Home Work 2:
balls and bins

Maksymilian Neumann

December 7, 2022

1 Wprowadzenie

Przprowadziłem symulacje jednej z klasycznych modeli kul i urn dla ilości urn $n \in \{1000, 2000, ..., 10000\}$ zbierając dane:

1. B_n = moment pierwszej kolizji
2. U_n = liczba pustych urn po wrzuceniu n kul
3. L_n = maksymalna liczba kul w urnie po wrzuceniu n kul
4. C_n = minimalna liczba rzutów, po której w kazdej z urn jest co najmniej jedna kula
5. D_n = minimalna liczba rzutów, po której w kazdej z urn są co najmniej dwie kule
6. $D_n - C_n$ = liczba rzutów od momentu C_n do momentu D_n

po 50 iteracji każdego n

2 Wyniki Zebranych Danych

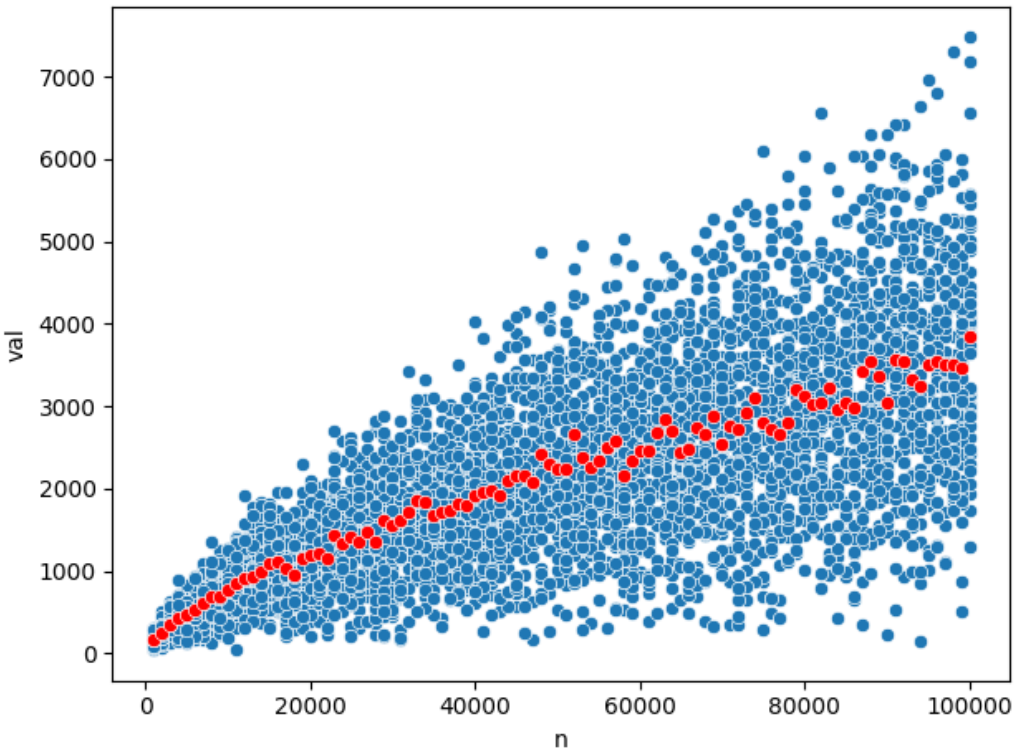


Figure 1: B_n

Zależność między n a val średnim wydaje się być w tym przypadku logarytmiczna. Natomiast koncentracja wyników wokół średnej jest coraz słabsza im n jest większe.

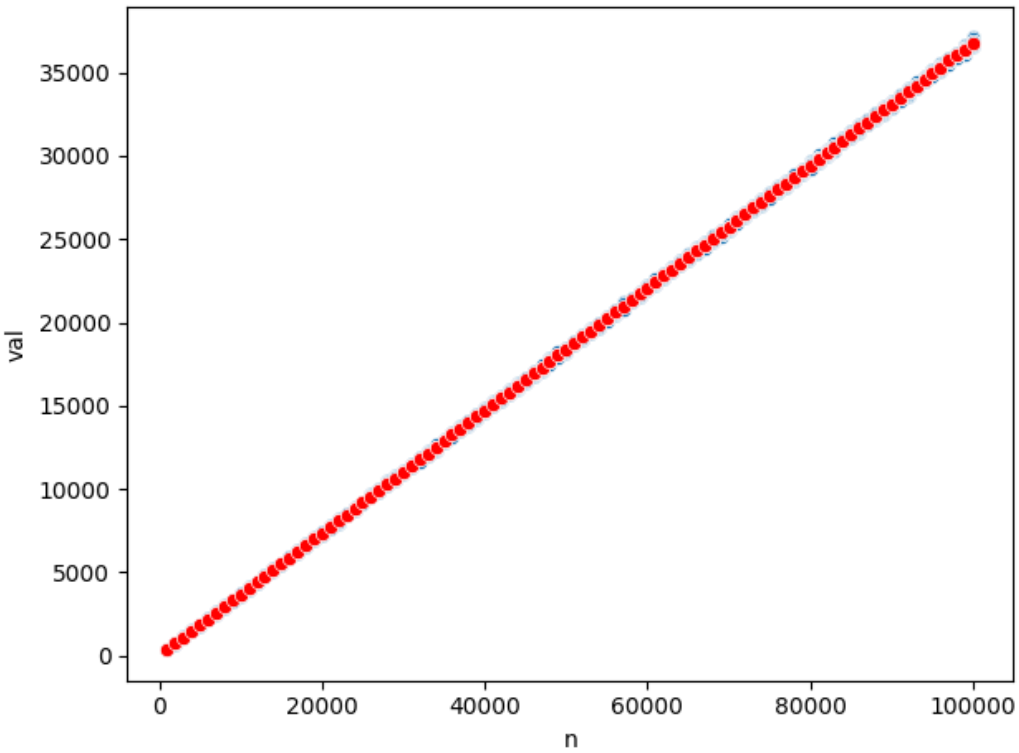


Figure 2: U_n

Można zwuważyć silną zależność liniową między n a średnim val. Koncentracja wyników wokół średniej jest mocna oraz w miare stała.,

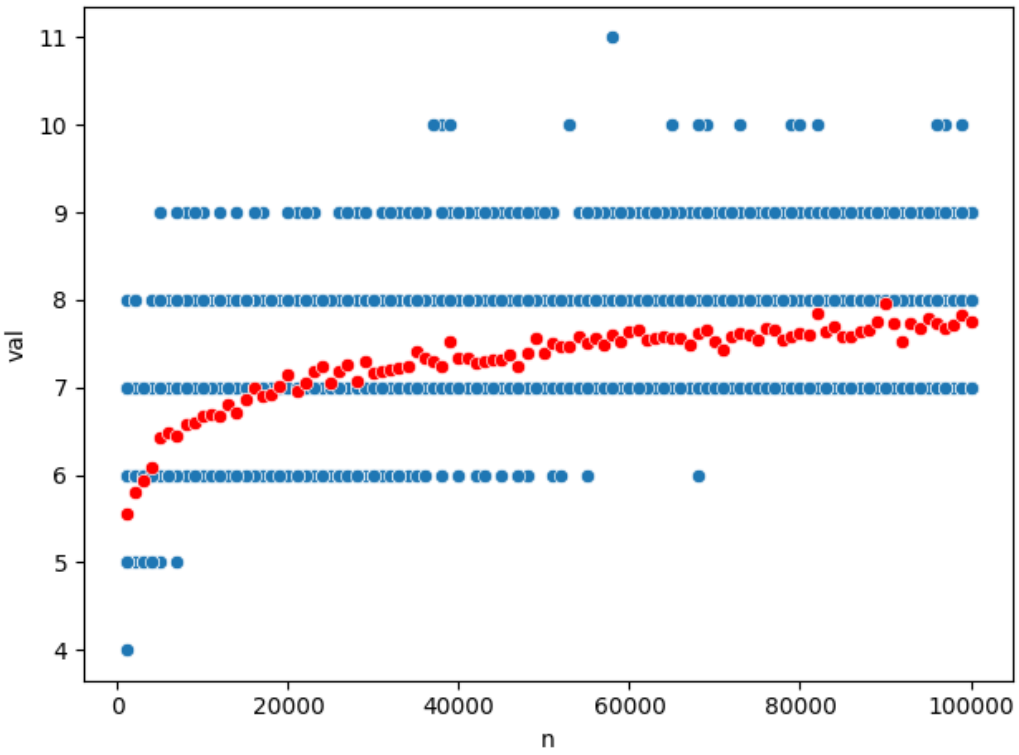


Figure 3: L_n

Zależność między n a val średnim wydaje się być w tym przypadku logarytmiczna. Poziom koncentracji wyników w okół średnej wydaje się być w miarę stały.

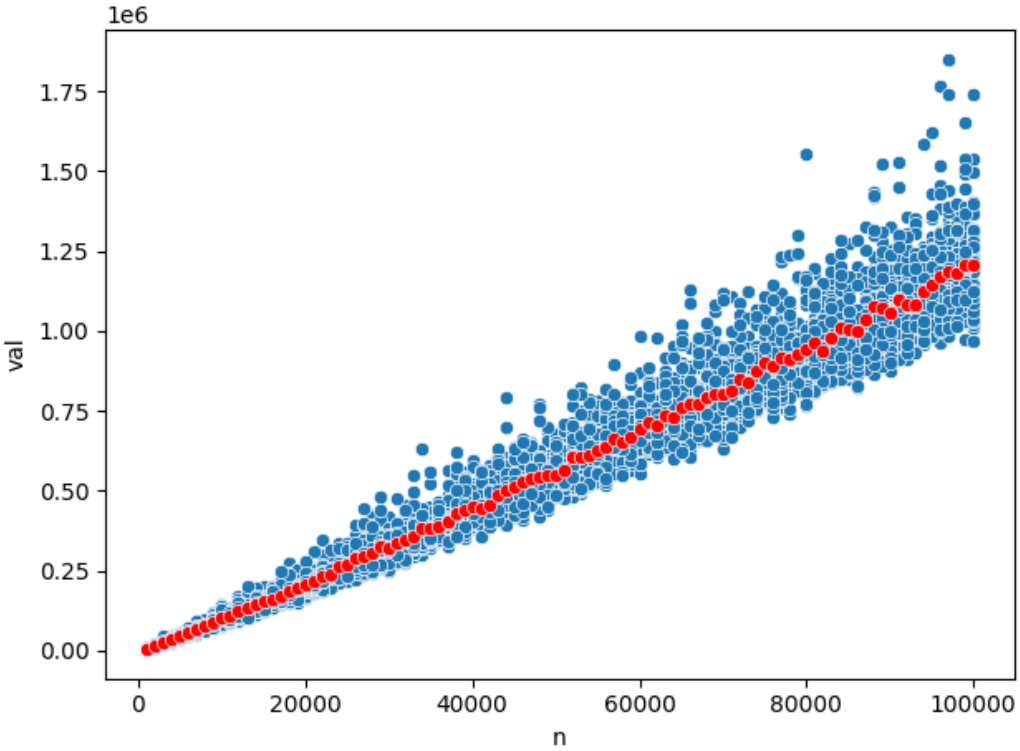


Figure 4: C_n

Widać zależność liniową miedzy n i średnim val. Podobnie jak przy statystyce B_n podczas zwiększana n wyniki oddalają się bardziej od średniej.

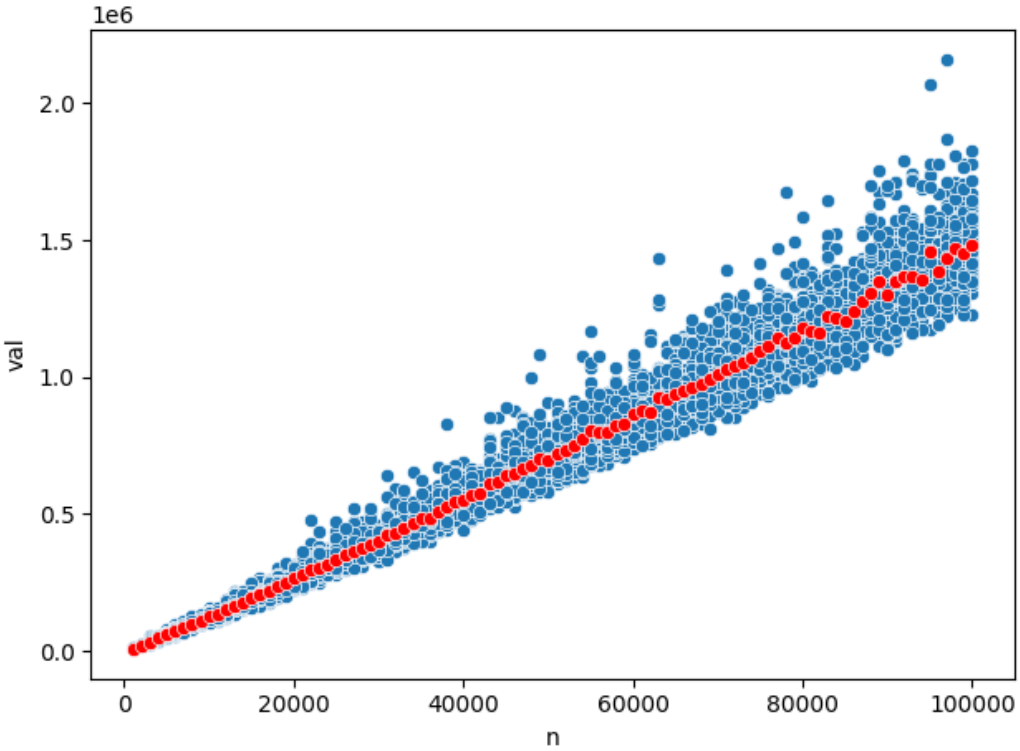


Figure 5: D_n

Widoczna jest silna liniowa zależność. Im większe n tym wieksze rozrzucenie wyników w okół średniej.

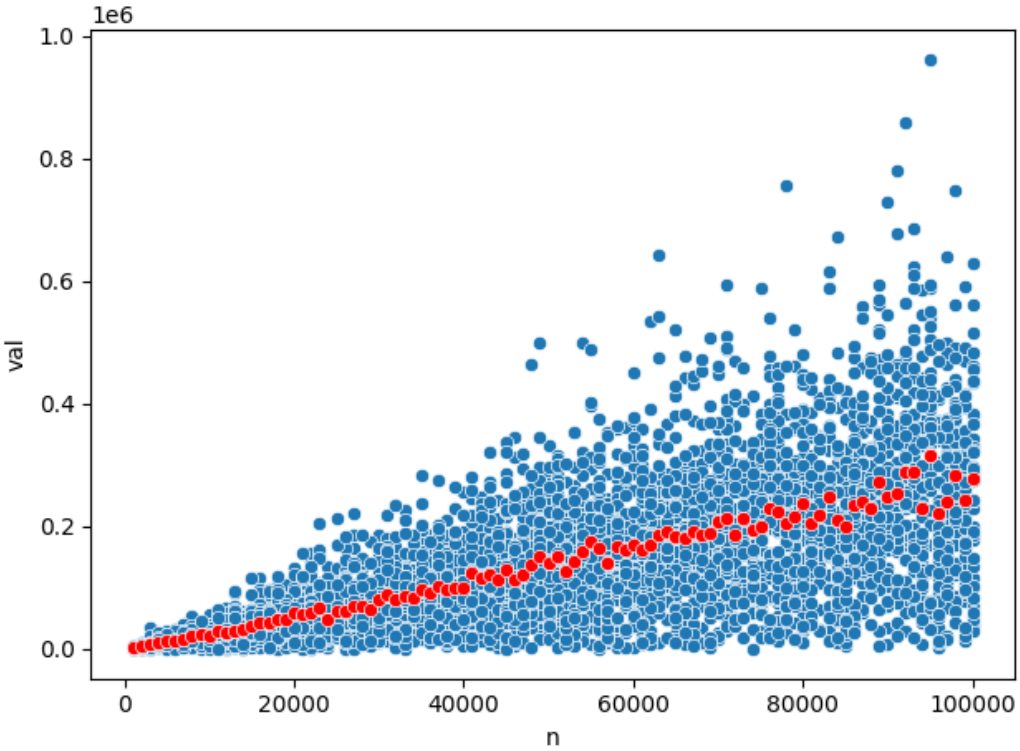
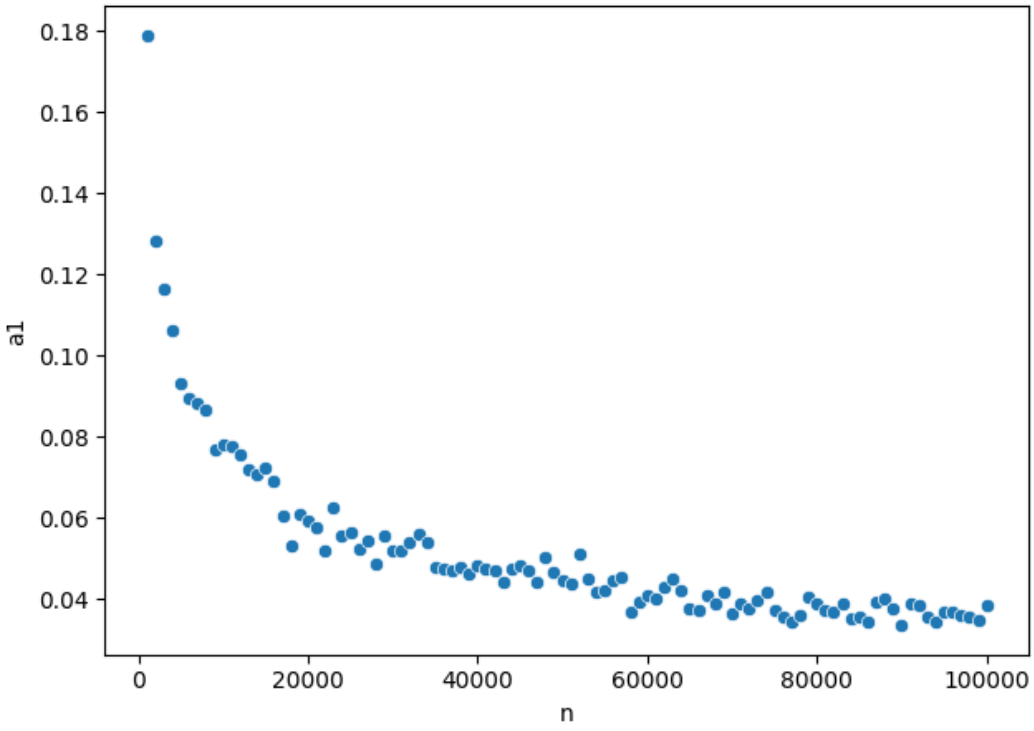


Figure 6: $D_n - C_n$

Widoczna jest silna liniowa zależność. Im większe n tym wieksze rozrzucenie wyników w okół średniej. a

3 Wykresy dodatkowe



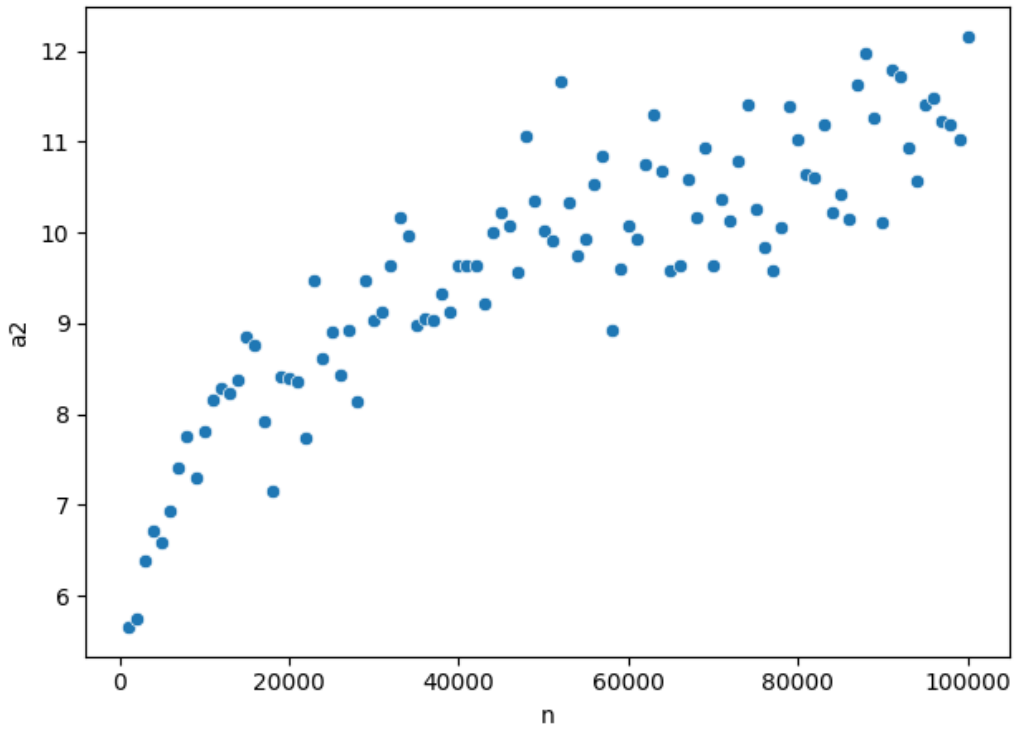


Figure 7: $\frac{b(n)}{n}$ i $\frac{b(n)}{\sqrt{n}}$

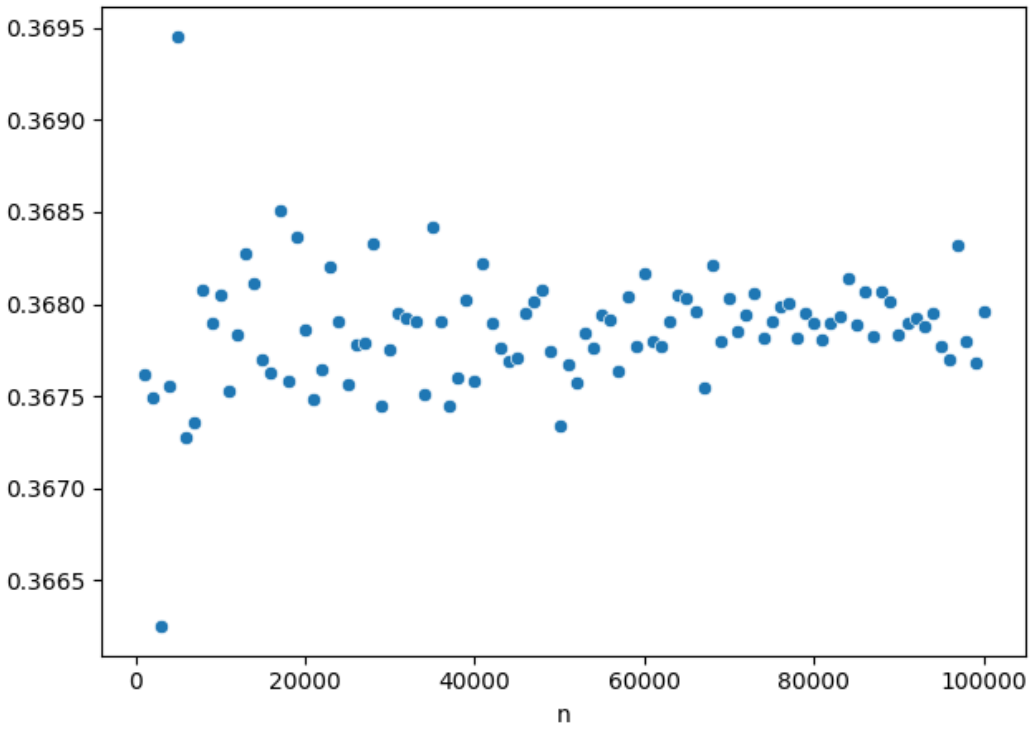
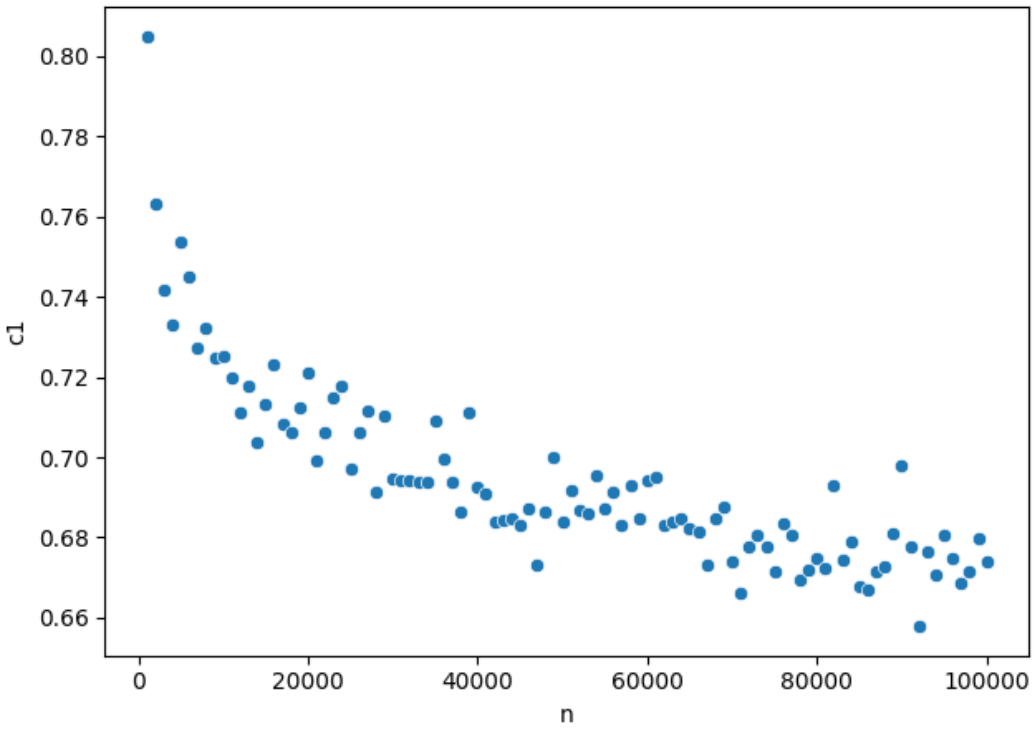


Figure 8: $\frac{u(n)}{n}$



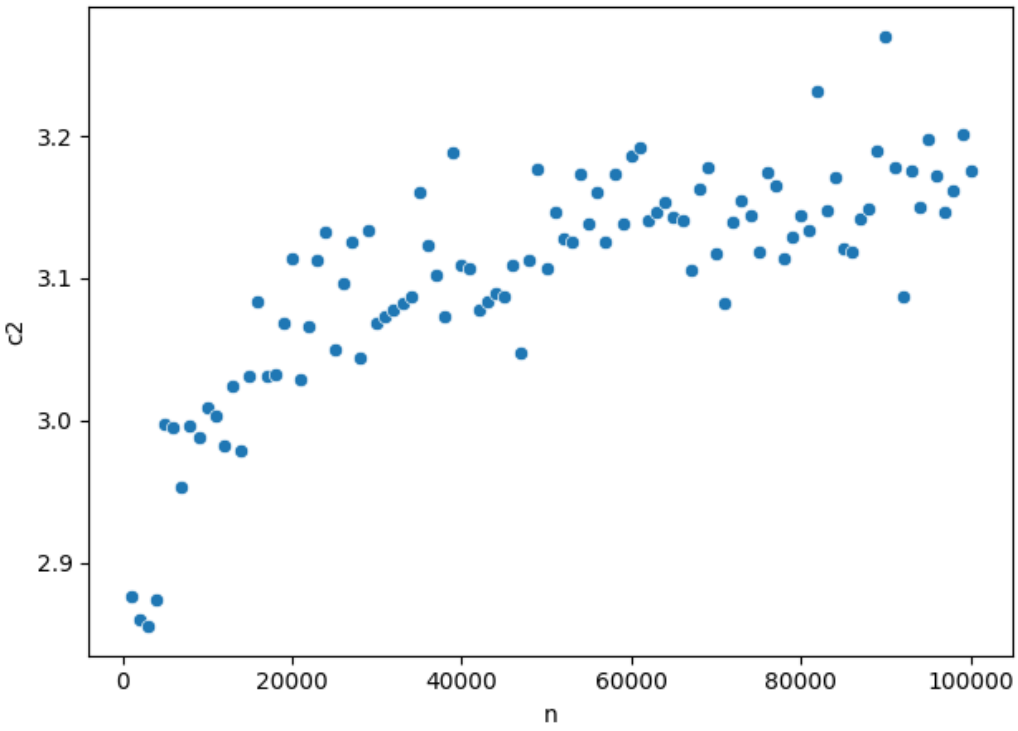


Figure 9: $\frac{l(n)}{ln(n)}$ i $\frac{l(n)}{(ln(n)/ln(ln(n)}$

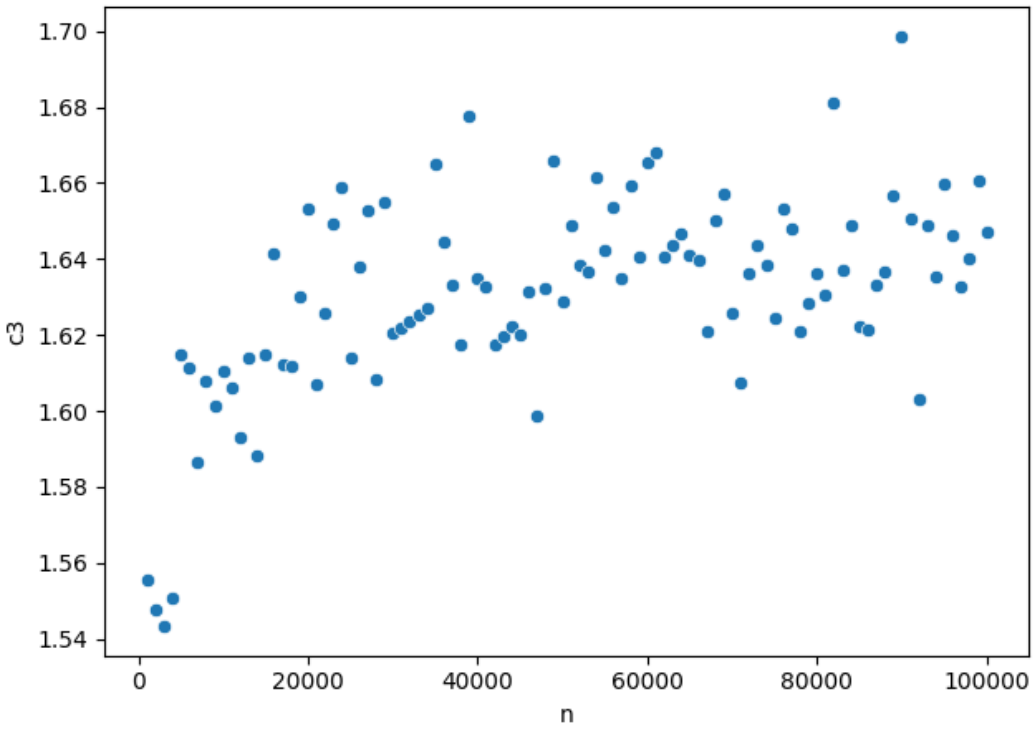
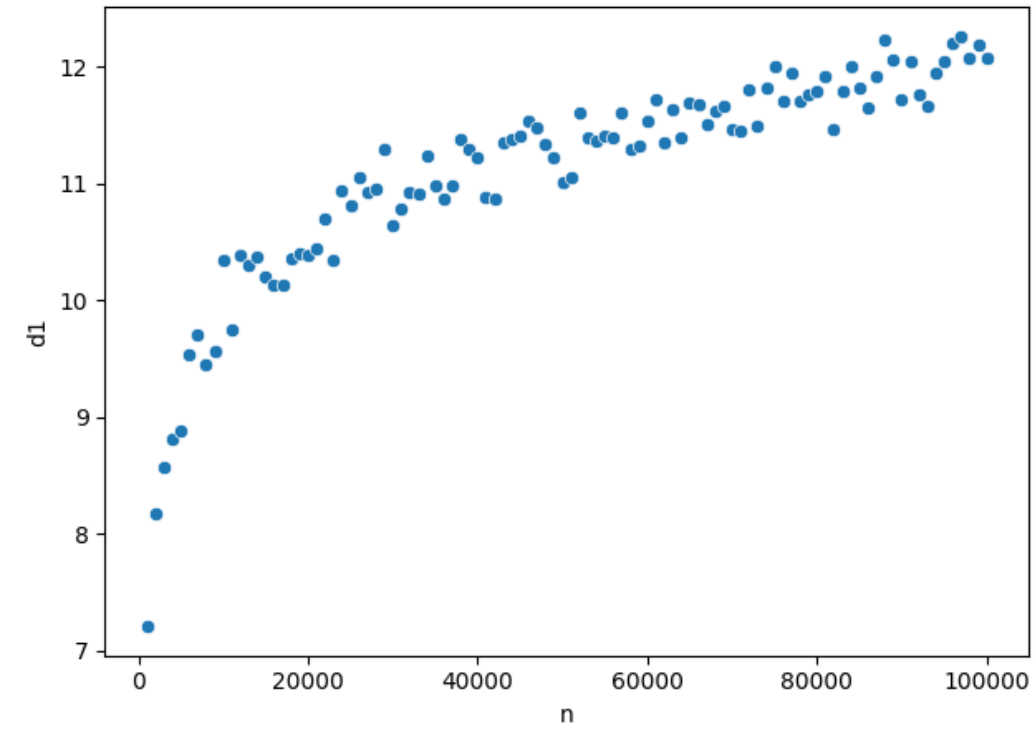


Figure 10: $\frac{l(n)}{ln(ln(n))}$



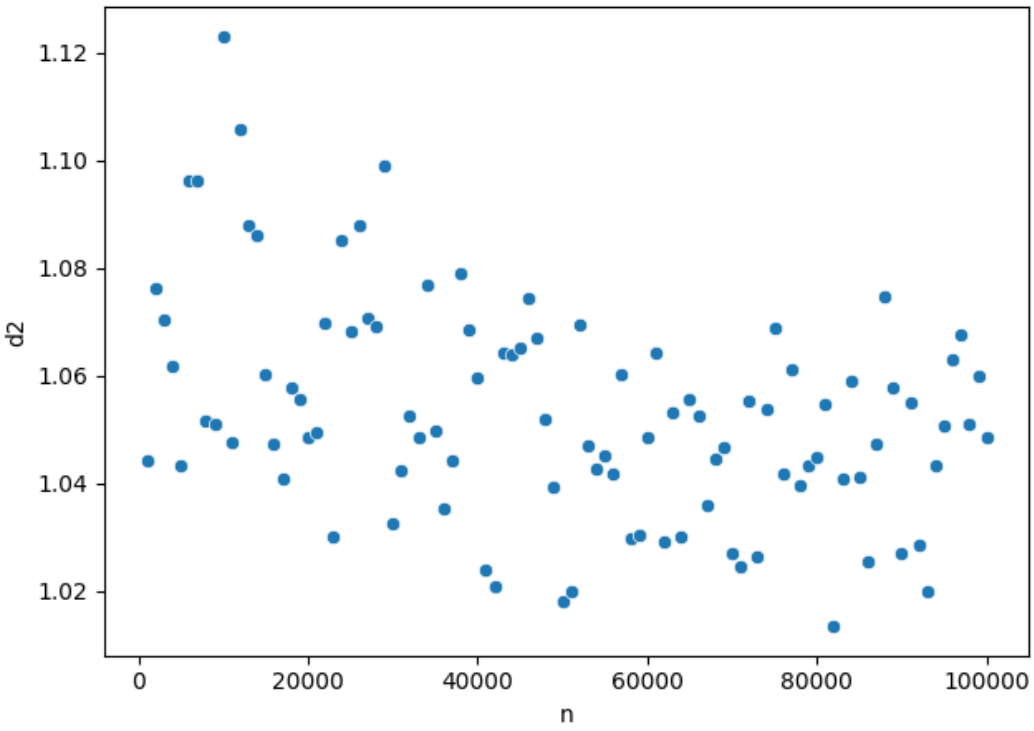


Figure 11: $\frac{c(n)}{n}$ i $\frac{c(n)}{n*ln(n)}$

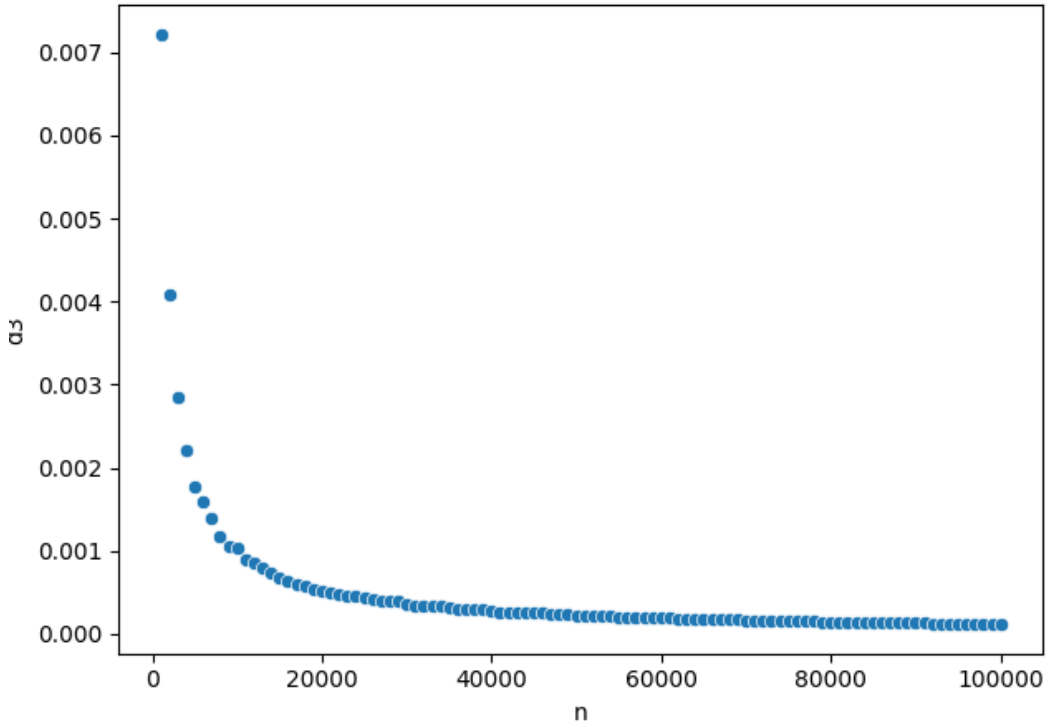


Figure 12: $\frac{c(n)}{n^2}$

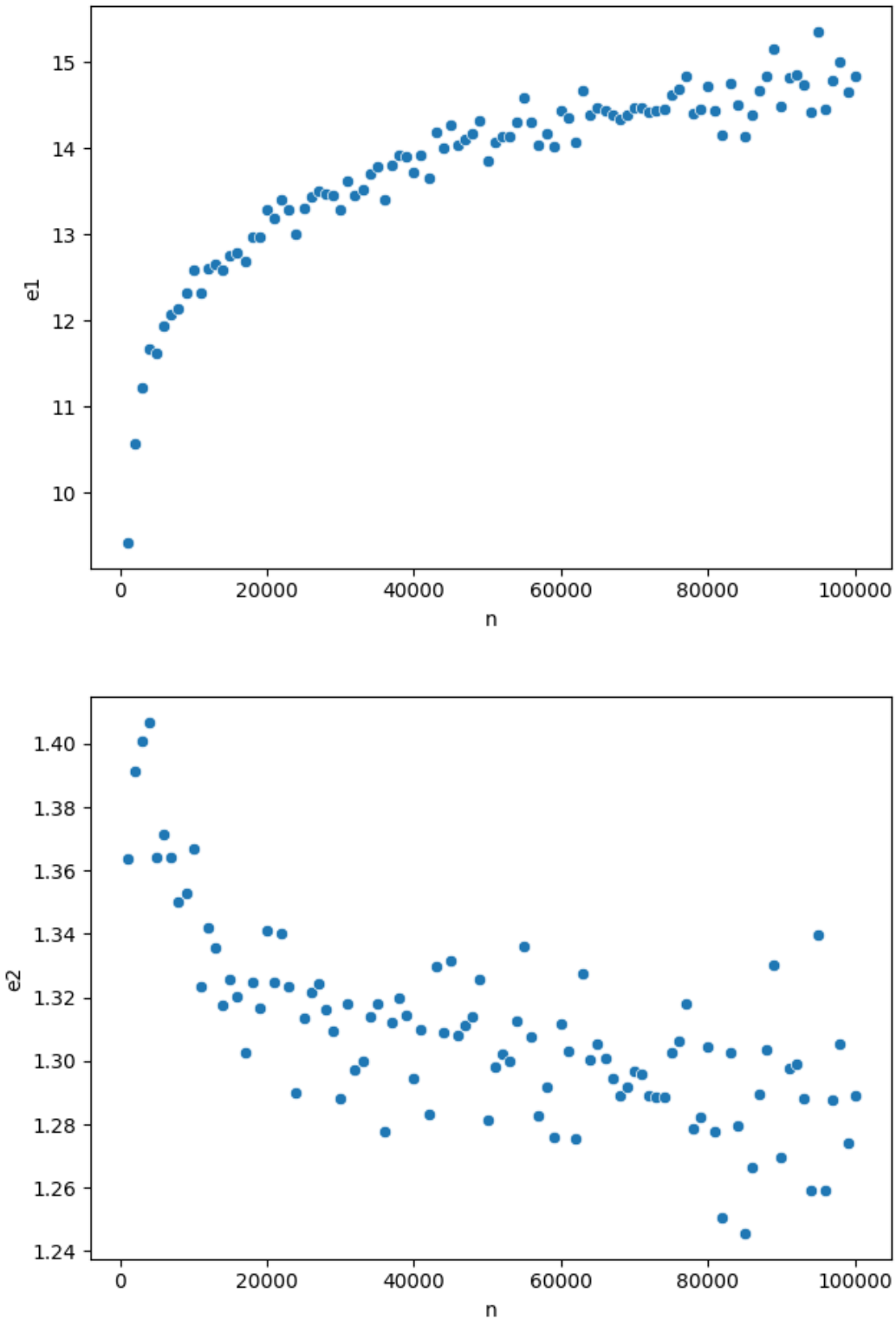


Figure 13: $\frac{d(n)}{n}$ i $\frac{d(n)}{n*ln(n)}$

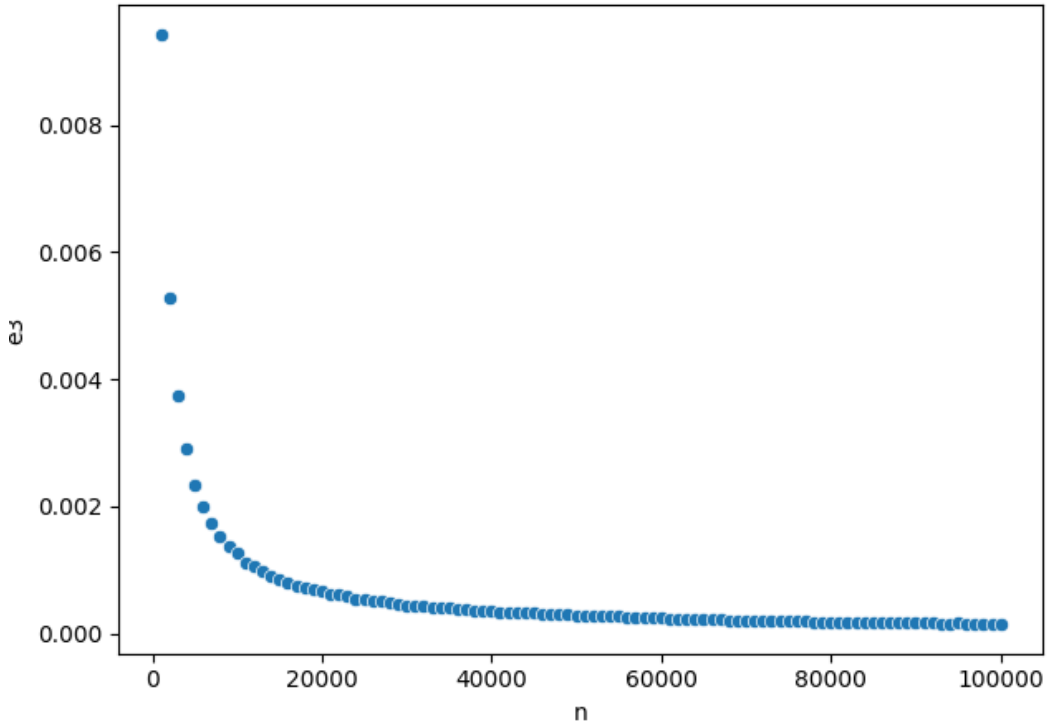
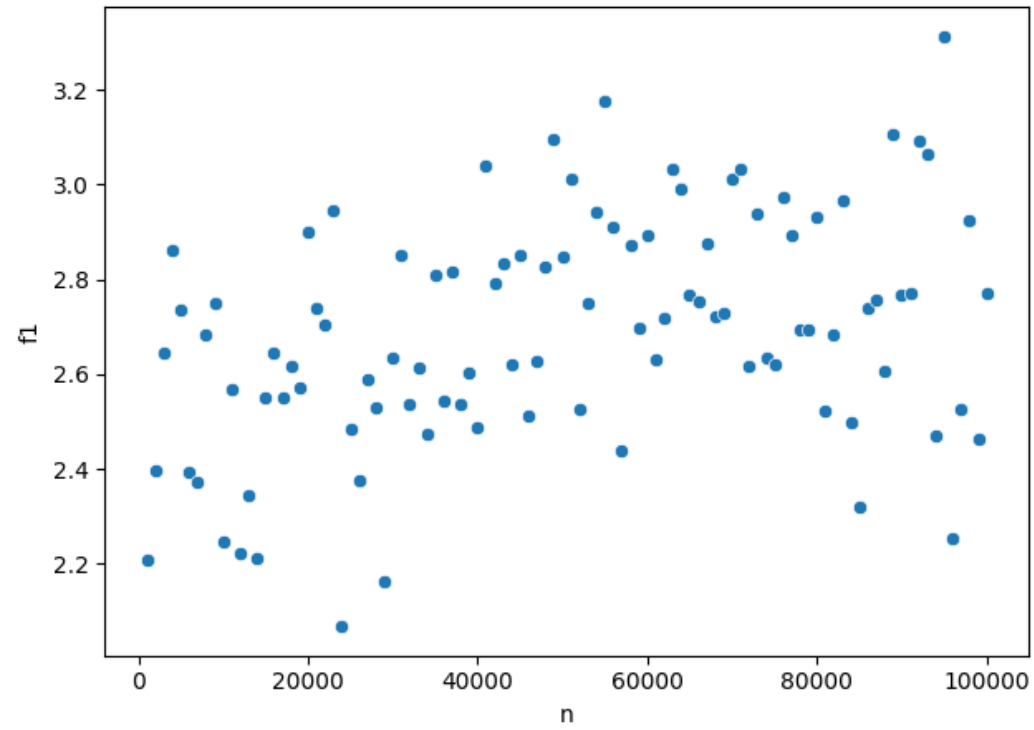


Figure 14: $\frac{c(n)}{n^2}$



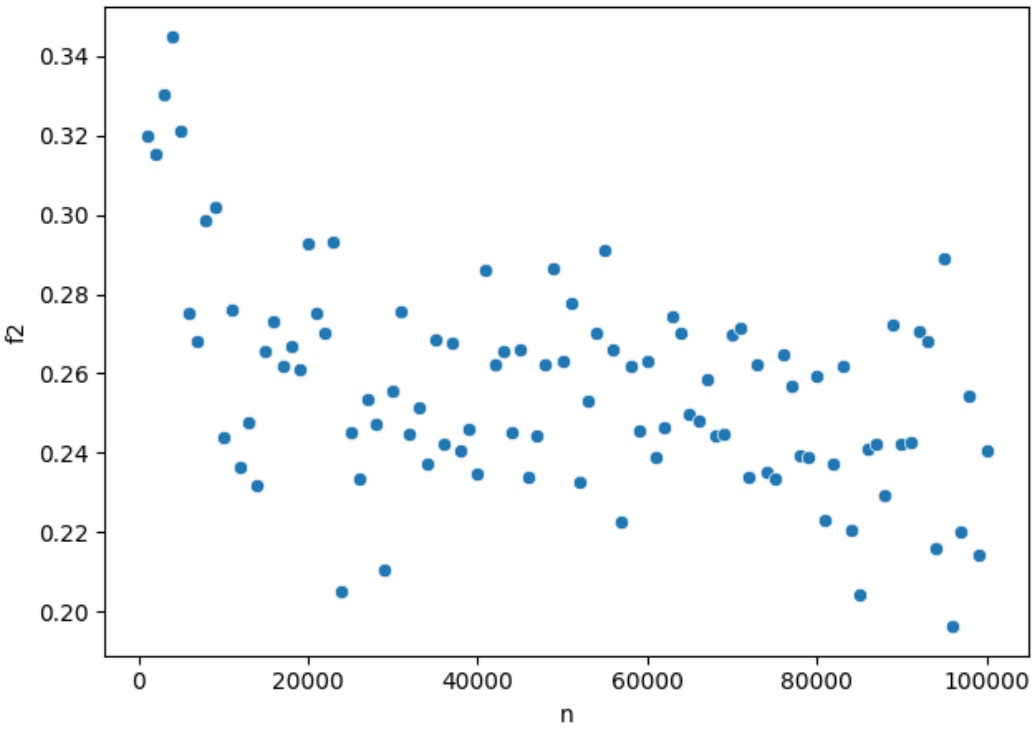


Figure 15: $\frac{d(n)-c(n)}{n}$ i $\frac{d(n)-c(n)}{n*ln(n)}$

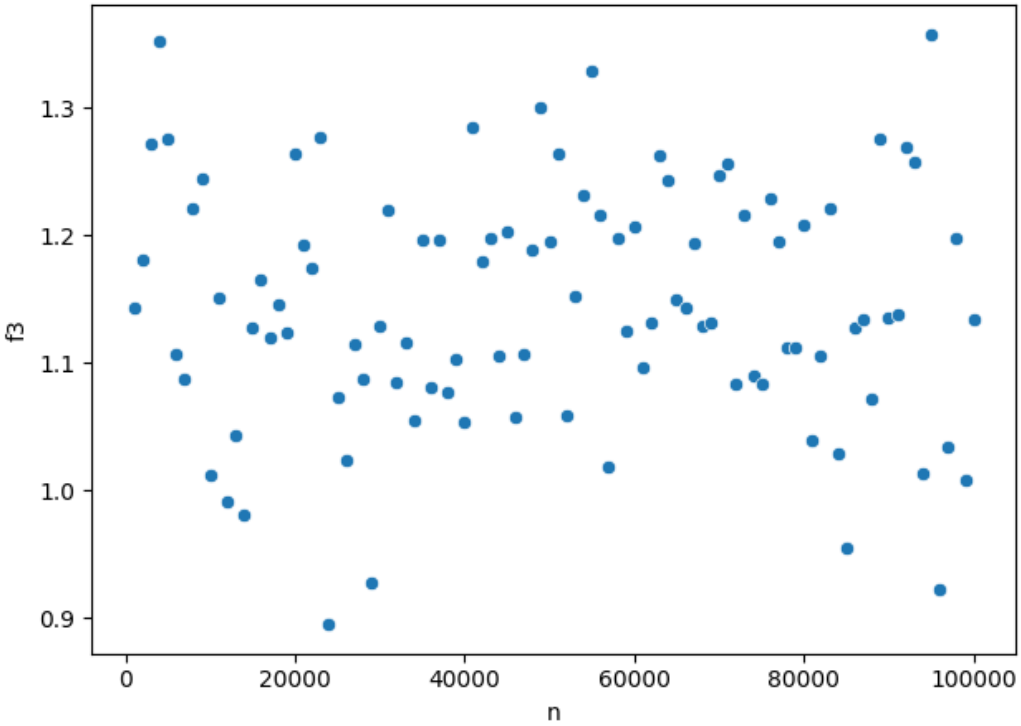


Figure 16: $\frac{d(n)-c(n)}{n*ln(ln(n))}$

4 Wnioski i Odpowiedzi

Statystyka B_n jest analogiczna do "birthday paradox" wystaczny pomyśleć tylko o urnach jako dniach w roku a kulach do nich wrzucanych jak osobach które się rodzą danego dnia. Pierwsze zdeże nie bedzie tu analogiczne do dwó0ch osob o tym samym dniu urodzin.

Statystyka C_n jest analogiczna natomiast do "Coupon collector's problem" który polega na losowaniu n rodzaji kuponów każdy z równą szansą wylosowania i zebreniu wszytkich rodzai kuponów. W naszym przykładzie urna będzie rodzajem kuponu a kula wrzucona do urny bedzie oznaczać wylosowanie tego rodzaju kuponu więć kiedy w każdej urnie będziemy mieli conajmniej 1 kulę wylosujemy. wszystkie rodzaje kuponów.

W kryptografii lekcje nauczone od "birthday paradox" wykożystywane są jako aktak kryptograficznny tak zwany "Birthday attack" który polega na znalezieniu kolizji w hashu. W porównaniu do "Preimage attack" które dla n bitowego hashu ma $O(2^n)$ "Birthday attack" ma $O(\sqrt{2^n})$ oraz jes szeroko przypuszczne że komputer kwantowy mogłby go wykonać z $O(\sqrt[3]{2^n})$