

Wstęp

Tematem projektu jest centralne twierdzenie graniczne. Jest to twierdzenie uzasadniające występowanie rozkładów bardzo zbliżonych do rozkładu normalnego w przyrodzie. Kod dostępny jest w repozytorium na **GitHub**([tutaj](#)).

Twierdzenie

Twierdzenie to mówi, że jeżeli X_i są niezależnymi zmiennymi losowymi, posiadającymi tę samą wartość oczekiwaną $\mathbb{E}X < \infty$ oraz wariancję $\sigma^2 < \infty$ (obie te wartości muszą być skończone), to ciąg zmiennych losowych postaci:

$$U_n = \frac{\frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}X}{\sigma} \sqrt{n}, \quad (1)$$

jest zbieżny według rozkładu do standardowego rozkładu normalnego $U \sim \mathcal{N}(0, 1)$, gdy $n \rightarrow \infty$.

Opis projektu

Program został napisany w języku **R** oraz wymaga paczek **shine**, **ggplot2**, **comprehenr** i **pracma**. Głównie opiera się na interaktywnej aplikacji w **shine**. Poniżej zawarte jest pole z opcjami symulacji:

Choose distribution:

Uniform

Sample Size:

10000

Average Over (n):

2

Figure 1: Zmienne symulacji.

Pozwala ono na wybranie zadanego rozkładu ciągłego spośród:

- Uniform - $U(0,1)$.
- Exponential - $\text{Exp}(1)$.
- Chi2 - $\chi^2(5)$.
- Beta - $\text{Beta}(2,5)$.
- MyFunction - $c \left((e^x \mathbb{1}(x)_{(0,1)} + (-x + e + 1) \mathbb{1}(x)_{(1,2)} + \frac{5}{x} \mathbb{1}(x)_{(2,3)}) \right)^*$

* Numerycznie wyznaczone $c = 0.16$, $\mathbb{E}X = 1.55$, $\sigma^2 = 0.64$.

Pole **Sample Size** pozwala na ilość na wybranie ilości zmiennych U_n z wzoru (1). Dla podanej domyślnie wartości 10000, tworzenie wykresów działa dostatecznie szybko, natomiast wykresy charakteryzuje całkiem spora dokładność. Pozwala to na obserwację wpływu ilości zmiennych na histogramy.

Zmienna **Average Over** jest wartością n ze wzoru (1). Zgodnie z założeniem centralnego twierdzenia granicznego, histogram U_n powinien dążyć do $\mathcal{N}(0, 1)$ wraz ze wzrostem n .

Poniżej zamieszczony został przykładowy wynik symulacji, która tworzona jest na bieżąco za każdą zmianą parametrów.

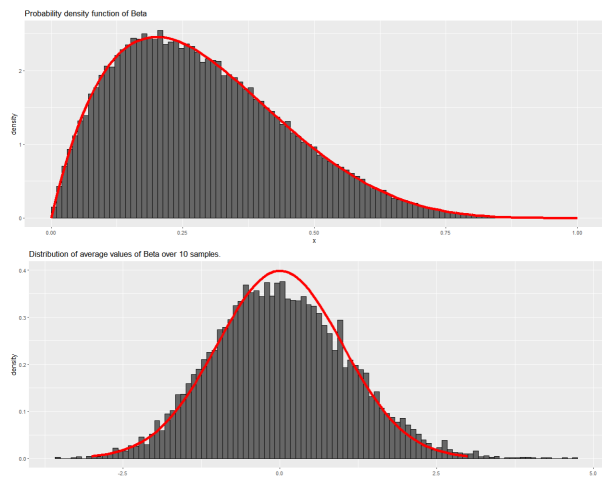


Figure 2: Wykresy dla Beta, 10000, 10.

Histogram na pierwszym wykresie przedstawia znormalizowany rozkład wszystkich zmiennych (jest ich $\text{SampleSize} * n$) wygenerowanych dla danej symulacji, natomiast czerwoną linią zaznaczona jest teoretyczna funkcja gęstości tego rozkładu.

Na drugim wykresie histogram przedstawia funkcję gęstości zmiennej U_n ze wzoru (1). Na czerwono zaznaczony jest rozkład $\mathcal{N}(0, 1)$, co pozwala na porównanie otrzymanego znormalizowanego rozkładu, z rozkładem, do którego dąży, zgodnie z centralnym twierdzeniem granicznym. Oba wykresy posiadają dynamiczną liczbę słupków równą $n * 10$.

Podsumowanie

Przeprowadzając symulacje dla różnych rozkładów oraz wartości `sampleSize` oraz n , możemy zauważyć, że rzeczywiście wykres drugi bardzo przypomina rozkład normalny dla $n > 10$. Symulacja pozwala empirycznie udowodnić poprawność tezy.