**Sprawozdanie z listy 2. – Symulacja komputerowa L**

Filip Antoniak (279929)

1. **Wstęp**

Celem tego sprawozdania jest stworzenie generatora liczb pseudolosowych dla rozkładu równomiernego 𝑈[0,1) i zbadanie jego właściwości i statystyk.

1. **Generator ciągów**

Aby powstałe ciągi liczb miały rozkład równomierny i spełniały założenie zakresu 𝑈[0,1), należy zadbać aby wśród liczb nie było wartości 1. Można zrobić to na szereg sposobów, np.:

**Obraz zawierający tekst, zrzut ekranu, Czcionka

Opis wygenerowany automatycznie**

Rysunek : Kod z funkcjami generującymi sekwencje z rozkładu U[0, 1)

Powyższy kod generuje sekwencje kolejno, 20 i 100 liczb z zakresu [0, 1).

1. **Najważniejsze statystyki ciągów:**

Tabela : Najważniejsze statystyki ciągów 20 i 100 elementowych z rozkładu U[0, 1) w zestawieniu z rozkładem U[a. b)

|  |  |  |  |
| --- | --- | --- | --- |
| **Statystyka** | **20 elementów** | **100 elementów** | **Rozkład U[a,b)** |
| Średnia | 0.5658014284566937 | 0.5030061965761424 |  |
| Mediana | 0.5888990157422684 | 0.5279573234092918 |  |
| Dominanta | 0.011304845366908567 | 0.0003780852239052 |  |
| Wariancja | 0.11489991921566682 | 0.07945409758392864 |  |
| Odchylenie standardowe | 0.338968905971723 | 0.2818760322977614 |  |
| Skośność | -0.2353565301636867 | 0.13687055091429873 |  |
| Kurtoza | 1.5032580617009816 | 1.7406611412588406 |  |
| Rozstęp | 0.9851399960361277 | 0.9833333264094991 |  |

* Średnia: w miarę wzrostu ilości elementów dąży do .
* Mediana: tak samo jak średnia, dąży do .
* Dominanta: jako że każda liczba ma równą szansa na bycie wylosowaną, dominantą może być każda dowolna liczba, więc jest niezdefiniowana.
* Wariancja: obliczana ze wzoru:

Mówi o tym jak bardzo wartość zmiennej losowej różni się od średniej.

Więc dla U[0, 1), gdy liczba elementów dąży do nieskończoności wariancja wyniesie . Widać, że dla U[0, 1) gdzie liczba elementów osiąga 100 wariancja zbliża się do tej wartości.

* Odchylenie standardowe: Pierwiastek z omówionej wariancji.
* Skośność: Rozkład jest symetryczny więc wartość wynosi 0. Wartości skośności dla 20 i 100 elementów różnią się od 0, ponieważ wielkość sekwencji jest za mała by uzyskać idealny rozkład równomierny.
* Kurtoza: Wartości kurtozy dla sekwencji 20 i 100 elementowych mocno odbiegają od tej dla U[a, b) wynoszącej . Możliwe, że zbieżność osiągana jest później, gdy ilości elementów są większe.
* Rozstęp: będący różnicą wartości maksymalnej przypomina ten oczekiwany już w przypadku 20 elementów.

1. **Zestawienie histogramów**

Obraz zawierający tekst, zrzut ekranu, diagram, Prostokąt

Opis wygenerowany automatycznieObraz zawierający tekst, zrzut ekranu, diagram, Wykres

Opis wygenerowany automatycznie

Rysunek : Histogramy (bins = 10) dla sekwencji 20 i 100 elementowych z rozkładu U[0, 1).

Czerwoną linią są oznaczone oczekiwane ilości występowań dla każdej z prób.

Obraz zawierający tekst, zrzut ekranu, linia, diagram

Opis wygenerowany automatycznie

Rysunek : Funkcja gęstości rozkładu równomiernego U[0, 1).

Kształty PDF i histogramów odbiegają od siebie. Wynika to jednak tylko z małej wielkości próby. Zgodnie z prawem wielkich liczb, gdy wielkość próby będzie zbliżała się do nieskończoności, histogram będzie swoim kształtem coraz bardziej przypominał funkcje gęstości prawdopodobieństwa. Różnić się będzie tylko oś Y. W PDF to natomiast w histogramach dla bins=10 wyniesie , gdzie n to liczba elementów.

1. **Test statystyczny średnich z prób**

**Wymagania testu:**

* **Normalność rozkładu**: Dane w każdej z porównywanych grup powinny pochodzić z rozkładów normalnych.
* **Jednorodność wariancji**: Wariancje w obu grupach powinny być do siebie zbliżone.
* **Niezależność obserwacji**: Obserwacje w każdej grupie powinny być niezależne od siebie.

O ile spełnione są założenia 2 i 3, co widać m.in. w tabeli zestawień najważniejszych statystyk, to dane nie pochodzą z rozkładu normalnego.

Z pomocą przychodzi tu *CTG*, które dla dużych prób (>30), ponieważ rozkład średnich z próby dąży do **rozkładu normalnego.**

**Hipoteza 0:** średnie z prób są równe 0.5

**Wyniki testu:**

20 elementów:

t: 0.39626742465189213

p: 0.34816140730017264

**Nie ma podstaw do odrzucenia hipotezy zerowej.**

100 elementów:

t: -1.2640813711441008

p: 0.8954154996305793

**Nie ma podstaw do odrzucenia hipotezy zerowej.**

Dla n=20 nie został spełniony warunek, że dane pochodzą z rozkładu normalnego, tutaj CTG nie pomaga, bo próba 20<30. Więc do weryfikacji wykorzystany zostanie dodatkowy test:

Dla testu **Wilcoxon`a:**

Statytyka = 98.0

p-value = 0.8124

wartość p jest znacznie większa niż przyjmowany poziom istotności 0.05, więc **nie ma podstaw co odrzucenia hipotezy zerowej.**

1. **Wnioski**

Testy statystyczne pokazały, że dla prób n=20 i n=100 nie ma podstaw do odrzucenia hipotezy zerowej: *średnia z próby jest równa 0.5*.

Nie było to jednak jednoznaczne dla próby n=20, w niektórych wywołaniach generowana sekwencja nie przechodziła testu, p value było niższe niż poziom istotności 0.05, więc czasem hipoteza 0 była odrzucana, co oznacza, że średnia była statystycznie istotnie inna niż 0,05.

Wynika to z prawa wielkich liczb, im bardziej zwiększymy wielkość próby w tej symulacji, tym bardziej jej rozkład zbliży się statystycznie do rozkładu równomiernego.