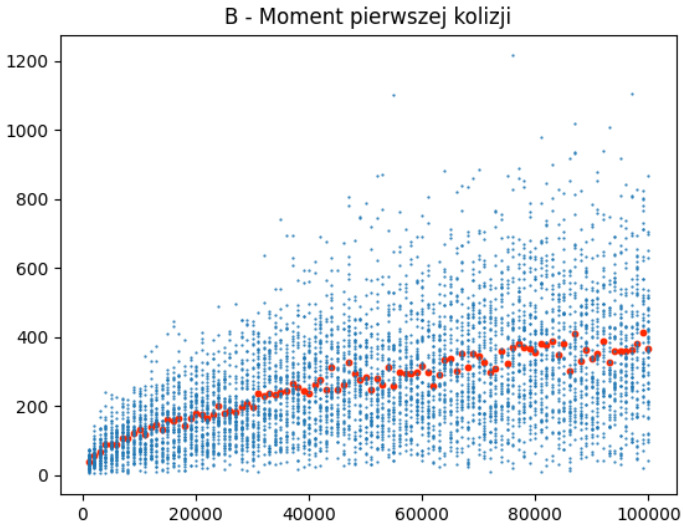
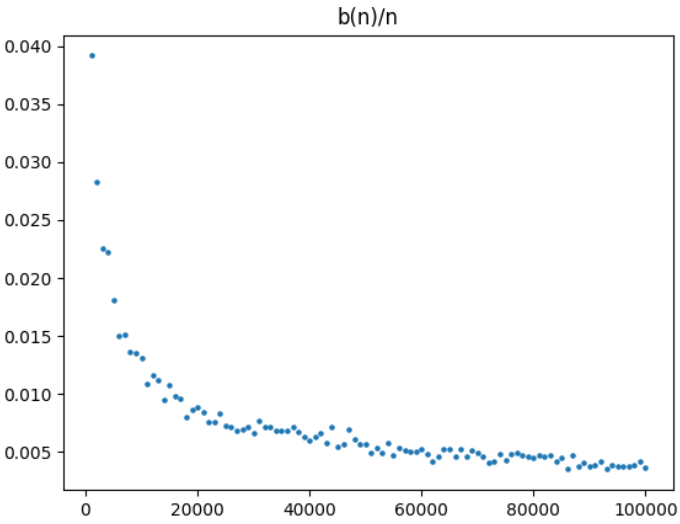
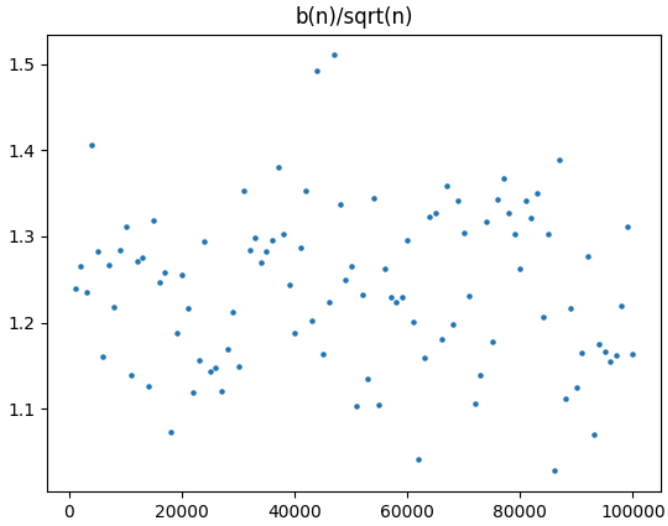
1. WYKRESY DLA PIERWSZEGO MOMENTU KOLIZJI







AD1.

Na wykresie nr.1 zaprezentowane są odczytane momenty pierwszej kolizji tzn. pierwszy moment, w którym dokładana kula trafia do niepustej urny. Czerwonym kolorem zaznaczono ich średnie wartości z 50 powtórzeń symulacji przy n-urnach. Można zauważyć, że wraz z rosnącą liczbą urn, maleje kurtoza (słabnie koncentracja uzyskanych wyników wokół średniej).

Dokładność aproksymacji rośnie wraz z zwiększoną liczbą urn, lecz z powodu malejącego skupienia wartości należałoby zwiększyć ilość powtorzeń symulacji przy danej liczbie, w celu dodatkowego zwiększenia dokładności aproksymacji.

FAKTY:

$$P(k) = \frac{k}{n} * \prod_{i=0}^{k} \frac{n-i}{n}$$

Powyżej umieszczony jest wzór na prawdopodobieństwo pierwszej kolizji przy dodawaniu k-tej kuli do losowej urny.

$max\{P(1), \cdots, P(n)\} = P(x)$, gdzie x to najbliższa liczba całkowita do $\sqrt[2]{n}$. (Najwyższe prawdopodobieństwo pierwszej kolizji jest dla kuli dodawanej jako $\sqrt[2]{n}$ w kolejności.)

Prawdopodobieństwo na wystąpienie pierwszej kolizji w powtórzeniu między 1, a k-tym losowaniem:

$$\sum_{i=1}^{k} P(i)$$

Prawdopodobieństwo na wystąpienie pierwszej kolizji przed lub podczas $\sqrt[2]{n}$ -tego dołożenia kuli będzie równe około 0,4 natomiast przed lub podczas $1{,}4 * \sqrt[2]{n}$ -tego dołożenia kuli około 0,6. Przy $1{,}2 * \sqrt[2]{n}$ jest to około 0,5. Dlatego na wykresie nr.3 średnie wartości $b(n)/\sqrt[2]{n}$ w większości występują właśnie pomiędzy 1.0 a 1.4, wokół 1.2.

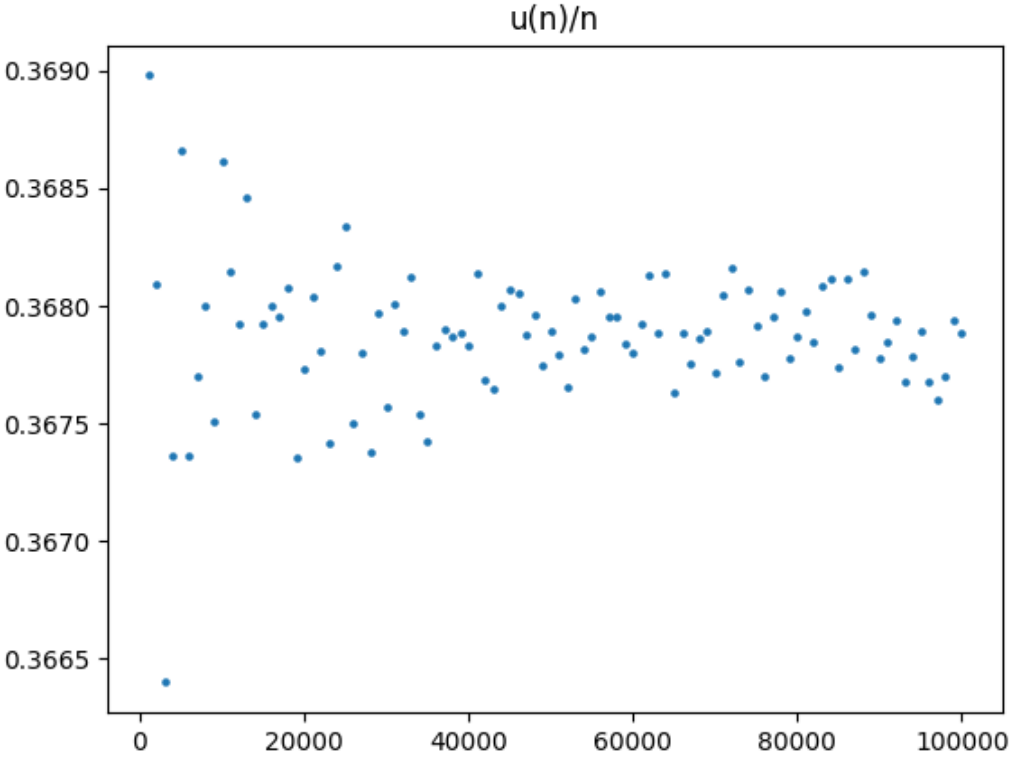
$$\lim_{n\to\infty} Avg\big(B(n)\big) = +\infty$$

$$E\big(B(n)\big) = \sqrt{\frac{n\pi}{2}} + \frac{2}{3} + \frac{1}{12}\sqrt{\frac{\pi}{2n}} + O\left(\frac{1}{n}\right)$$

WNIOSEK:

$$B(n) \sim O(\sqrt[2]{n})$$

2. WYKRESY DLA LICZBY PUSTYCH URN PO WRZUCENIU N KUL



AD2.

$$U_n \sim O(n)$$

Na wykresie nr.1 widoczne jest, że skupienie wartości pobieranych w k-tych powtórzeniach symulacji wokół ich średniej wartości jest bardzo wysokie. Wykres asymptotycznie przypomina wykres rosnącej funkcji liniowej.

$$\lim_{n\to\infty} Avg(U_n) = +\infty$$

Wartość oczekiwana dla zmiennej losowej X_i równej 1 gdy i-ta urna jest pusta lub 0 w przeciwnych wypadku wynosi:

$$E(X_i) = \left(1 - \frac{1}{n}\right)^n \to \frac{1}{e}$$

Na tej podstawie możemy wyznaczyć wartość oczekiwaną dla średniej wartości U_n:

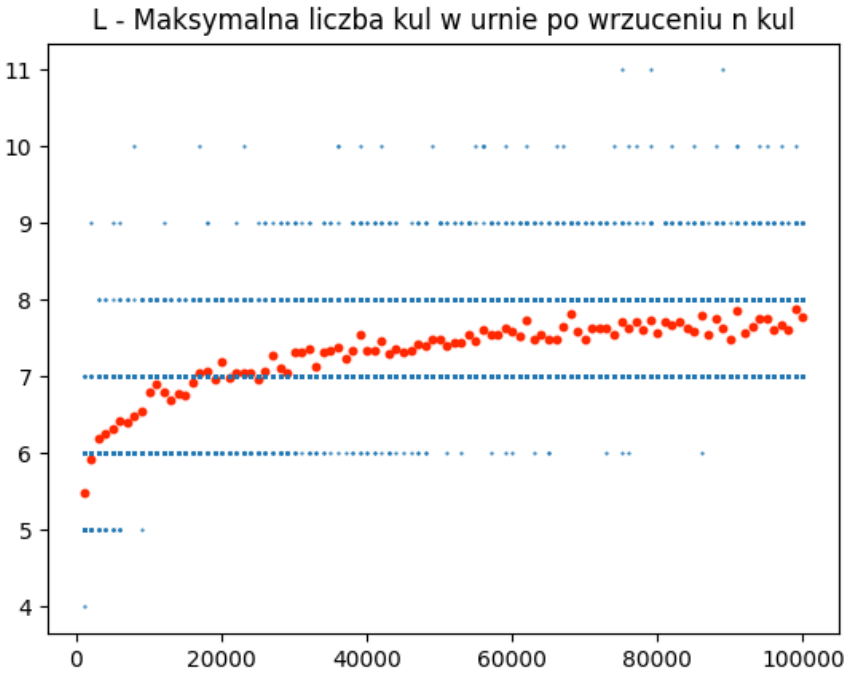
$$E\big(Avg(U_n)\big) = n * \left(1 - \frac{1}{n}\right)^n$$

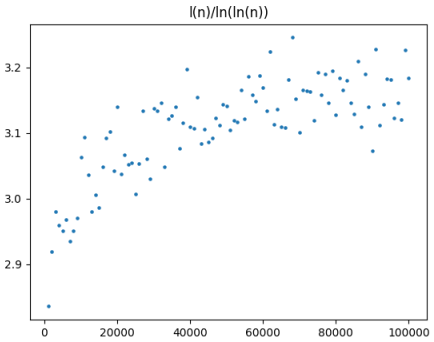
Odczytując wykres nr.1 widzimy, że uzyskane w aproksymacji wartości potwierdzają poprawność powyższego wzoru.

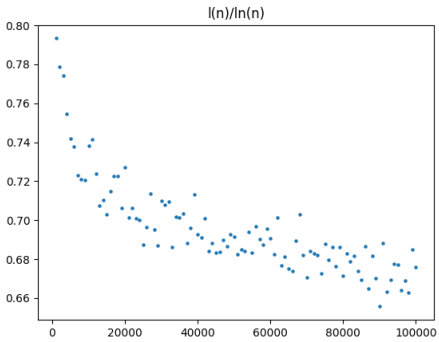
Wykres nr.2 ukazuje nam aproksymacje wartości oczekiwanej wyżej zdefiniowanej zmiennej losowej X_i.

$$\lim_{n\to\infty} \frac{Avg(U_n)}{n} = \frac{1}{e}$$

3. WYKRESY DLA MAKSYMALNEJ LICZBY KUL W URNIE PO WRZUCENIU N KUL







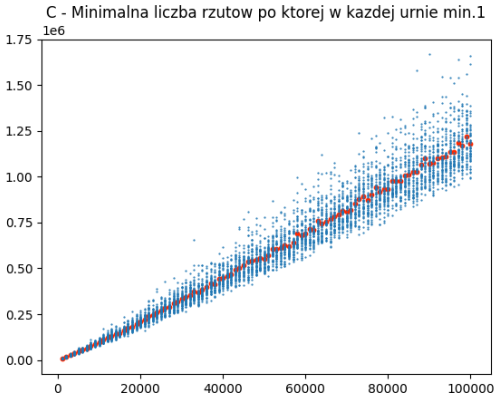
AD3.

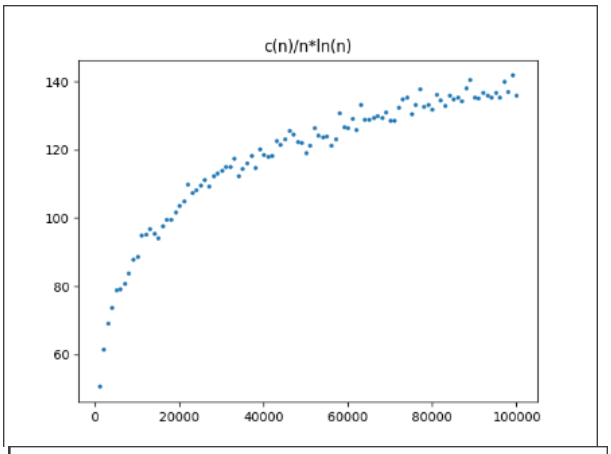
$$L_n \sim O(\ln(n)/\ln(\ln(n)))$$

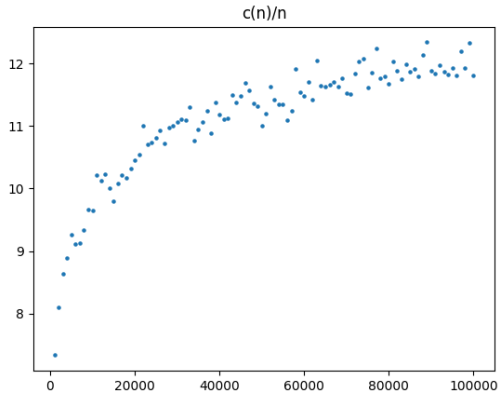
Wartość oczekiwana maksymalnej liczby kul w jednej urnie po wrzuceniu losowo n kul do n urn rośnie wraz z wzrostem liczby urn lecz znacznie wolniej.

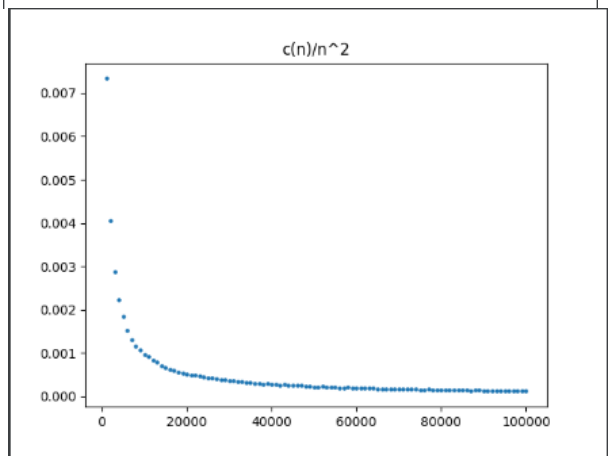
$$\lim_{n\to\infty} Avg(L_n) = +\infty$$

4. WYKRESY DLA MINIMALNEJ LICZBY RZUTÓW PO KTÓREJ W KAŻDEJ URNIE ZNAJDUJE SIĘ MIN. 1 KULA









AD4.

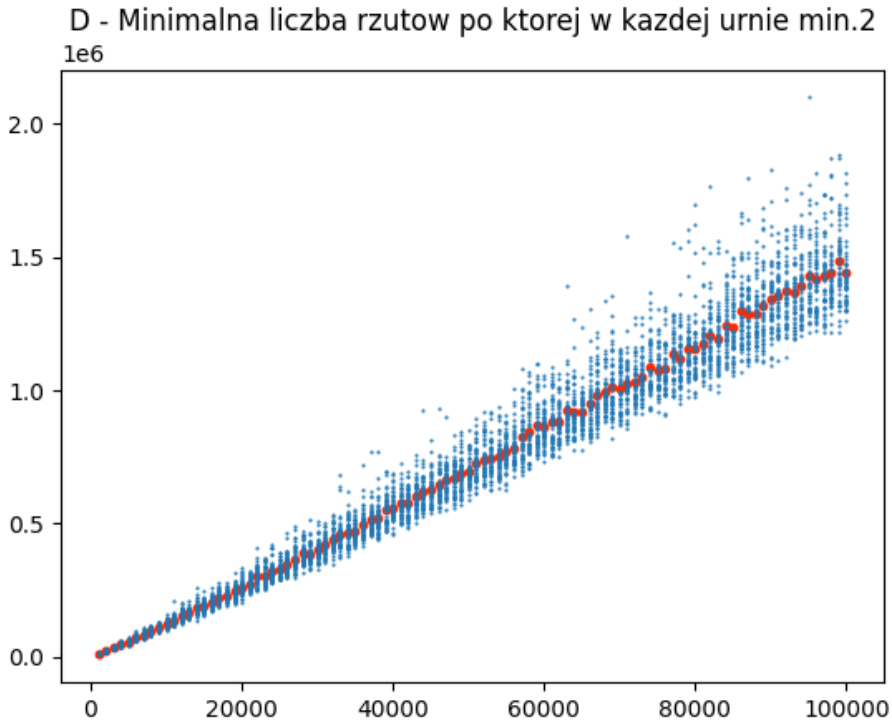
$$\lim_{n\to\infty} \boldsymbol{Avg(C_n)} = +\infty$$

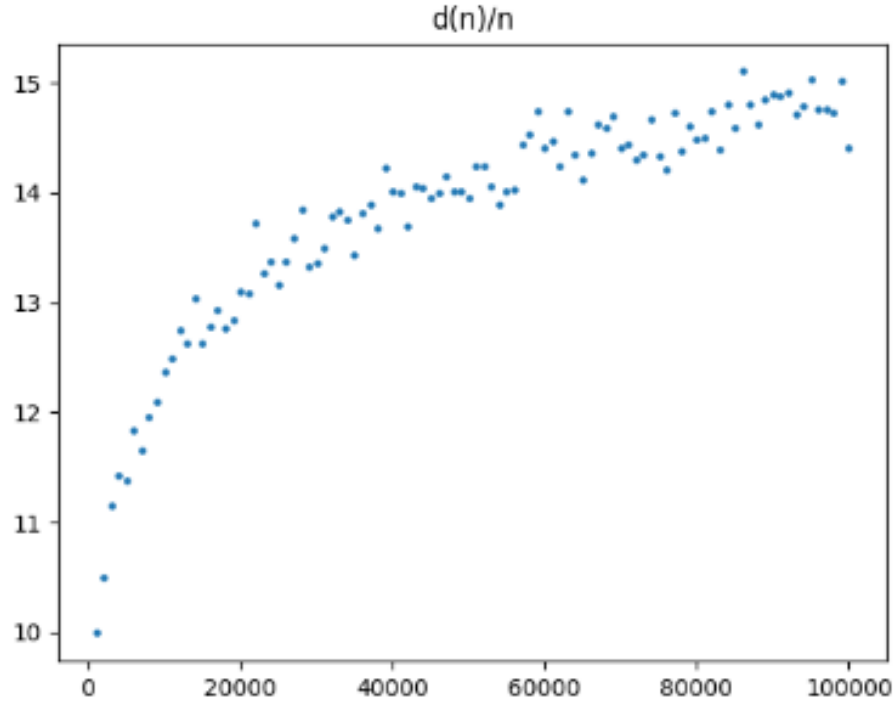
Asymptotyka: $\theta(n * log(n))$

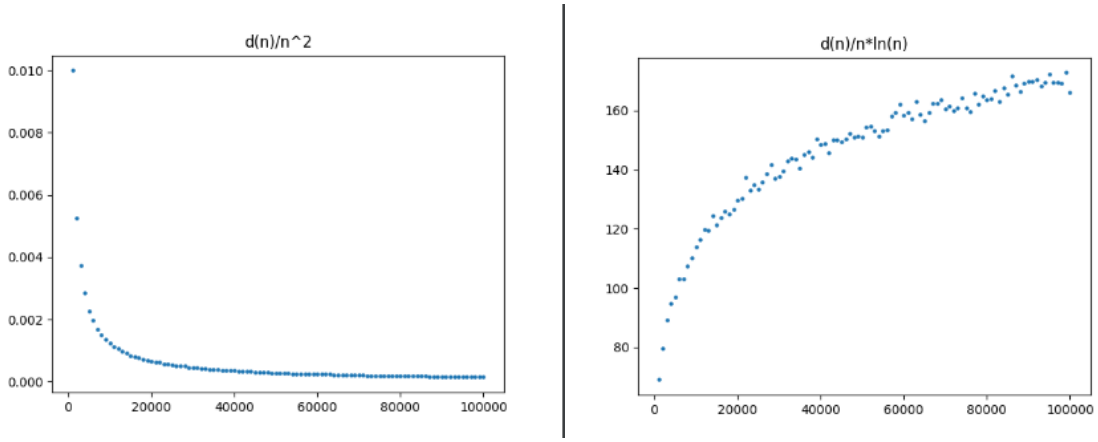
$$E(Avg(C_n)) = n * H_n = n * log\, n + \gamma * n + \frac{1}{2} + O(1/n)$$

Z wykresu możemy odczytać, że średnia wartość C rośnie wraz z rosnącą liczbą urn. Średnia wartość C zwiększa się szybciej, asymptotycznie przypomina wykres funkcji logarytmicznej.

5. WYKRESY DLA MINIMALNEJ LICZBY RZUTÓW PO KTÓREJ W KAŻDEJ URNIE ZNAJDUJĄ SIĘ MIN. 2 KULE







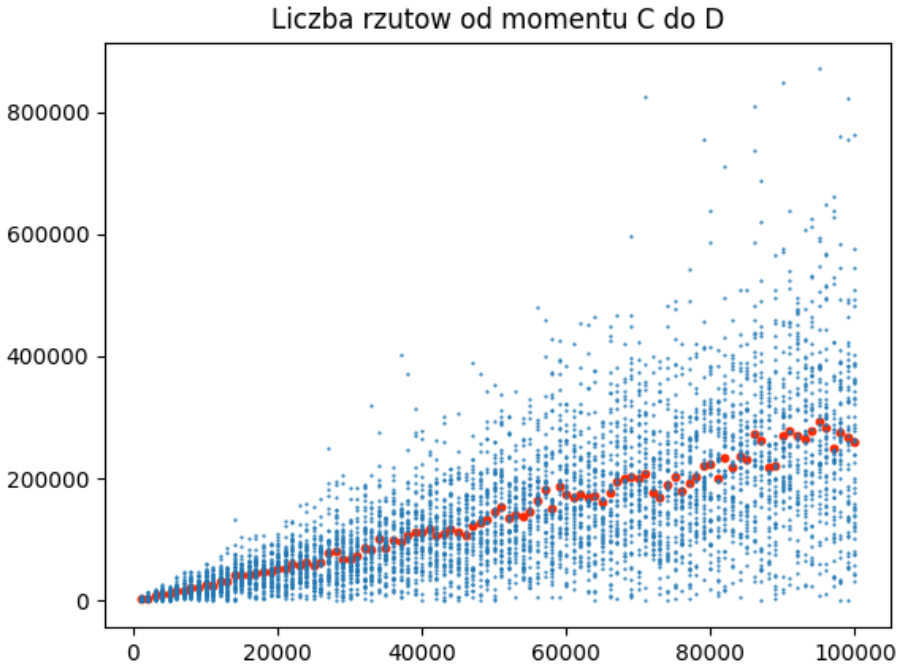
AD5.

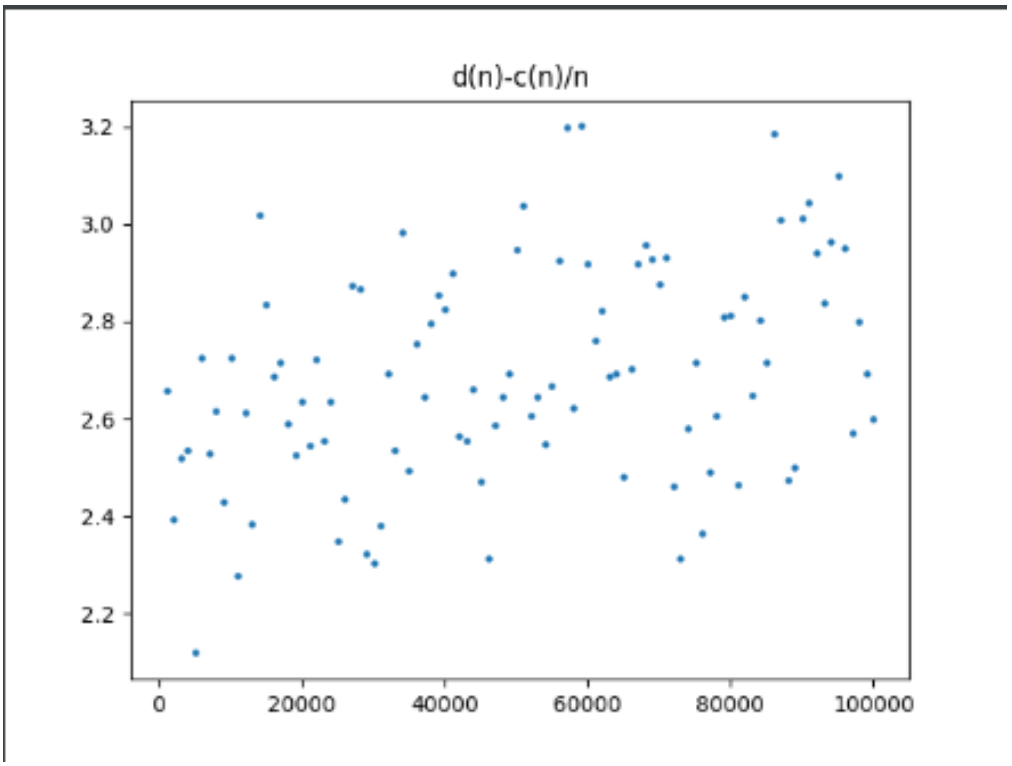
$$\boldsymbol{D_n \sim O(n * ln(n))}$$

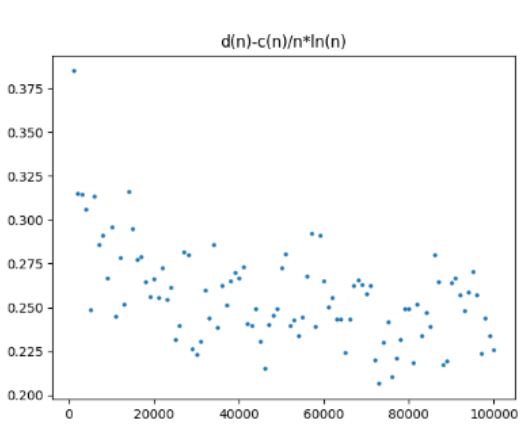
$$\boldsymbol{\lim_{n\to\infty} Avg(D_n)} = +\infty$$

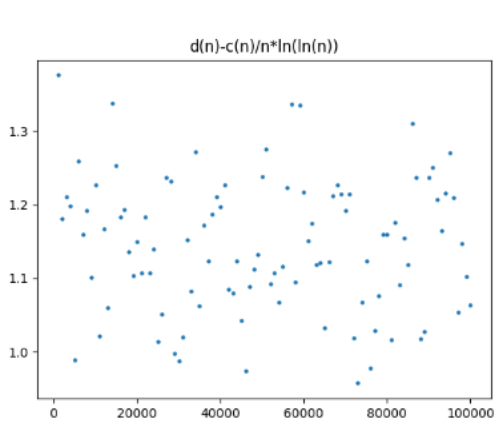
Z wykresów możemy odczytać że średnia minimalna liczba rzutów potrzebna aby w każdej urnie rośnie znacznie szybciej niż zwiększa się liczba urn. Przyrost ten jednak jest logarytmiczny. Wykresy asymptotycznie przypominają wykresy poprzedniej statystyki tzn. minimalna liczba wrzuconych kul po których w każdej z n urn znajdzie się min.1 kula.

6. WYKRESY DLA LICZBY RZUTÓW MIĘDZY MOMENTEM C A MOMENTEM D









AD6.

$$(\boldsymbol{D_n - C_n}) \sim \boldsymbol{O(n * ln(ln(n)))}$$

$$\boldsymbol{\lim_{n\to\infty} Avg(D_n - C_n)} = +\infty$$

Z wykresu nr.1 można odczytać, że różnica rzutów między momentem w którym w każdej urnie są minimum dwie kula a momentem w którym było w nich po min. jednej jest ponad dwukrotnie większa od liczby urn. Na wykresie nr.2 widać że dla niektórych n była ona nawet trzykrotnie większa. Średnia wartość jest przeważnie bliska $n * e$ i delikatnie większa od $n * ln(ln(n))$

POWIĄZANIE Z BIRTHDAY PARADOX ORAZ COUPON COLLECTOR'S PROBLEM

W teorii prawdopodobieństwa „birthday problem" dotyczy prawdopodobieństwa, że w zbiorze n losowo wybranych osób co najmniej dwie będą obchodzić urodziny w tym samym dniu. Sprzecznie z intuicją już w grupie 23 osób prawdopodobieństwo takiego zjawiska wynosi aż ponad 50%. Jest to znacznie mniejsza liczba niż można by pomyśleć. Na tym polega „birthday paradox". W ten sam sposób możemy spojrzeć na statystkę nr.1 dotyczącą prawdopodobieństwa kolizji po dorzuceniu n kul. Dla przykładu przy 1000 urn sprzecznie z intuicją prawdopodobieństwo kolizji osiąga aż 51% przy 38 dorzucanej kuli.

W teorii prawdopodobieństwa problem kolekcjonera kuponów opisuje konkursy „zbierz wszystkie kupony i wygraj". Zadaje następujące pytanie: Jeśli każde pudełko płatków zbożowych zawiera kupon, a istnieje n różnych rodzajów kuponów, jakie jest prawdopodobieństwo, że trzeba będzie kupić więcej niż t pudełek, aby zebrać wszystkie n kuponów? Możemy powiązać to z naszym problemem: „Jakie jest prawdopodobieństwo, że trzeba będzie dorzucić więcej niż t kul, aby w każdej urnie znajdowała się co najmniej jedna kula?"

BIRTHDAY PARADOX W FUNKCJACH HASHUJĄCYCH

Funkcje hashujące transformują dokumenty w numery hash. Numer ten jest następnie łączony z tajnym kluczem osoby podpisującej dokument w celu utworzenia podpisu. Ktoś czytający dokument mógłby następnie „odszyfrować" podpis za pomocą klucza publicznego osoby podpisującej, co dowodziłoby, że osoba ta podpisała dokument cyfrowo. Używając Birthday Paradox możemy kogoś wrobić w podpisanie fałszywych dokumentów bez jego wiedzy. Potrzebne są do tego dwa dokumenty (fałszywy i prawdziwy), które będą generowały tą samą wartość hash. Znając omawiany paradoks wiemy, że prawdopodobieństwo wystąpienia kolizji wartości hash między prawdziwym dokumentem, a którymś z fałszywych jest znacznie większe niż mogło by się wydawać. W tym przypadku jest to około pierwiastek kwadratowy z liczby możliwych wyjść funkcji hashującej.