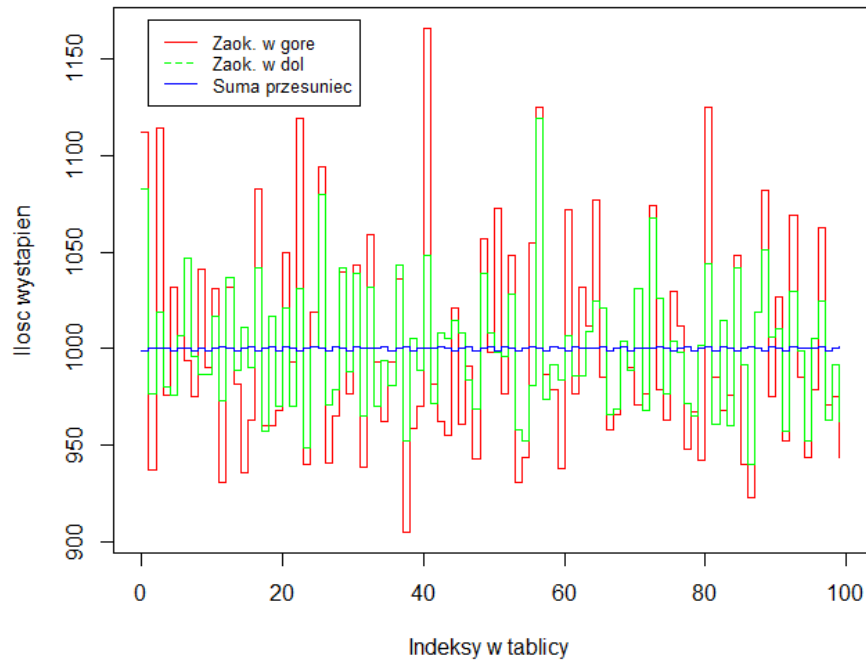


# 1 Porównanie funkcji haszującej

Porównanie dotyczy zbioru liczb całkowitych z zakresu  $x \in \{0, 1, 2, \dots, 10^4 - 1\}$  oraz wielkości tablicy haszującej równej  $a = 100$ . Porównywane są 3 metody z czego dwie między sobą różnią się sposobem zaokrąglania wyniku. Zaokrąglanie w metodzie środka kwadratu jest wykorzystywane w celu jednoznacznego określenia początku sekwencji wynikowej. Metody:

1. metoda środka kwadratu(zaokrąglanie przedziału w górę),
2. metoda środka kwadratu(zaokrąglanie przedziału w dół),
3. metoda 'sumowania przesunięć bitowych'.



Rysunek 1: Porównanie funkcji haszujących dla  $x \in \{0, 1, 2, \dots, 10^4 - 1\}$  oraz rozmiaru tablicy haszującej  $a = 100$ .

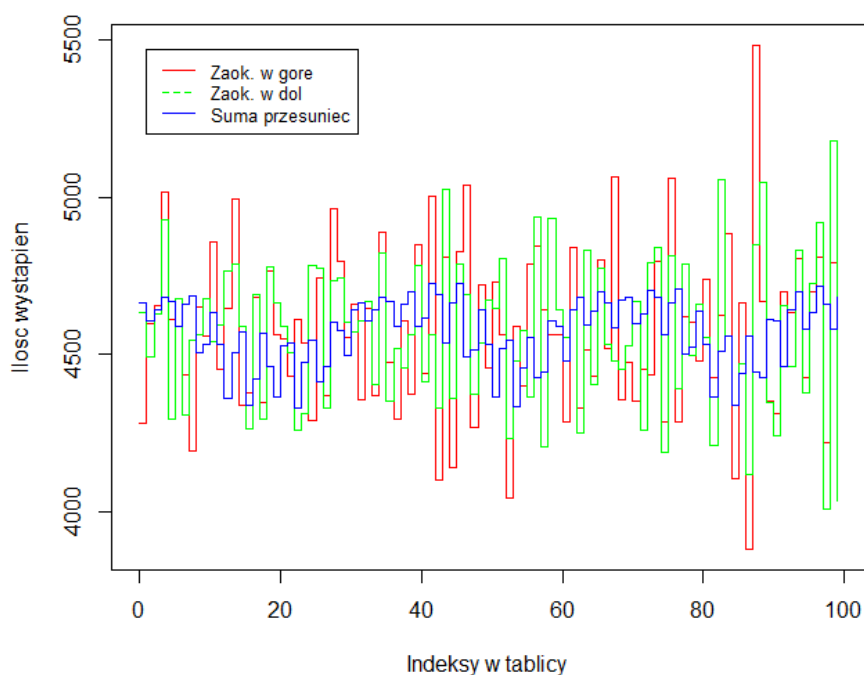
**Wnioski** Metoda środka kwadratu osiąga bardzo niejednolite wyniki. Odchylenia standardowe dla takiej funkcji haszującej wynosi odpowiednio 53.76999 oraz 32.64347 dla zaokrąglania w górę i w dół. Natomiast metoda sumowania przesunięć bitowych posiada bardzo małe odchylenie równe zaledwie 0.7385489.

## 2 Porównanie funkcji haszującej dla wariancji z powtórzeniami

W poniższym paragrafie opisane zostaną wyniki wcześniej wspomnianych funkcji dla wariancji cztero-elementowych ( $k = 4$ ) z powtórzeniami dla alfabetu łacińskiego ( $x \in \{A, \dots, Z\}$ ).

### 2.1 I Sposób wyznaczania liczby z ciągu symboli

Wyznaczanie liczby z ciągu symboli odbywa się poprzez tworzenie liczby w zapisie dziesiętnym, gdzie współczynnikami kolejnych potęg są wartości kodów ASCII poszczególnych symboli —  $L = 'A' * 10^3 + 'B' * 10^2 + 'C' * 10 + 'D'$ .



Rysunek 2: Porównanie f. haszujących dla I sposobu.

Tabela 2.1 zawiera odchylenia standardowe dla uzyskanych histogramów:

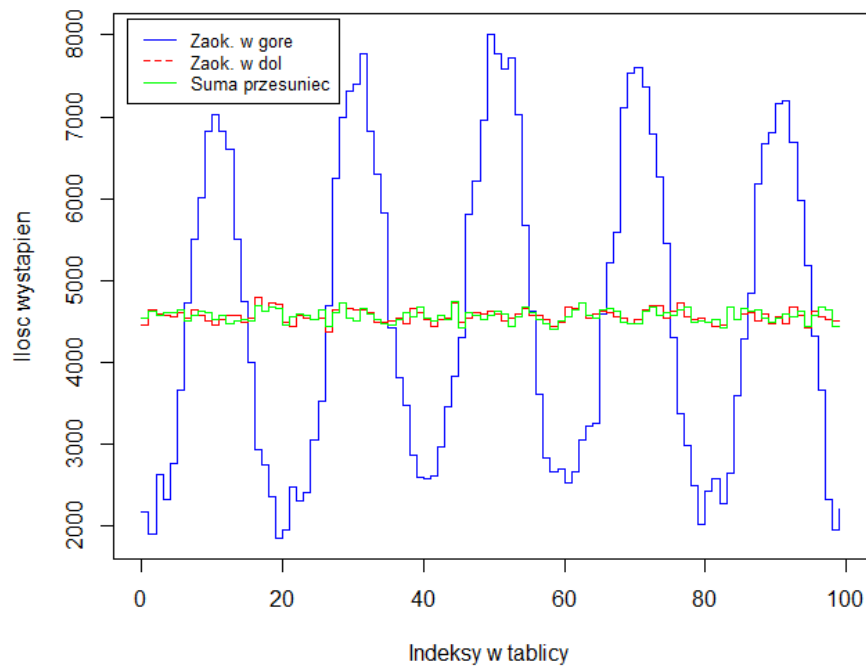
**Wnioski** Metoda sumowania przesunień bitowych okazała się podobnie jak w poprzednim przypadku dużo lepsza.

Tablica 1: Odchylenia standardowe dla I sposobu

Metoda	Odchylenie standardowe
środka kwadratu(góra)	258.9582
środka kwadratu(dół)	233.916
sumowania przesunięć bitowych	103.8099

## 2.2 II Sposób wyznaczania liczby z ciągu symboli

Ciąg symboli wyznaczany jest poprzez dodawanie oraz przesuwanie bitowe kolejnych wartości kodów ASCII ciągu symboli do wartości początkowej równej zero. Tabela 2.2 zawiera odchylenia standardowe dla uzyskanych histogramów:



Rysunek 3: Porównanie f. haszujących dla II sposobu.

**Wnioski** Zmiana sposób wyznaczania liczby z ciągu symboli wpłynęła znacząco na zachowanie się poszczególnych funkcji haszujących. Tym razem prosta metoda środka kwadratu uzyskała dużo lepsze wyniki.

Tablica 2: Odchylenia standardowe dla II sposobu

Metoda	Odchylenie standardowe
środek kwadratu(góra)	80.8298
środek kwadratu(dół)	77.59492
sumowania przesunięć bitowych	1926.456

### 3 Podsumowanie

W pierwszym paragrafie(1) metoda sumowania przesunięć bitowych dała najlepszy rezultat ponieważ wartości cechowały się jednostajnym rozkładem prawdopodobieństwa. Jeżeli mamy do czynienia z niejednostajnym rozkładem prawdopodobieństwa jak w przypadku wariancji z powtórzeniami dla alfabetu bardzo istotne jest dobranie dwóch elementów: funkcji przekształcającej(w przykładzie omówionym w paragrafie 2.1 – ciąg symboli w liczbę) oraz funkcji haszującej. Zmiana funkcji omówiona w paragrafie 2.2 przekształcania ciągu symboli bardzo wpłynęła na jakość funkcji haszującej ponieważ generowane indeksy zbyt często się powtarzały, dlatego wykorzystanie tablicy haszującej nie było równomierne. W rezultacie metoda środek kwadratu dla zmodyfikowanej funkcji przekształcania okazała się lepsza i posiadała mniejsze odchylenie od wartości średniej.

### 4 Użyte programy

- Java - jako język programowania do generowania wyników
- R - język i środowisko do generowania wykresów i wyznacznia odchyleń