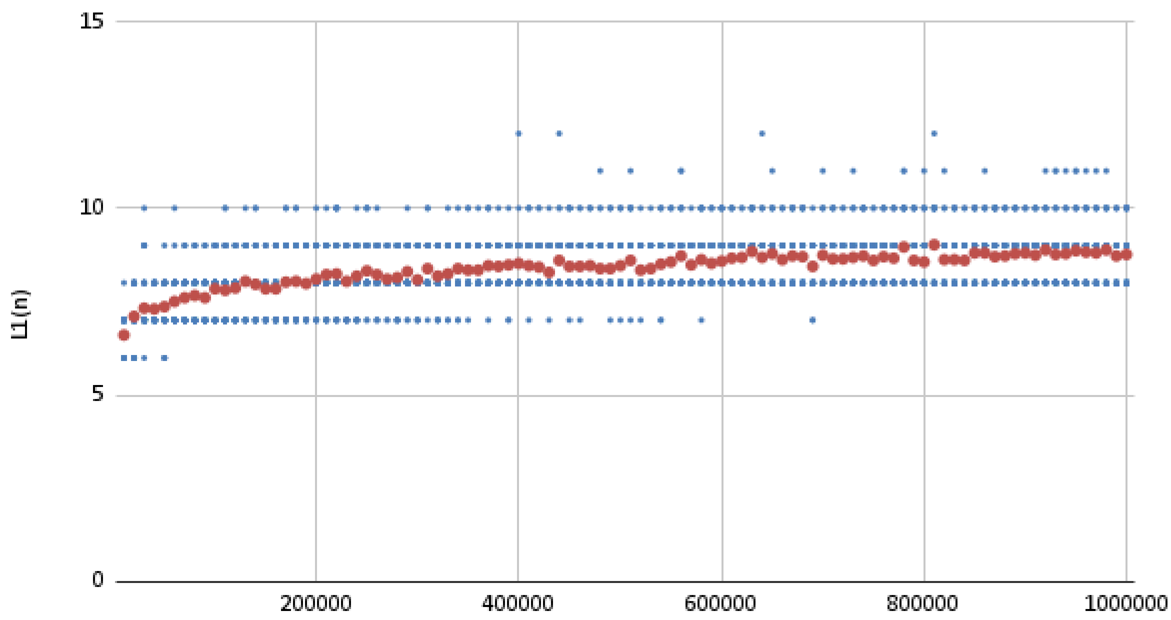
Wiktor Bachta, 272 353

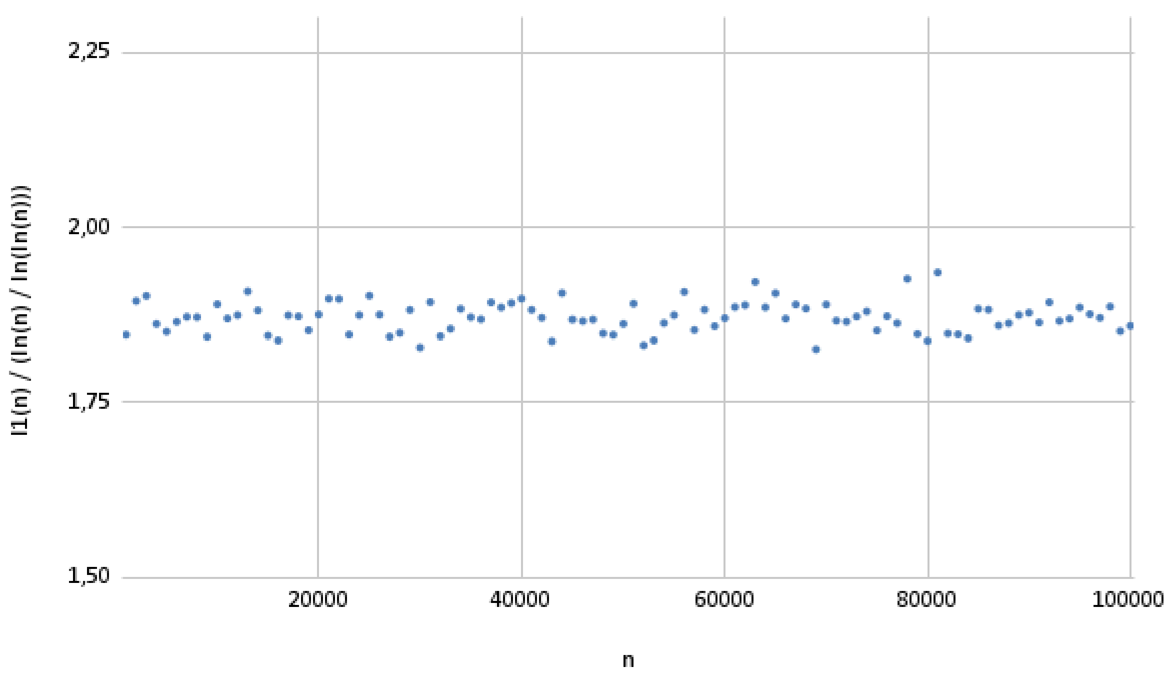
ZADANIE DOMOWE 3

OPIS:
Zastosowane przeze mnie oznaczenia są takie same jak w treści zadania domowego. Czerwone punkty na wykresach są wartościami średnimi. Najlepszą hipotezę dla asymptotyki daje wykres najbliższy funkcji stałej. Ponadto, gdzie to możliwe, starałem się "wcisnąć" e. W obu zadaniach wykonałem 50 powtórzeń eksperymentu dla każdego n. W zadaniu 1. $n \in \{10\,000,\ 20\,000,\ \ldots,\ 1\,000\,000\}$, natomiast w zadaniu 2. $n \in \{100,\ 200,\ \ldots,\ 10\,000\}$.

Zadanie 1. The Power of Two Choices

L1(n) - maksymalna liczba kul w urnie po wrzuceniu n kul do n urn (dla każdej kuli wybieramy niezależnie i jednostajnie losowo jedną z n urn, w której umieszczamy kulę)



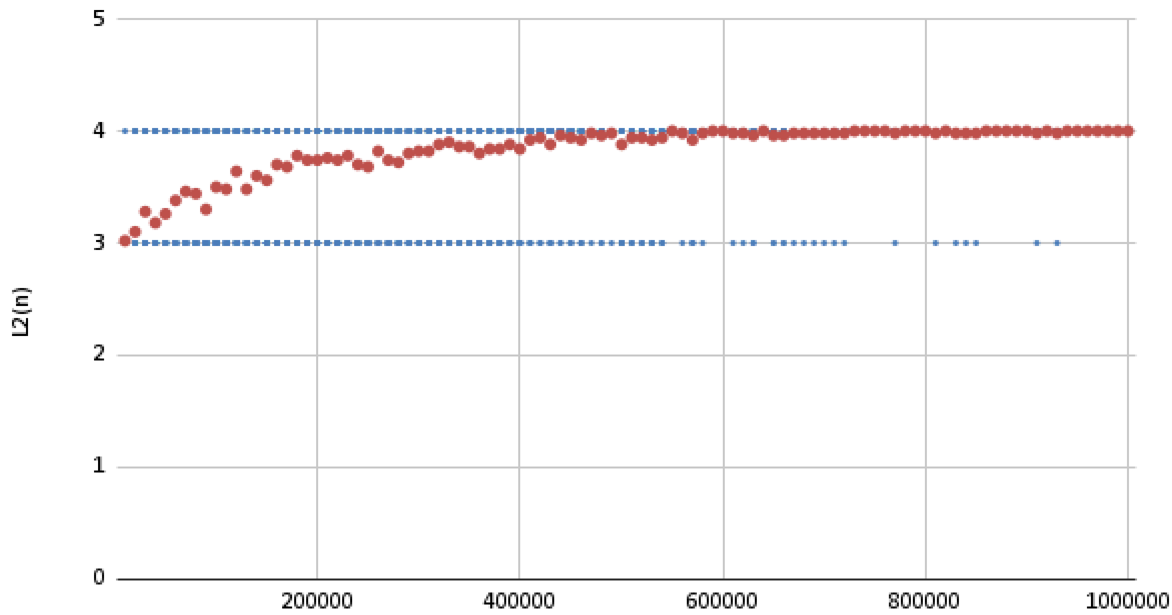


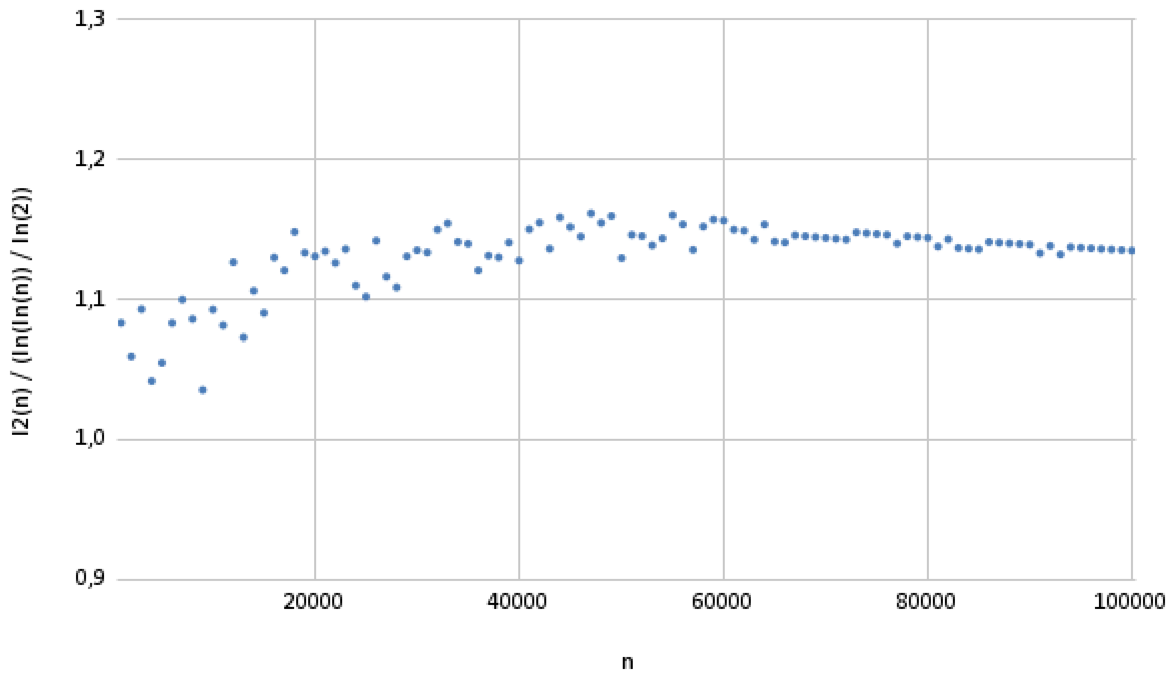
Widać, że L1(n) przyjmuje mało różnych wartości (tutaj {6, …, 12}). Dlatego też, skoncentrowanie wokół średniej jest silne.

Na podstawie wykresu $\frac{l1(n)}{f1(n)}$, gdzie $f1(n) = \frac{ln(n)}{ln(ln(n))}$, można wnioskować, że $l1(n) = O(f1(n))$.

Hipoteza: $l1(n) \sim \left(\frac{e}{2}\right)^2 \cdot \frac{ln(n)}{ln(ln(n))}$

L2(n) - maksymalna liczba kul w urnie po wrzuceniu n kul do n urn (dla każdej kuli wybieramy niezależnie i jednostajnie losowo z powtórzeniami jedną z dwóch urn, a następnie umieszczamy kulę w najmniej zapełnionej z wybranych urn)





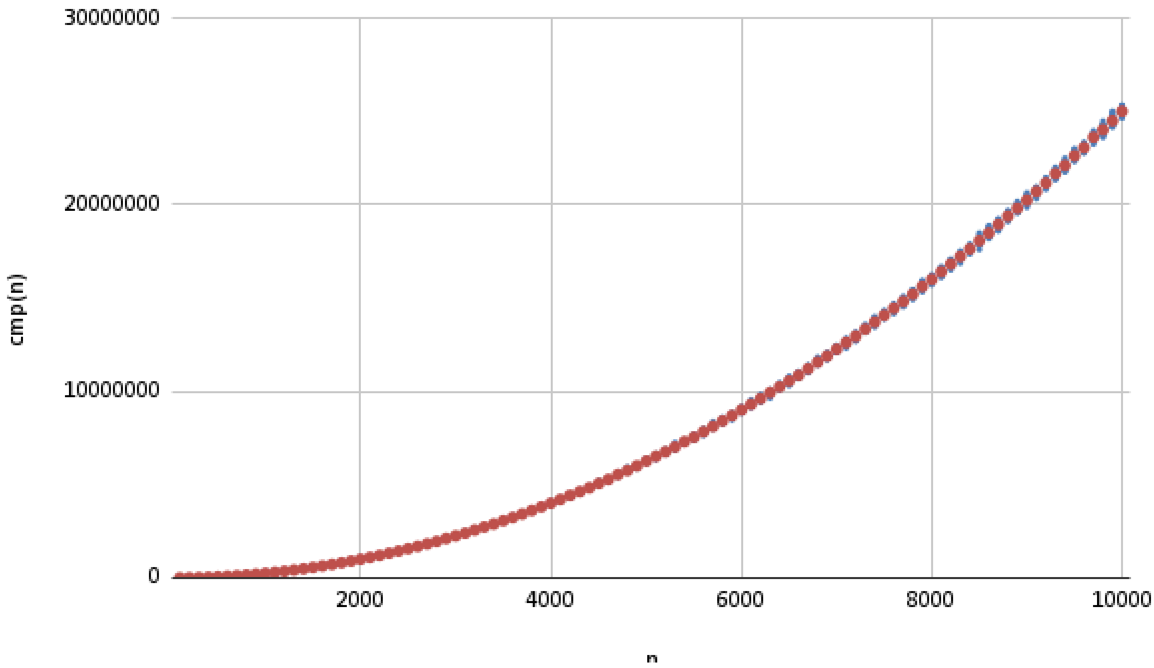
L(2) przyjmuje tutaj już tylko dwie wartości (3 i 4). Z tego względu, skoncentrowanie wokół średniej jest tutaj także bardzo silne. Losowanie dwóch urn sprawiło, że średnia maksymalna liczba kul w urnie zmalała ponad dwukrotnie.

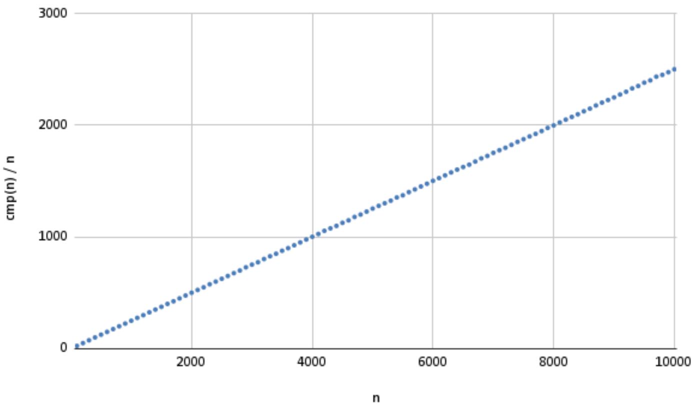
Na podstawie wykresu $\frac{l2(n)}{f2(n)}$, gdzie $f2(n) = \frac{ln(ln(n))}{ln(2)}$, można wnioskować, że $l2(n) = O(f2(n))$.

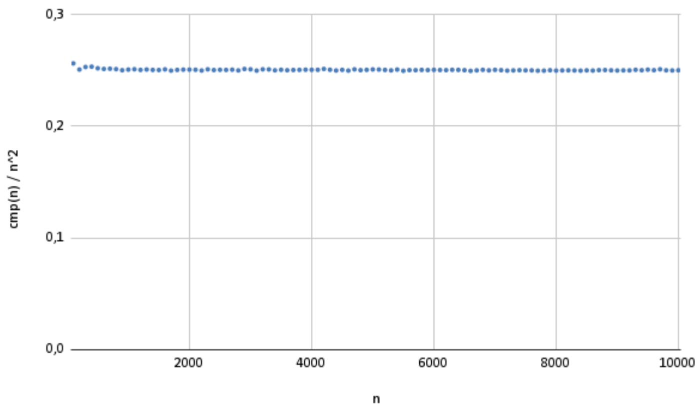
Hipoteza: $l2(n) \sim \frac{e}{2} \cdot \frac{ln(ln(n))}{ln(2)}$

Zadanie 2. Sortowanie przez wstawianie losowych danych

Liczba wykonanych porównań w poszczególnych powtórzeniach





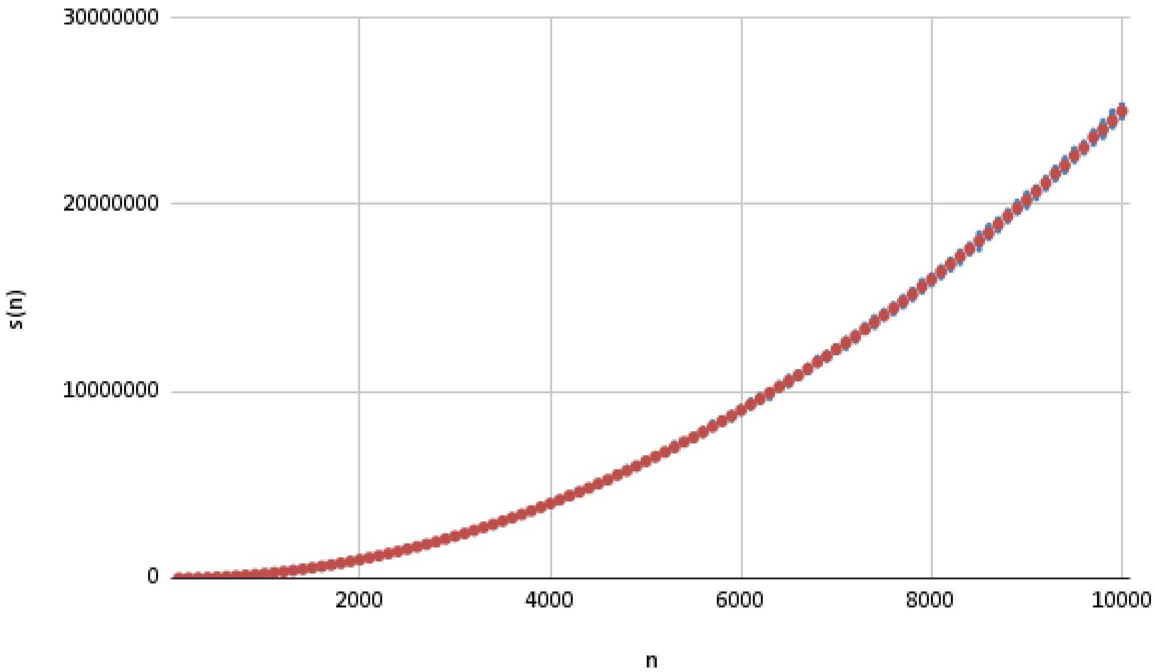


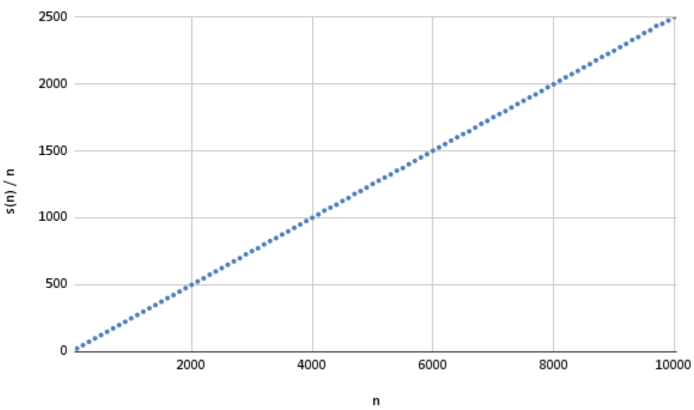
Skoncentrowanie wokół wartości średniej jest tutaj bardzo silne. Oznacza to, że większość losowych permutacji tablicy wymaga podobnej liczby porównań.

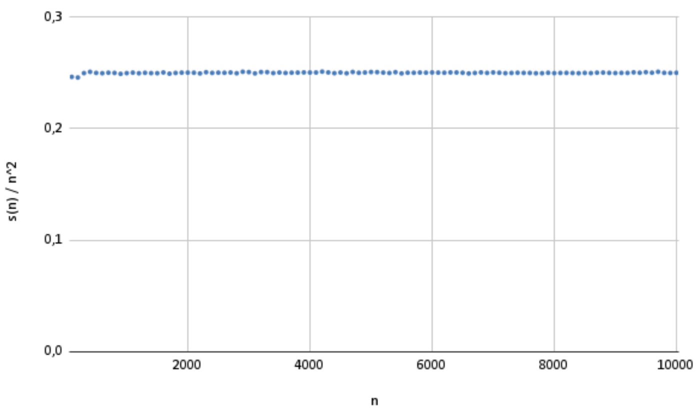
Na podstawie wykresu widać, że $cmp(n) = O(n^2)$.

Hipoteza: $cmp(n) \sim \frac{n^2}{4}$

Liczba przestawień w poszczególnych powtórzeniach





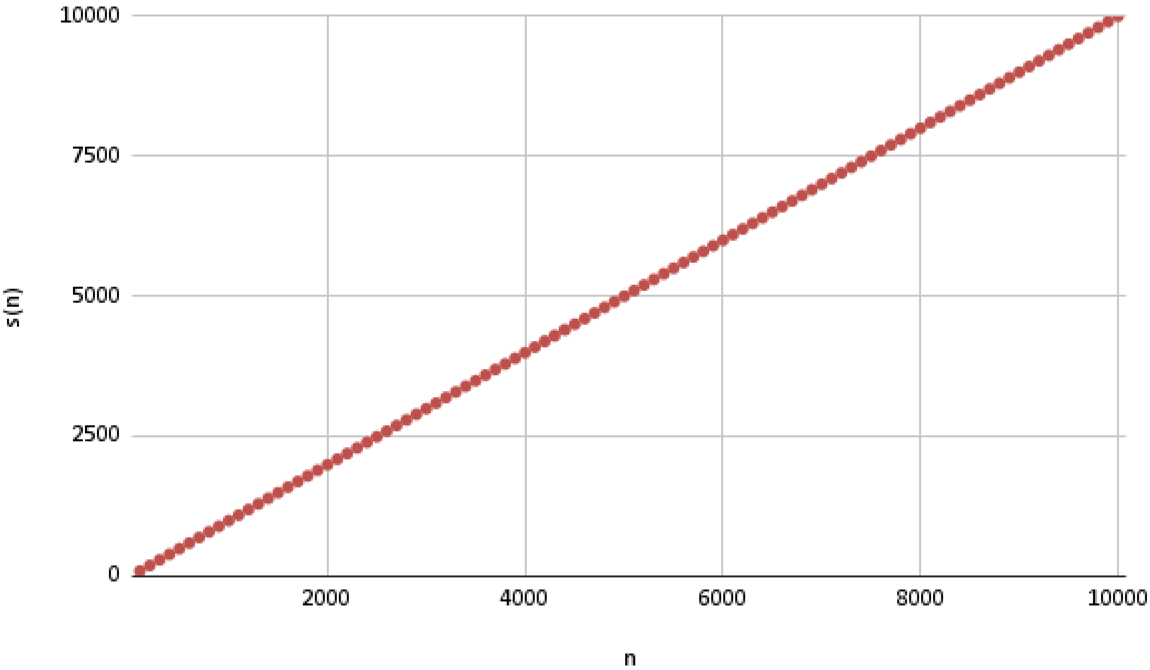


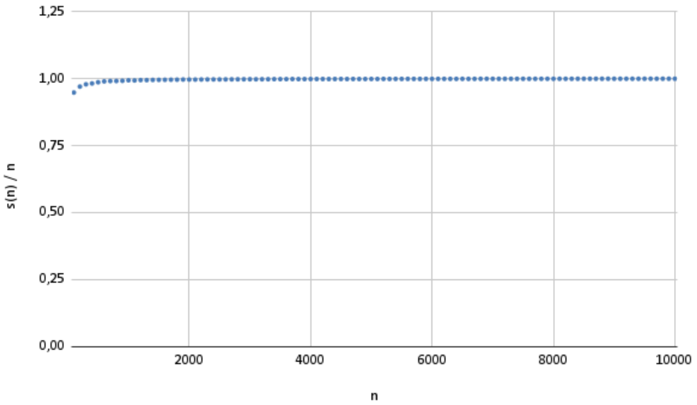
Wykres s(n) oraz cmp(n) jest bardzo zbliżony, ponieważ $cmp(n) = s(n) + n - 1$. Każde porównanie wywołuje przestawienie, oprócz tych porównań, dla których algorytm kończy pracę na pewnej wartości (każdej z n-1).

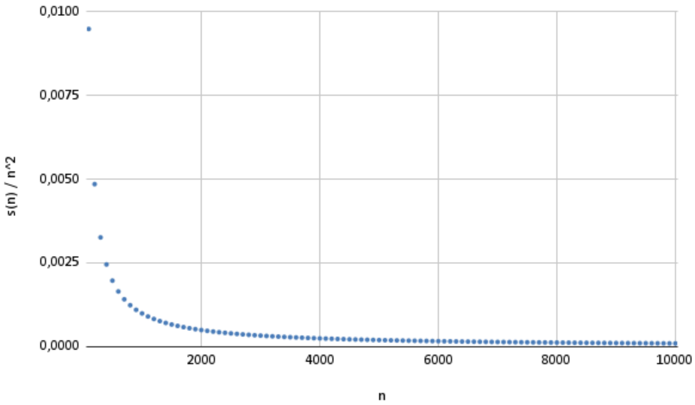
Dlatego też $s(n) = O(n^2)$

Oraz, na podstawie cmp(n), moja hipoteza to: $s(n) \sim \frac{n^2}{4}$

Liczba przestawień w poszczególnych powtórzeniach (nie licząc przestawień “pośrednich”, tzn. wstawienie elementu liczone jest jako 1 przestawienie)







Ze względu na ścisłe powiązanie obliczonych cmp(n) oraz s(n), postanowiłem policzyć s(n) jeszcze raz - tym razem traktując każdy ciąg przestawień jest liczony jako jedno. Opierając się na tym kryterium, s(n) jest praktycznie równe n, ponieważ tylko kilka wartości nie wymaga przestawiania. Innymi słowy, $s(n) = n - 1 - h$, gdzie h to liczba “poukładanych” największych wartości w tablicy.

Wówczas $s(n)$ ~ n