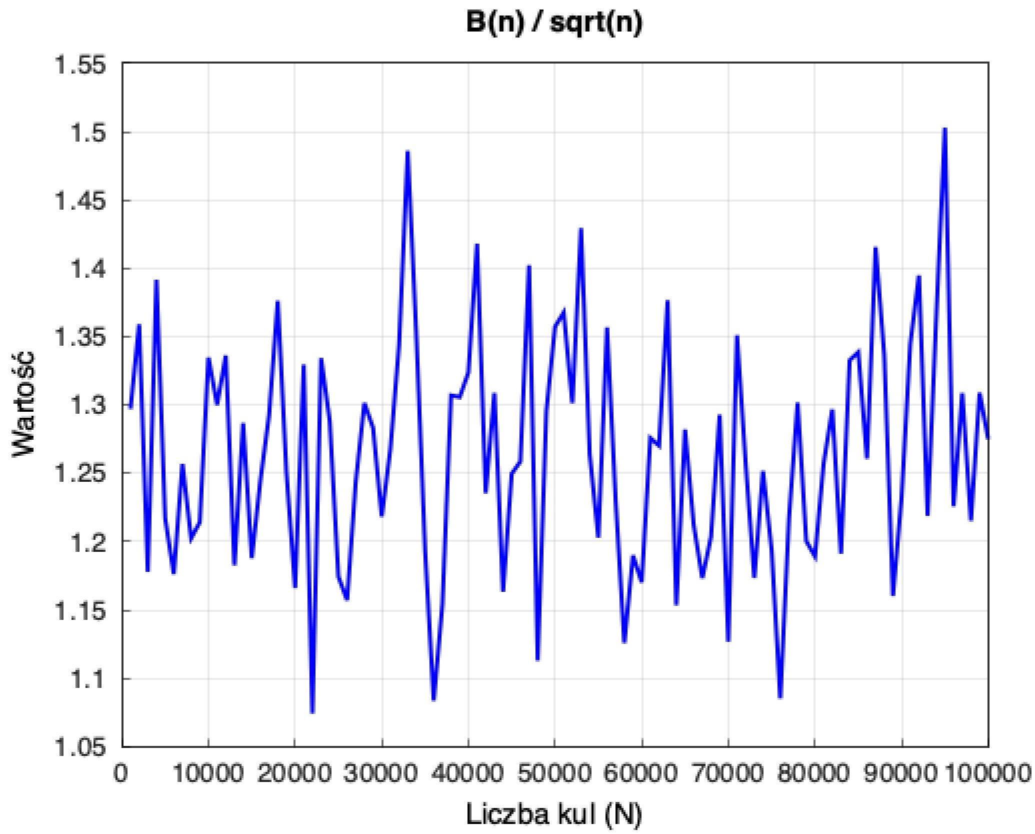


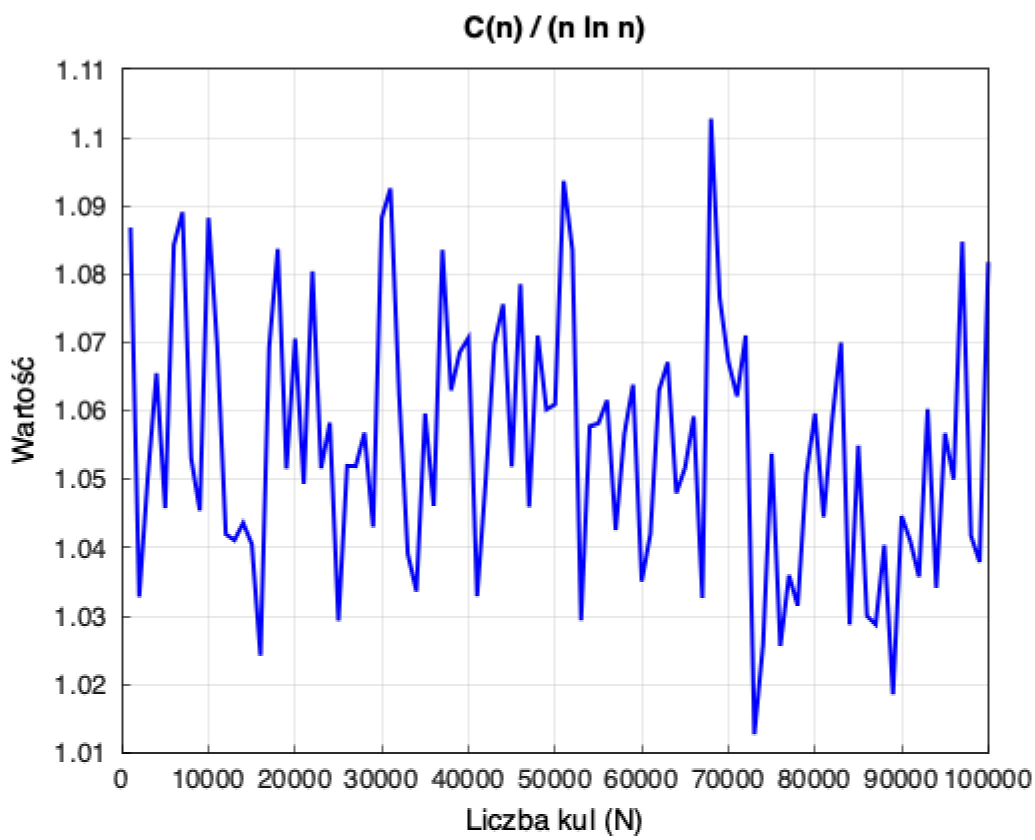

U(n) / n
0.3705
0.37
0.3695
0.369
0.3685
0.368
0.3675
0.367
0.3665
Wartość
0
10000
20000
30000
40000
50000
60000
70000
80000
90000
100000
Liczba kul (N)

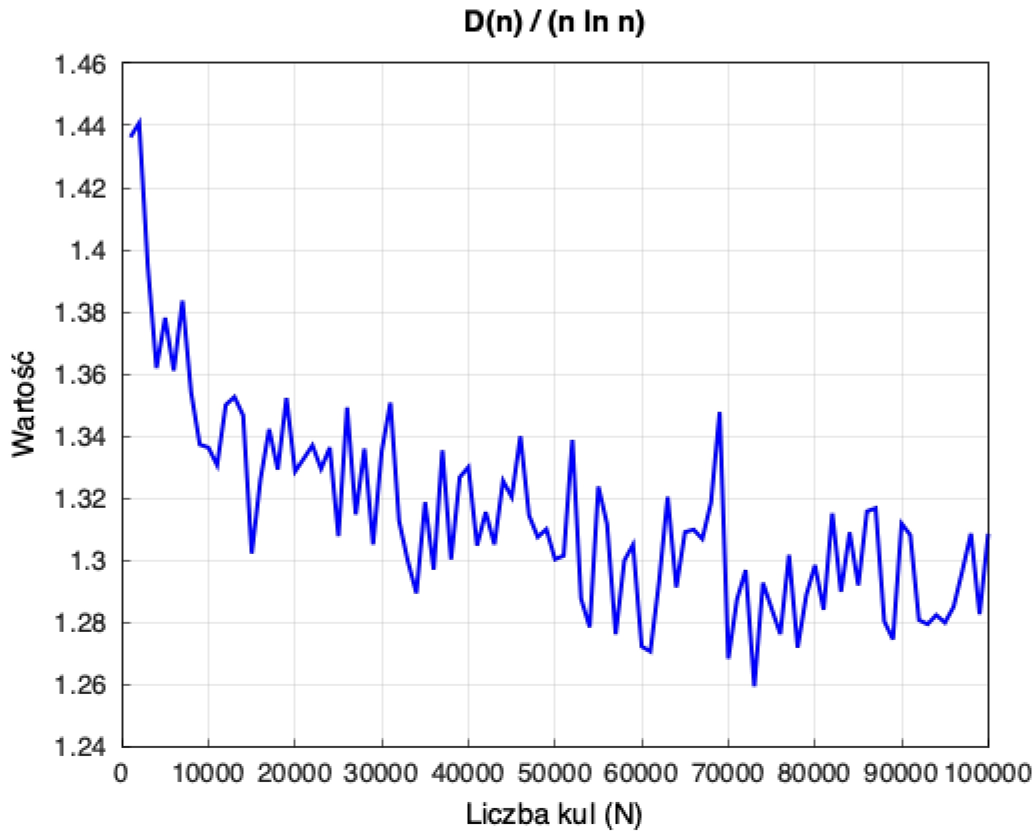

C(n) / n
12.5
12
11.5
11
10.5
10
9.5
9
8.5
8
7.5
Wartość
0
10000
20000
30000
40000
50000
60000
70000
80000
90000
100000
Liczba kul (N)

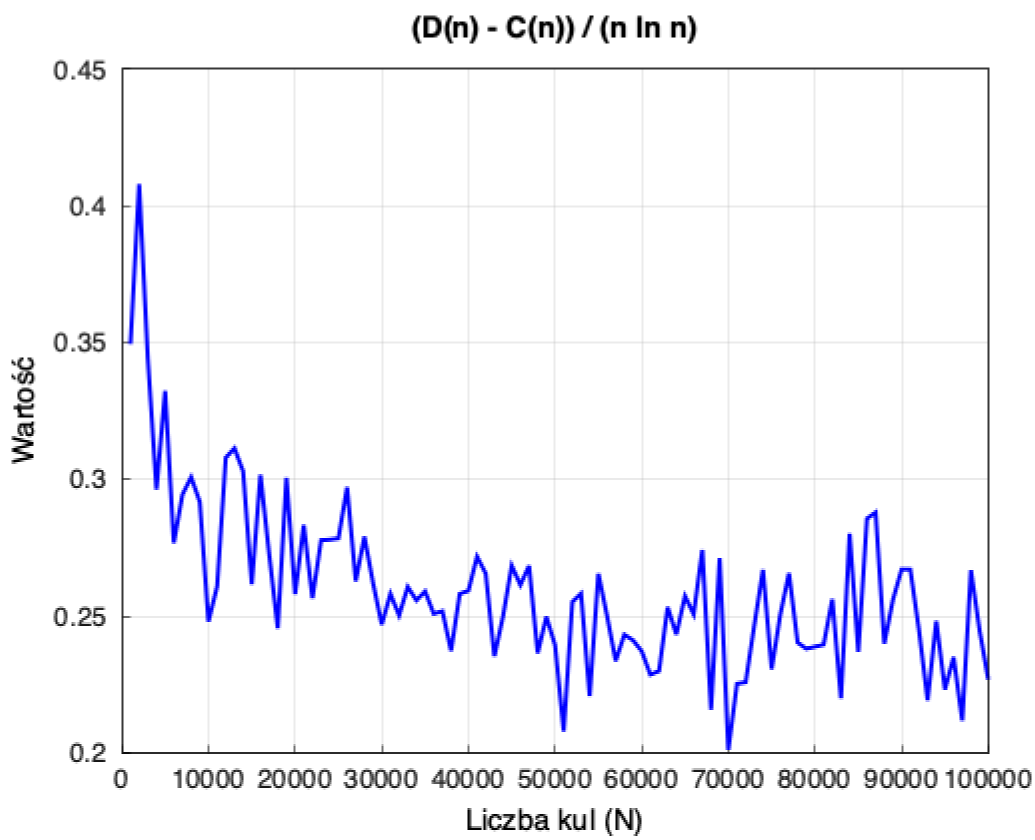

D(n) / n
Wartość
Liczba kul (N)
16
15
14
13
12
11
10
9
0
10000
20000
30000
40000
50000
60000
70000
80000
90000
100000

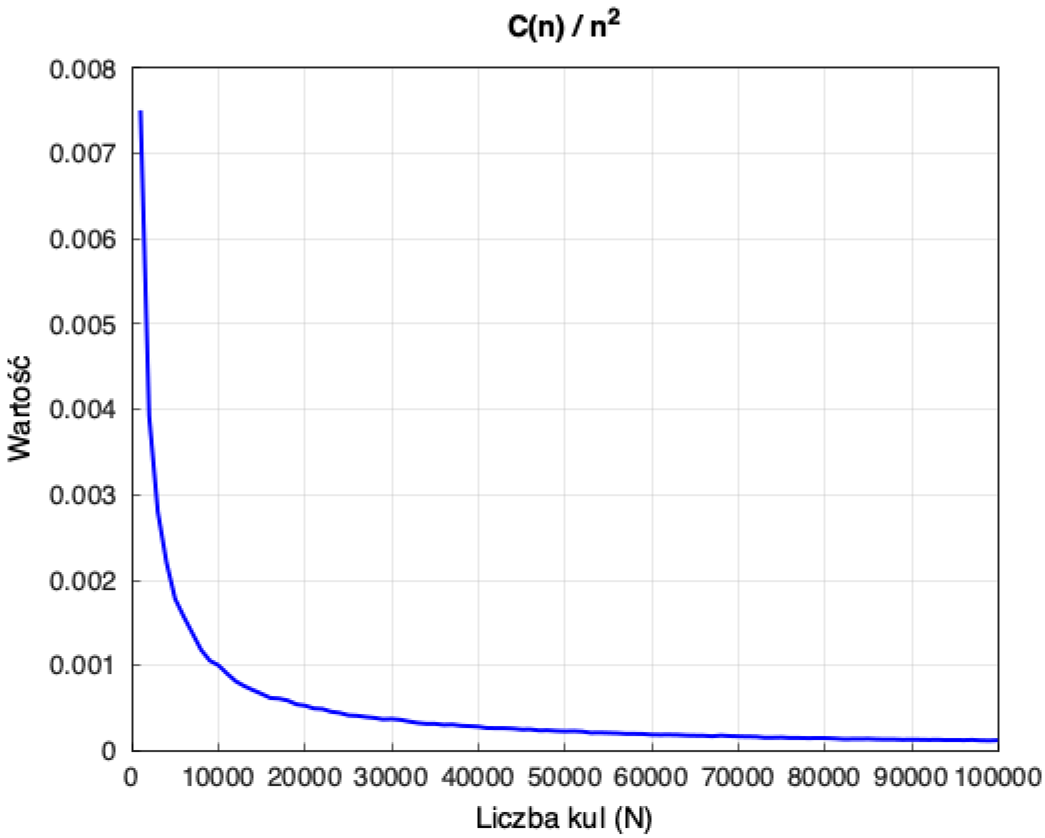

(D(n) - C(n)) / n
Wartość
Liczba kul (N)
3.4
3.2
3
2.8
2.6
2.4
2.2
0
10000
20000
30000
40000
50000
60000
70000
80000
90000
100000

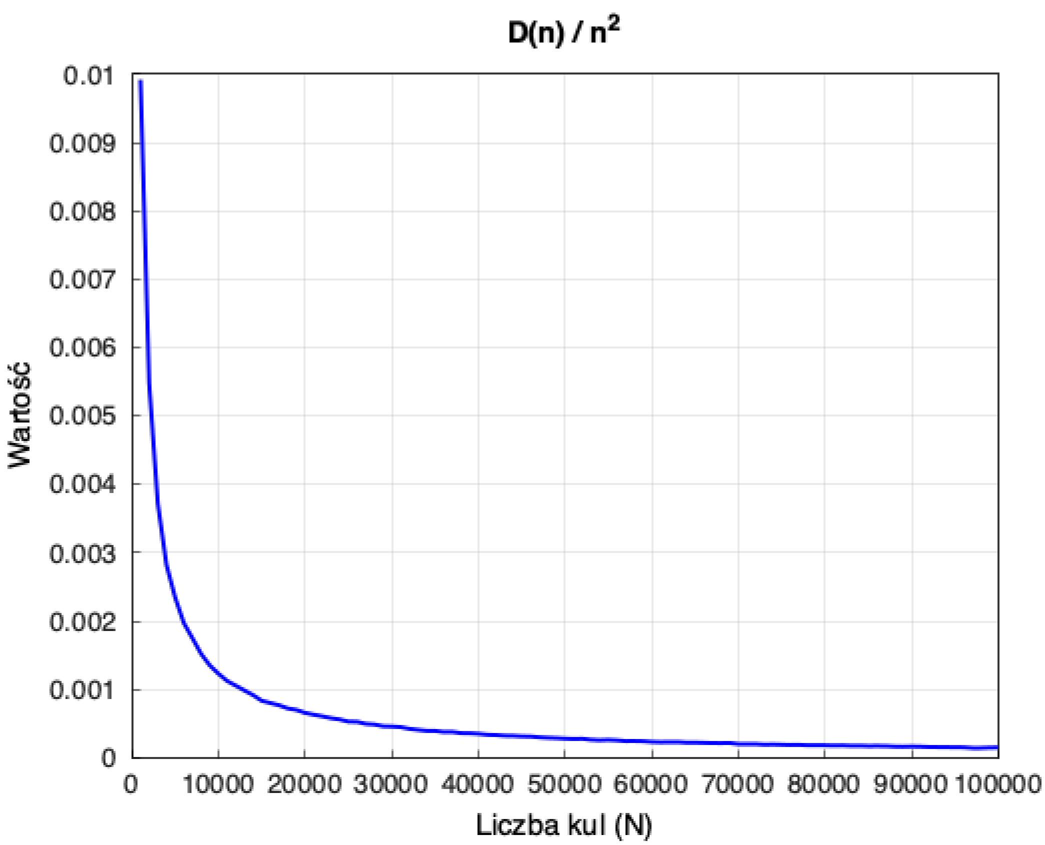

B(n) / sqrt(n)
Wartość
Liczba kul (N)
1.55
1.5
1.45
1.4
1.35
1.3
1.25
1.2
1.15
1.1
1.05
0
10000
20000
30000
40000
50000
60000
70000
80000
90000
100000

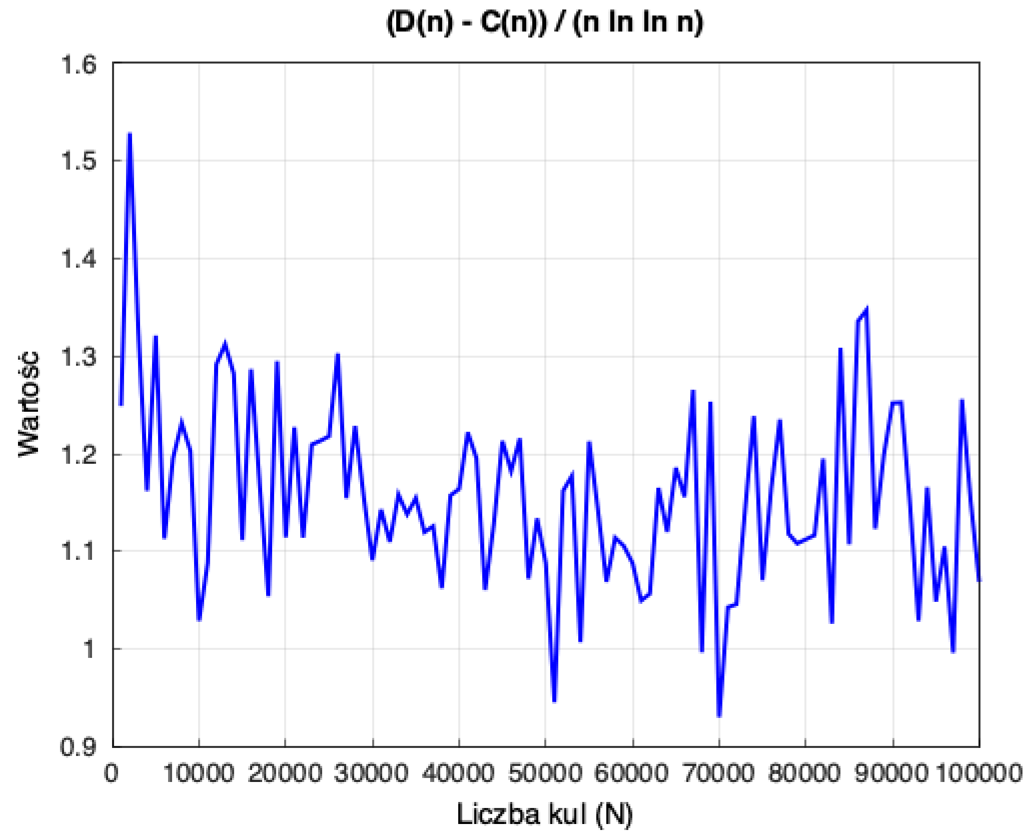

C(n) / (n ln n)
Wartość
Liczba kul (N)
1.11
1.1
1.09
1.08
1.07
1.06
1.05
1.04
1.03
1.02
1.01
0
10000
20000
30000
40000
50000
60000
70000
80000
90000
100000

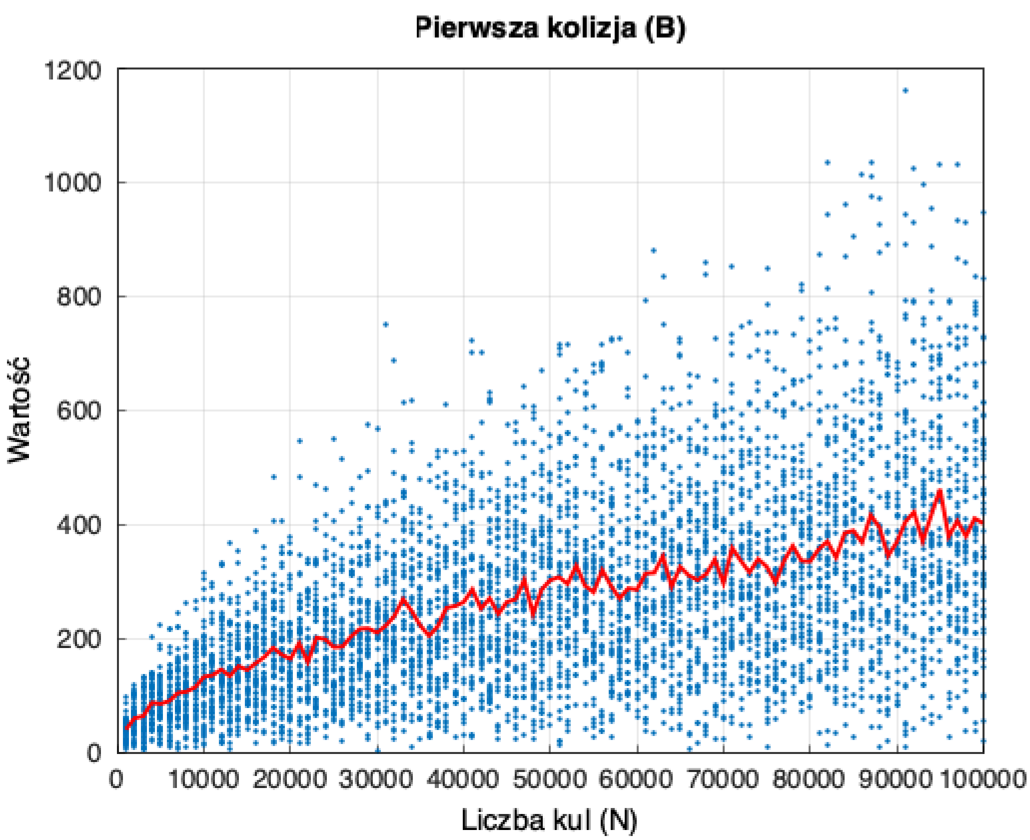

D(n) / (n ln n)
1.46
1.44
1.42
1.4
1.38
1.36
1.34
1.32
1.3
1.28
1.26
1.24
Wartość
0
10000
20000
30000
40000
50000
60000
70000
80000
90000
100000
Liczba kul (N)

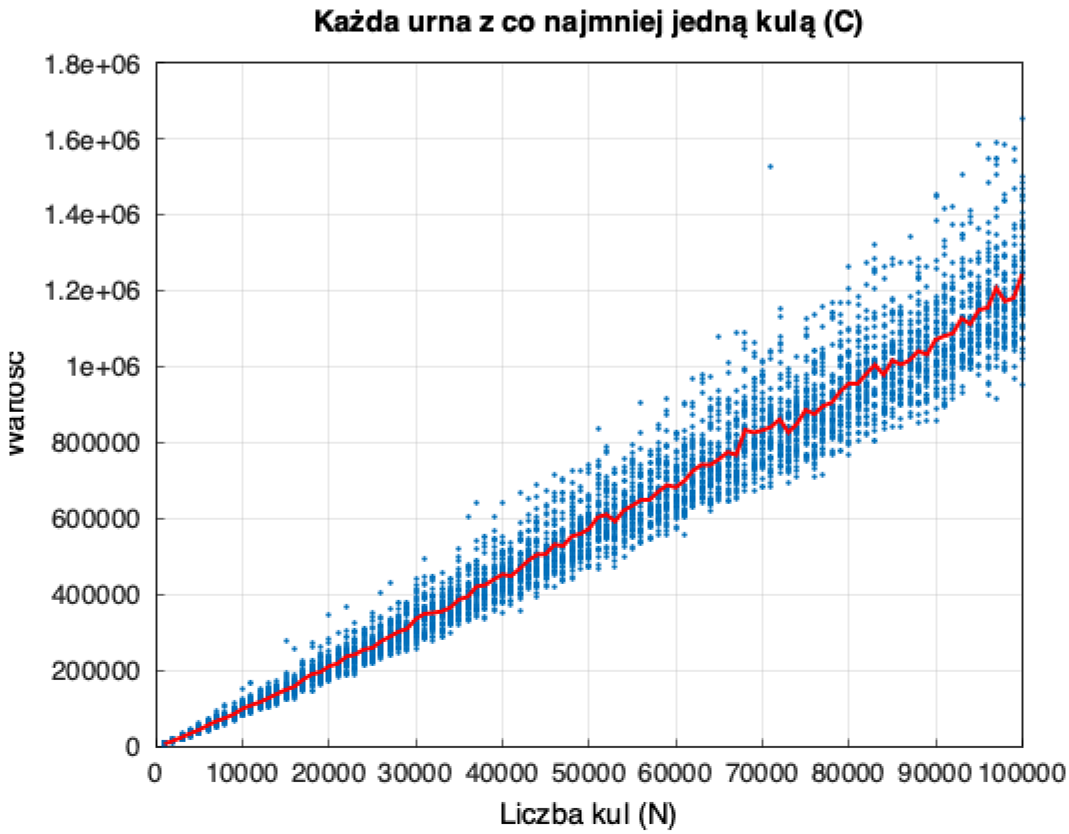

(D(n) - C(n)) / (n ln n)
0.45
0.4
0.35
0.3
0.25
0.2
Wartość
0
10000
20000
30000
40000
50000
60000
70000
80000
90000
100000
Liczba kul (N)

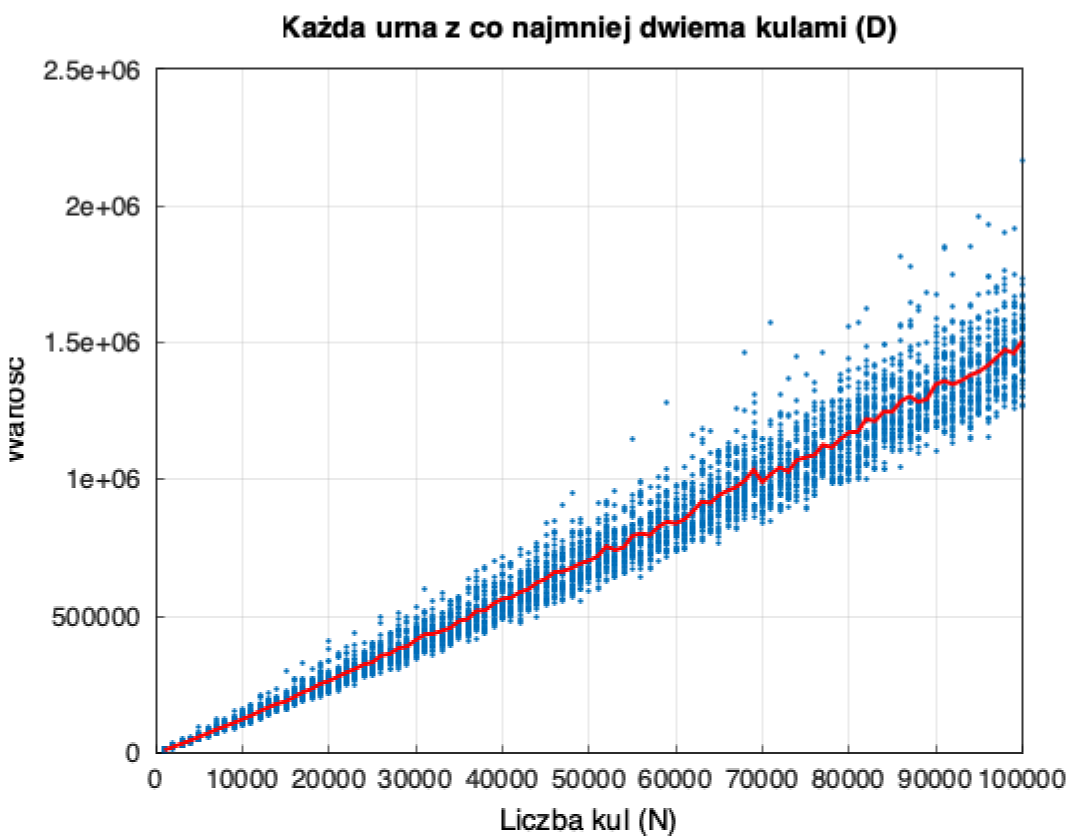

C(n) / n2
Wartość
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0
0
10000
20000
30000
40000
50000
60000
70000
80000
90000
100000
Liczba kul (N)

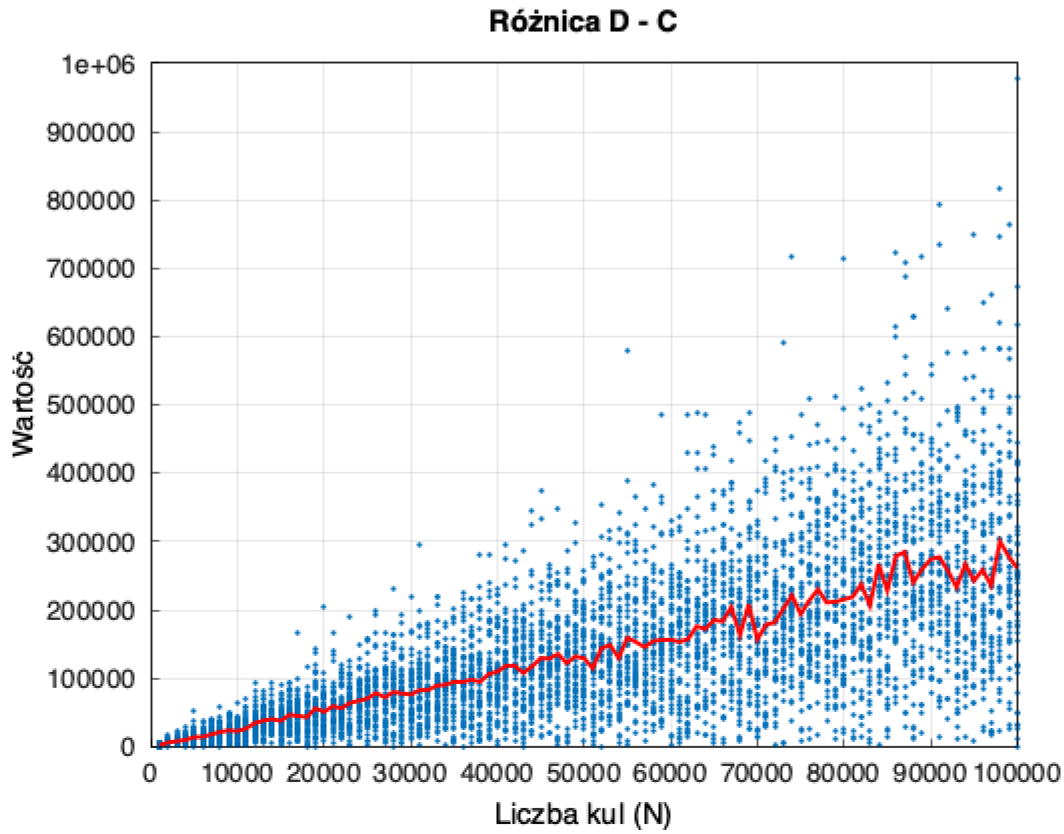

D(n) / n2
Wartość
0.01
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0
0
10000
20000
30000
40000
50000
60000
70000
80000
90000
100000
Liczba kul (N)

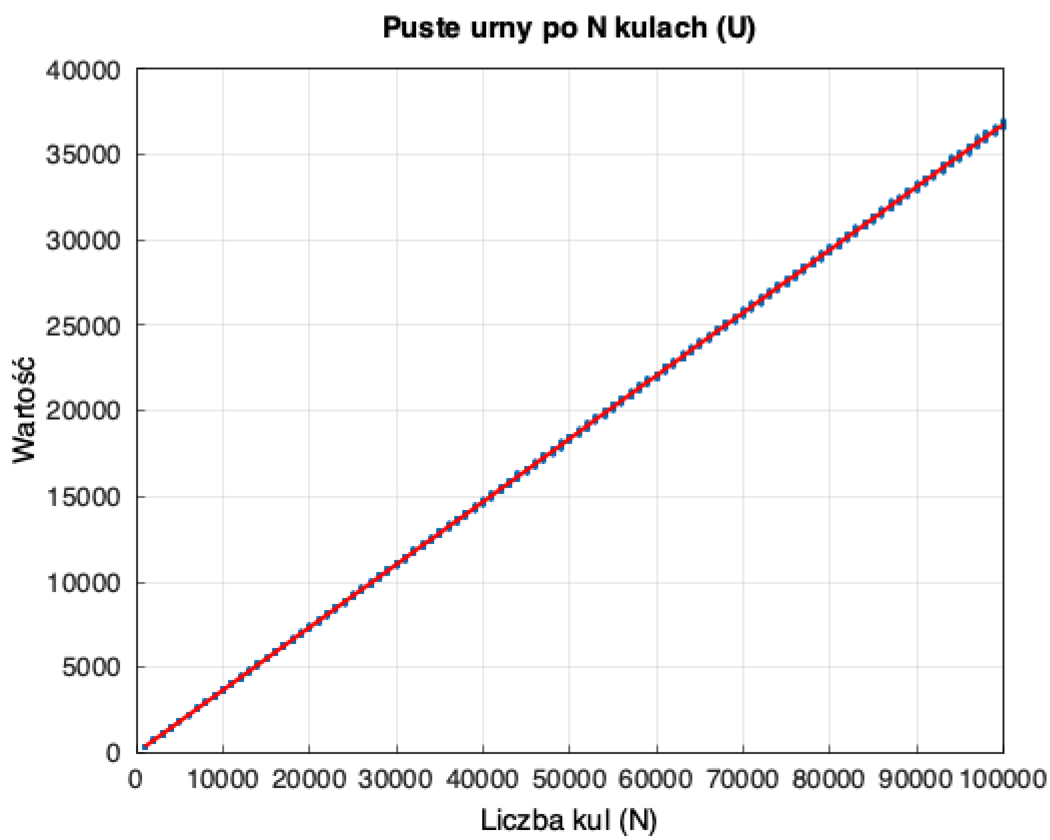

(D(n) - C(n)) / (n ln ln n)
Wartość
Liczba kul (N)
1.6
1.5
1.4
1.3
1.2
1.1
1
0.9
0
10000
20000
30000
40000
50000
60000
70000
80000
90000
100000


Pierwsza kolizja (B)
Wartość
Liczba kul (N)
1200
1000
800
600
400
200
0
0
10000
20000
30000
40000
50000
60000
70000
80000
90000
100000


Każda urna z co najmniej jedną kulą (C)
wartosc
Liczba kul (N)
0
200000
400000
600000
800000
1e+06
1.2e+06
1.4e+06
1.6e+06
1.8e+06
0
10000
20000
30000
40000
50000
60000
70000
80000
90000
100000


Każda urna z co najmniej dwiema kulami (D)
wartosc
Liczba kul (N)
0
500000
1e+06
1.5e+06
2e+06
2.5e+06
0
10000
20000
30000
40000
50000
60000
70000
80000
90000
100000


Różnica D - C
Wartość
Liczba kul (N)
1e+06
900000
800000
700000
600000
500000
400000
300000
200000
100000
0
10000
20000
30000
40000
50000
60000
70000
80000
90000
100000


Puste urny po N kulach (U)
Wartość
Liczba kul (N)
40000
35000
30000
25000
20000
15000
10000
5000
0
10000
20000
30000
40000
50000
60000
70000
80000
90000
100000

Celem zadania było zbadanie klasycznego modelu probabilistycznego kul i urn, polegającego na losowym wrzucaniu kul do urn. Na podstawie przeprowadzonych symulacji miały zostać wyznaczone następujące wielkości:

1. Bn – moment pierwszej kolizji (ang. birthday paradox),
2. Un – liczba pustych urn po wrzuceniu n kul,
3. Cn – minimalna liczba rzutów potrzebna do zapełnienia wszystkich urn (ang. coupon collector's problem),
4. Dn – minimalna liczba rzutów potrzebna do tego, aby każda urna zawierała co najmniej dwie kule,
5. Dn - Cn – różnica między Dn a Cn.

Symulacje zostały wykonane dla $n \in \{1000, 2000, ..., 100000\}$, a każdy eksperyment powtarzany był 50 razy.

Analiza wyników

Bn – Moment pierwszej kolizji

Pierwsza kolizja zachodzi bardzo szybko. Wykres $B(n) / \sqrt{n}$ pokazuje stabilizację wokół 1, co oznacza, że $Bn \sim \sqrt{n}$. Wyniki są zgodne z paradoksem urodzinowym, który przewiduje szybkie wystąpienie kolizji. Pojedyncze pomiary Bn mogą być bardzo odległe od średniej – nawet dziesięciokrotnie mniejsze lub większe.

Un – Liczba pustych urn

Liczba pustych urn rośnie liniowo wraz z n. Wykres U(n) / n stabilizuje się w przedziale 0.365 - 0.370, co pokazuje, że asymptotycznie $Un \sim 0.37 \cdot n$. Poszczególne wyniki są bardzo blisko średniej, co czyni je niemal niewidocznymi na wykresach.

Cn – Liczba rzutów do wypełnienia urn

Cn rośnie logarytmicznie z n. Na wykresie $C(n) / (n \ln n)$ widać oscylacje wokół 1, co potwierdza, że $Cn \sim n \ln n$. Wyniki te dobrze ilustrują problem kolekcjonera kuponów. Warto zauważyć, że wyniki są bliższe średniej niż w przypadku Bn, ale z czasem odchylenie od średniej rośnie.

Dn – Liczba rzutów do wypełnienia urn co najmniej dwiema kulami

Dn zachowuje się podobnie jak Cn, ale osiąga większe wartości. Iloraz $D(n) / (n \ln n)$ oscyluje wokół 1, co wskazuje na proporcję $Dn \sim n \ln n$. Różnice między poszczególnymi wynikami są większe niż w przypadku Cn, szczególnie dla mniejszych wartości n.

Dn - Cn – Różnica między Dn a Cn

Różnica Dn - Cn dla mniejszych n wykazuje większą zmienność, ale z czasem stabilizuje się. Na wykresie $(D(n) - C(n)) / (n \ln \ln n)$ wartości zbliżają się do 1, co sugeruje, że $Dn - Cn \sim n \ln \ln n$. Jest to związane z tym, że już w momencie osiągnięcia Cn wiele urn zawiera więcej niż jedną kulę.

Asymptotyka wyników
1. Bn – Stabilizacja wokół B(n) / √n potwierdza asymptotykę Θ(√n). Wyniki są rozproszone, ale bliskie linii średniej.
2. Un – Wyniki U(n) / n wskazują na asymptotykę Θ(n) i bardzo dokładnie odzwierciedlają teoretyczne przewidywania.
3. Cn – Wykres C(n) / (n ln n) pokazuje, że asymptotyka tej wielkości to Θ(n ln n). Punkty na wykresie są blisko linii średniej.
4. Dn – Wyniki D(n) / (n ln n) również potwierdzają asymptotykę Θ(n ln n), choć z większymi odchyleniami niż w przypadku Cn.
5. Dn - Cn – Funkcja (D(n) - C(n)) / (n ln ln n) sugeruje asymptotykę Θ(n ln ln n), co jest zgodne z oczekiwaniami teoretycznymi.

Wnioski
- Wyniki symulacji dobrze ilustrują badane wielkości i wykazują zgodność z teoretycznymi przewidywaniami.
- Koncentracja wyników wokół średnich jest różna w zależności od wielkości – Un są niemal idealnie skupione, podczas gdy Bn i różnica Dn - Cn wykazują największe odchylenia.
- Wszystkie badane wielkości zachowują asymptotyczne zależności: Bn ~ √n, Un ~ 0.37 · n, Cn ~ n ln n, Dn ~ n ln n, Dn - Cn ~ n ln ln n.

Znaczenie nazw:
- Paradoks urodzinowy: Pokazuje, jak szybko dochodzi do kolizji. Wyniki dla Bn ilustrują, że prawdopodobieństwo kolizji rośnie szybciej, niż wynika to z intuicji.
- Problem kolekcjonera kuponów: Liczba rzutów Cn, potrzebna do wypełnienia wszystkich urn, odpowiada liczbie losowań wymaganych do zebrania wszystkich elementów w zestawie, co jest analogiczne do zbierania kuponów.

Zastosowanie w kryptografii:
- Paradoks urodzinowy pokazuje, że kolizje w funkcjach hashujących mogą wystąpić z większym prawdopodobieństwem, niż intuicyjnie się wydaje. Jest to kluczowe w projektowaniu funkcji hashujących oraz systemów kryptograficznych, gdzie unikanie kolizji ma fundamentalne znaczenie.