Metody Probabilistyczne i Statystyka

Zadanie domowe 2

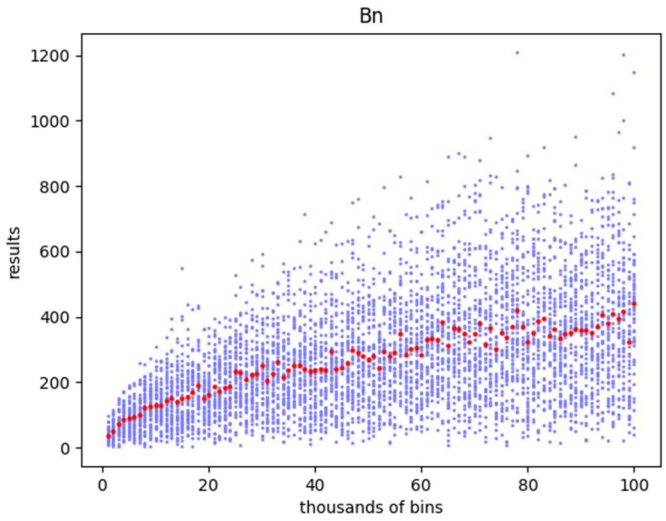
Arina Lazarenko

257259

Eksperymentalnie wyznaciłam następujące wielkości:

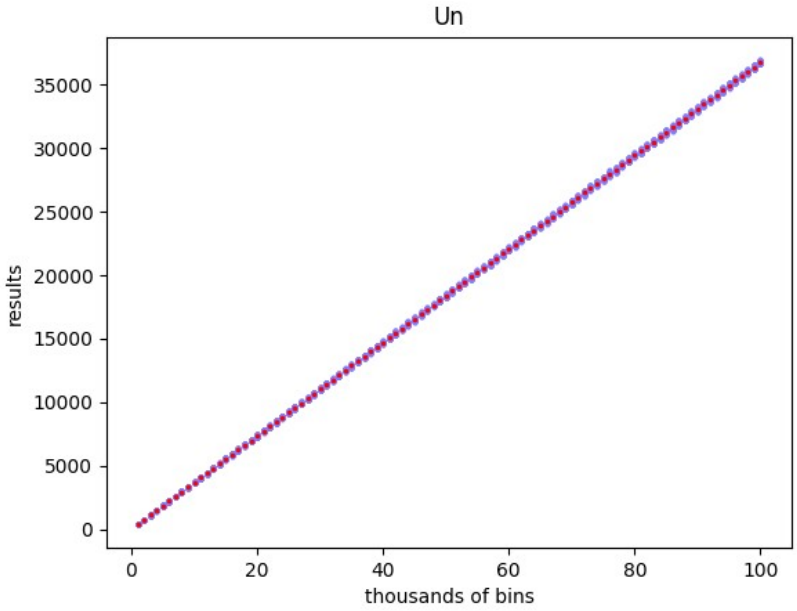
a) Bn– moment pierwszej kolizji; Bn=k, jeśli k-ta z wrzucanych kul jest pierwszą, która trafiła do niepustej urny (paradoks urodzinowy, ang.birthday paradox),
b) Un– liczba pustych urn po wrzuceniu n kul,
c) Ln– maksymalna liczba kul w urnie po wrzuceniu n kul (maximum load),
d) Cn– minimalna liczba rzutów, po której w każdej z urn jest co najmniej jedna kula (pierwszy moment, po którym nie ma już pustych urn; problem kolekcjonera kuponów, ang.coupon collector's problem),
e) Dn– minimalna liczba rzutów, po której w każdej z urn są co najmniej dwie kule (thesiblings of the coupon collector / coupon collector's brother),
f) Dn−Cn – liczba rzutów od momentu Cn potrzeba do tego, żeby w każdej urnie były conajmniej dwie kule.

Korzystając z zebranych danych, dla każdej z badanych statystyk wygenerowałam wykres z wynnikami poszczególnych powtórzeń oraz wartość śriednią.



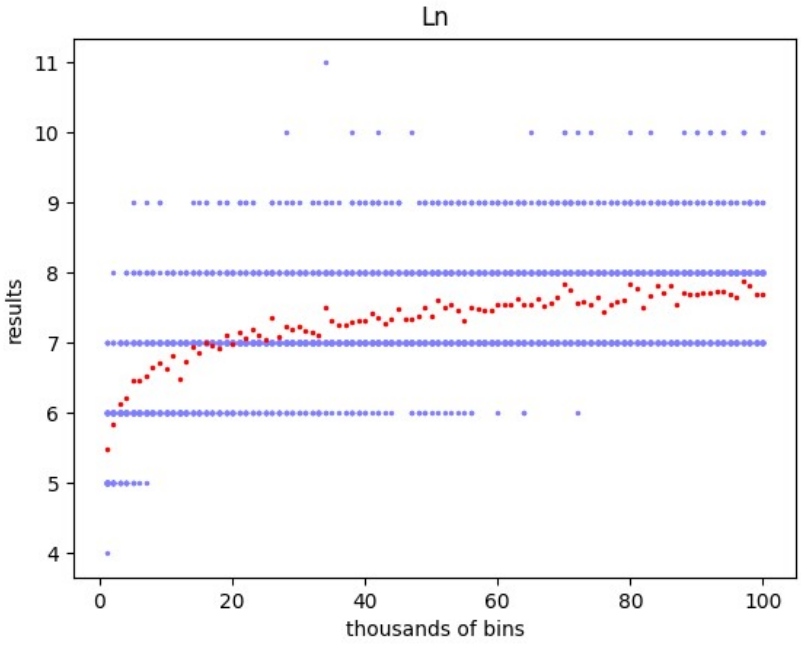
Wykres 1: Wyniki eksperymentu Bn. Niebieskie kropki – wyniki pościególnych prób, czerwone – wartości śriednie.

Na wykresie widać, że wyniki rosną logarytmicznie



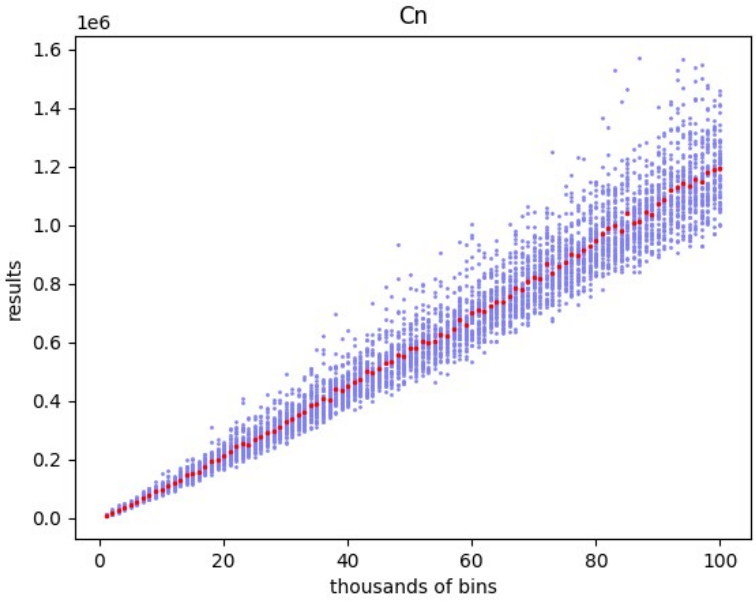
Wykres 1: Wyniki eksperymentu Un. Niebieskie kropki – wyniki pościególnych prób, czerwone – wartości śriednie.

Na wykresie widać, że wyniki rosną lsniowo



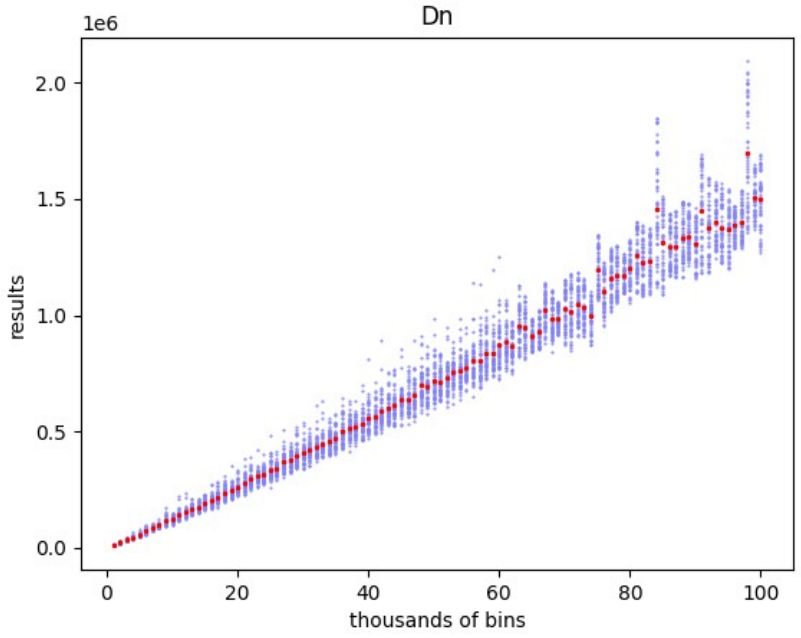
Wykres 1: Wyniki eksperymentu Ln. Niebieskie kropki – wyniki pościególnych prób, czerwone – wartości śriednie.

Na wykresie widać, że wyniki rosną logarytmicznie



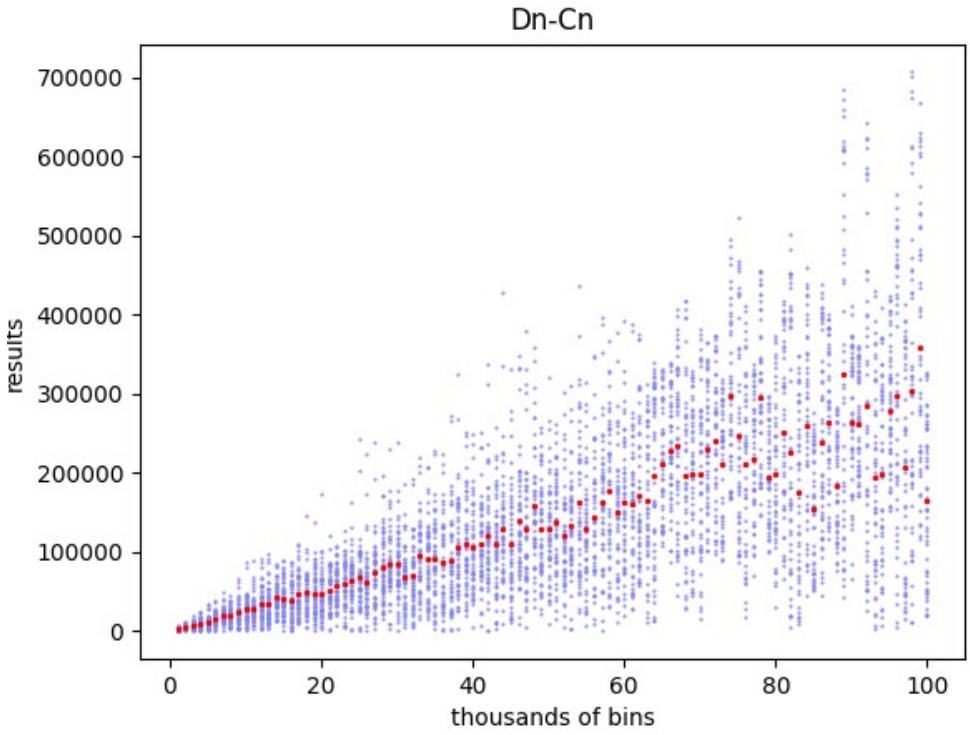
Wykres 1: Wyniki eksperymentu Cn. Niebieskie kropki – wyniki pościególnych prób, czerwone – wartości śriednie.

Na wykresie widać, że wyniki rosną lsniowo



Wykres 1: Wyniki eksperymentu Dn. Niebieskie kropki – wyniki pościególnych prób, czerwone – wartości śriednie.

Na wykresie widać, że wyniki rosną lsniowo

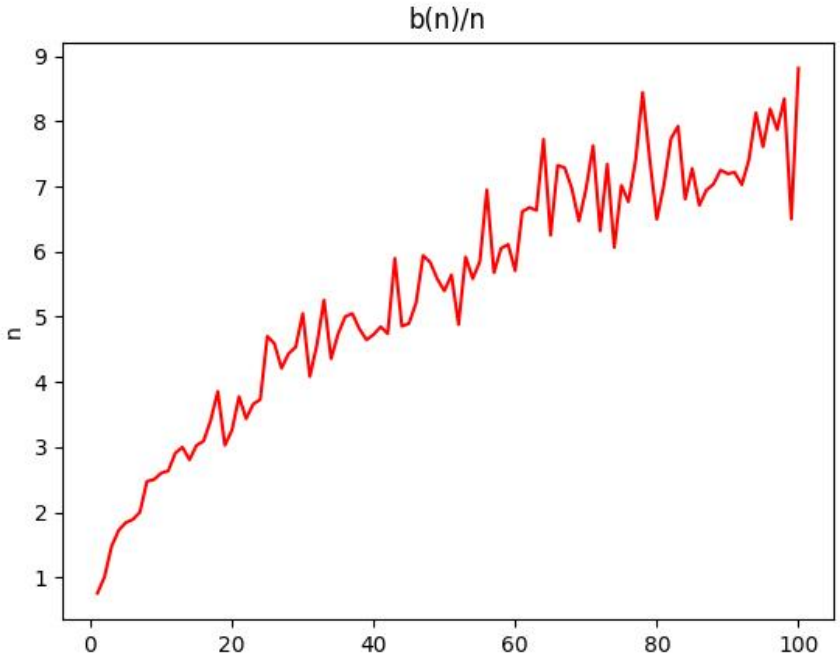


Wykres 1: Wyniki eksperymentu Dn-Cn. Niebieskie kropki – wyniki pościególnych prób, czerwone – wartości śriednie.

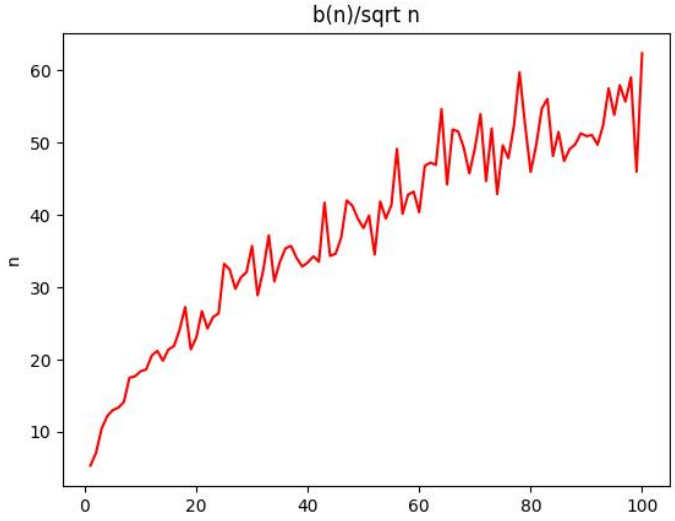
Na wykresie widać, że wyniki rosną liniowo, ale przy dużych liczbach traci sie dokładność.

Dodatkowe wykresy:

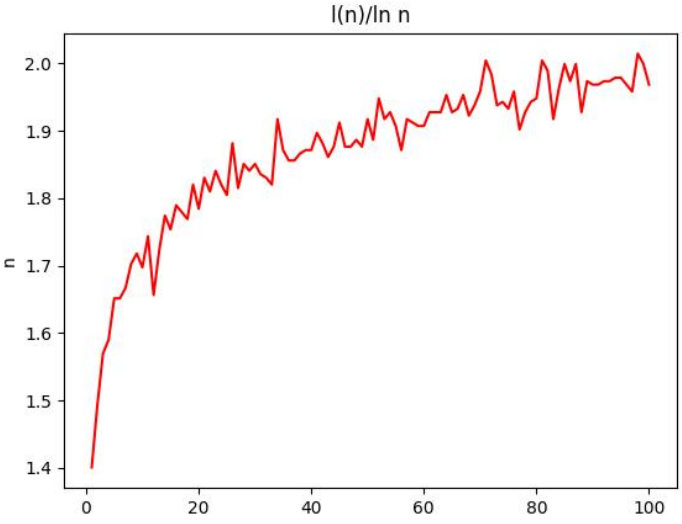
- b(n)/n



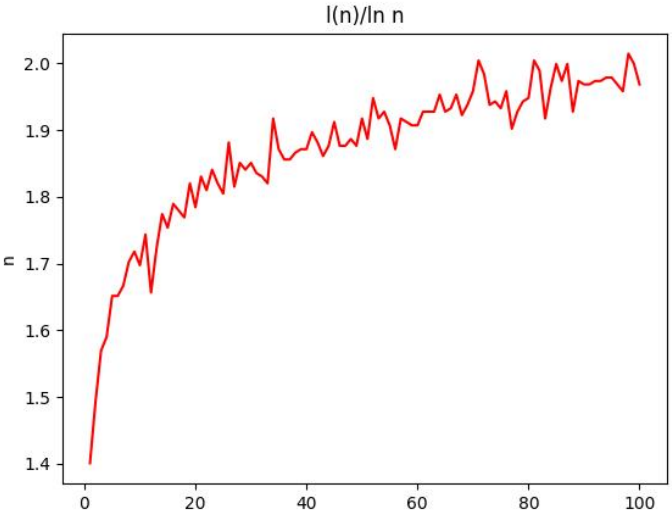
- b(n)/sqrt(n)



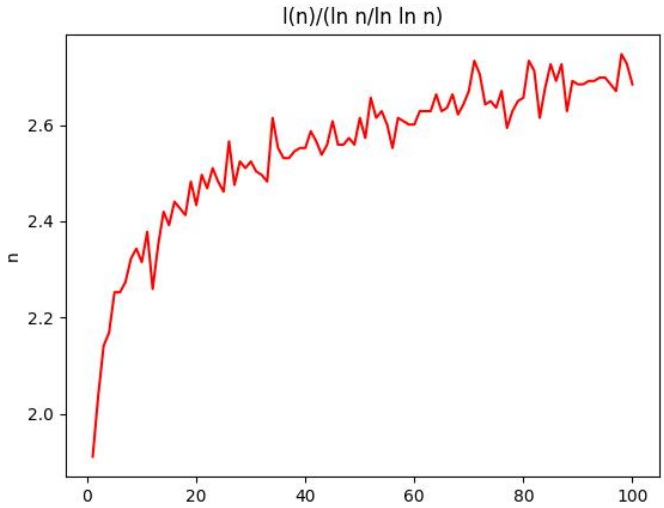
- u(n)/n



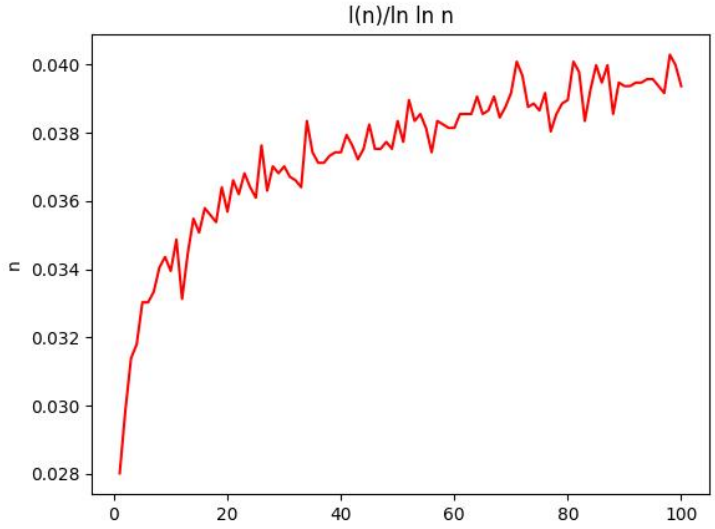
- l(n)/ln n



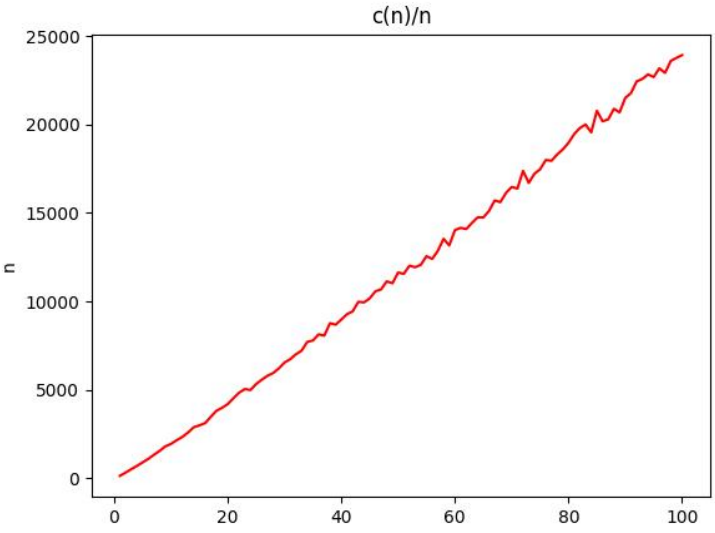
- l(n)/((ln n)/(ln ln n))



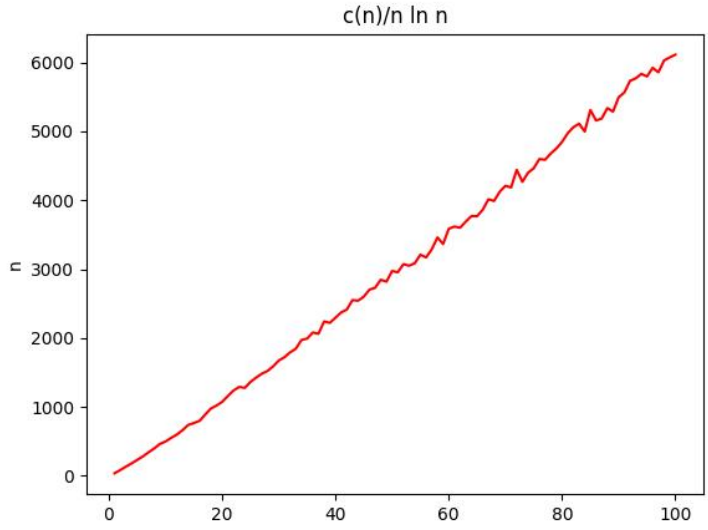
- l(n)/(ln ln n)



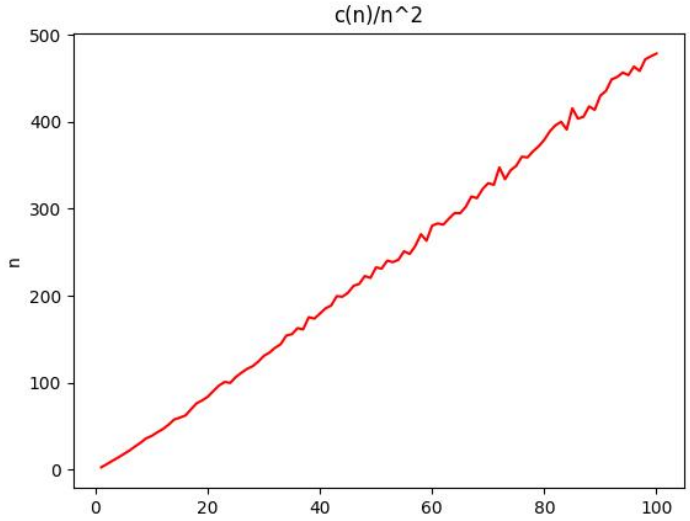
- c(n)/n



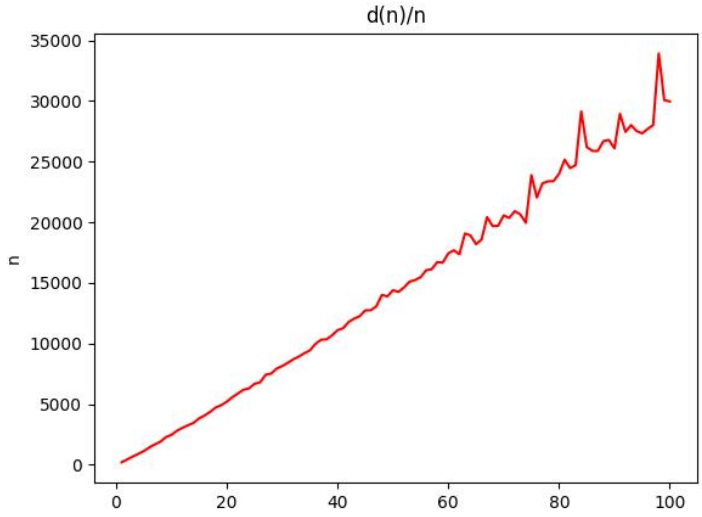
- c(n)/(n ln n)



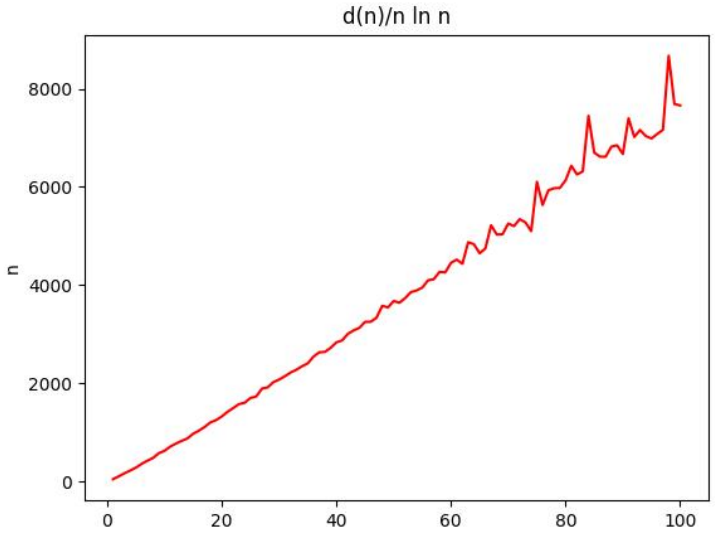
- $c(n)/n^2$



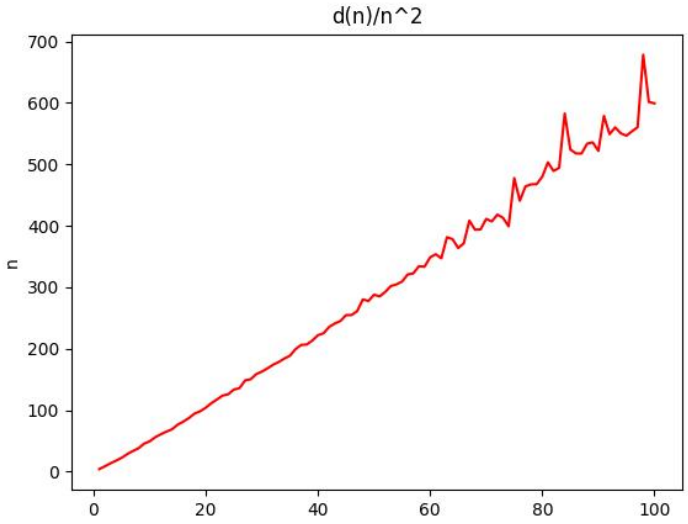
- d(n)/n



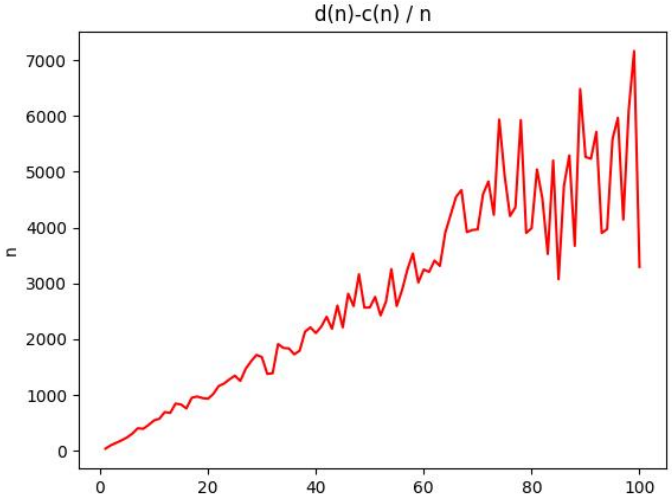
- d(n)/(n ln n)



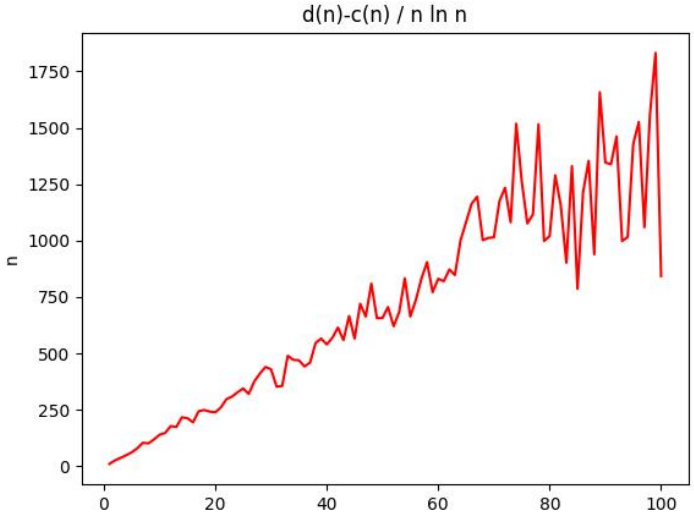
- $d(n)/n^2$



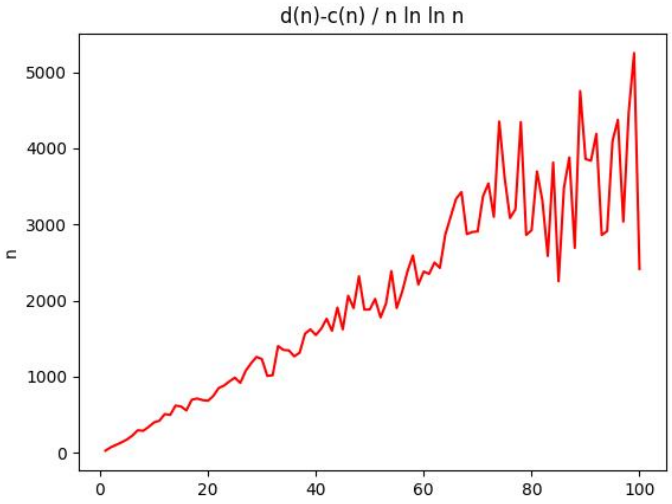
- d(n)-c(n)/n



- d(n)-c(n)/(n ln n)



- d(n)-c(n)/(n ln ln n)



Moje intuicje na rachunek nazw:

birthday paradox – mamy tylko 365 dni na rok i aż 7 milliardów ludzi, czyli około 19 000 000 ludzi mają urodziny w tym samym dniu. I to jest dość nieoczekiwane, że w grupie z 23 rondomowo wziętych osób ppb tego, że ktoś ma urodziny w ten samy dzień jest aż 50%, a dla grupy z 62 osób – 100%.

coupon collector – im więcej czegoś z jednego naboru chcemy otrzymać, tym więcej prób musimy pobic. w najlepczym przypadku, gdy chcemy n czegoś – dostaniemy to w n prób.

Znaczenie birthday paradox w:

- funkcji hashujcych – oblicza przybliżone ryzyko kolizji haszów istniejącej w haszach danej wielkości populacji.

- kryptograficznych funkcjacji hashujcych –

Urodzinowy paradoks można wykorzystać w ataku kryptograficznym na podpisy cyfrowe. Podpisy cyfrowe opierają się na czymś, co nazywa się funkcją hash f(x), która przekształca wiadomość lub dokument w bardzo dużą liczbę. Ktoś czytający dokument mógłby następnie „odszyfrować” podpis za pomocą klucza publicznego sygnatariusza, co dowodziłoby, że sygnatariusz podpisał dokument cyfrowo.