MPiS - Zadanie Domowe 2

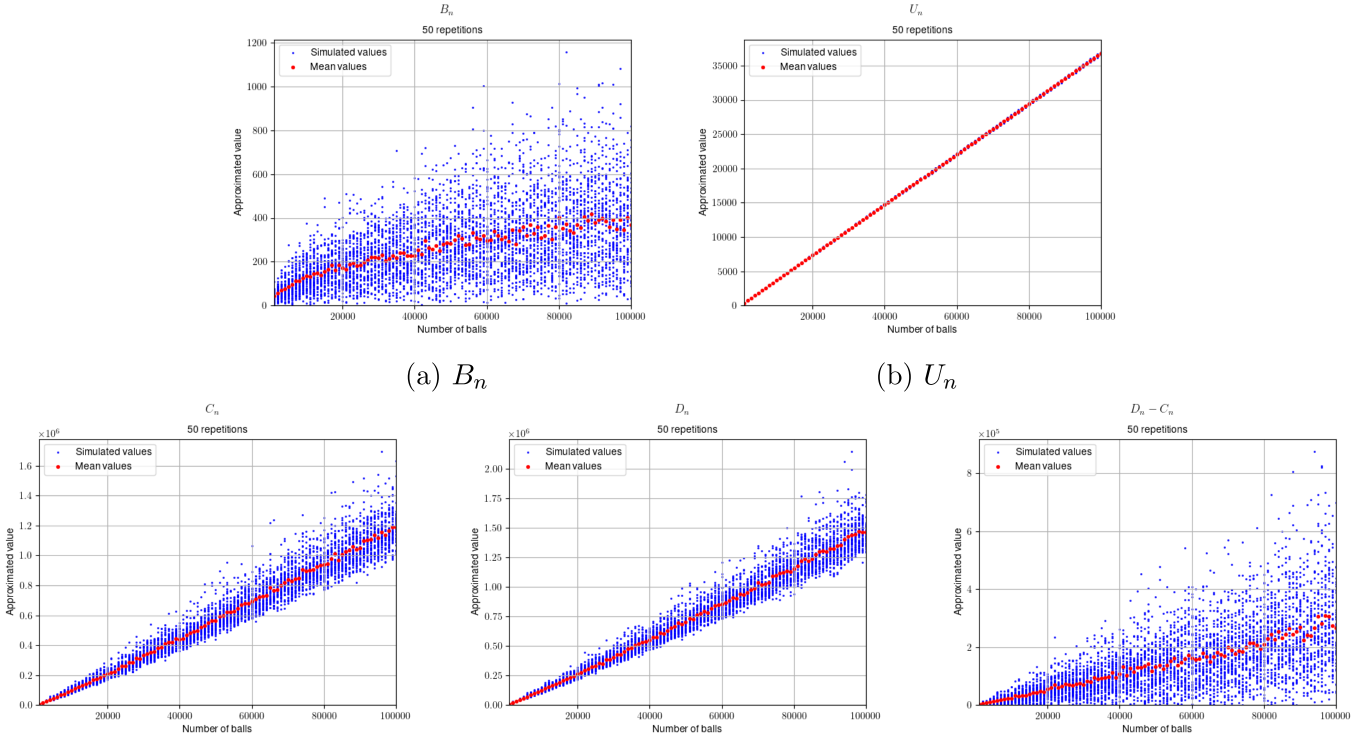
Michał Waluś

27 listopada 2024

Implementacja składa się z pliku main.cpp, który jest odpowiedzialny za model kul i urn oraz pliku main.py, w którym generowane są wykresy.

Zadanie 2

a) Wykresy uzyskanych rezultatów:

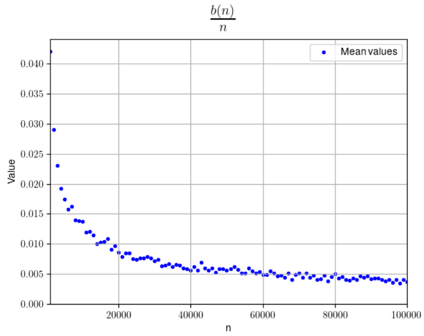


(a) B_n (b) U_n (c) C_n (d) D_n (e) $D_n - C_n$

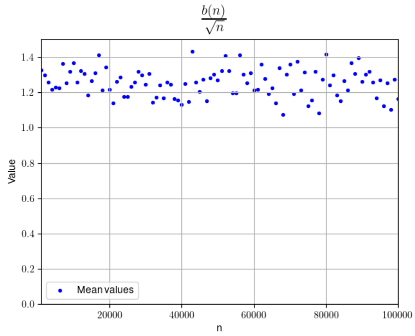
Rysunek 1: Wykresy wielkości wyznaczonych za pomocą symulacji

Analizując powyższe wykresy, wyraźnie poszczególne momenty pierwszej kolizji (B_N) są najbardziej oddalone od średniej, co jest zgodne z intuicją w przypadku wielkości, która po pierwszej kuli zależy tylko i wyłącznie od jednej 'szczęśliwej' kuli.
Wykres liczby pustych urn (U_n) z kolei przedstawia linię niemalże prostą, a poszczególne wyniki są prawie niewidoczne na wykresie. Liczba pustych urn po n rzutach, wynosi około 37%.
Minimalne liczby rzutów by wszystkie urny były niepuste (C_n) oraz by zawierały co najmniej dwie, są dość zbliżone, a różnica między nimi jest bliższa 25% niż 100%, co jest spowodowane tym, że w momencie osiągnięcia pierwszej sytuacji wiele urn zawiera co najmniej dwie kule. Skutkuje to ich różnicą będącą wyraźnie mniejszą, mającą poszczególne wyniki relatywnie dalej od średniej, co wynika także z potencjalnego sumowania się oddalenia od wartości oczekiwanej tych dwóch funkcji.

Wykresy funkcji średnich wartości:



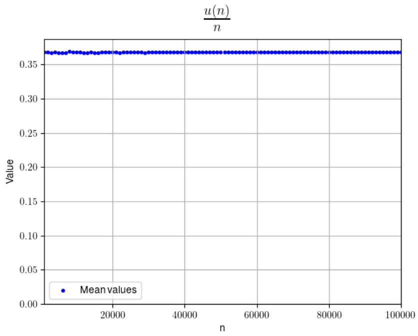
(a) $\frac{b(n)}{n}$



(b) $\frac{b(n)}{\sqrt{n}}$

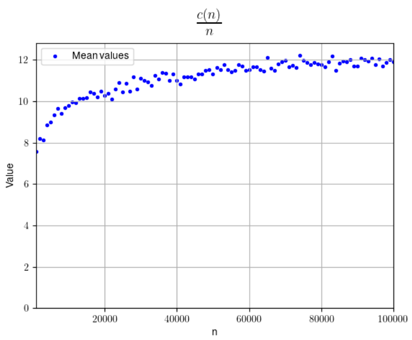
Rysunek 2: Funkcje powiązane z B_n

Powyższe wykresy wyraźnie prezentują dużą tendencję do zmian funkcji B_n.

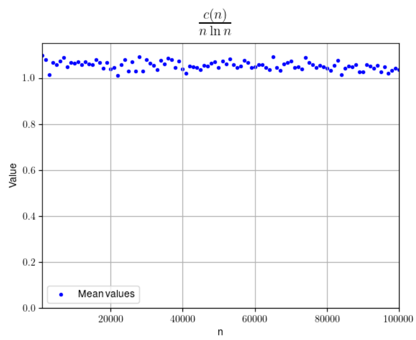


Rysunek 3: Funkcja powiązana z U_n: $\frac{u(n)}{n}$

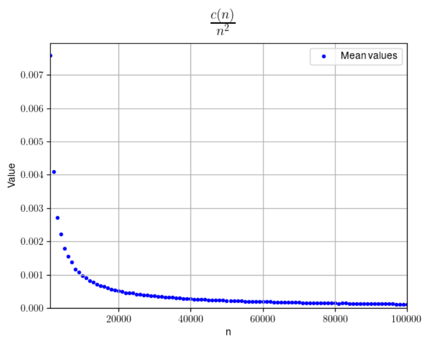
Wykres przedstawiony powyżej przypomina funkcje stałą, pokazując, że po wrzuceniu n kul, najczęściej około 73% urn zostanie zapełnione.



(a) $\frac{c(n)}{n}$



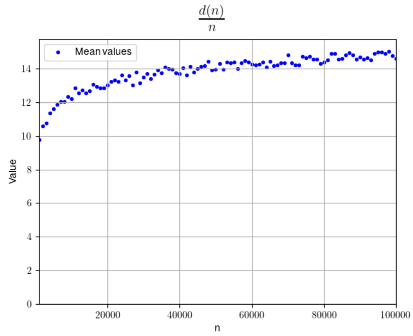
(b) $\frac{c(n)}{n \ln n}$



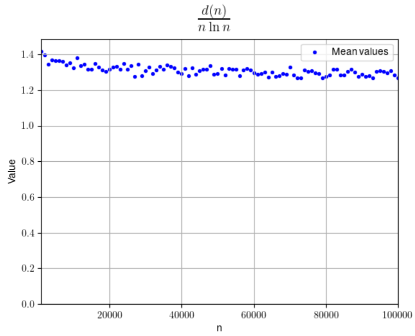
(c) $\frac{c(n)}{n^2}$

Rysunek 4: Funkcje powiązane z C_n

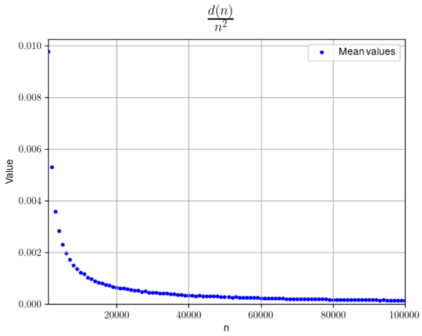
Wykresy zarówno średnich C_n i D_n uwidoczniają ich niemalże takie samo tempo wzrostu oraz pokazują, że mimo relatywnie szybkiego wzrostu rosną one wolniej od n^2.



(a) $\frac{d(n)}{n}$

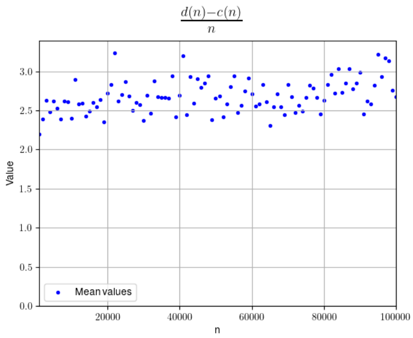


(b) $\frac{d(n)}{n \ln n}$

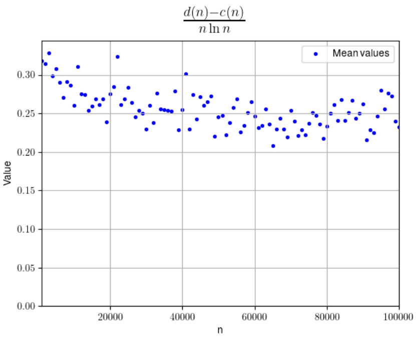


(c) $\frac{d(n)}{n^2}$

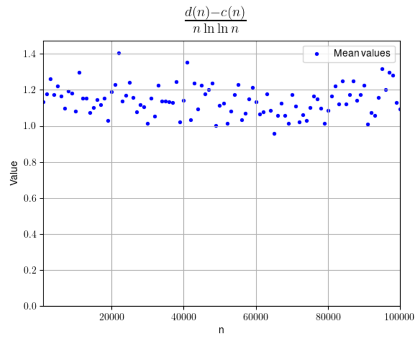
Rysunek 5: Funkcje powiązane z D_n



(a) $\frac{d(n)-c(n)}{n}$



(b) $\frac{d(n)-c(n)}{n \ln n}$



(c) $\frac{d(n)-c(n)}{n \ln \ln n}$

Rysunek 6: Funkcje powiązane z $D_n - C_n$

Z powyższych wykresów możemy wywnioskować, że asymptotyka różnicy C_n i D_n jest mniejsza od asymptotyki tych dwóch wielkości.

b) Charakteryzacja koncentracji wyników

Wyniki U_n są bardzo zbliżone do wartości średniej, do tego stopnia, że ciężko je dostrzec na wykresie. Wielkości C_n i D_n są dość zliżone do wartości średnich, jednak z wzrostem n co raz bardziej od nich się oddalają, a różnica tych dwóch wielkości, dla niektórych przypadków jest dwukrotnie większa od wartości średniej. Poszczególne pomiary B_n z kolei często są bardzo oddalone od wartości średniej, niekiedy będąc trzykrotnie większe lub ponad dziesiąciokrotnie mniejsze.

c) Asymptotyka

B_n - Wzorując się na wykresach asymptotyka jest zdecydowanie mniejsza od liniowej i wielkość zdaje się mieć asymptotykę $\Theta(\sqrt{n})$, ponieważ mimo stosunkowo dużej różnicy w wartościach, na wykresie (b), punkty zdają się być skupione około $y = 1.25$.
U_n - Wyniki podzielone przez n są bardzo zbliżone do stałej i silnie sugerują asymptotykę $\Theta(n)$.
C_n i D_n - Wykresy obu funkcji wyglądają bardzo podobnie i wykres $\frac{c(n)}{n}$ (oraz $\frac{d(n)}{n}$) sugeruje tempto wzrostu szybsze od liniowego, lecz zdecydowanie jest ono mniejsze od $\Theta(n^2)$ i wygląda jak $\Theta(n \ln n)$.
$D_n - C_n$ - Mimo relatywnie dużych różnic w wartościach, funkcja ilorazu średniej z $n \ln n$ ma tendencję malejącą i skupienie punktów na wykresie (c) sugeruje asymptotykę $\Theta(n \ln \ln n)$.

d) Intuicje za nazwami

Birthday paradox - Podejrzewam, że nazwa pochodzi od rozważań nad prawdopodobieństwem znajdowania się dwóch osób z urodzinami tego samego dnia (tygodnia, miesiąca, roku) w danej grupie osób. Na przykład z Zasady Szufladkowej Dirichleta wiemy, że w ośmio-osobowej grupie co najmniej dwie osoby mają urodziny tego samego dnia tygodnia, a wielkość B_n pomaga nam oszacować, kiedy możemy oczekiwać, że w danej grupie, dwie osoby będą miały urodziny tego samego dnia.

Coupon collector's problem - Ponieważ opisuje to przypadek gdy żadna urna nie jest pusta, podejrzewam, że przedstawia to problem zebrania pełnej kolekcji kuponów, kiedy każdy z nich możemy dostać z równym pradopodobieństwem i nie wiemy na jaki trafiliśmy dopóki go nie dostaniemy. Opisuje to ile średnio kuponów taki kolekcjoner musi zdobyć by uzyskać pełną kolekcje.

e) Birthday paradox w kontekście funkcji hashujących

Funkcje hashujące polegają na przypisaniu wiadomości ciągu znaków (hash) i w więkości przypadków chcemy uzyskać jak najmniej kolizji między hashami. Jest określona liczba hashy i z zewnątrz są one przypisywane pseudo-losowo, więc birthday paradox może pomóc nam obliczyć prawdopodobieństwo powtórzenia się hasha i jest on nazywany paradoksem, ponieważ liczba prób do znalezienia kolizji jest znacznie niższa niż oczekiwana.