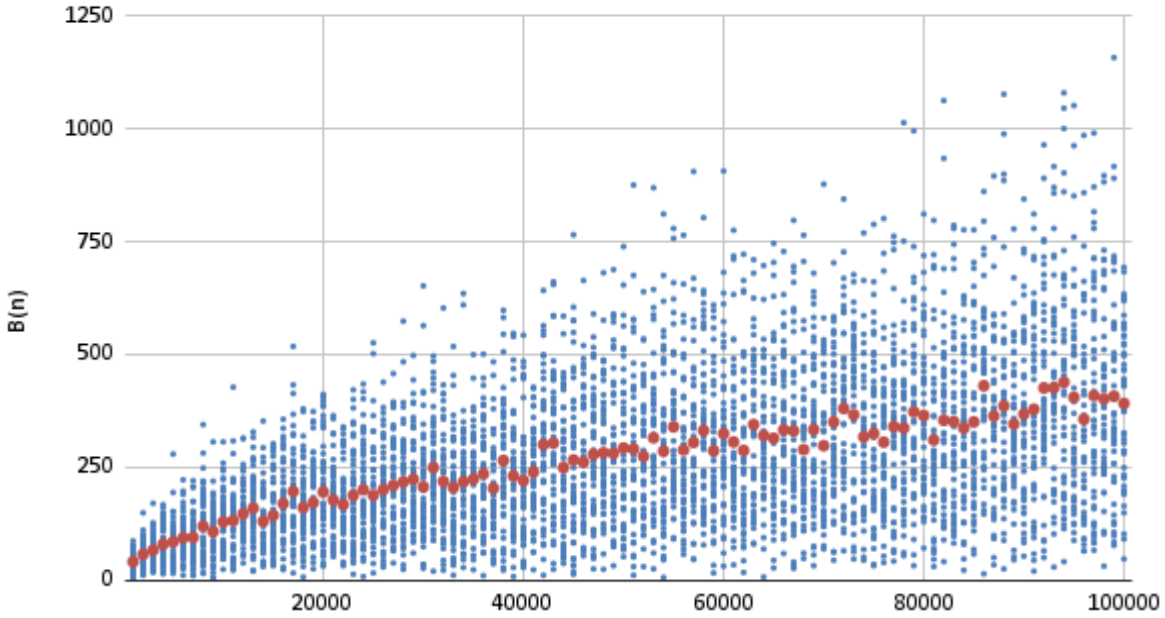
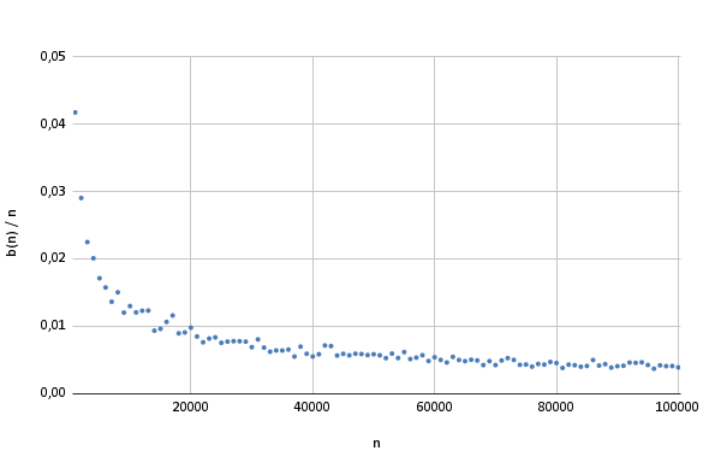
Wiktor Bachta, 272 353

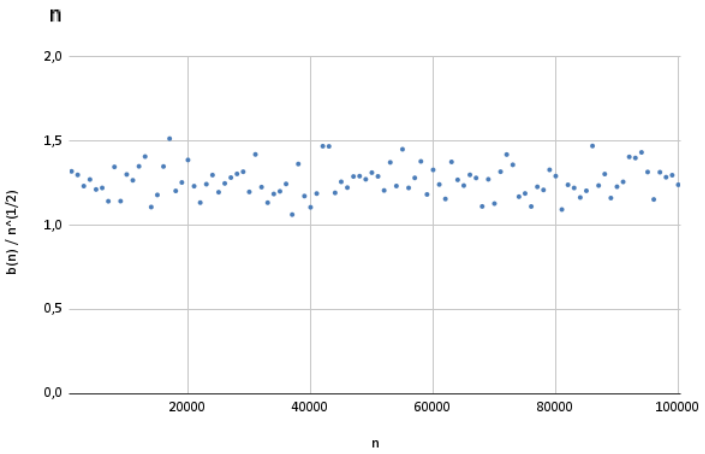
ZADANIE DOMOWE 2

OPIS:
Zadanie polega na symulacji losowego wrzucania kul do urn i obserwacji pewnych parametrów eksperymentu. Zastosowane przeze mnie oznaczenia są takie same jak w treści zadania domowego.
Najlepszą hipotezę dla asymptotyki daje wykres najbliższy funkcji stałej.

B(n) - moment pierwszej kolizji

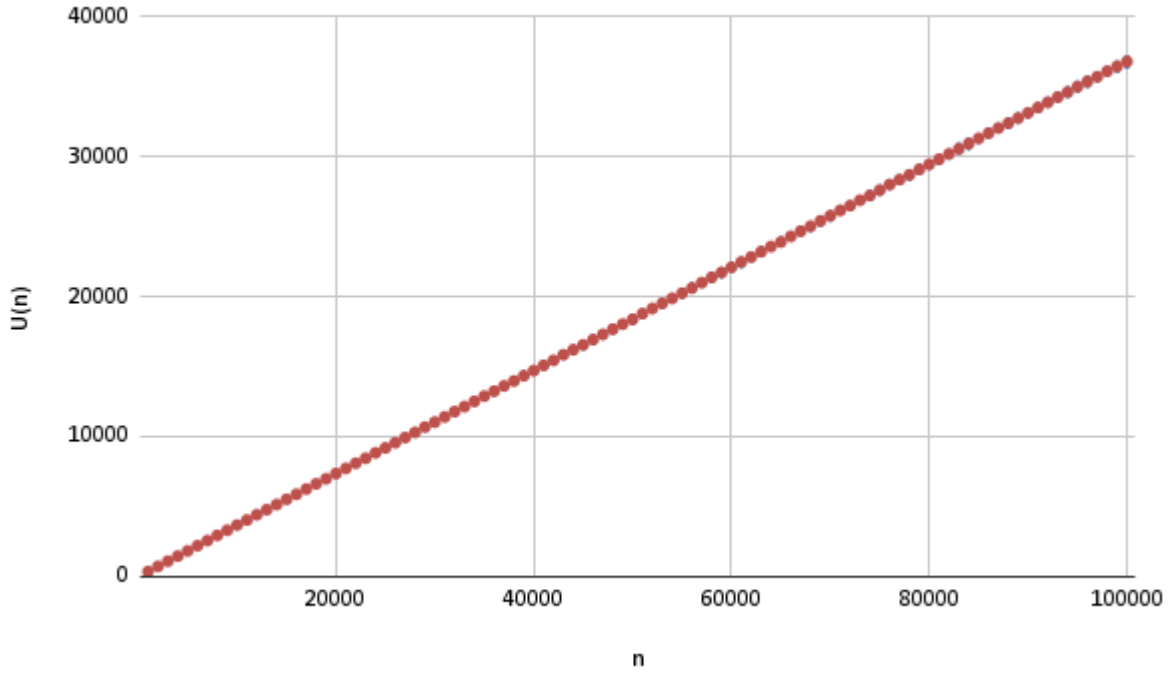


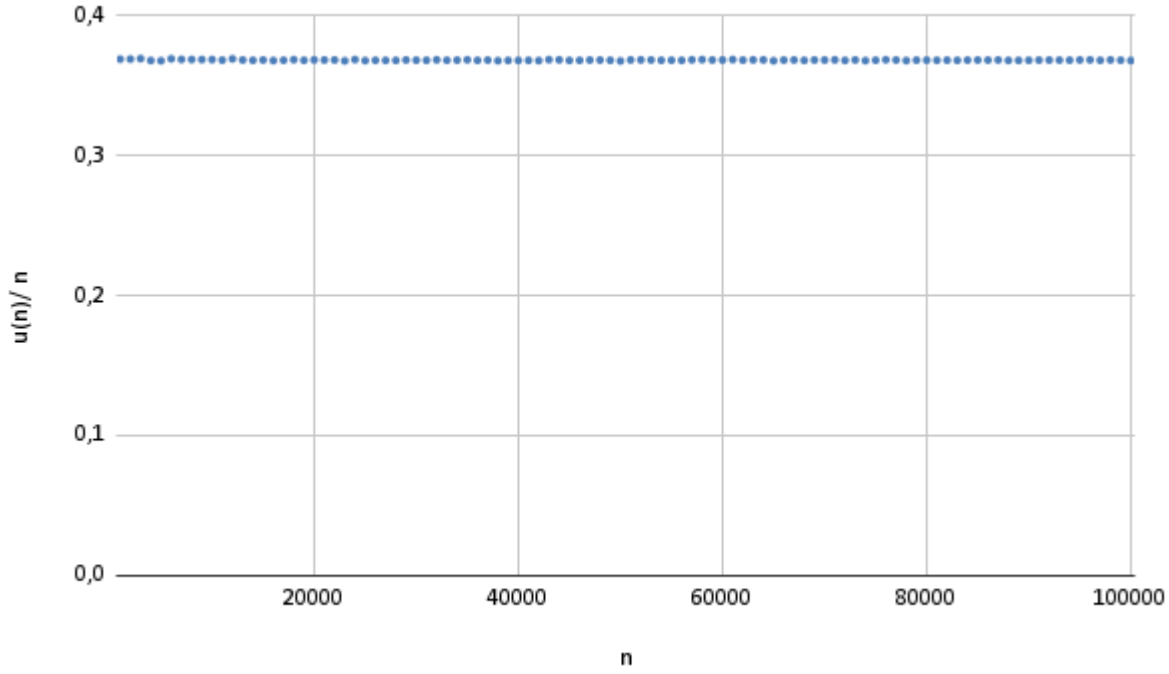




Wyniki eksperymentu mogą wydawać się nieintuicyjne, ze względu na niskie wartości b(n). Jednak po namyśle, ma to sens. Prawdopodobieństwo, że k kul wpadnie do różnych urn to $\frac{n!}{(n-k)!\cdot n^k}$ i szybko zbiega do zera.
Jak widać na wykresie, wyniki eksperymentów dla danego n nie są silnie skoncentrowane wokół średniej. Większość z nich należy do przedziału $[2,\ 2b(n)]$
Hipoteza: $b(n) = O(\sqrt{n})$

U(n) - liczba pustych urn po wrzuceniu n kul

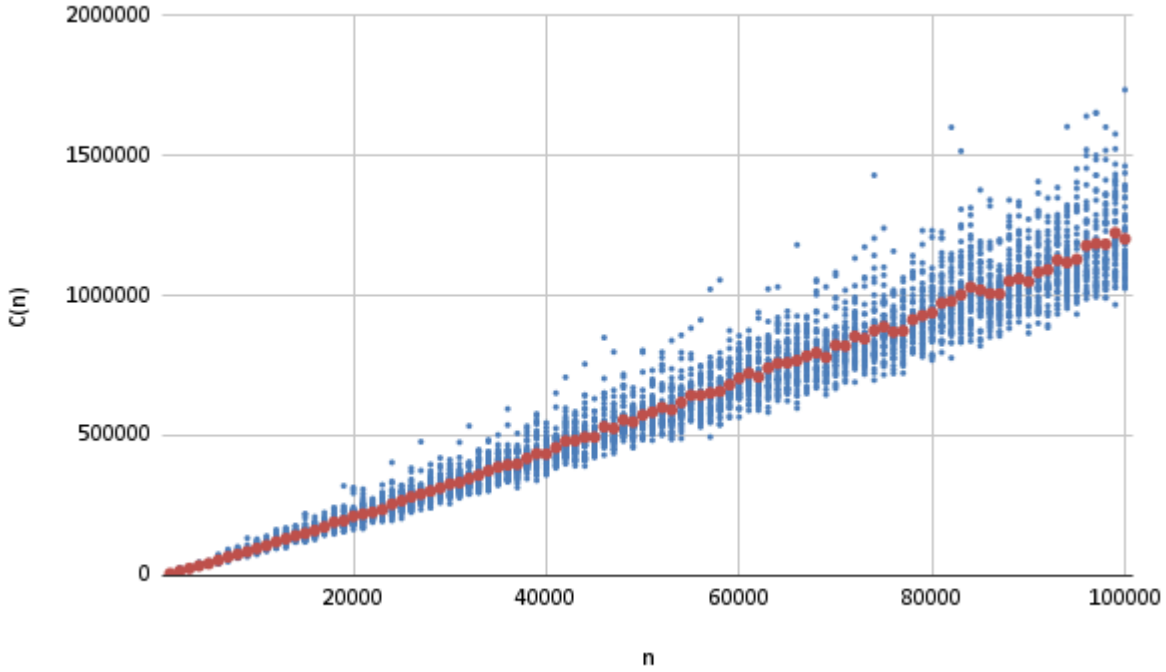


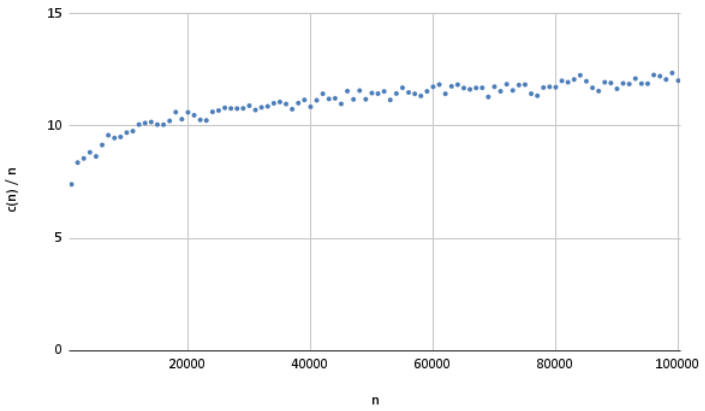


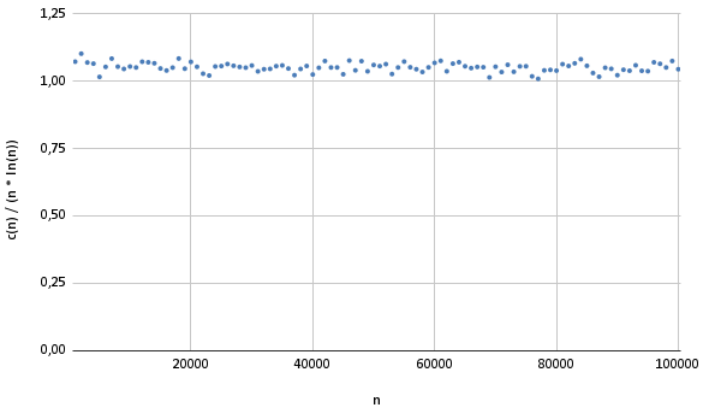
W przypadku U(n) skoncentrowanie wokół wartości średniej jest bardzo mocne. Dla każdego powtórzenia eksperymentu wyniki są prawie takie same.

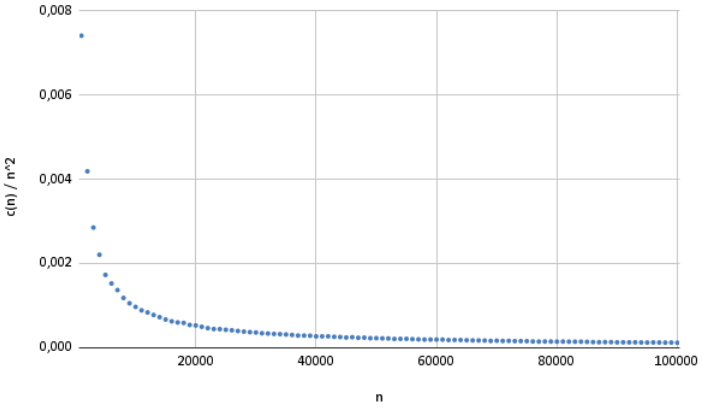
Bez wątpienia można stwierdzić, że $u(n) = O(n)$. Hipoteza: $u(n) = \frac{3}{8}n$

C(n) - minimalna liczba rzutów, po której w każdej z urn jest co najmniej jedna kula





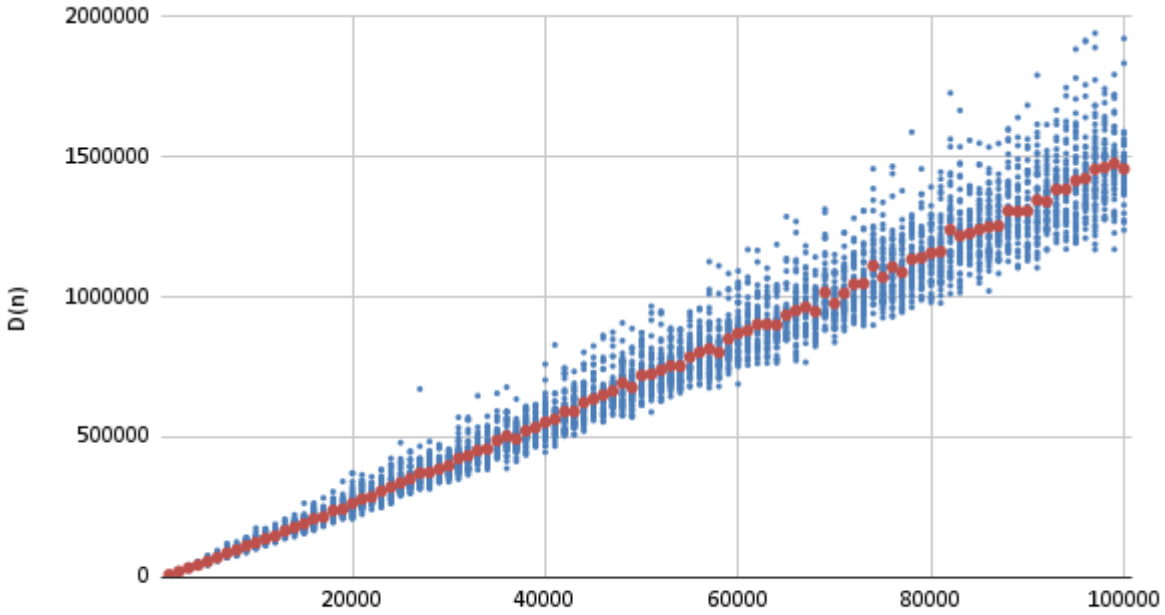


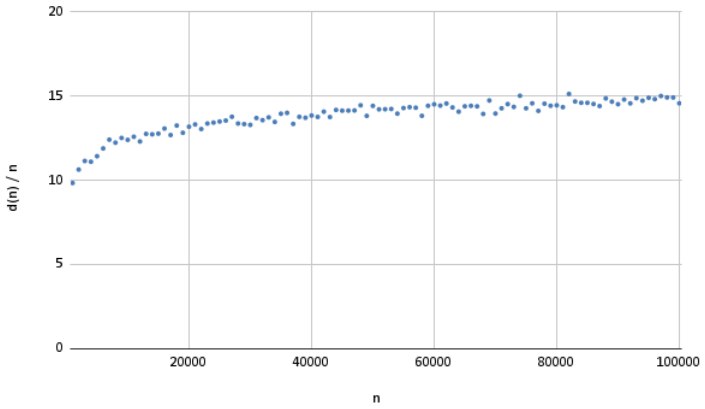


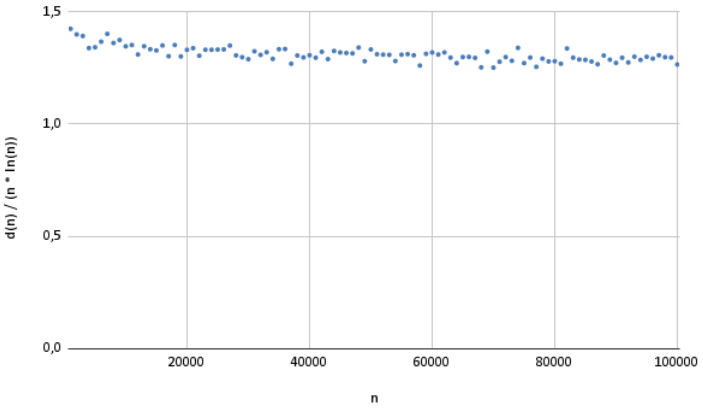
Duże wyniki eksperymentu pokrywają się z intuicją, bo dla ostatnich pustych urn prawdopodobieństwo wrzucenia do nich kuli jest bardzo małe.
Skoncentrowanie wokół średniej jest dość silne. Wyniki należą do przedziału $[0,8c(n),\ 1,2c(n)]$

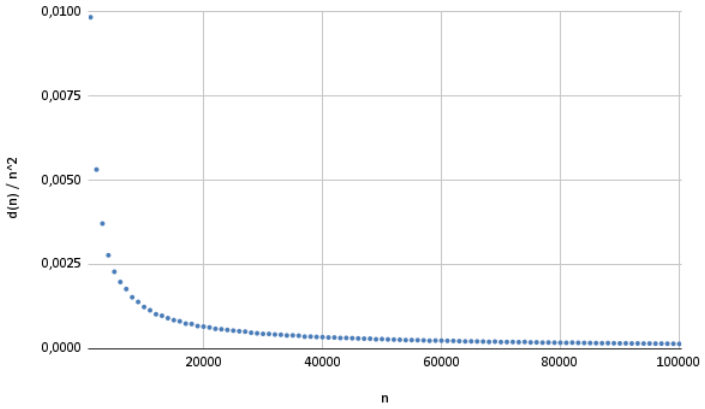
Hipoteza: $c(n) = O(n \cdot ln(n))$

D(n) - minimalna liczba rzutów, po której w każdej z urn są co najmniej dwie kule





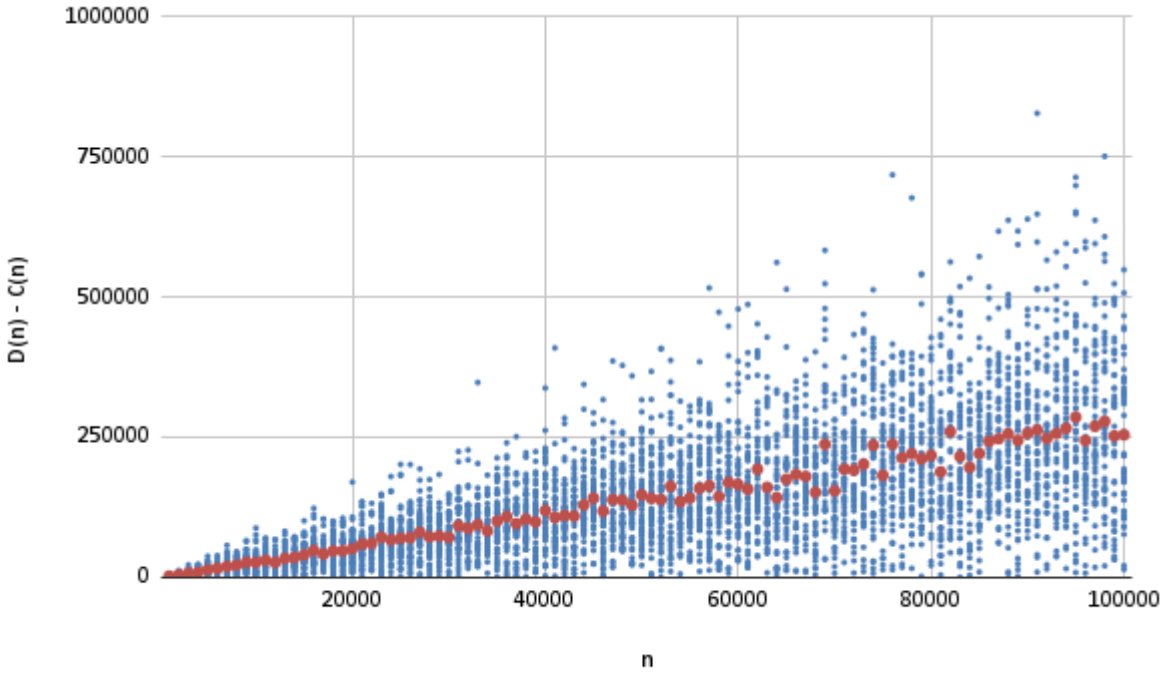


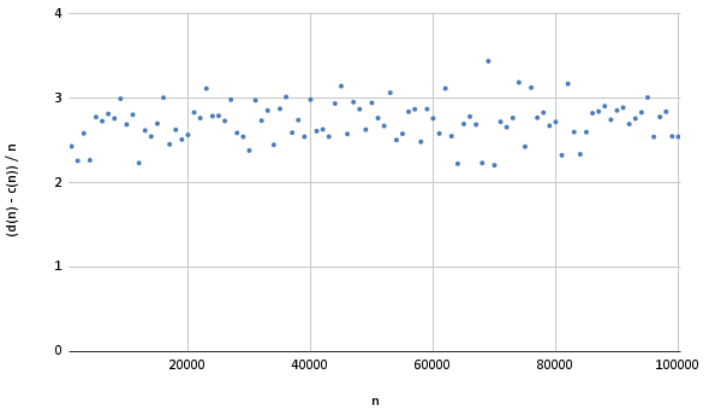


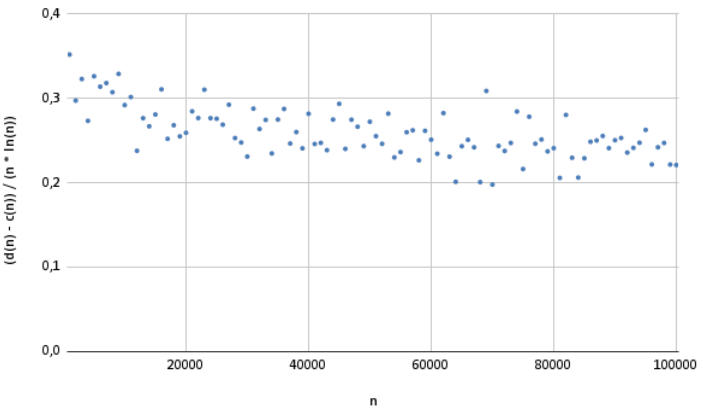
Wykres jest bardzo zbliżony do poprzedniego. Tutaj skoncentrowanie wokół średniej także jest silne. Wyniki należą do przedziału $[0,8d(n),\ 1,2d(n)]$

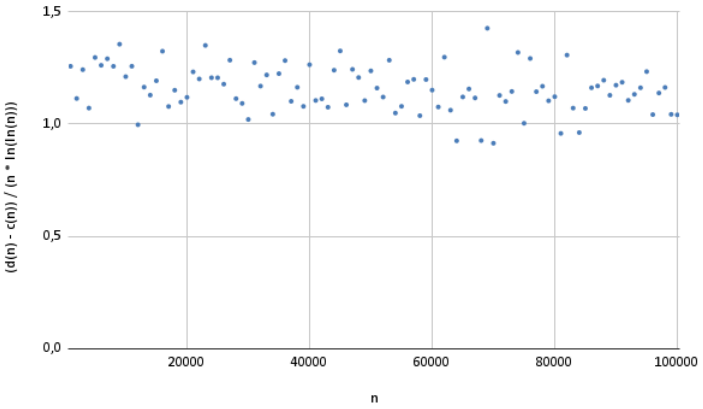
Hipoteza: $d(n) = O(n \cdot ln(n))$

D(n) - C(n) - liczba rzutów od momentu C(n) potrzeba do tego, żeby w każdej urnie były co najmniej dwie kule









To, że d(n) - c(n) << c(n) pokrywa się z intuicją. W momencie, kiedy do ostatniej pustej urny wpada kula, w wielu urnach jest już więcej niż jedna kula.
Skoncentrowanie wokół średniej jest dość słabe, a większość wyników zawiera się w przedziale $[1,\ 2(d(n)\ -\ c(n))]$
Na podstawie wykresów trudno stwierdzić asymptotykę funkcji. Opierając się na hipotezach dla c(n) oraz d(n) na pewno nie będzie to O(n).
Hipoteza: $d(n)\ -\ c(n)\ =\ O(n\ \cdot\ ln(ln(n)))$

B(n) jako birthday paradox:
Wyobraźmy sobie, że w sali jest 365 miejsc dla każdego dnia roku (nieprzestępnego).
B(365) = Ile ludzi musi wejść do pokoju, żeby zabrakło dla kogoś miejsca?
Paradoks nie jest prawdziwym paradoksem, ale objawia się tym, że odpowiedź na to pytanie jest zaskakująco mała.

C(n) jako coupon collector's problem:
Kolekcjoner zbiera karty baseballowe. Pełen zestaw składa się z n kart. Każda kolejna zebrana karta jest losową z zestawu. Powtórki nie mają znaczenia.
C(n) = Ile kart musi zebrać, aby skompletować cały zestaw?

Jakie znaczenie ma birthday paradox w kontekście funkcji haszujących i kryptograficznych funkcji haszujących?

W kontekście funkcji haszujących, birthday paradox jest istotny w analizie bezpieczeństwa kryptograficznych funkcji haszujących. Główne pytanie brzmi: jak dużo wyników musiałoby być wygenerowanych przez funkcję haszującą, zanim nastąpiłoby zdarzenie, że dwie różne dane wejściowe zostaną odwzorowane na ten sam skrót?

Dla przykładu, załóżmy, że mamy funkcję haszującą, która przekształca dowolne dane wejściowe na unikalny skrót o stałej długości. Birthday paradox sugeruje, że nawet wtedy, gdy skróty są relatywnie długie, istnieje duże prawdopodobieństwo, że dwie różne dane wejściowe wygenerują ten sam skrót.

W kryptografii istotne jest zapewnienie, aby funkcje haszujące były odporne na kolizje, czyli aby było trudno znaleźć dwie różne dane wejściowe, które generują ten sam skrót. Jeśli znajdziemy plik, który generuje taki sam skrót jak pewien plik służący do np. podpisu cyfrowego, to możemy wykorzystać go do sfałszowania tego podpisu.