# Raport z Symulacji Urn

Jan Ryszkiewicz

## Opis wyników

W ramach zadania przeprowadziłem symulację wrzucana kul do n urn po 50 razy śledząc następujące informacje:

- $B$ moment pierwszej kolizji,
- $C$ minimalna liczba rzutów, po której w kazdej z urn jest co najmniej jedna kula,
- $D$ minimalna liczba rzutów, po której w kazdej z urn są co najmniej dwie kule,
- $U$ liczba pustych urn po wrzuceniu n kul.
- $D-C$ liczba rzutów od momentu Cn potrzebna do tego, zeby w każdej urnie były conajmniej dwie kule.

Na wykresach (pliki .png w folderze /plots) przedstawiam wyniki pojedynczych symulacji oraz ich średnie wartości dla danych n.

## Wnioski na temat wartości śledzonych danych

- $B$ warto zauważyć, że pierwsza kolizja ma mniejsce bardzo szybko
- $C$ bardzo duże wartości w stosunku do n, odchylenie rośnie wraz z n
- $D$ podobnie do C(n)
- $U$ rośnie stale liniowo, odhylenie od średniej jest minimalne
- $D-C$ bardzo chaotyczne wyniki, średnia wizualizuje ogólną tendencję

## Wnioski na temat asymptotyki

- $\frac{b(n)}{n}$ - zbiega do 0
- $\frac{b(n)}{\sqrt{n}}$ - wydaje się oscylować wokół 1 czyli te funkcje uzyskują podobne wartości
- $\frac{u(n)}{n}$ - między 0.365 a 0.370, tak samo jak poprzednia, wydaje się stałą więc asymptotycznie funkcje są równe
- $\frac{c(n)}{n}$ - wydaje się rosnąć, kształt przypomina $\sqrt{x}$, możliwe że c(n) jest podobna do n$\sqrt{n}$
- $\frac{c(n)}{n^2}$ - zbiega do 0
- $\frac{c(n)}{n \ln(n)}$ - oscyluje nieco powyżej 1 co sugeruje asymptotyczną równość
- $\frac{d(n)}{n}$ - wydaje się rosnąć, kształt przypomina $\sqrt{x}$, możliwe że d(n) jest podobna do n$\sqrt{n}$
- $\frac{d(n)}{n^2}$ - zbiega do 0
- $\frac{d(n)}{n \ln(n)}$ - wydaje się bardzo powolnie zbiegać do 0, ciężko określić

- $\frac{d(n)-c(n)}{n}$ - oscyluje nieco powyżej 2 wydaje się stałą więc prawdopodobnie asymptotycznie funkcje są równe
- $\frac{d(n)-c(n)}{n\ln(n)}$ - wydaje się bardzo powolnie zbiegać do 0, ciężko określić
- $\frac{d(n)-c(n)}{n\ln\ln(n)}$ - funkcje są sobie równe, iloraz jest asymptotycznie $= 1$

# *Birthday paradox* oraz *coupon collector's problem*

## 0.1 Użycie nazw

- *Paradoks urodzinowy* - jak szybko dojdziemy do kolizji (ten sam dzień urodzin) w pewnej grupie ludzi jest dobrym obrazem prezentowanych danych, szansa na kolizję rośnie znacznie szybciej niż mogłoby się wydawać stąd nazwa "paradoks"
- *Problem zbieracza kuponów* - załóżmy że mamy do zebrania 10 kuponów i w każdej paczce jest losowy kupon, jeżeli chcemy zebrać wszystkie kupony to musimy średnio otworzyć własnie taką liczbę paczek, stąd nazwa.

## 0.2 *Birthday paradox* a funkcje hashujące

"Birthday paradox" pokazuje jak łatwo może dojść do kolizji w funkcji hashującej. Sam paradoks mówi że jeśli w pokoju znajdują się 23 osoby to szansa na to że mają urodziny tego samego dnia wynosi $>$50% co pokazuje jak szybko wbrew intuicji szansa na kolizję (tutaj daty urodzin) wzrasta.

**Jan Ryszkiewicz**