# Centralne Twierdzenie Graniczne - symulacja

Katarzyna Bęben, styczeń 2025

## Cele

Pokazać za pomocą symulacji, że zmienna losowa $Z_n$ zdefiniowana w założeniach centralnego twierdzenia granicznego zbiega według dystrybuanty do zmiennej losowej $Z$ rozkładu normalnego standardowego:

$$Z_n \xrightarrow[n \to \infty]{wg\ dystrybuanty} Z \sim N(0{,}1)$$

## Założenia Centralnego Twierdzenia Granicznego (CTG)

Niech $X_n$ będzie ciągiem zmiennych losowych niezależnych o tym samym rozkładzie ze skończoną średnią $\mu$ oraz wariancją $\sigma^2 > 0$. Przyjmujemy:

$$Z_n = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

gdzie: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ , $n - liczba\ obserwacji$

## Warunki początkowe symulacji

Pod rozważania weźmiemy zmienne z rozkładu $\chi^2(chi\ kwadrat)$. Gęstość tego rozkładu dla k stopni swobody wyraża się wzorem:

$$f(x) = \begin{cases} \frac{1}{\left(\sqrt{2}\right)^n \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & dla\ x > 0 \\ 0, & dla\ x \leq 0 \end{cases}$$

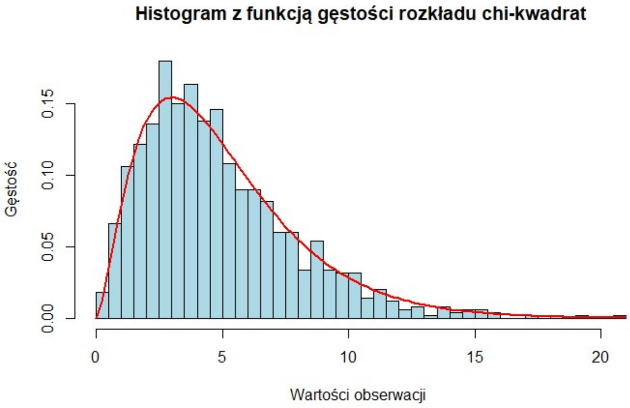
wartość średnia dla rozkładu $\chi^2$: $\mu = k$
wariancja dla rozkładu $\chi^2$: $\sigma^2 = 2k$
Przyjmujemy:

$$k = 5\ -\ liczba\ stopni\ swobody$$
$$n = 1000\ -\ liczba\ obserwacji\ w\ pojedycznym\ podejściu$$
$$s = 10\,000\ -\ liczba\ powtórzeń\ symulacji$$

Z użyciem języka R generujemy ciąg 1000 obserwacji $(x_i)$ (realizacje zmiennej losowej $X_i$), które mogłyby pochodzić z rozpatrywanego rozkładu $\chi^2$

$$x\ < -\ rchisq(1000, df\ =\ 5)$$

I przedstawiamy na histogramie:




*Rysunek 1 histogram i funkcja gęstości rozkładu chi-kwadrat*

Kształt histogramu wykazuje wyraźną asymetrię charakterystyczną dla tego rozkładu. Zgodność histogramu z funkcją gęstości rozkładu $\chi^2$ wskazuje, że dane zostały wygenerowane poprawnie.

## Przebieg symulacji

Realizujemy symulację. 10 000 razy za pomocą generatora liczb pseudolosowych tworzymy potencjalne obserwacje $x_i$ (realizacje zmiennej losowej $X_i$) jak powyżej i na ich podstawie (korzystając ze wzoru z CTG) wyznaczamy elementy ciągu realizacji zmiennej losowej $Z_j$ - $z_j$ gdzie $j \in \{1,2, \dots ,1000\}$. Dostajemy w ten sposób macierz 10 000 x 1 000 przedstawiającą wyniki dla każdego podejścia.

```
k <- 5                #liczba stopni swobody
m <- 5                # Wartość oczekiwana
sigma <- sqrt(10)      # Odchylenie standardowe
n <- 1000              # Liczba obserwacji w pojedynczej próbie
s <- 10000            # Liczba powtórzeń

# Inicjalizacja macierzy wyników: każdy wiersz to ciąg z_j dla jednej symulacji
z_matrix <- matrix(0, nrow = s, ncol = n)

# Symulacja s prób i obliczanie z_j
for (i in 1:s) {
  x <- rchisq(n, df = k)  # Generowanie ciągu obserwacji
  for (j in 1:n) {
    x_mean <- mean(x[1:j])  # Średnia z pierwszych j obserwacji
    z_matrix[i, j] <- (x_mean - m) / (sigma / sqrt(j))  # Obliczenie z_j
  }
}
```

*Rysunek 2 kod symulacji w R*

Dla wybranych j (3, 5, 10, 100, 500, 1000) przedstawiamy rozkłady zmiennej losowej $Z_j$ na histogramie i porównujemy jego kształt z funkcją gęstości rozkładu normalnego standardowego.

```r
# Wybór wartości j do analizy histogramów
j_values <- c(3, 5, 10, 100, 500, 1000)

# Rysowanie histogramów dla wybranych j
for (j in j_values) {
  z_j_values <- z_matrix[ ,j] # Pobranie j-tej kolumny z macierzy

  # Rysowanie histogramu
  hist(z_j_values, probability = TRUE, breaks = 30, col = "lightblue",
       main = paste("Histogram dla j =", j),
       xlab = paste("Realizacje zmiennej losowej Z_", j, sep = ""), ylab = "Gęstość")

  # Naniesienie funkcji gęstości N(0, 1)
  curve(dnorm(x, mean = 0, sd = 1), col = "red", lwd = 2, add = TRUE)

  # Dodanie legendy
  legend("topright", legend = c("N(0,1)"),
         col = c("red"), lwd = 2)
}
```

*Rysunek 3 kod generowania histogramów w R*

Dla lepszego zobrazowania generujemy także wykresy dystrybuant.

```r
for (j in j_values) {
  z_j_values <- z_matrix[ ,j] # Pobranie j-tej kolumny z macierzy

  empirical_cdf <- ecdf(z_j_values)  # Empiryczna dystrybuanta dla Z_j

  # Wykres empirycznej i teoretycznej dystrybuanty
  plot(empirical_cdf, col = "blue", lwd = 2, main = paste("Porównanie dystrybuant dla j =", j),
       xlab = "z", ylab = "Dystrybuanta", xlim = c(-4, 4), ylim = c(0, 1))
  curve(pnorm(x, mean = 0, sd = 1), col = "red", lwd = 2, add = TRUE)  # Dystrybuanta N(0,1)

  # Legenda
  legend("bottomright", legend = c("Dystrybuanta rozkładu zmiennej Z_j", "Dystrybuanta N(0,1)"),
         col = c("blue", "red"), lwd = 2)
}
```

*Rysunek 4 kod generowania wykresu dystrybuant w R*

# Analiza wyników i wnioski

Otrzymujemy następujące histogramy:

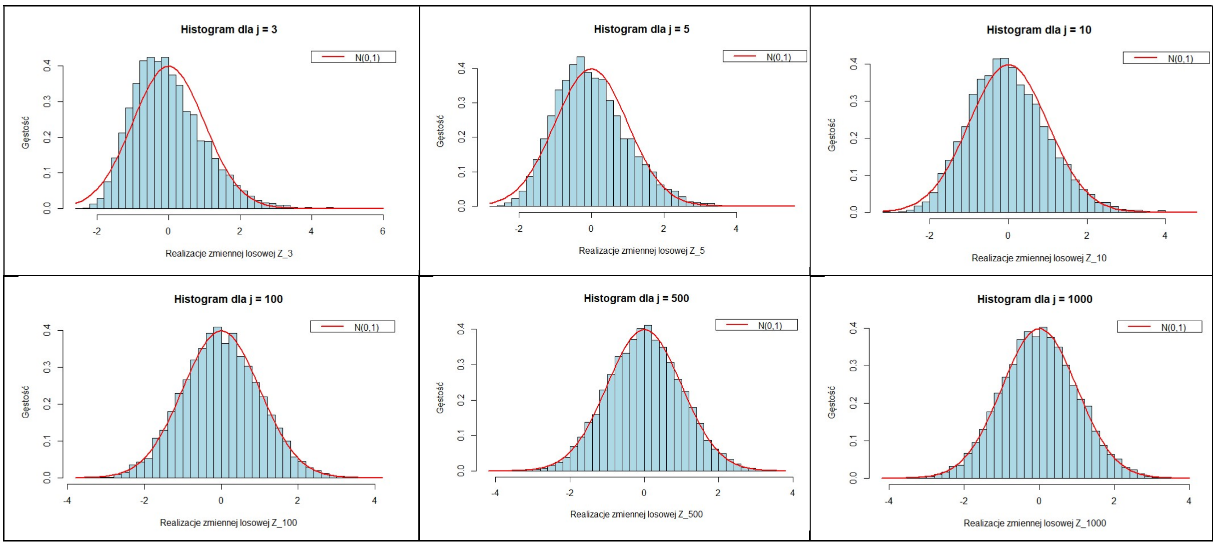



*Tabela 1 Porównanie histogramów otrzymanego rozkładu z funkcją gęstości rozkładu N (0,1)*

Histogramy pokazują, że w miarę zwiększania liczby obserwacji j rozkład zmiennej losowej $Z_j$ jest coraz bardziej zbliżony do rozkładu normalnego standardowego (kształt histogramu zgodny z kształtem naniesionej funkcji gęstości rozkładu N (0,1)).

Dodatkowym potwierdzeniem zbiegania zmiennej losowej $Z_j$ według dystrybuanty do zmiennej rozkładu normalnego standardowego jest coraz większe podobieństwo dystrybuanty rozkładu zmiennej $Z_j$ i N (0,1) dla coraz większych j (dla j=1000 na wykresie niezauważalne):

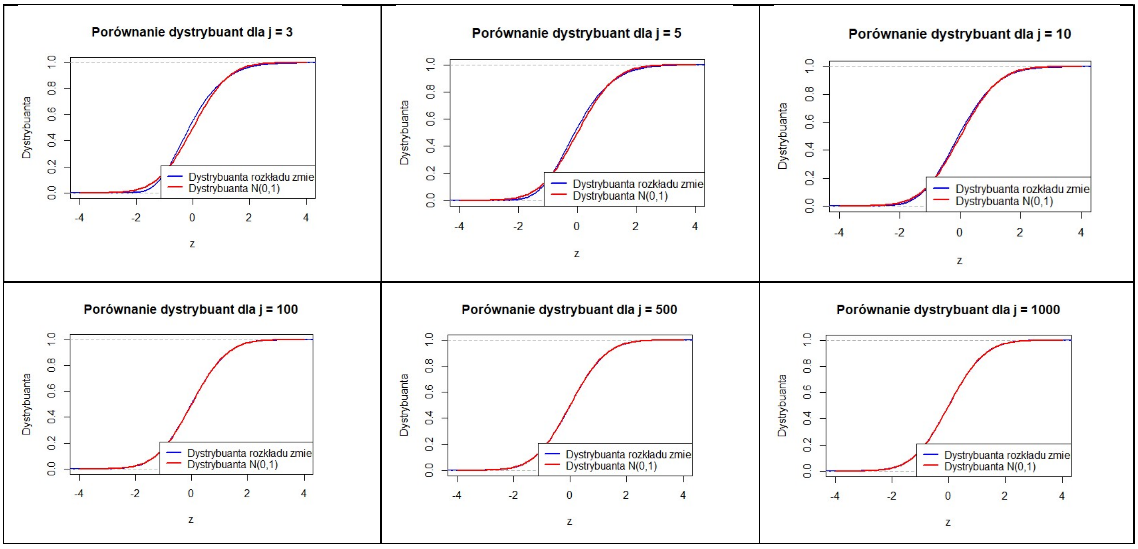



*Tabela 2 Porównanie dystrybuanty otrzymanego rozkładu z dystrybuantą rozkładu N (0,1)*

Biorąc pod uwagę powyższe, można stwierdzić, że przeprowadzona symulacja potwierdza słuszność centralnego twierdzenia granicznego.