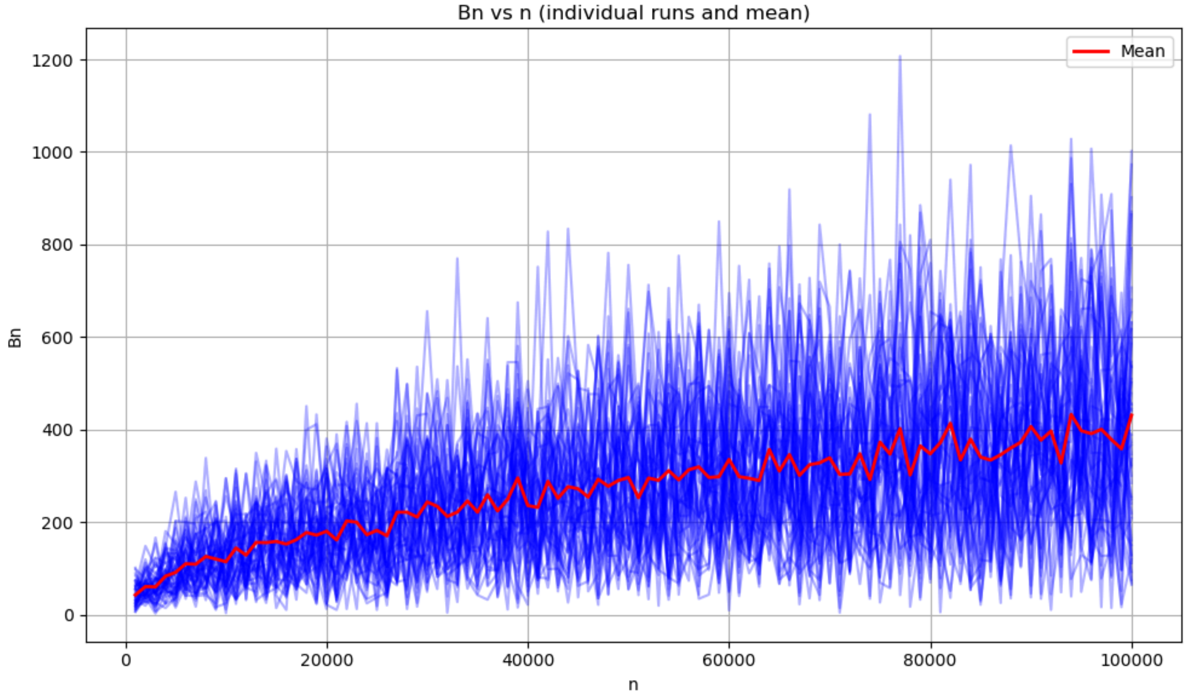
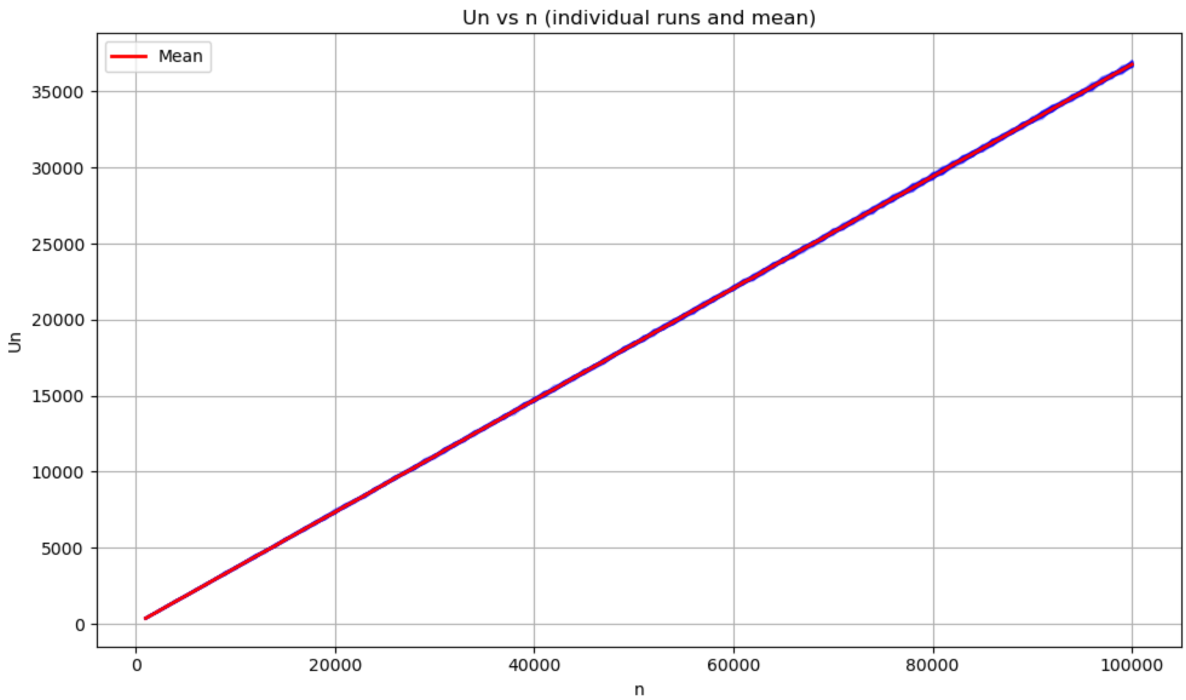
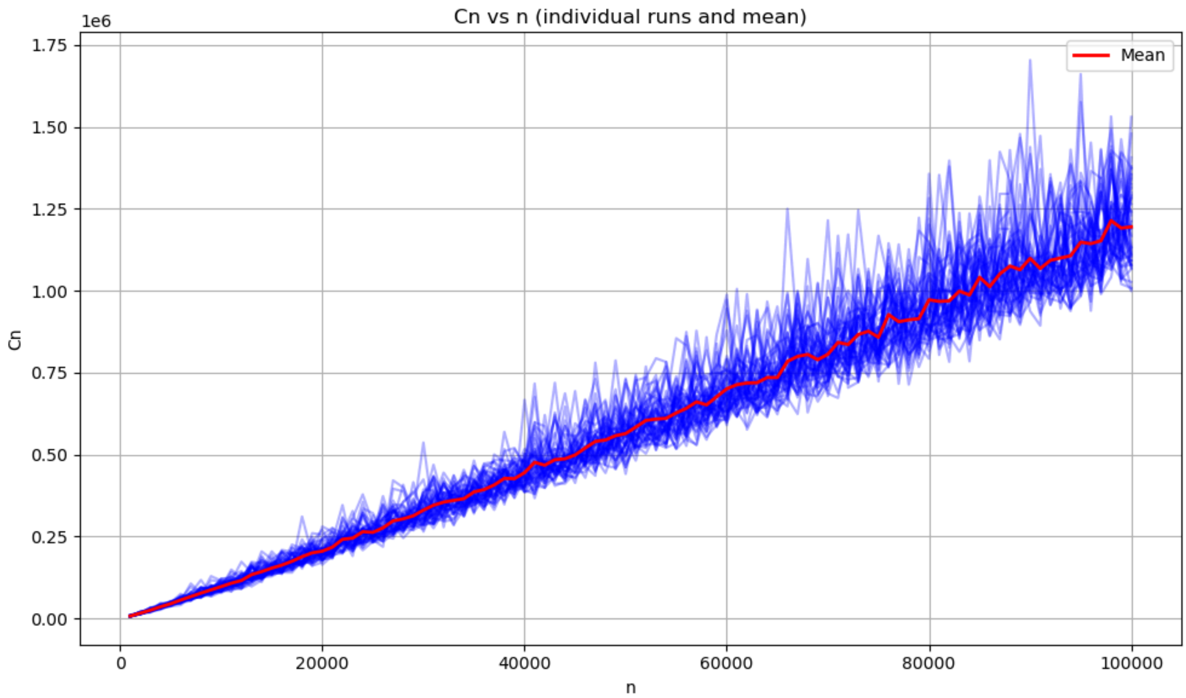
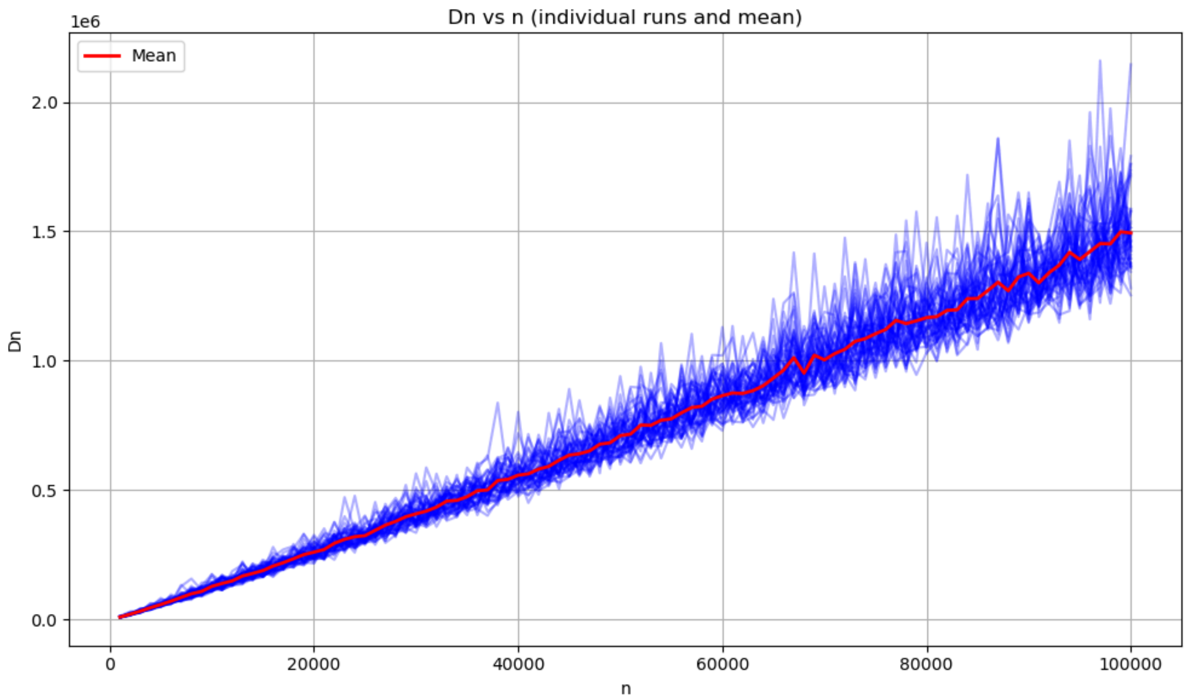
Zadanie 2

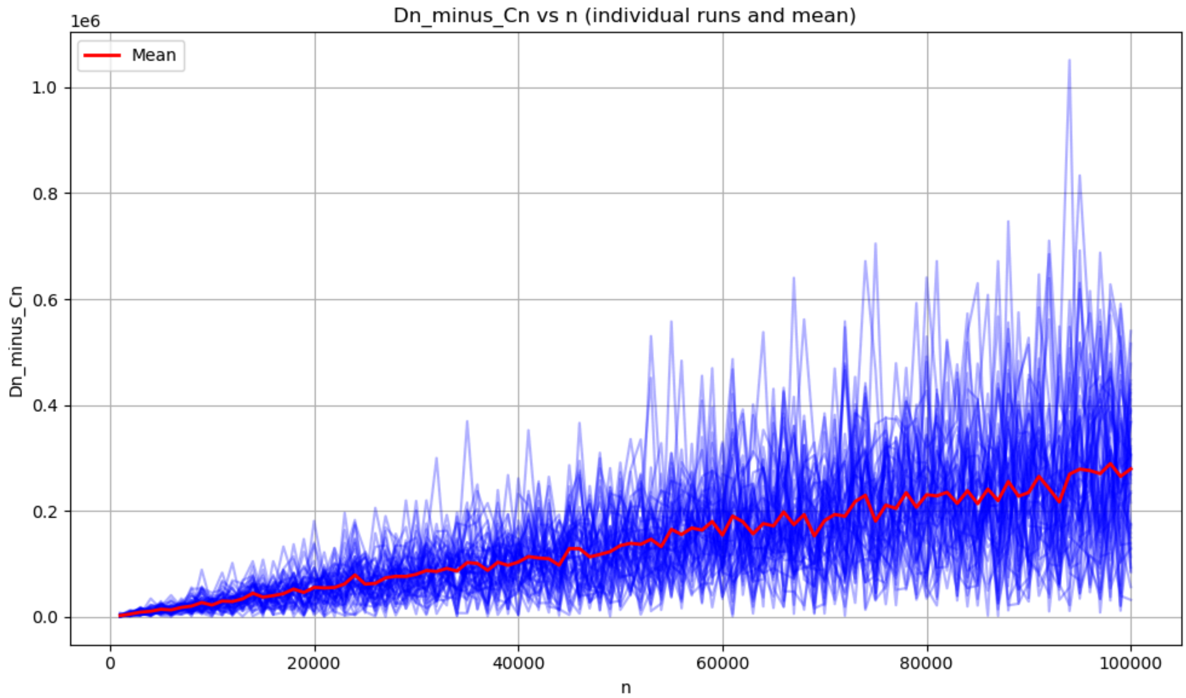
Wykresy główne






Cn vs n (individual runs and mean)
1e6
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Cn
0
20000
40000
60000
80000
100000
n
Mean


Dn vs n (individual runs and mean)
1e6
2.0
1.5
1.0
0.5
0.0
Dn
0
20000
40000
60000
80000
100000
n
Mean



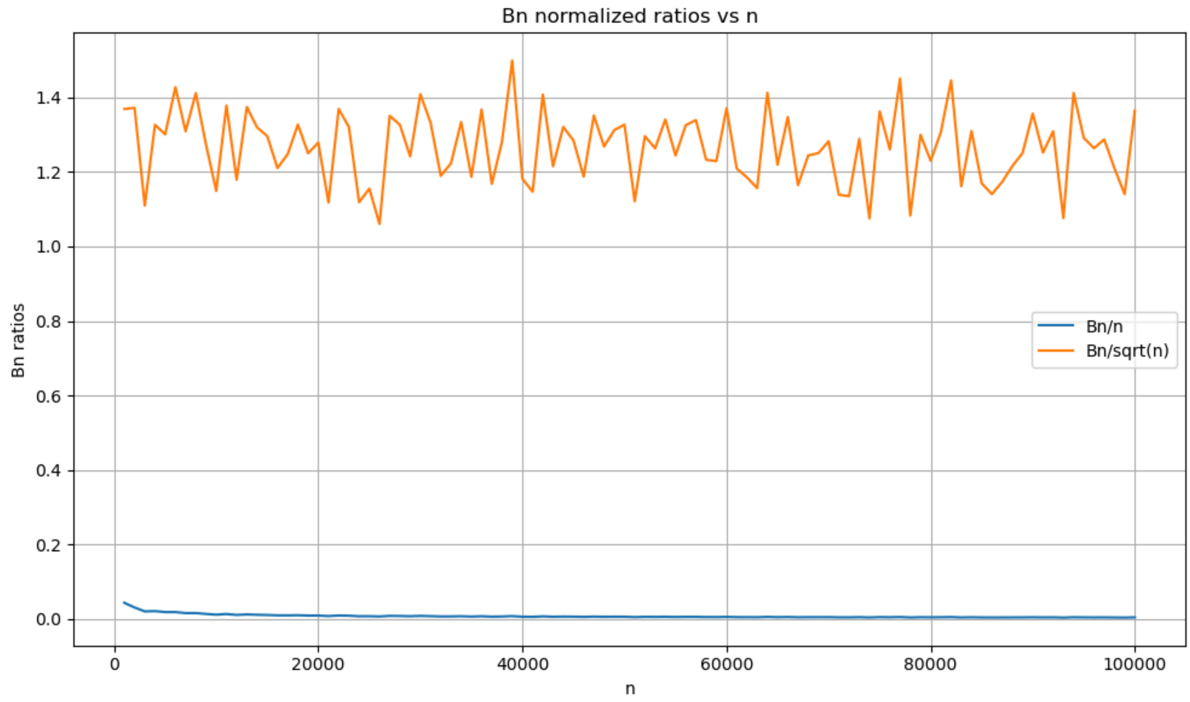
(a) W ramach eksperymentów przeprowadzono 50 niezależnych
powtórzeń dla każdego $n \in \{1000, 2000, ..., 100000\}$, mierząc wartości zmiennych:

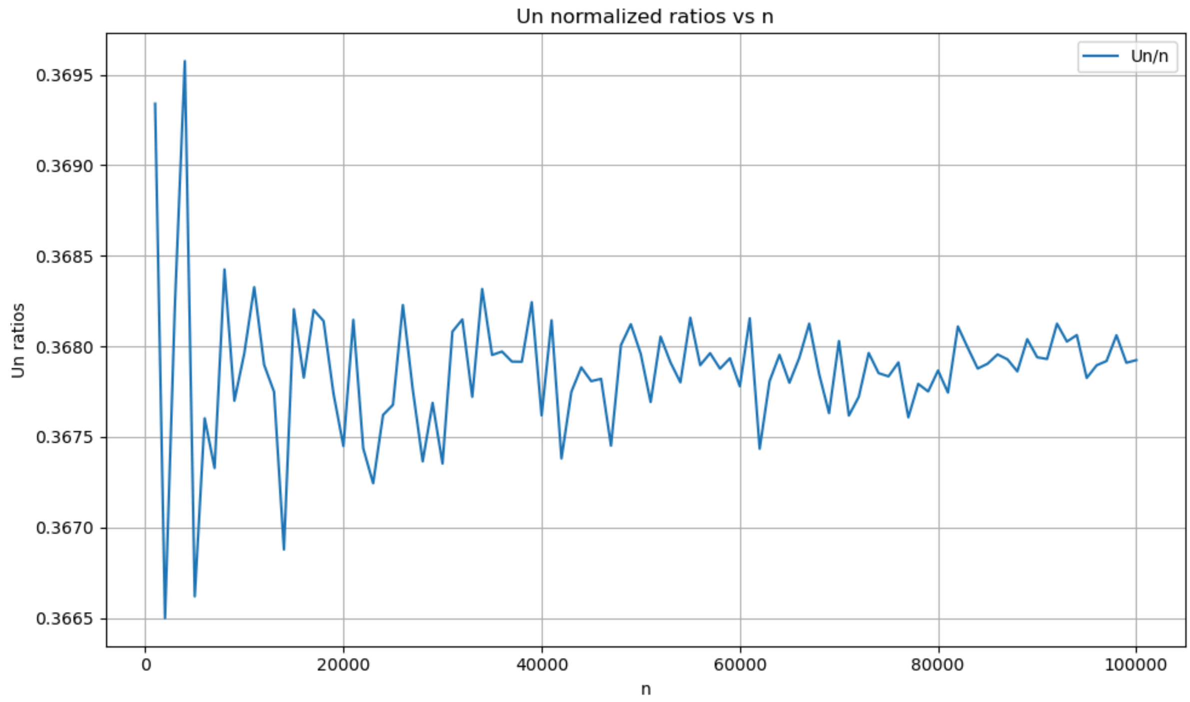
- B_n – moment pierwszej kolizji,
- U_n – liczbę pustych urn po wrzuceniu n kul,
- C_n – liczbę rzutów do momentu pierwszego pełnego pokrycia (coupon collector),
- D_n – liczbę rzutów do momentu, gdy każda urna zawiera co najmniej dwie kule,
- $D_n - C_n$ – dodatkową liczbę rzutów potrzebną od chwili C_n do osiągnięcia 2 kul w każdej urnie.

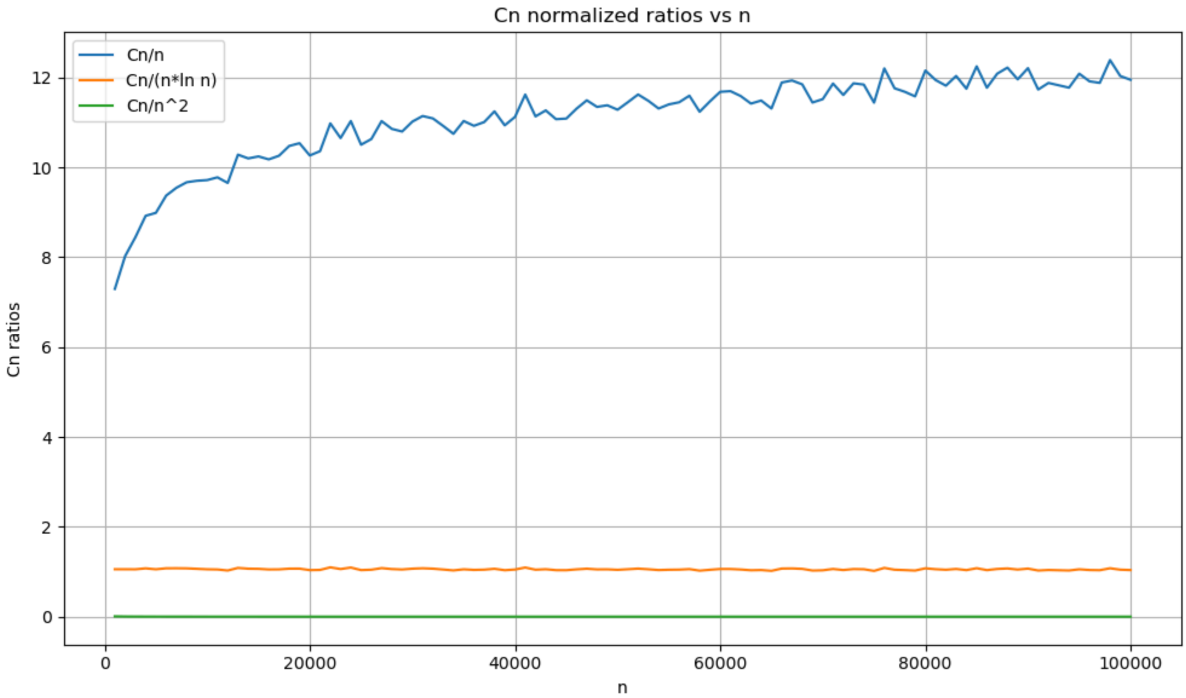
Dla każdej z tych zmiennych wyraźnie obserwujemy znane asymptotyczne trendy:

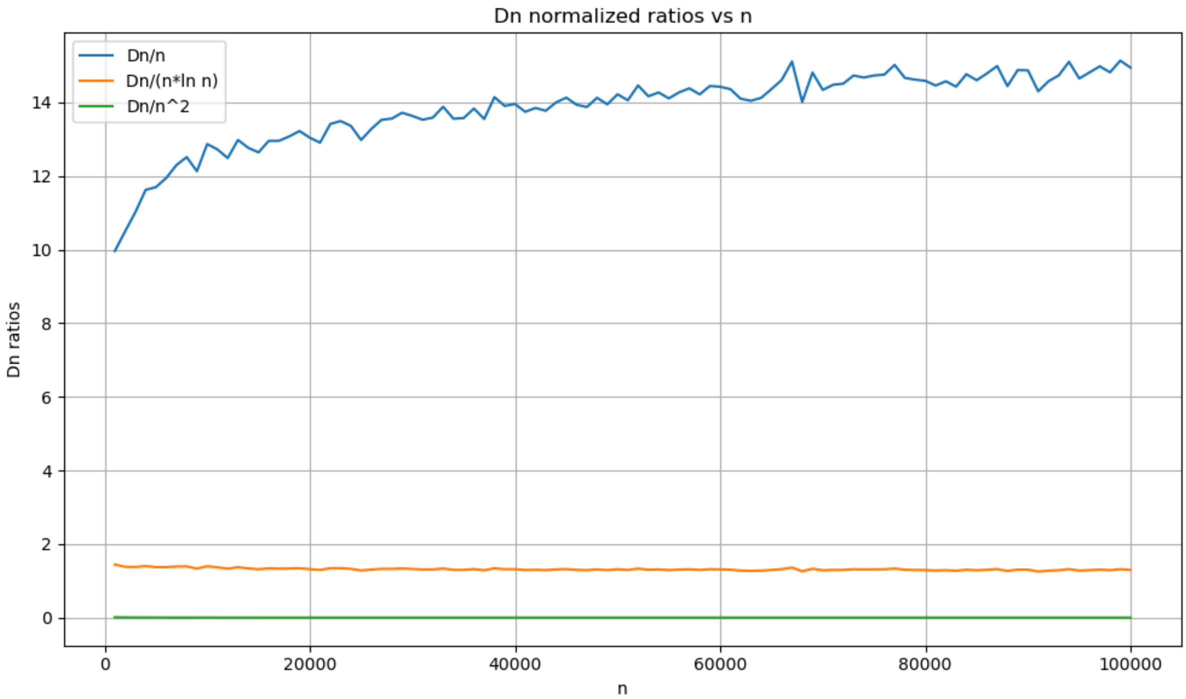
- B_n rośnie jak $\Theta(\sqrt{n}$. Kolizja pojawia się bardzo wcześnie.
- U_n jest liniowo powiązana z n, z nachyleniem około $\frac{1}{e}$.
- C_n rośnie jak $n \log n$, co jest zgodne z klasycznym problemem zbierania kuponów.
- D_n jest większe – rośnie jak $n \log n + n \log \log n$.
- $D_n - C_n$ rośnie wolniej, około $n \log \log n$.

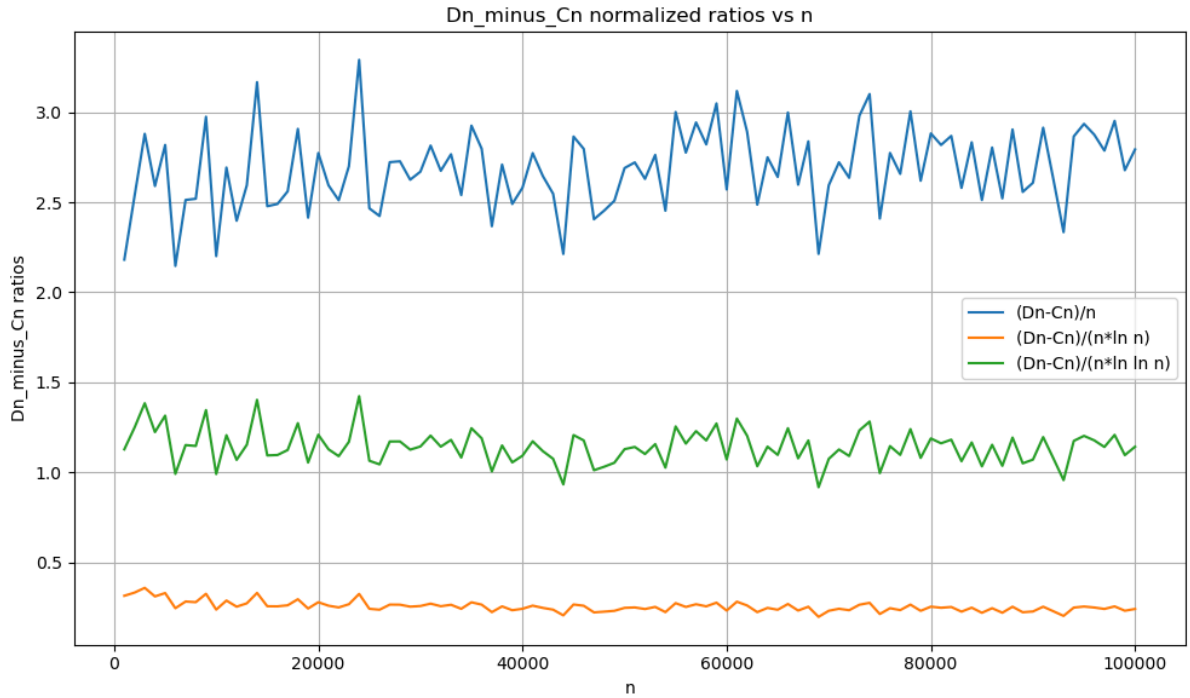
Wykresy ilorazowe






Cn normalized ratios vs n
Cn/n
Cn/(n*ln n)
Cn/n^2
Cn ratios
n
0
2
4
6
8
10
12
0
20000
40000
60000
80000
100000


Dn normalized ratios vs n
Dn/n
Dn/(n*ln n)
Dn/n^2
Dn ratios
n
0
2
4
6
8
10
12
14
0
20000
40000
60000
80000
100000



(b) Koncentracja wyników wokół średnich:

- Wszystkie zmienne są wyraźnie skoncentrowane wokół wartości średnich.
- B_n i U_n wykazują minimalne odchylenia od średniej.
- C_n i D_n mają większy rozrzut, ale wyniki nadal dobrze skupione.
- $D_n - C_n$ pokazuje największą zmienność, jednak średnia jest wyraźnie określona.

(c) Hipotezy asymptotyczne:

- $\frac{B_n}{n} \to 0$, $\frac{B_n}{\sqrt{n}} \to c$, więc $B_n = \Theta(\sqrt{n})$.
- $\frac{U_n}{n} \to \frac{1}{e}$, stąd $U_n \to \frac{n}{e}$.
- $\frac{C_n}{n \log n} \to 1$, czyli $C_n \approx n \log n$.
- $\frac{D_n}{n \log n} \to 1$ i $\frac{D_n}{n^2} \to 0$, stąd $D_n \approx n(\log n + \log \log n)$.
- $\frac{D_n - C_n}{n \log \log n} \to c$, stad $D_n - C_n = \Theta(n \log \log n)$.

(d) Intuicja nazw:

- Birthday paradox – pierwsza kolizja pojawia się znacznie szybciej niż intuicyjnie, podobnie jak w paradoksie urodzin.
- Coupon collector's problem – liczba rzutów potrzebna, aby „zebrać wszystkie kupony" (niepuste urny) rośnie logarytmicznie z n.

(e) Znaczenie dla funkcji hashujących:

- Birthday paradox pokazuje, że kolizje w funkcjach hashujących mogą wystąpić znacznie wcześniej niż intuicja podpowiada.
- W kryptografii wymaga się dużych zakresów hashy, aby zmniejszyć prawdopodobieństwo kolizji.

Podsumowanie

Eksperymenty potwierdziły teoretyczne zależności między zmiennymi losowymi a liczbą urn n. Średnie wartości i wykresy ilorazowe pokazują stabilność i koncentrację wyników. Raport stanowi pełne podsumowanie badania kul i urn w kontekście birthday paradox oraz problemu kolekcjonera kuponów.