

Symulowanie Procesów Losowych - Kule i Urny

Jakub Kogut

1 Wstęp

Sprawozdanie do zadania domowego 2.

2 Opis Zadania

Należało zaprojektować oraz zaimplementować probabilistyczny model kul i urn. Kolejne m kul wrzucanych jest niezależnie oraz jednostajnie do n ponumeryowanych urn. Celem zadania było eksperymentalne wyznaczenie następujących wartości:

1. B_n - moment pierwszej kolizji
2. U_n - liczba urn, które pozostają puste po wrzuceniu n kul
3. C_n - minimalna liczba rzutów, po której w każdej z urn znajduje się co najmniej jedna kula
4. D_n - minimalna liczba rzutów, po której w każdej z urn znajduje się co najmniej dwie kule (równocześnie jest to warunek końcowy symulacji)
5. $D_n - C_n$ - ilość rzutów potrzebna do uzyskania drugiej kuli w każdej z urn, po tym jak w każdej jest już minimum jedna kula

3 Metodologia

Podana była ustalona wartość $k = 50$ niezależnych symulacji dla każdego n z zakresu $n \in \{1000, 2000, \dots, 100000\}$. W trakcie wykonywania jej, na bieżąco zapisywane były szukane wartości. Następnie wypisywano wyniki do pliku, liczone średnie oraz generowano wykresy.

Przy generowaniu liczb losowych użyłem generatora *Mersenne Twister*

4 Wnioski

Na podstawie przeprowadzonych symulacji i wygenerowanych wykresów można wyciągnąć następujące wnioski dotyczące badanych zmiennych:

1. Moment pierwszej kolizji (B_n)

Wartości momentu pierwszej kolizji rosną w miarę zwiększania liczby urn, co jest zgodne z intuicją wynikającą z paradoksu urodzinowego. Wykres średnich wartości B_n względem n wskazuje na niską koncentrację wyników wokół wartości średnich, zmniejszającą się wraz ze wzrostem n .

Wykres ciągu B_n podzielonego przez podane funkcje n oraz \sqrt{n} sugeruje, że asymptotyka B_n jest zbliżona do $O(n)$.

2. Liczba pustych urn (U_n)

Liczba pustych urn po wrzuceniu n kul zwiększa się wraz ze wzrostem n , dobrze odzwierciedlając wykres $y = n$. Wykres U_n przedstawia wartości poszczególnych symulacji mocno skoncentrowane wokół średniej. Wykresy ilorazu U_n/n wskazują, że wartości te są asymptotycznie zbliżone do stałej wartości $O(n)$.

3. Minimalna liczba rzutów do zapełnienia urn (C_n)

Na wykresie C_n widoczna jest dobra koncentracja wyników poszczególnych symulacji do wartości średniej, choć zmniejszającą się wraz z n . Wykresy ilorazów $\frac{C_n}{n}$, $\frac{C_n}{\ln n}$ wskazują, że wartości te są asymptotycznie zbliżone do stałej wartości $O(n)$.

4. Minimalna liczba rzutów do dwóch kul w każdej urnie (D_n)

Aby każda urna zawierała przynajmniej dwie kule, potrzebna jest większa liczba rzutów niż w przypadku C_n . Wyniki symulacji pokazują, że wartości D_n rosną szybciej niż C_n . Wykres D_n pokazuje, że wartości D_n są podobnie skoncentrowane wokół wartości średnich jak C_n . Z Wykresów ilorazów $\frac{D_n}{n}$, $\frac{D_n}{n \ln n}$, $\frac{D_n}{n \ln \ln n}$ można jednoznacznie stwierdzić, że wartości D_n są asymptotycznie zbliżone do $O(n \ln n)$.

5. Różnica między D_n a C_n

Wartości $D_n - C_n$ reprezentują dodatkowy wysiłek wymagany, aby każda urna zawierała co najmniej dwie kule po osiągnięciu pełnego zapełnienia przynajmniej jedną kulą. Wykres $D_n - C_n$ pokazuje, że wartości te rosną w miarę zwiększania n . Wykresy ilorazów $\frac{D_n - C_n}{n}$, $\frac{D_n - C_n}{n \ln n}$, $\frac{D_n - C_n}{n \ln \ln n}$ nie jednoznacznie sugerują, że wartości te są asymptotycznie zbliżone do $O(n \ln n)$ lub $O(n \ln \ln n)$.

6. Intuicje związane z nazwami problemów

1. **Paradox urodzinowy (B_n):** Intuicja opiera się na zaskakująco małej liczbie kul potrzebnych do pierwszej kolizji, co jest analogiczne do problemu dwóch osób dzielących tę samą datę urodzin.

2. **Problem kolekcjonera kuponów (C_n):** Minimalna liczba rzutów do zapelnienia urn przypomina problem zebrania pełnego zestawu kuponów w przypadku losowego wybierania.

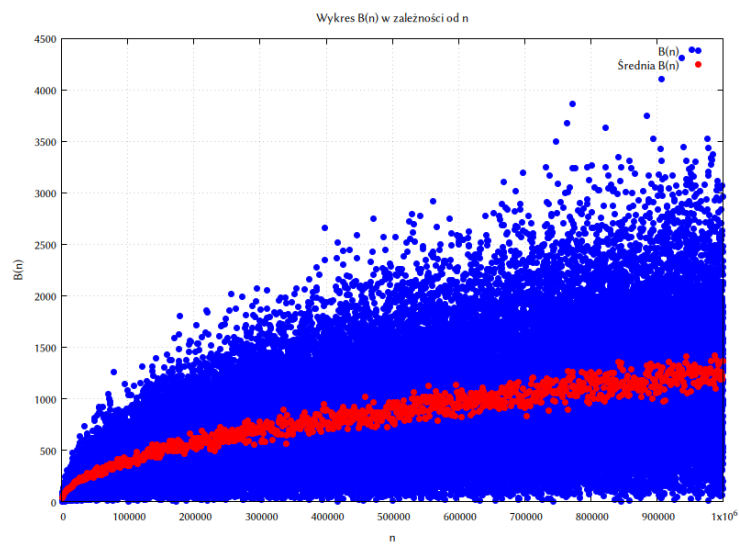
7. Znaczenie w kontekście funkcji haszujących

Paradox urodzinowy ma istotne zastosowanie w kryptografii, szczególnie w kontekście funkcji haszujących. Mała liczba rzutów prowadząca do kolizji podkreśla konieczność stosowania funkcji haszujących o dużej przestrzeni wyjściowej w celu minimalizacji ryzyka kolizji.

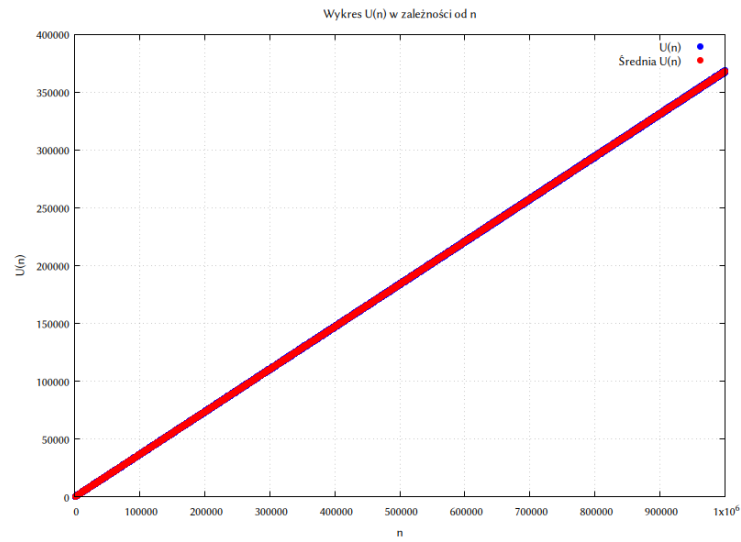
5 Podsumowanie

Przeprowadzone symulacje potwierdzają teoretyczne właściwości modelu kul i urn, w tym asymptotykę badanych zmiennych oraz intuicje związane z ich nazewnictwem. Dodatkowo, zastosowanie generatora Mersenne Twister zapewniło stabilność i wiarygodność wyników.

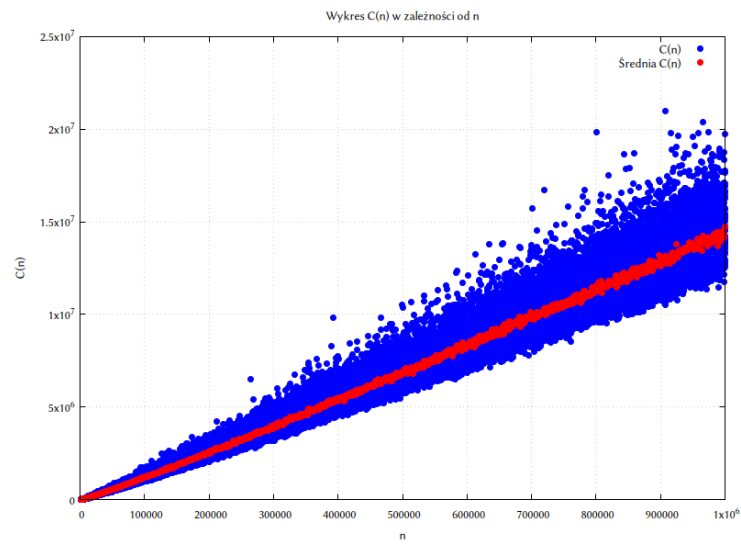
6 Wykresy



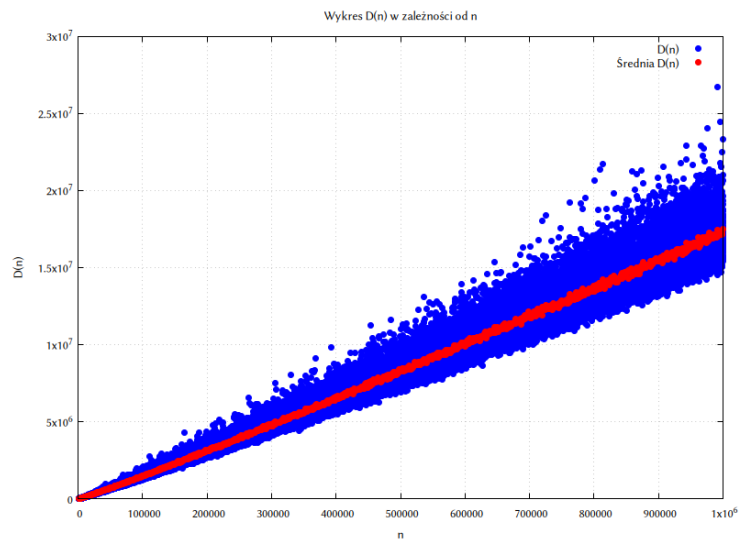
Rysunek 1: Wykres Wykreswartości B_n



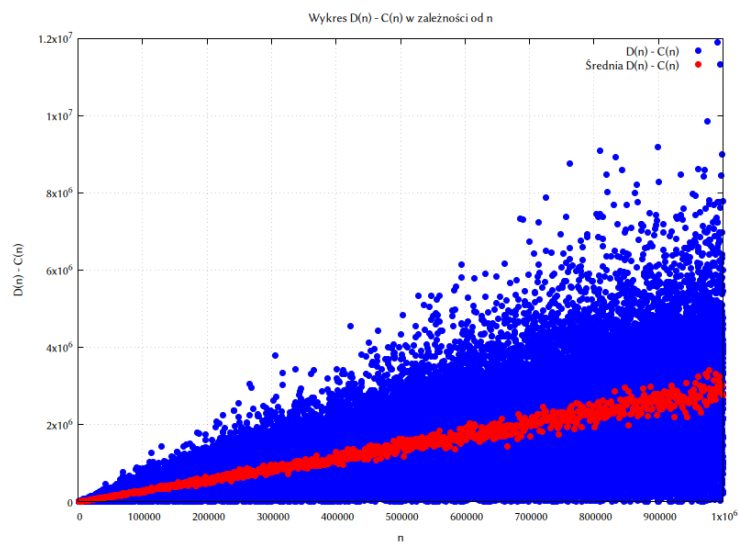
Rysunek 2: Wykres wartości U_n



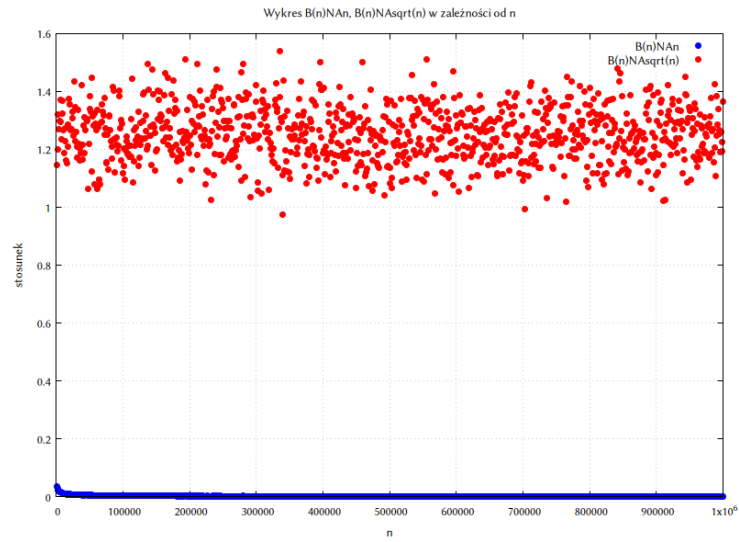
Rysunek 3: Wykres wartości C_n



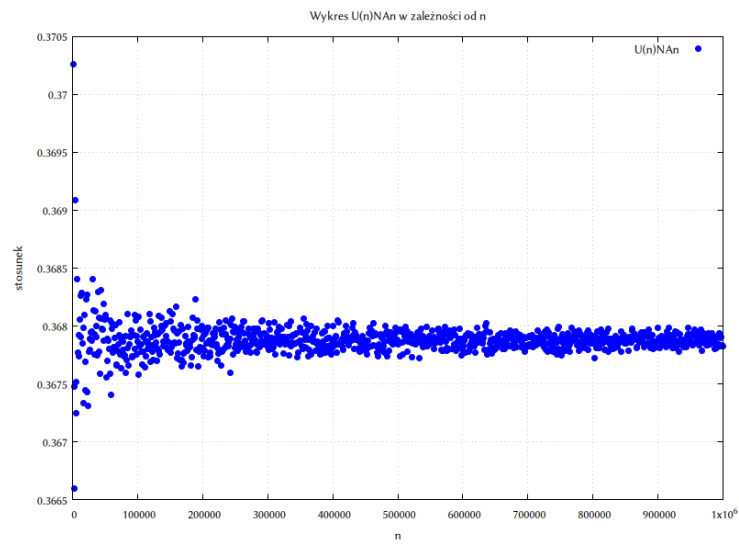
Rysunek 4: Wykres wartości D_n



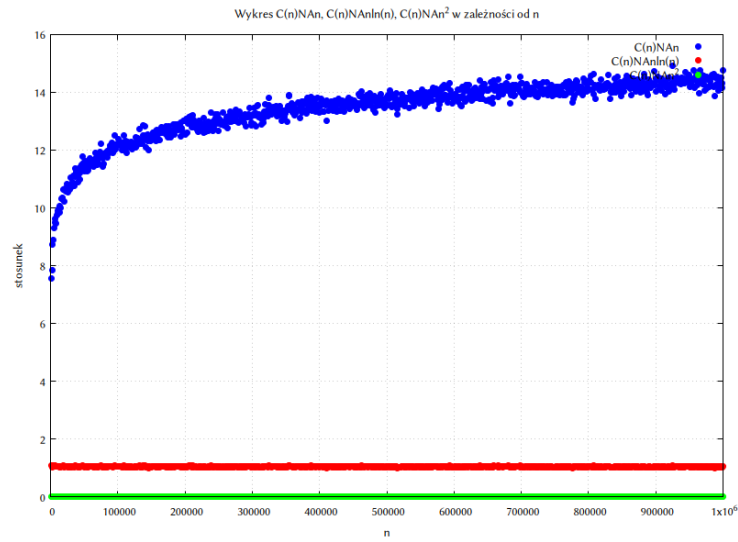
Rysunek 5: Wykres wartości $D_n - C_n$



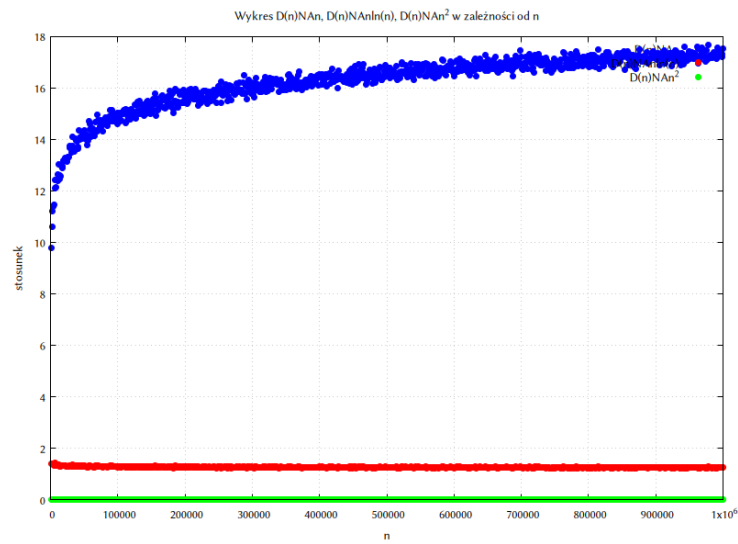
Rysunek 6: Wykres wartości B_n podzielone przez n oraz \sqrt{n}



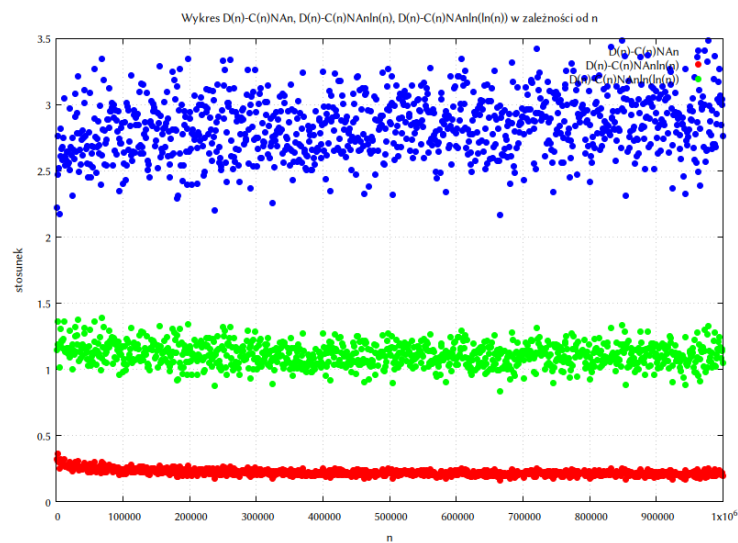
Rysunek 7: Wykres wartości U_n podzielone przez n



Rysunek 8: Wykres wartości C_n podzielone przez n , $n \ln n$ oraz n^2



Rysunek 9: Wykres wartości D_n podzielone przez n , $n \ln n$, $n \ln \ln n$



Rysunek 10: Wykres wartości $D_n - C_n$ podzielone przez n , $n \ln n$, $n \ln \ln n$