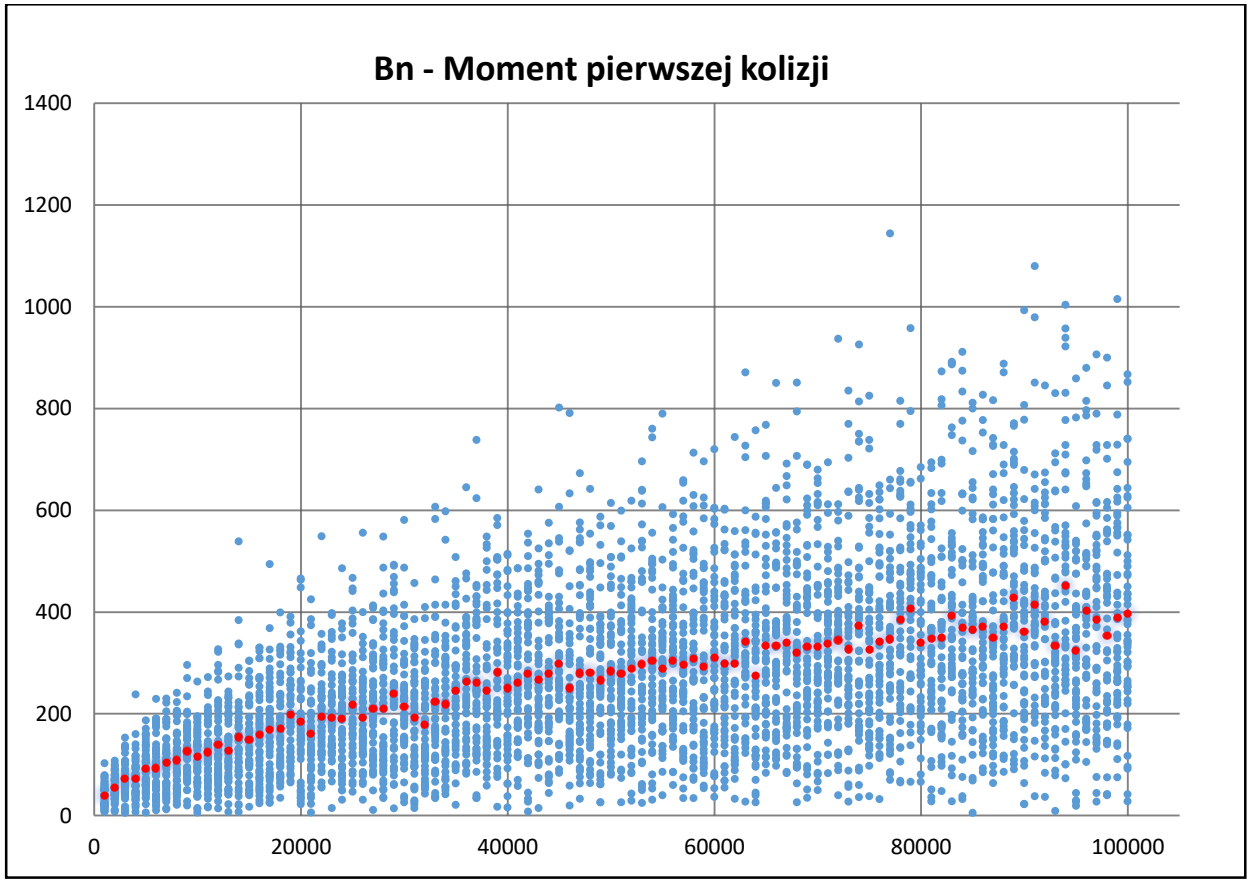
METODY PROBABILISTYCZNE I STATYSTYKA

ZADANIE DOMOWE 2

(Valeriia Loichyk)

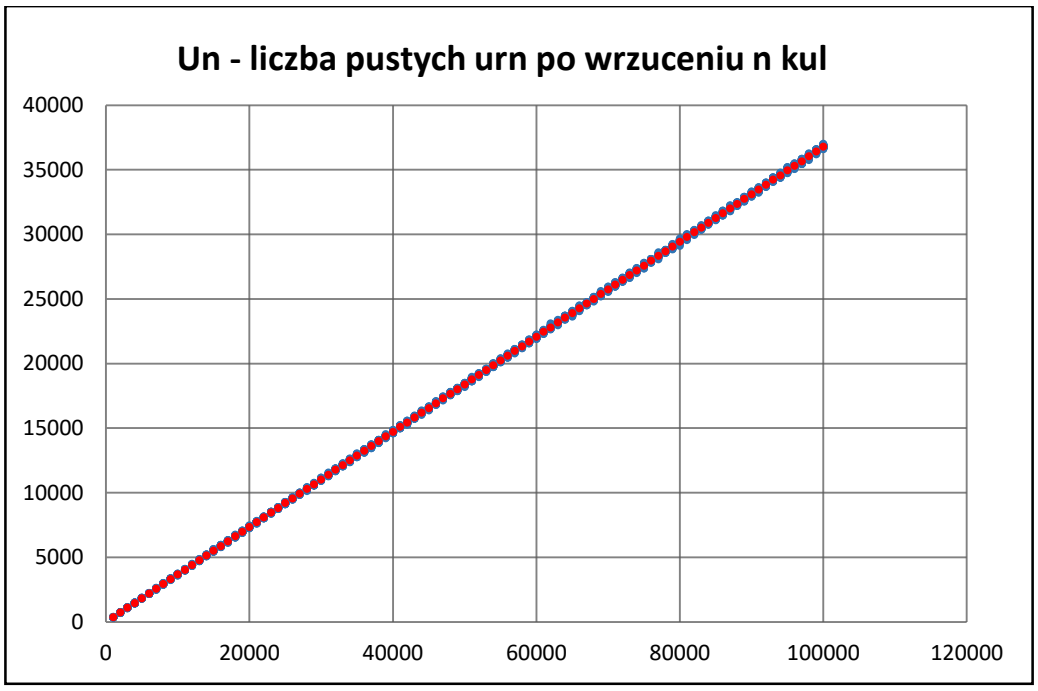
Zadanie 2.1:

(a) Bn – moment pierwszej kolizji; Bn = k, jeśli k-ta z wrzucanych kul jest pierwszą, która trafiła do niepustej urny. Niebieskie punkty przestawiają wyniki poszczególnych powtórzeń, czerwone punkty odpowiadają wartości średniej dla każdego n ∈ {1000, 2000, ... , 100 000}. Wykonano po k=50 niezależnych powtórzeń algorytmu.



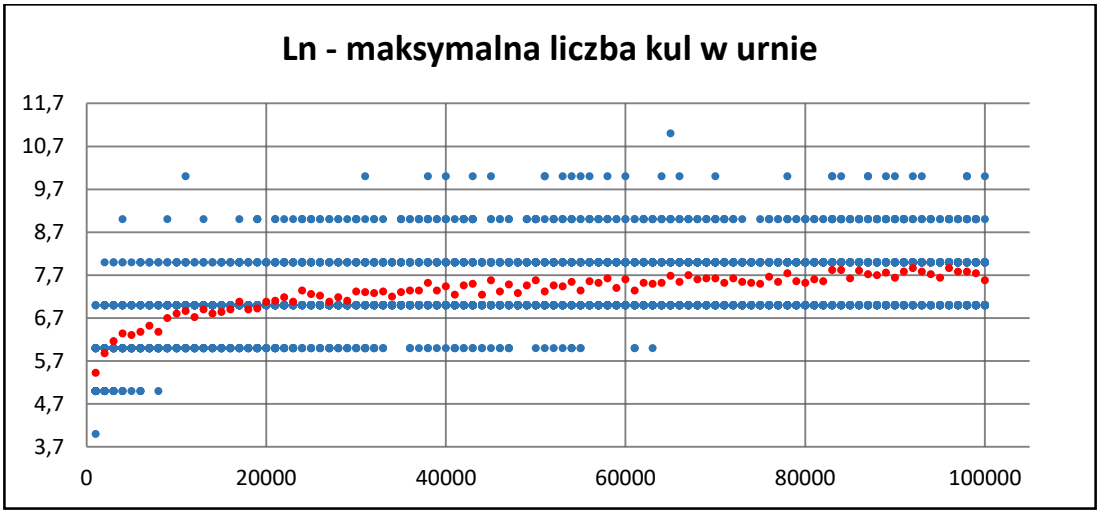
Wniosek: Patrząc na wyniki symulacji można zauważyć bardzo rozbieżne wartości dla poszczególnych prób momentów kolizji.

(b) Un – liczba pustych urn po wrzuceniu n kul. Niebieskie punkty przestawiają wyniki poszczególnych powtórzeń, czerwone punkty odpowiadają wartości średniej dla każdego n ∈ {1000, 2000, ... , 100 000}. Wykonano po k=50 niezależnych powtórzeń algorytmu.



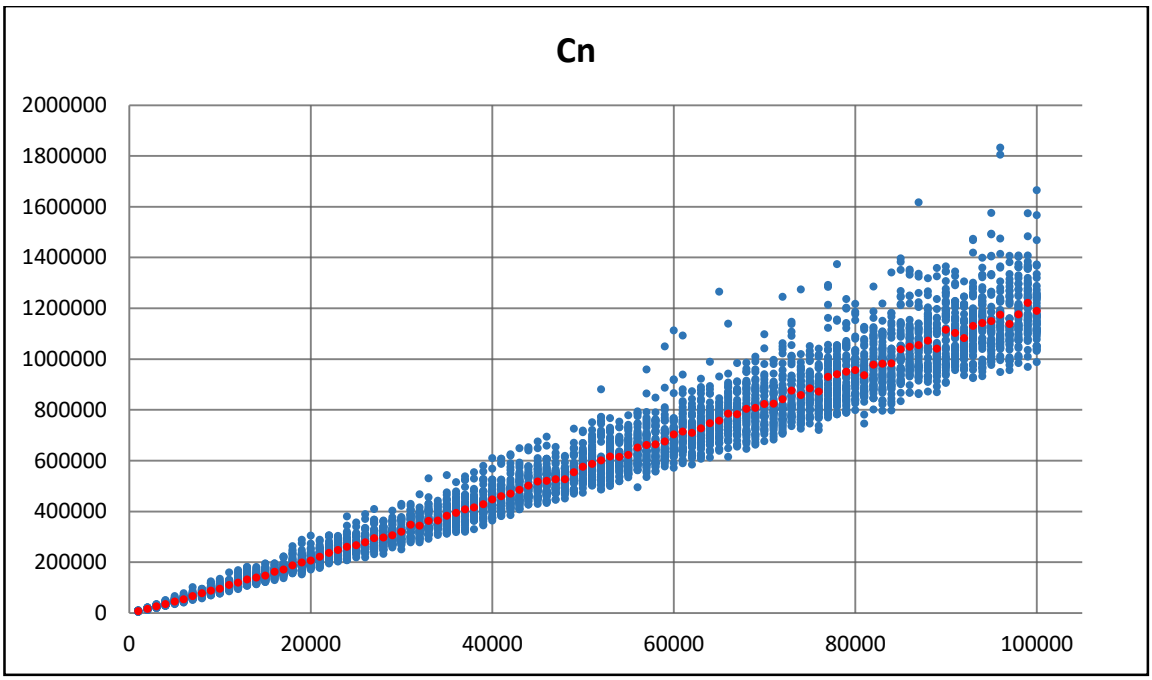
Wniosek: Możemy zauważyć, że liczba pustych urn po wrzuceniu n kul, rośnie liniowo i wartości poszczególnych powtórzeń są blisko średniej.

(c) Ln – maksymalna liczba kul w urnie po wrzuceniu n kul. Niebieskie punkty przestawiają wyniki poszczególnych powtórzeń, czerwone punkty odpowiadają wartości średniej dla każdego n ∈ {1000, 2000, ... , 100 000}. Wykonano po k=50 niezależnych powtórzeń algorytmu.



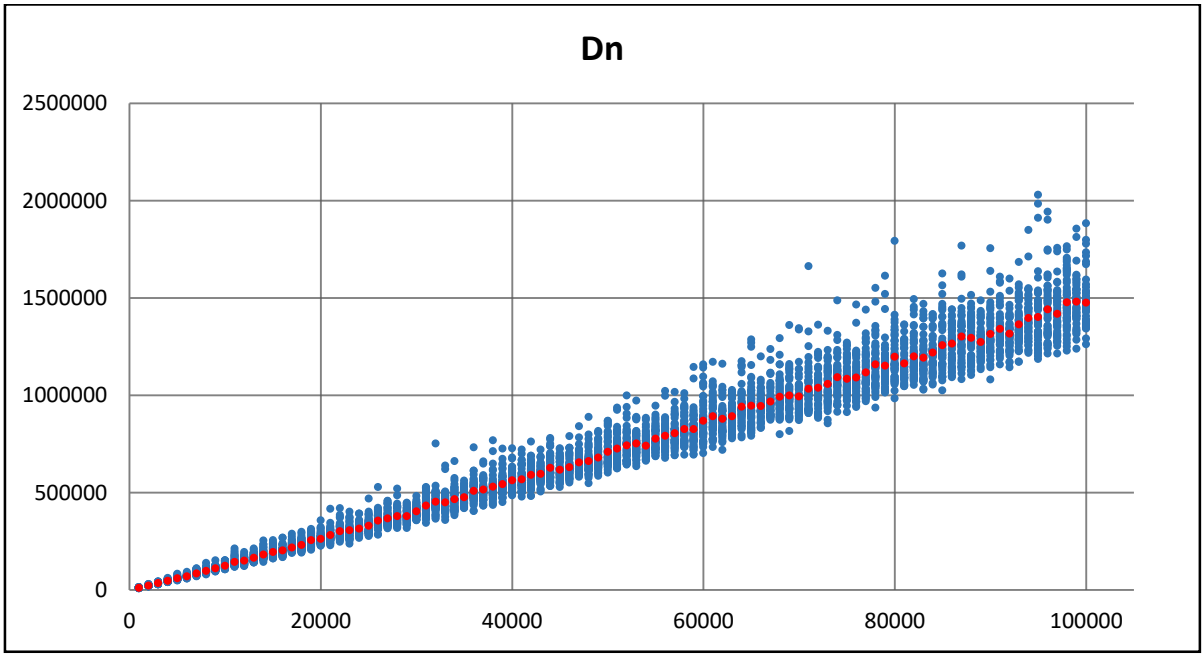
Wniosek: Możemy zauważyć, że wartości funkcji za każdym razem rosną wolniej.

(d) Cn – minimalna liczba rzutów, po której w każdej z urn jest co najmniej jedna kula. Niebieskie punkty przestawiają wyniki poszczególnych powtórzeń, czerwone punkty odpowiadają wartości średniej dla każdego n ∈ {1000, 2000, ... , 100 000}. Wykonano po k=50 niezależnych powtórzeń algorytmu.

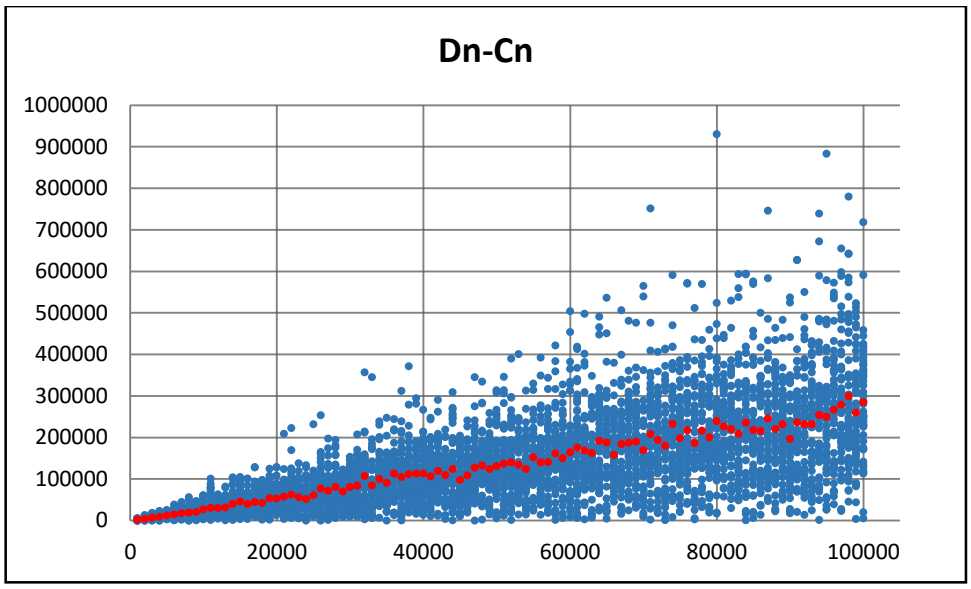


Wniosek: Możemy zauważyć, że wyniki symulacji rozbieżno rosną.

(e) Dn – minimalna liczba rzutów, po której w każdej z urn są co najmniej dwie kule. Niebieskie punkty przestawiają wyniki poszczególnych powtórzeń, czerwone punkty odpowiadają wartości średniej dla każdego n ∈ {1000, 2000, ... , 100 000}. Wykonano po k=50 niezależnych powtórzeń algorytmu.



(f) Dn−Cn – liczba rzutów od momentu Cn. Niebieskie punkty przestawiają wyniki poszczególnych powtórzeń, czerwone punkty odpowiadają wartości średniej dla każdego n ∈ {1000, 2000, ... , 100 000}. Wykonano po k=50 niezależnych powtórzeń algorytmu.



Zadanie 2.2:

Bn – uzyskiwane wyniki dla poszczególnych powtorzeń są dość rozbieżne.

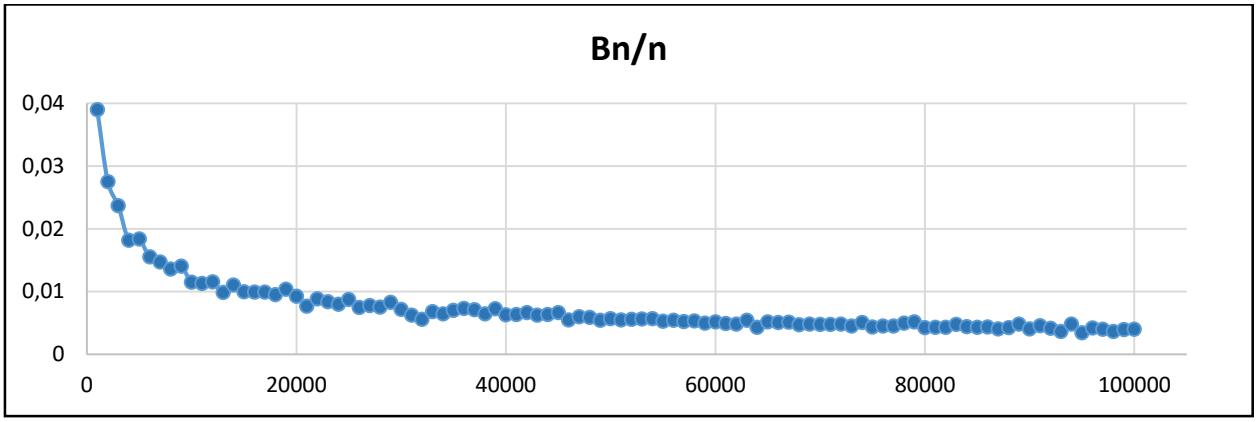
Un - odchylenia od średniej są minimalne.

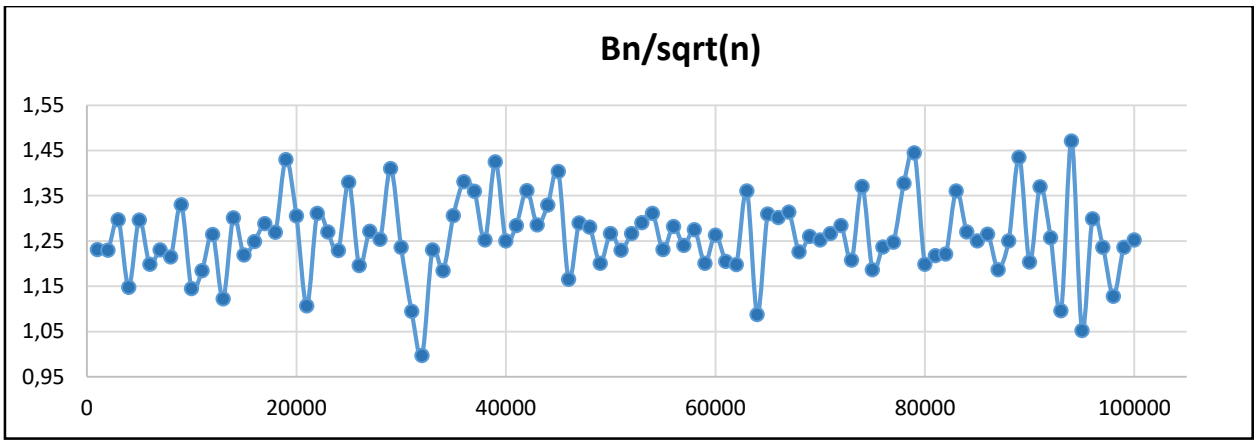
Ln – odchylenia od średniej są małe.

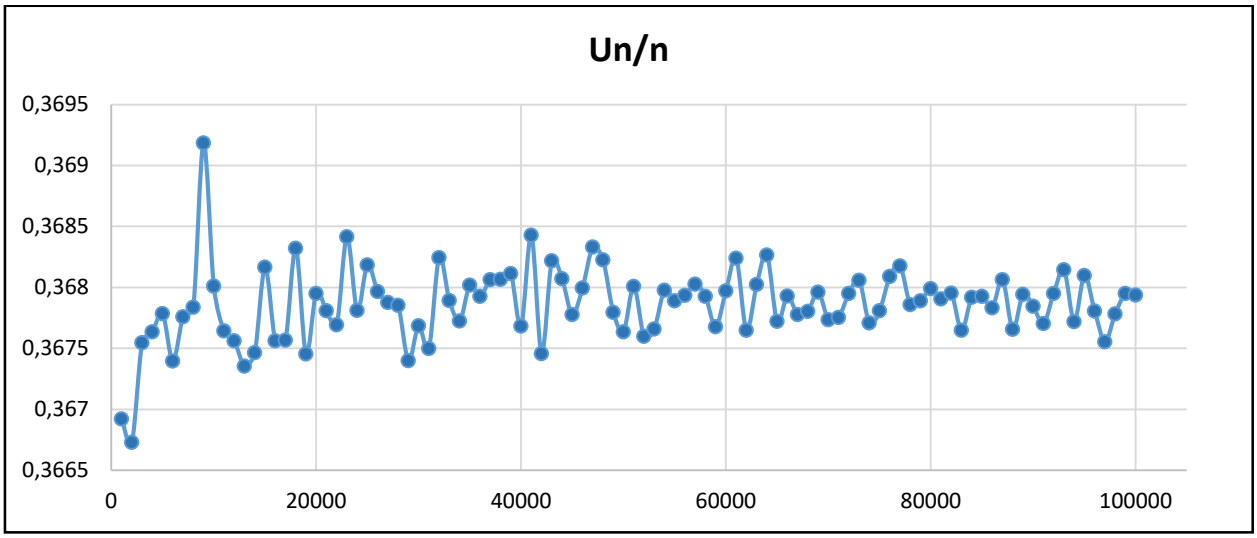
Cn i Dn – mają dość podobne wyniki i są rozbieżne.

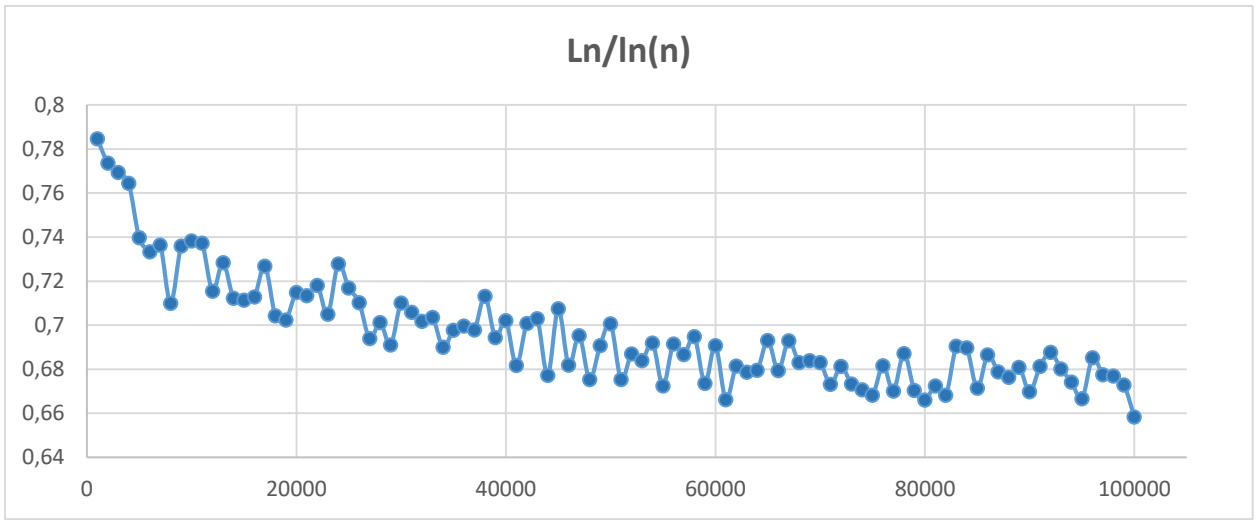
Dn-Cn – odchylenia od średniej są bardzo wielkie.

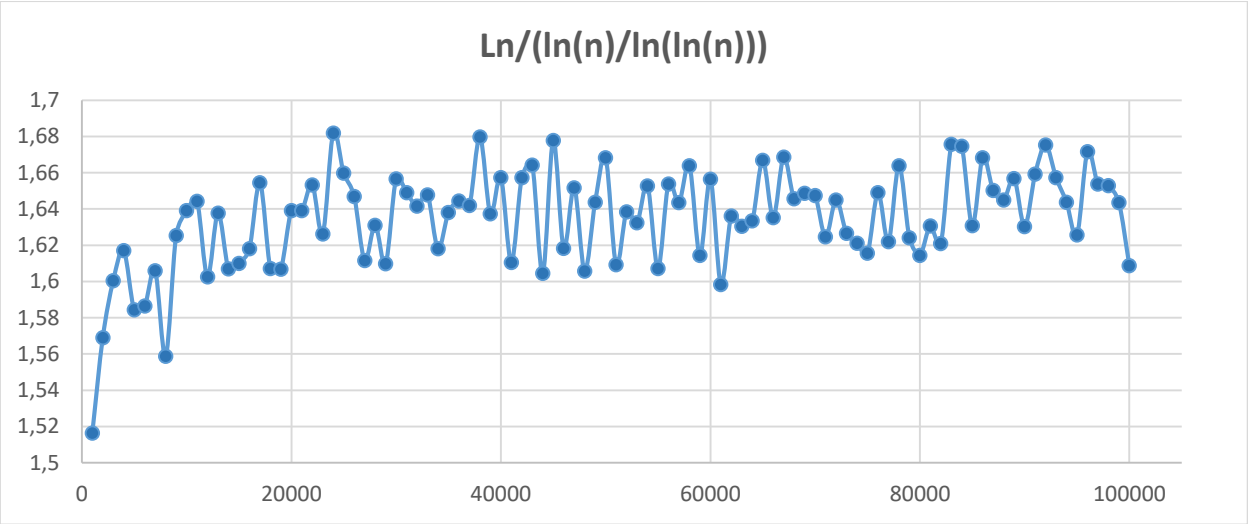
Zadanie 2.3:

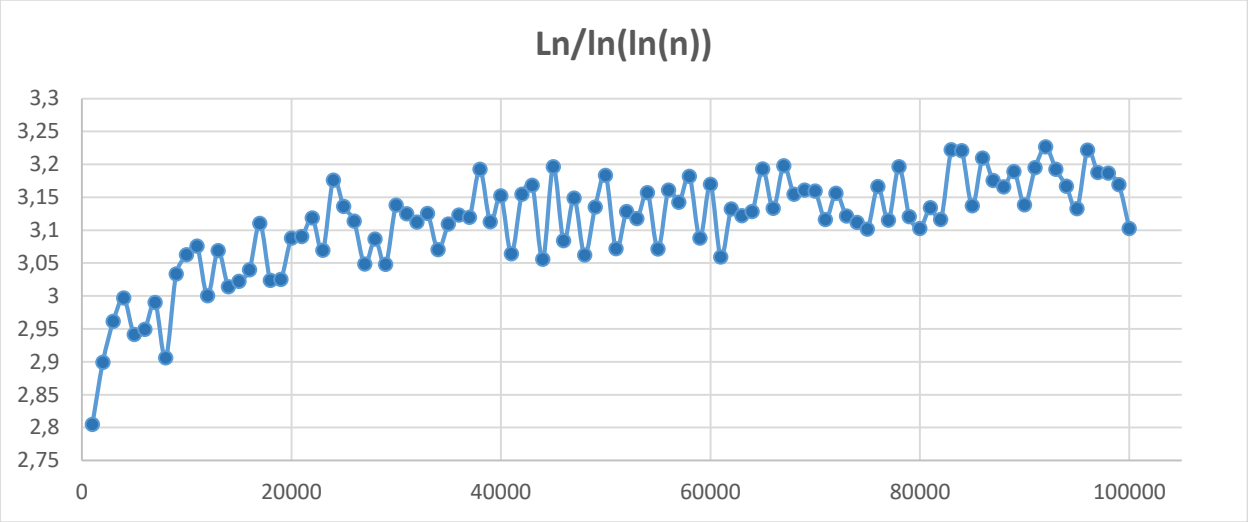


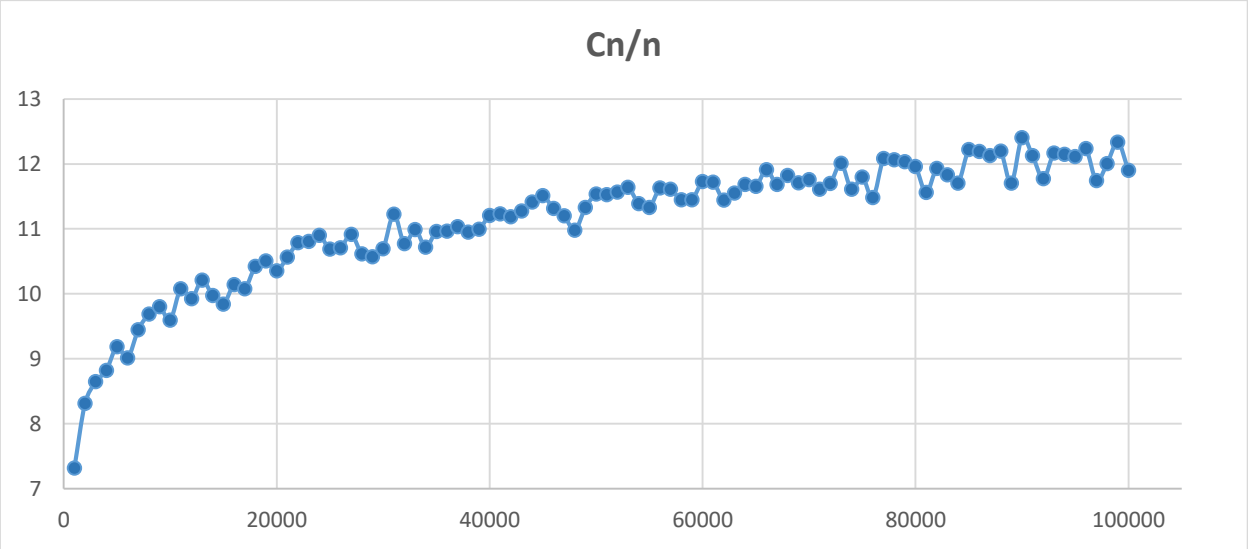

Bn/sqrt(n)
1,55
1,45
1,35
1,25
1,15
1,05
0,95
0
20000
40000
60000
80000
100000

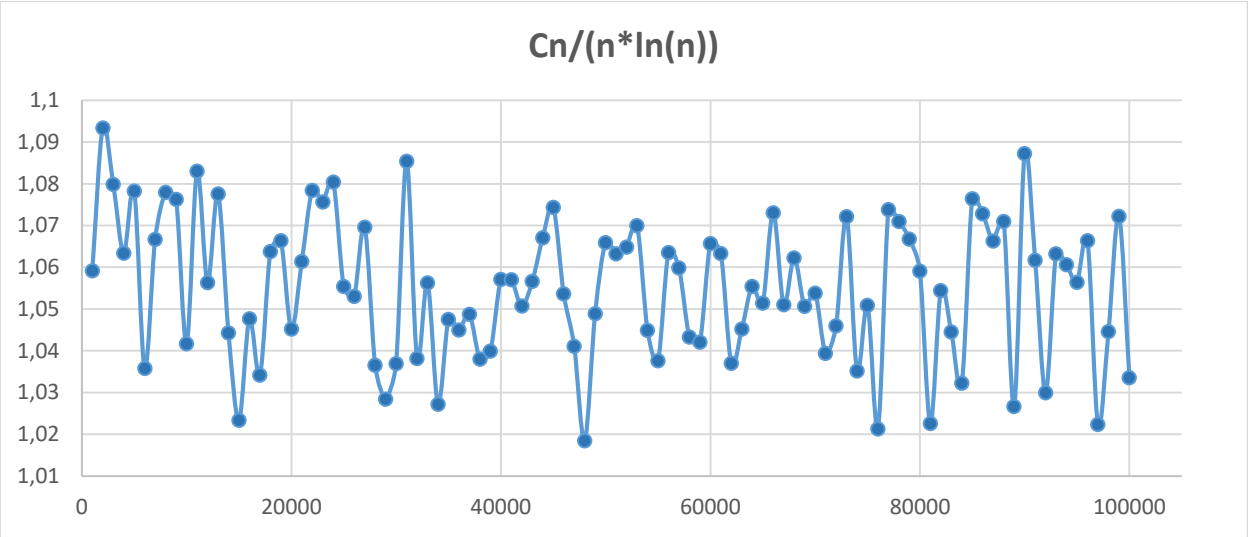

Un/n
0,3695
0,369
0,3685
0,368
0,3675
0,367
0,3665
0
20000
40000
60000
80000
100000

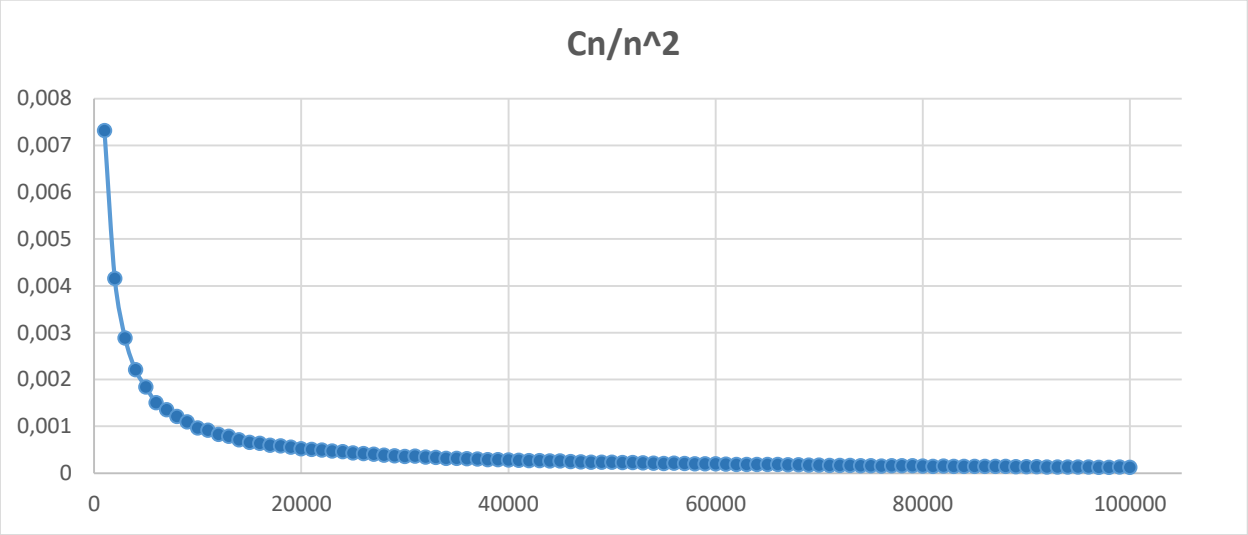

Ln/ln(n)
0,8
0,78
0,76
0,74
0,72
0,7
0,68
0,66
0,64
0
20000
40000
60000
80000
100000

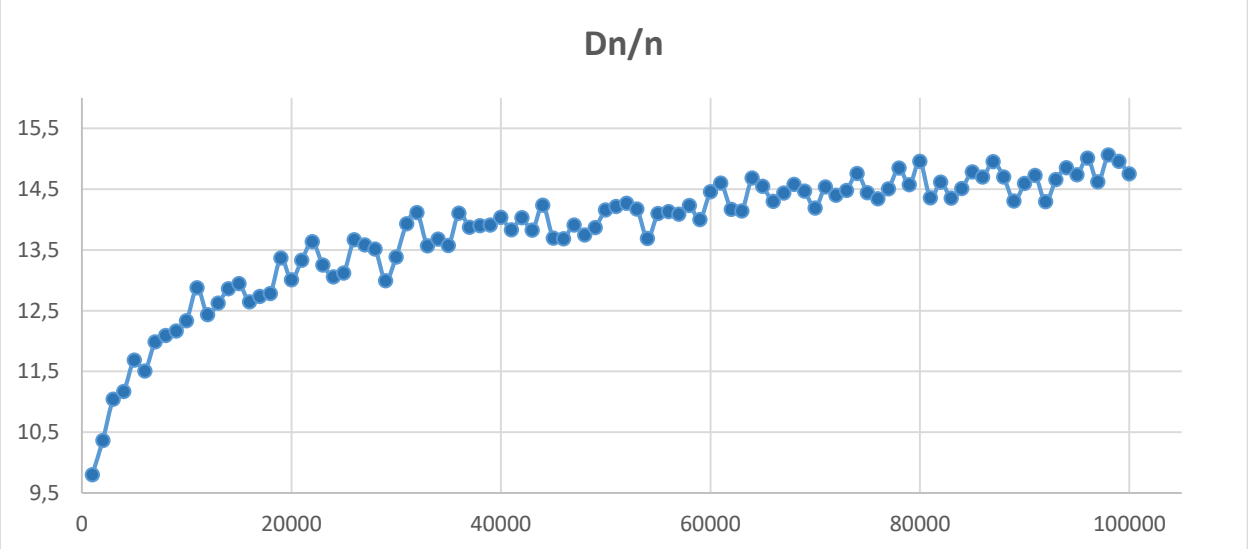

Ln/(ln(n)/ln(ln(n)))
1,7
1,68
1,66
1,64
1,62
1,6
1,58
1,56
1,54
1,52
1,5
0
20000
40000
60000
80000
100000

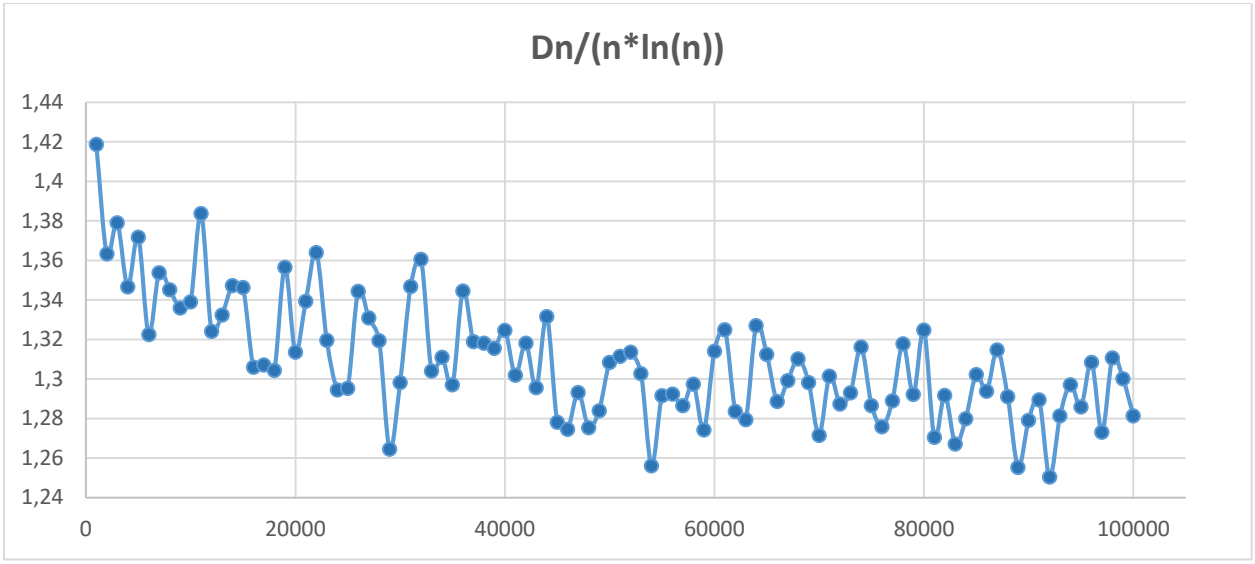

Ln/ln(ln(n))
3,3
3,25
3,2
3,15
3,1
3,05
3
2,95
2,9
2,85
2,8
2,75
0
20000
40000
60000
80000
100000

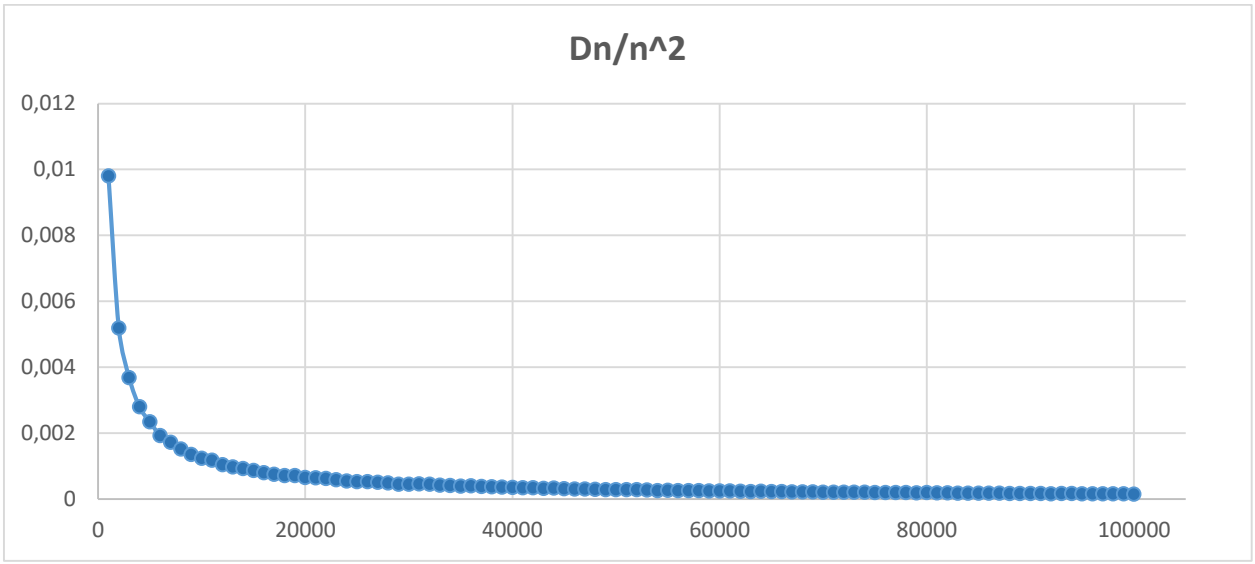

Cn/n
13
12
11
10
9
8
7
0
20000
40000
60000
80000
100000

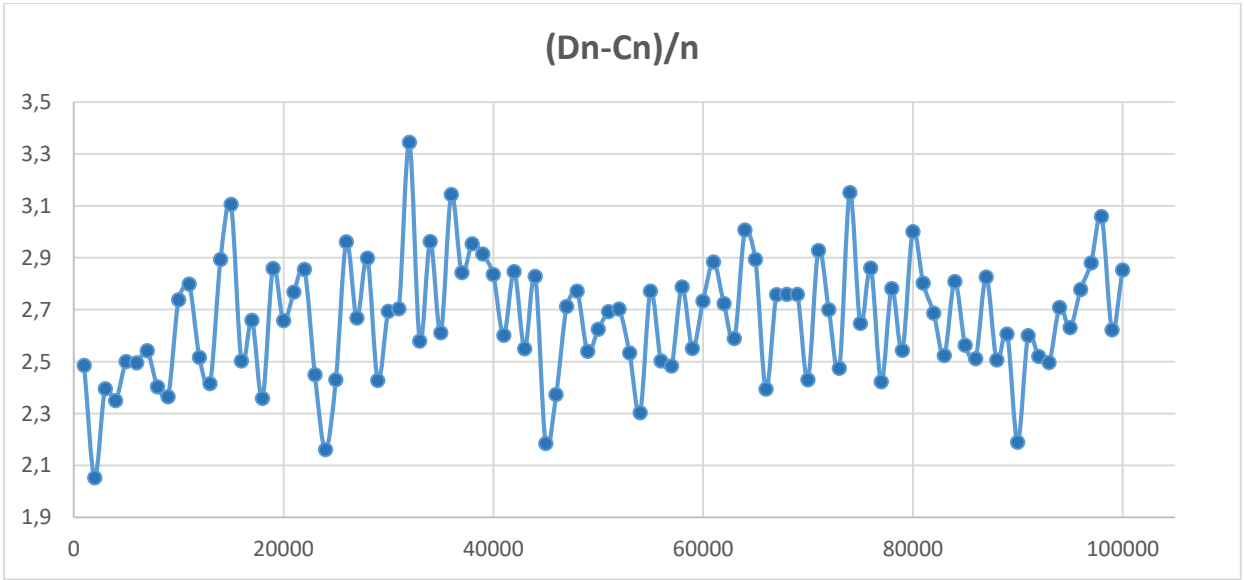

Cn/(n*ln(n))
1,1
1,09
1,08
1,07
1,06
1,05
1,04
1,03
1,02
1,01
0
20000
40000
60000
80000
100000

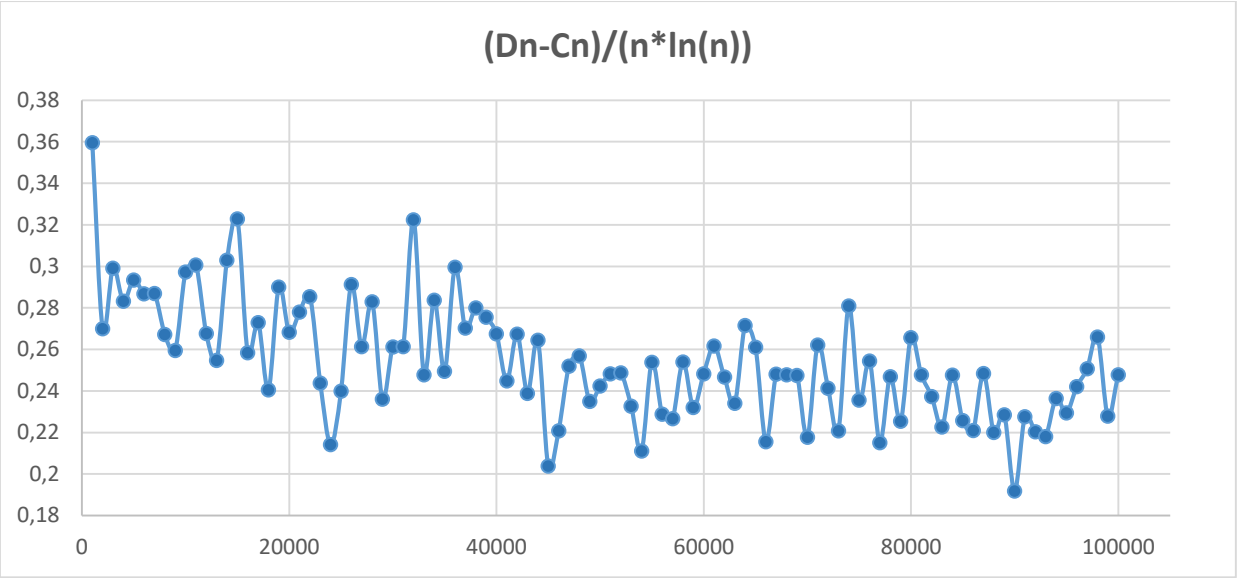

Cn/n^2
0,008
0,007
0,006
0,005
0,004
0,003
0,002
0,001
0
0
20000
40000
60000
80000
100000

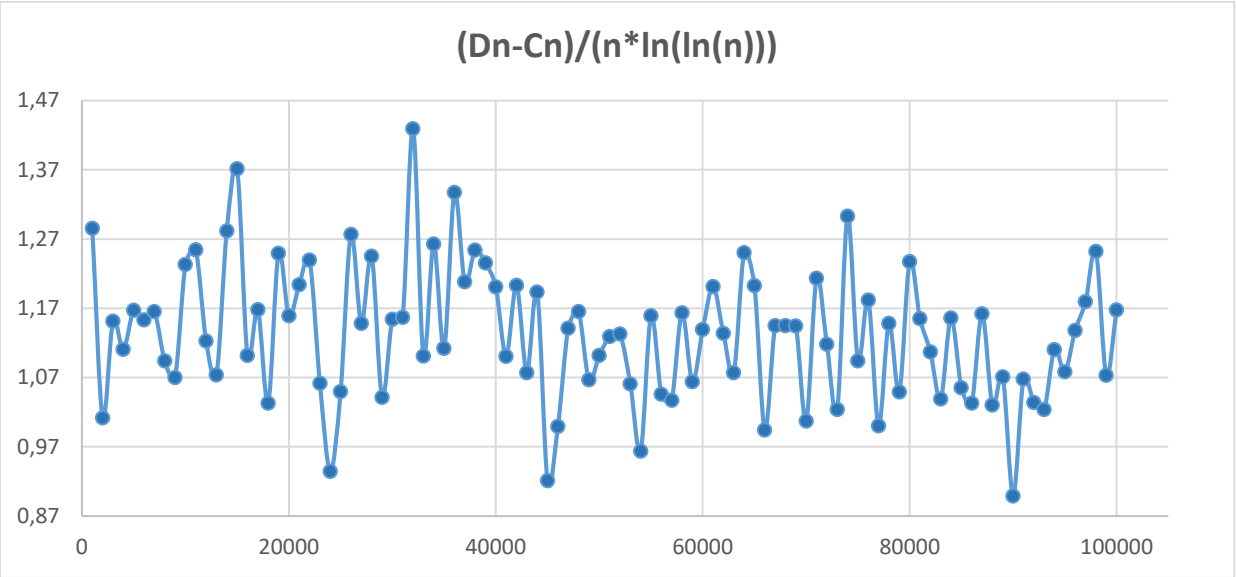

Dn/n
15,5
14,5
13,5
12,5
11,5
10,5
9,5
0
20000
40000
60000
80000
100000


Dn/(n*ln(n))
1,44
1,42
1,4
1,38
1,36
1,34
1,32
1,3
1,28
1,26
1,24
0
20000
40000
60000
80000
100000


Dn/n^2
0,012
0,01
0,008
0,006
0,004
0,002
0
0
20000
40000
60000
80000
100000


(Dn-Cn)/n
3,5
3,3
3,1
2,9
2,7
2,5
2,3
2,1
1,9
0
20000
40000
60000
80000
100000





Zrobiłam dodatkowo wykresy i na ich podstawie mogę zauważyć, że asymptoty wartości średnich wyglądają tak:

Bn =O($\sqrt{n}$)

Un = O(n)

Ln = O(ln(n)/ln(ln(n)))

Cn = O(n*ln(n))

Dn = O(n*ln(n))

Dn-Cn = O(n*ln(ln(n))

Zadanie 2.4:

Birthday Paradox – to stwierdzenie, że w grupie 23 lub więcej osób prawdopodobieństwo zbieżności urodzin (dzień i miesiąc) dla co najmniej dwóch osób przekracza 50%. Na przykład, jeśli w klasie jest 23 lub więcej uczniów, bardziej prawdopodobne jest, że para kolegów z klasy będzie miała urodziny tego samego dnia, niż że każdy z nich będzie miał urodziny w inny sposób. Dla 57 i więcej osób prawdopodobieństwo takiego zbiegu okoliczności przekracza 99%, choć sięga 100%, zgodnie z zasadą Dirichleta, tylko wtedy, gdy w grupie jest co najmniej 367 osób (dokładnie o 1 więcej niż liczba dni w rok przestępny; biorąc pod uwagę lata przestępne).

Cupon Collector's Problemzjawisko – to jest badane w teorii prawdopodobieństwa i kombinatoryce. Kolekcjoner dąży do posiadania wszystkich naklejek z serii, ale w momencie zakupu nie jest znana liczba naklejek (np. zabawki w torebkach na chleb): dlatego wzór jest przeceniony. Powstaje pytanie: ile trzeba kupić, aby mieć kompletną kolekcję? Badanie tego problemu i jego uogólnień znajduje zastosowanie w szczególności w inżynierii telekomunikacyjnej.

Zbiór n szkiców wymaga średnio około n*ln(n) zakupów.

Zadanie 2.5:

Efektem pojawiania się każdej nowej metody kryptoanalizy jest rewizja szacunków bezpieczeństwa szyfrów, co z kolei pociąga za sobą konieczność tworzenia silniejszych szyfrów. Istnieje możliwość realizacji wyboru wiadomości do zastąpienia elektronicznym podpisem cyfrowym metodą opartą na wspomnianym „urodzinowym paradoksie". Ta technika kryptoanalizy doprowadziła do powstania wymagań dotyczących odporności na kolizje dla funkcji mieszających. Urodzinowy paradoks wywarł znaczący wpływ na rozwój kryptosystemów i protokołów kryptograficznych i doprowadził do powstania ataku opartego na „paradoksie urodzin" (birthday attack) lub pierwiastka kwadratowego (square-roof attack) – w celu wykryć z prawdopodobieństwem większym niż 50% w pokoju wypełnionym przypadkowymi osobami, dwie osoby urodzone tego samego dnia wystarczą dla 23 osób w tym pokoju. Należy zauważyć, że atak na funkcje skrótu oparty na paradoksie urodzinowym jest jednym z najczęstszych ataków. Paradoks urodzin jest również wspomniany w ataku slajdów.