# Metody Probabilistyczne i Statystyka

## LISTA 10 do wykładu z metod Monte Carlo)

Informatyka algorytmiczna (I st.) WIiT – 2021/2022

**Zadanie 1.** **(zadanie programistyczne)** Zdefiniujmy zmienną losową $S_N = X_1 + \cdots + X_N$, gdzie $X_1, \ldots, X_N$ są niezależne oraz każde $X_i$ przyjmuje wartości $1, -1$ z ppb $0.5$.

- wyznacz (numerycznie) dystrybuantę dla $S_N$ dla kilku wartości $N$, np. $5, 10, 15, 20, 25, 30$.
- porównaj ją z dystrybuantą rozkładu normalnego, która miałaby aproksymować $S_N$.

Skorzystaj z Matlaba (w tym funkcji `normcdf` dla obliczania dystrybuanty rozkładu normalnego). Jeśli starczy Ci czasu, narysuj histogramy. Jakie są wnioski obliczeń?
Dla porównania, wykorzystaj ten sam kod dla sprawdzenia co się dzieje dla $N = 100$.

**Zadanie 2.** Na wykładzie omawialiśmy metodę szacowania wielkości $|p - \hat{p}|$, gdzie $p$ to rzeczywiste prawdopodobieństwo zdarzenia $A$, zaś $\hat{p}$ to częstość wystąpienia $A$ w $N$ niezależnych próbach. Gdy $N$ jest duże, to jak widać z zadania 1, można użyć przybliżenia przez rozkład normalny. Opracuj alternatywne podejście dla małych $N$ oparte na nierówności Czebyszewa.

**Zadanie 3.** Samorząd gminny musi przyjąć lub odrzucić uchwałę $X$. Sprawa jest trudna, nie jesteśmy w Szwajcarii (gdzie w takich sprawach robi się referenda lokalne), więc postanowiono spytać się $N$ losowych osób i postąpić zgodnie z wolą większości z nich. Jednak zanim podejmie się decyzję, to trzeba mieć gwarancję, że decyzja odzwierciedla wolę całej lokalnej społeczności z prawdopodobieństwem co najmniej 0.9. Jak dobrać liczbę $N$? Liczba ta powinna być jak najmniejsza (dotarcie do losowych mieszkańców kosztuje dużo pracy).
Warunki: nie wiemy jaka jest frakcja zwolenników $X$ w całej populacji.

**Zadanie 4.** Rozważamy sumę $N$ niezależnych zmiennych losowych $X_i$, każda o wartości oczekiwanej $\mu$ i wariancji $\sigma$. Niech $X = \frac{1}{N}\sum_{i=1}^{N} X_i$. Wariancja $X$ to $\frac{\sigma^2}{N}$. Na podstawie eksperymentu chcielibyśmy wyznaczyć $\sigma$. Jak słyszeliście, estymator nieobciążony dla wariancji $\sigma$ dany jest wzorem:

$$\frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2$$

gdzie $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ jest zaobserwowaną średnią z pomiarów.
Zbadać dla małych $N$ jaka jest wartość tego estymatora dla sumy $N$ zmiennych $X_i$ z rozkładem $\Pr(X_i = 0) = \Pr(X_i = 1)$. Dla porównania policzyć to samo dla wzoru

$$\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2$$