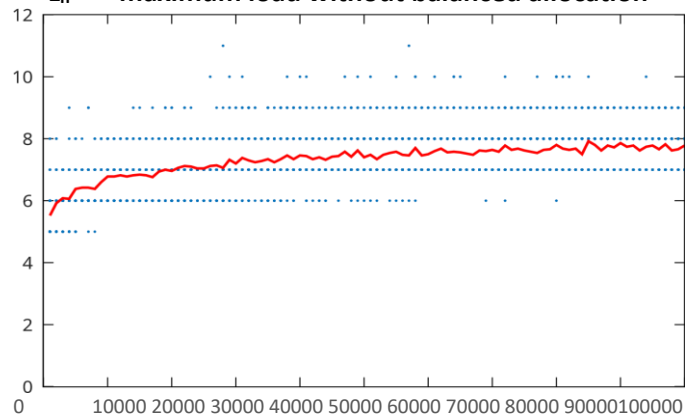


Zadanie domowe 3 – Nel Skowronek 279679

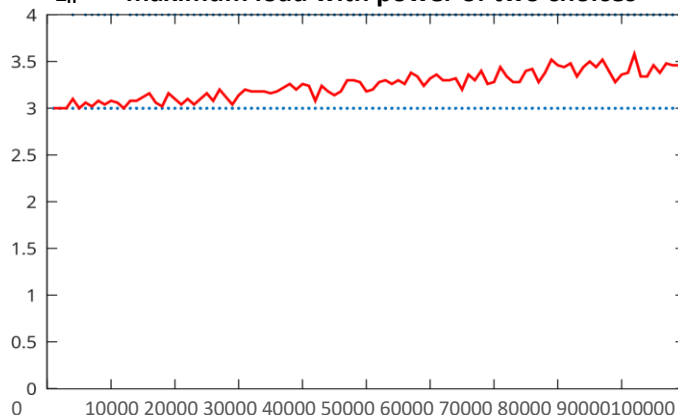
Zadanie 1

Wykresy maksymalnego wypełnienia urn po n-tym rzucie

$L_n^{(1)}$ – maximum load without balanced allocation



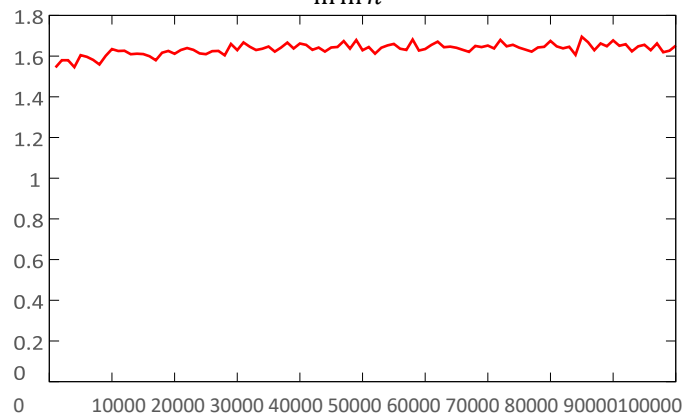
$L_n^{(2)}$ – maximum load with power of two choices



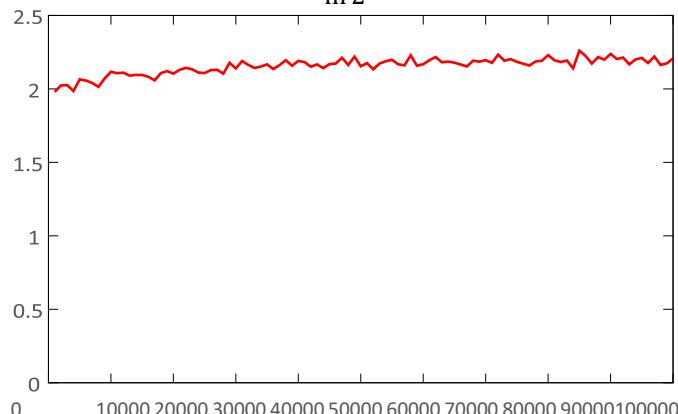
Zgodnie z intuicją - maksymalne wypełnienie urn jest o wiele mniejsze w przypadku wybierania dwóch urn i umieszczania kuli w tej mniej zapełnionej, gdyż daje nam to możliwość odciążenia urn. Dla większych n wartość $L_n^{(d)}$ będzie oczywiście coraz mniejsza, gdyż mamy wtedy większy wpływ na umiejscowienie kulek.

Eksperymentalne wyznaczanie asymptotyki

$$\frac{l_n^{(1)}}{\frac{\ln n}{\ln \ln n}}$$

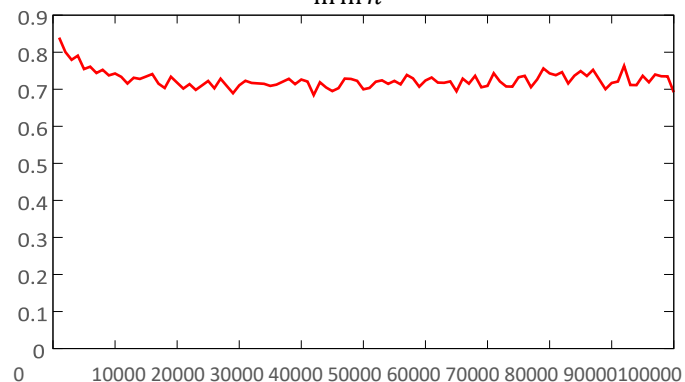


$$\frac{l_n^{(1)}}{\frac{\ln \ln n}{\ln 2}}$$

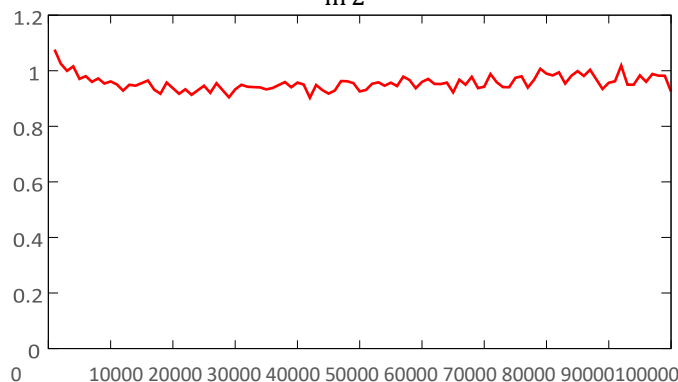


Powyższe wykresy sugerują, że asymptotyka wartości oczekiwanej $L_n^{(1)}$ jest bliska $\frac{\ln n}{\ln \ln n}$.

$$\frac{l_n^{(2)}}{\frac{\ln n}{\ln \ln n}}$$



$$\frac{l_n^{(2)}}{\frac{\ln \ln n}{\ln 2}}$$

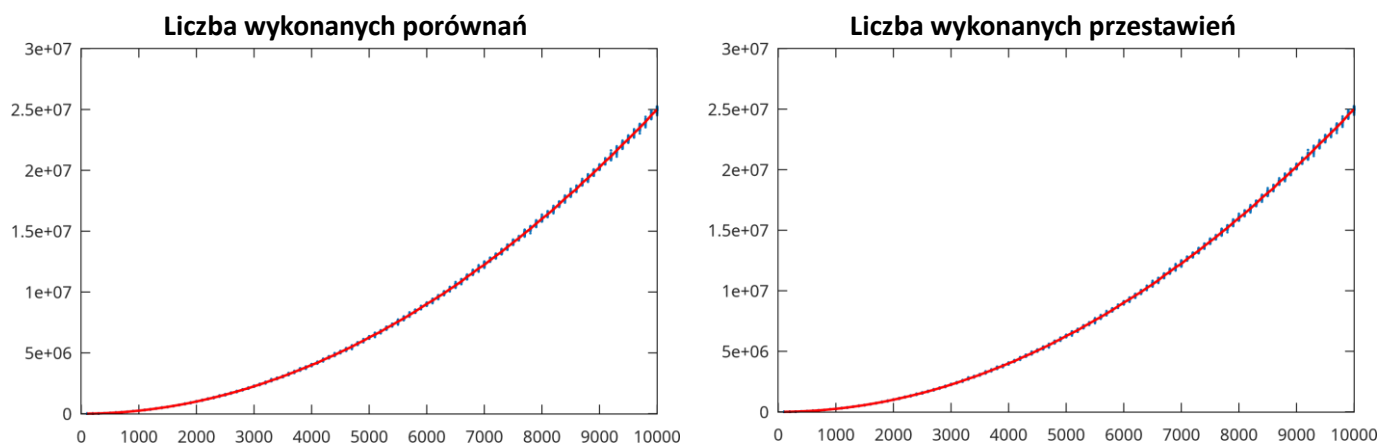


Tutaj sytuacja jest mniej klarowna, natomiast logicznym jest, że asymptotyka $L_n^{(2)}$ będzie mniejsza niż $L_n^{(1)}$, oraz powinna w jakiś sposób zależeć od d , co nie ma miejsca w $\frac{\ln n}{\ln \ln n}$. Zatem najprawdopodobniej $L_n^{(2)}$ jest rzędu $\frac{\ln \ln n}{\ln 2}$.

Dodatkowo, wartości badane w poprzednim zadaniu, których wykresy umieściłam [tutaj](#), również widocznie zmniejszyły się przy użyciu balanced allocation – poza, oczywiście, momentem pierwszej kolizji, który znacznie wzrósł. To może sugerować przydatność balanced allocation dla tworzenia funkcji hashujących.

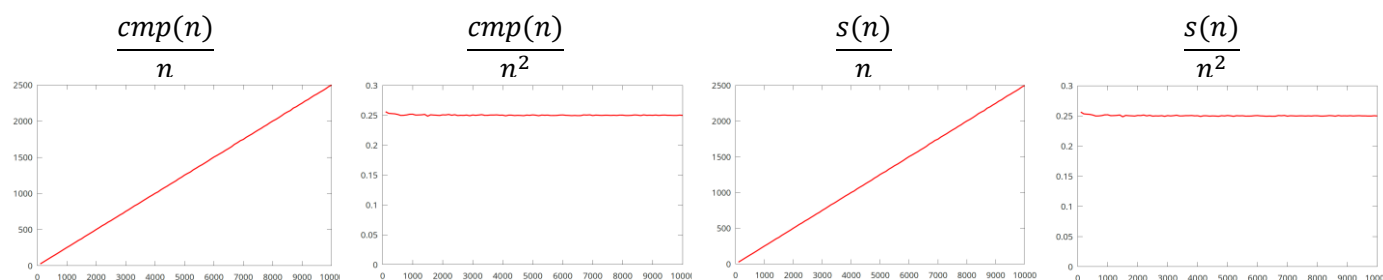
Zadanie 2

Wykresy badanych statystyk



W insertion sort dla prawie każdego porównania występuje także przestawienie, więc wykresy są bardzo podobne. Można też zaobserwować bardzo duże skupienie wartości wokół średniej, co oznacza, że niestety w prawie każdym przypadku czas działania insertion sort będzie tak samo wolny.

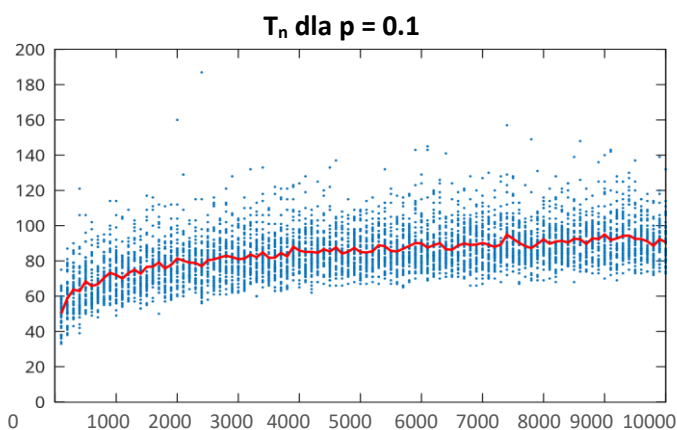
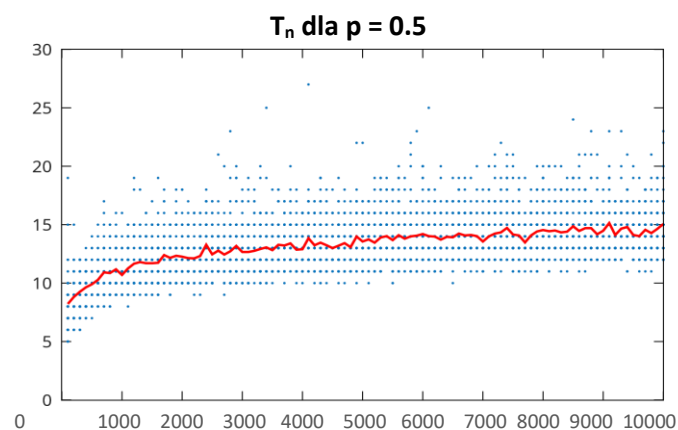
Eksperymentalne wyznaczanie złożoności obliczeniowej



Powyższe wykresy są o wiele bardziej klarowne, niż te z poprzedniego zadania. Możemy na nie spojrzeć tylko pobieżnie. Od razu widać, że zarówno $cmp(n)$, jak i $s(n)$ są asymptotyki n^2 – na szczęście, ponieważ gdyby algorytm sortujący oparty na porównaniach działał w czasie liniowym, byłoby to co najmniej podejrzane.

Zadanie 3

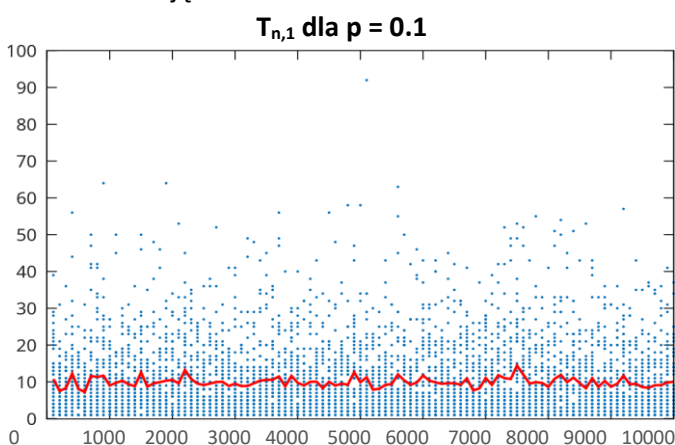
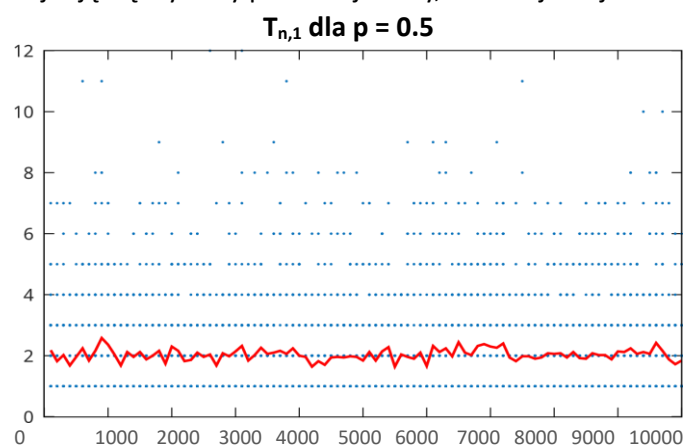
Wykresy minimalnej liczby rund, potrzebnych do rozesłania informacji



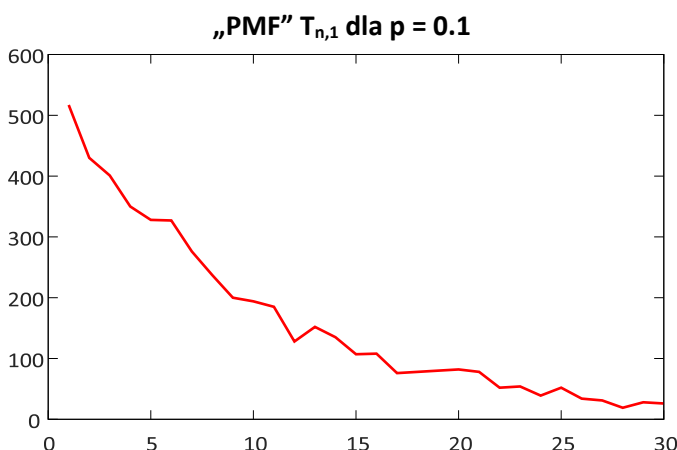
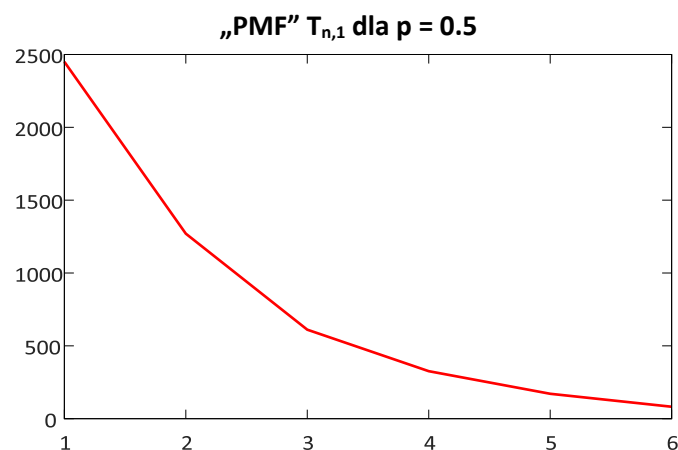
Na pierwszy rzut oka $t(n)$ wygląda na rosnące logarytmicznie. W następnej części zadania, przekonamy się, że to słuszna obserwacja.

Czym jest $T_{n,i}$?

Jako, że wszystkie zmienne losowe $T_{n,i}$ są od siebie niezależne, wystarczy przyrzeć się tylko jednej z nich. Poniżej znajdują się wykresy pierwszej rundy, w której stacja 1. odebrała informację.



Oczywiście wartość ta nie zależy od n . Z wykresów możemy natomiast odczytać, że wartość oczekiwana $T_{n,1}$ dla obu przypadków jest równa $\frac{1}{p}$, co w podejrzany sposób przypomina wartość oczekiwaną rozkładu geometrycznego z parametrem p . Jest to zgodne z intuicją, ponieważ w każdej rundzie stacja odbiera sygnał niezależnie od tego, czy udało jej się to w poprzedniej rundzie, czy nie – a wiemy, że „brak pamięci” jest cechą rozpoznawczą rozkładu Geo.



Powyższe wykresy przedstawiają liczbę występowania konkretnych wartości przyjmowanych przez $T_{n,1}$. Jak widać kształt krzywej jest bardzo podobny do wykresu funkcji masy prawdopodobieństwa rozkładu Geo(p), co potwierdza nasz poprzedni domysł.

Czym jest T_n ?

T_n „czeka” na stację, która najpóźniej odbierze informację, a więc oczywistym jest, że T_n można wyrazić jako:

$$T_n = \max\{T_{n,i} : i \in \{1, \dots, n\}\}$$

A więc szukamy asymptotyki wartości oczekiwanej maksimum niezależnych zmiennych losowych o rozkładzie Geo(p). W swojej pracy „[On the expectation of the maximum of IID geometric random variables](#)” Bennett Eisenberg pisze:

[...]

”

There is no such simple expression for $E(M_n^*)$, the expected value of the maximum of n IID geometric random variables with mean $1/p$. There is the infinite series expression from the tail probabilities of the maximum of the random variables and even a finite sum expression, but these expressions are not so useful.

It is also easily seen that

$$\frac{1}{\lambda} \sum_{k=1}^n \frac{1}{k} \leq E(M_n^*) < 1 + \frac{1}{\lambda} \sum_{k=1}^n \frac{1}{k}$$

Where $q = (1 - p) = e^{-\lambda}$. This is better, but not good enough.

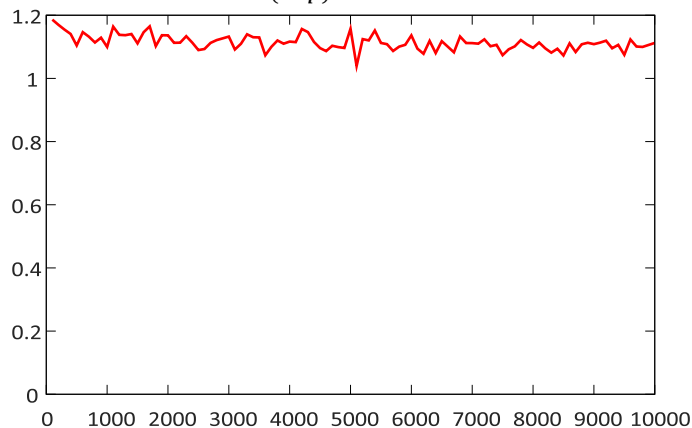
[...]

But it is good enough for us:

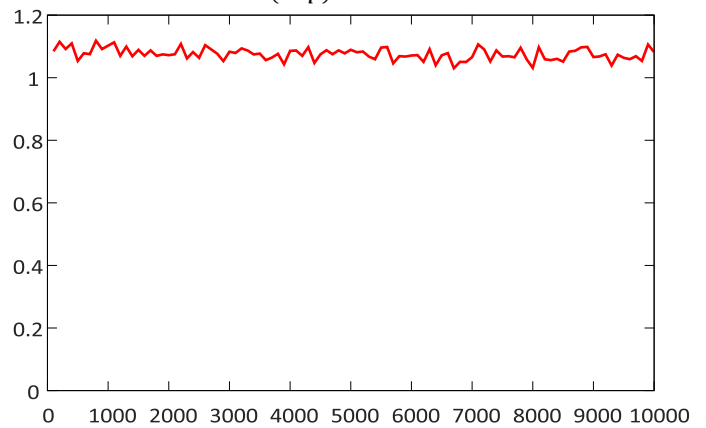
$\sum_{k=1}^n \frac{1}{k}$ to n -ta liczba harmoniczna H_n i wiemy, że ma asymptotykę $\ln n$.

$$(1 - p) = e^{-\lambda} \rightarrow \lambda = -\ln(1 - p)$$

$$\frac{t(n)}{\ln n} \text{ dla } p = 0.5$$



$$\frac{t(n)}{\ln n} \text{ dla } p = 0.1$$



Z powyższych wyników przeprowadzonych eksperymentów wyraźnie wynika, że asymptotyka maksimum niezależnych zmiennych losowych o rozkładzie Geo(p) najprawdopodobniej faktycznie jest bliska $\frac{\ln n}{-\ln(1-p)}$.

Kod źródłowy: <https://github.com/nskowron/Statistics2024/tree/main/Exc3>

Szczęśliwego Nowego Roku:)