SPRAWOZDANIE – LABOLRATORIUM NR 14

Generowanie ciągu liczb pseudolosowych o rozkładzie normalnym

metodą eliminacji.11.06.2021r.

Przemysław Rodzik

1. Wstęp teoretyczny

Pseudolosowość polega na tym że w ciągu liczby nie da się znaleźć wzoru w którym są rozłożone wyniki. Więc zakładamy że są losowe. Liczby takie są często potrzebne w wielu algorytmach, gdyż nie wszystkie zdarzenia są doskonale opisane przez prawa fizyki (np. kwantowej), w związku z czym niemożliwym staje się określenie rezultatu. Generatory liczb (pseudo-) losowych często „wybierają” na podstawie zegara systemowego (czas zapisany w komputerze jest przecież „zmienny w każdej chwili”) za pomocą ziarna startowego lub innych sposobów. Generatory te mogą tworzyć różne rozkłady.

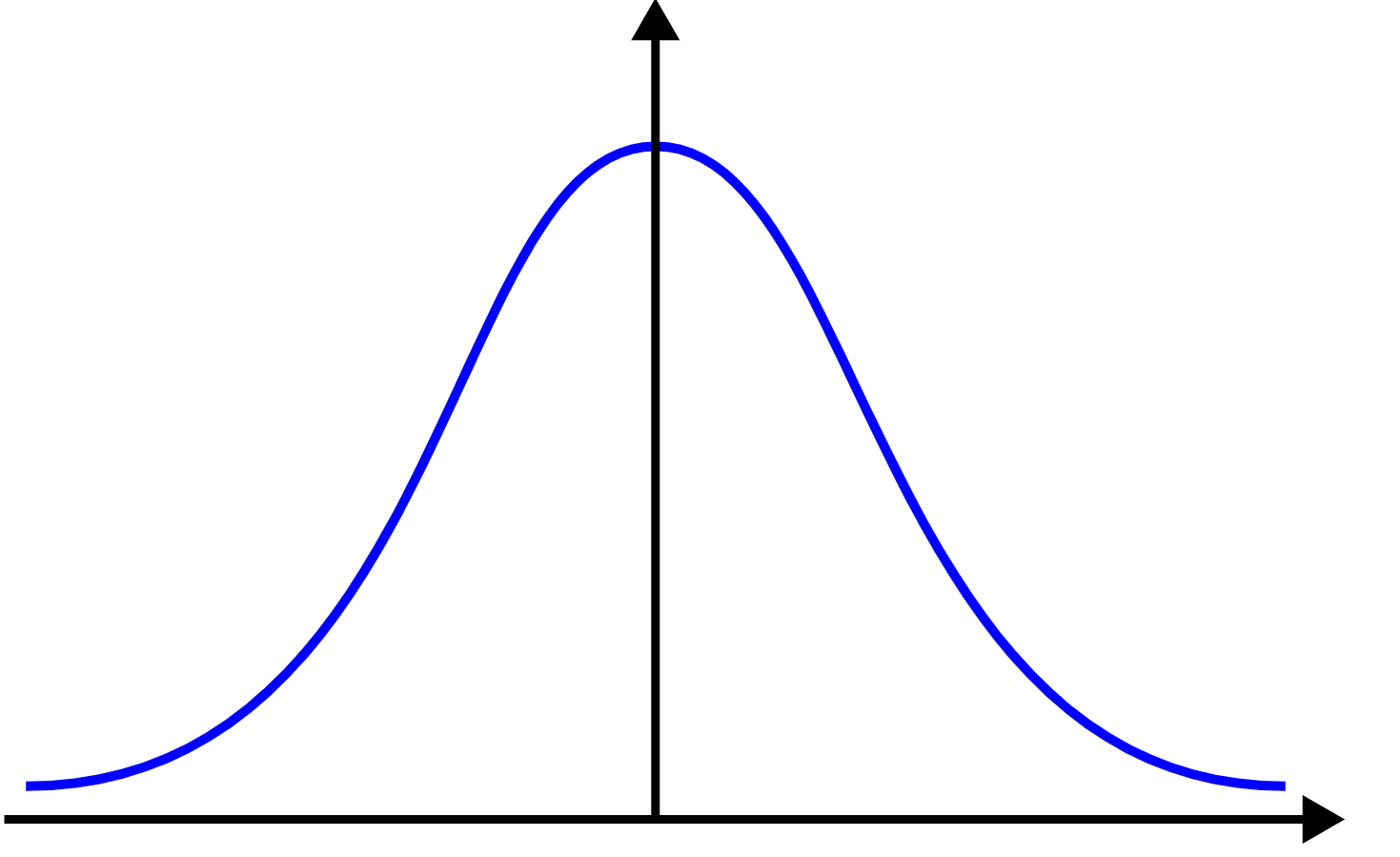
Rozkład jednostajny (też: jednorodny) to taki rozkład, w którym w danym określonym przedziale gęstość prawdopodobieństwa jest stała i niezerowa.

Liczby w ciągu rozkładu normalnego można wyznaczyć z pomocą generatora mieszanego:

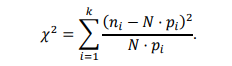


Rozkład normalny, inaczej zwany rozkładem Gaussa, krzywą Gaussa jest najważniejszym rozkładem teoretycznym prawdopodobieństwa w statystyce. Rozkład normalny jest też najbardziej intuicyjnym rozkładem statystycznym. W wielkim skrócie opisuje on sytuacje w świecie, gdzie większość przypadków jest bliska średniemu wynikowi, a im dany wynik bardziej odchyla się od średniej tym jest mniej reprezentowany.

Wykres funkcji prawdopodobieństwa tego rozkładu jest krzywą w kształcie dzwonu (tak zwaną krzywą dzwonową).



Test 𝝌2 (chi kwadrat). Polega na badaniu czy wylosowane liczby z rozkładu są wynikiem takiego rozkładu. Aby to sprawdzić dzieli się przedział na kilka podprzedziałów a następnie przydziela się dane wyniki do podanej grupy która odpowiada danemu przedziałowi. Zlicza się wszystkie podane liczby z danego przedziału, a następnie oblicza się statystykę testową:



1. Opis problemu

## Rozkład jednorodny

Startując od *x*0 = 10 należy wygenerować *n* = 104 liczb pseudolosowych przy użyciu generatora mieszanego

*xn*+1 = (*axn* + *c*) *mod m*

o parametrach (**typu long**):

1. *a* = 123, *c* = 1, *m* = 215
2. *a* = 69069, *c* = 1, *m* = 232

Proszę w obu przypadkach sporządzić rysunek *Xi*+1 = *f*(*Xi*) (*Xi* = *xi/*(*m* + 1*.*0) **z warunku normalizacji do rozkładu U(0,1)**). Czy porównując oba rysunki można stwierdzić, który generator ma lepsze własności statystyczne? W sprawozdaniu proszę uzasadanić odpowiedź. W sprawozdaniu proszę także zamieścić histogram (dla *k* = 12 podprzedziałów) rozkładu gętości prawdopodobieństwa dla *n* = 104 liczb pseudolosowych o rozkładzie równomiernym (oba przypadki). Proszę także podać obliczone wartości *µ* i *σ* i porównać je z wartościami teoretycznymi.

## Rozkład normalny

Wykorzystując generator mieszany z podpunktu (b) należy wygenerować ciąg *n* = 104 liczb pseudolosowych o rozkładzie normalnym z parametrami *µ* = 0*.*2 i *σ* = 0*.*5 metodą eliminacji. Liczby pseudolosowe mają zawierać się w przedziale *x ∈* [*µ −* 3*σ,µ* + 3*σ*].

## Testowanie generatora o rozkładzie *N*(*µ,σ*) - test *χ*2

Zadania do wykonania:

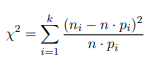
1. Obliczyć średnią arytmetyczną uzyskanego rozkładu normalnego:
2. Obliczyć wariancję



i odchylenie standardowe.

Obliczone wartości *µn* i *σn* zapisać do pliku.

1. Podzielić przedział [*µ−*3*σ,µ*+3*σ*] na *k* = 12 rozłącznych podprzedziałów o identycznej długości.
2. W każdym z podprzedziałów określić ilość liczb pseudolosowych (*ni*), która do niego trafia. Wartości *ni* zapisać do pliku.
3. Wyznaczyć wartość statystyki testowej



gdzie: n jest całkowitą ilością liczb pseudolosowych, *ni* ilość liczb w i-tym podprzedziale, *pi* teoretyczne prawdopodobieństwo wylosowania liczby z i-tego podprzedziału. Aby wyznaczyć wartości *pi* w każdym z podprzedziałów należy skorzystać z wzoru (8). Wartości: *pi* oraz *n · pi* dla każdego z podprzedziałów zapisać do pliku. Do obliczenia *pi* proszę użyć założonych na początku wartości *µ* i *σ*.

1. Testujemy hipotezę *H*0: wygenerowany rozkład jest rozkładem *N*(*µ,σ*) wobec *H*1 że nie jest to prawdą. Korzystając z odpowiednich tabel statystycznych proszę sprawdzić czy nasza hipoteza jest prawdziwa na poziomie istotności *α* = 0*.*05 (*α* jest prawdopodobieństwem pierwszego rodzaju czyli prawdopodobieństwem odrzucenia hipotezy *H*0 gdy ta jest prawdziwa). W tym celu definiujemy obszar krytyczny testu:

*K* = *{X* : *χ*2(*X*) *> ε}*

gdzie: *X* = *{x*1*,x*2*,...,xn}* jest ciągiem liczb pseudolosowych, *χ*2(*X*) wartością statystyki dla danego ciągu *X*, *ε* jest poziomem krytycznym danego rozkładu dla określonej liczby stopni swobody i założonego poziomu istotności. Liczbę stopni swobody określamy jako *ν* = *k − r −* 1, gdzie: k jest liczbą podprzedziałów, a *r* = 2 jest liczbą parametrów testowanego rozkładu (*µ* i *σ*). Jeśli *χ*2 *< ε* to stwierdzamy że dla zadanego poziomu istotności hipoteza *H*0 jest prawdziwa - nasz rozkład jest typu *N*(*µ,σ*).

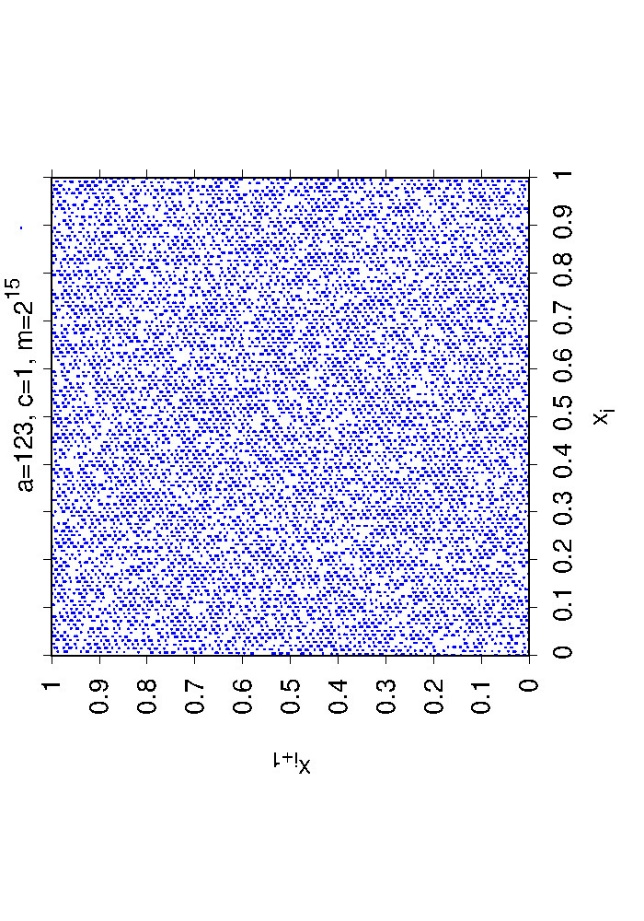
1. Określić poziom ufności dla obliczonej statystyki *χ*2:

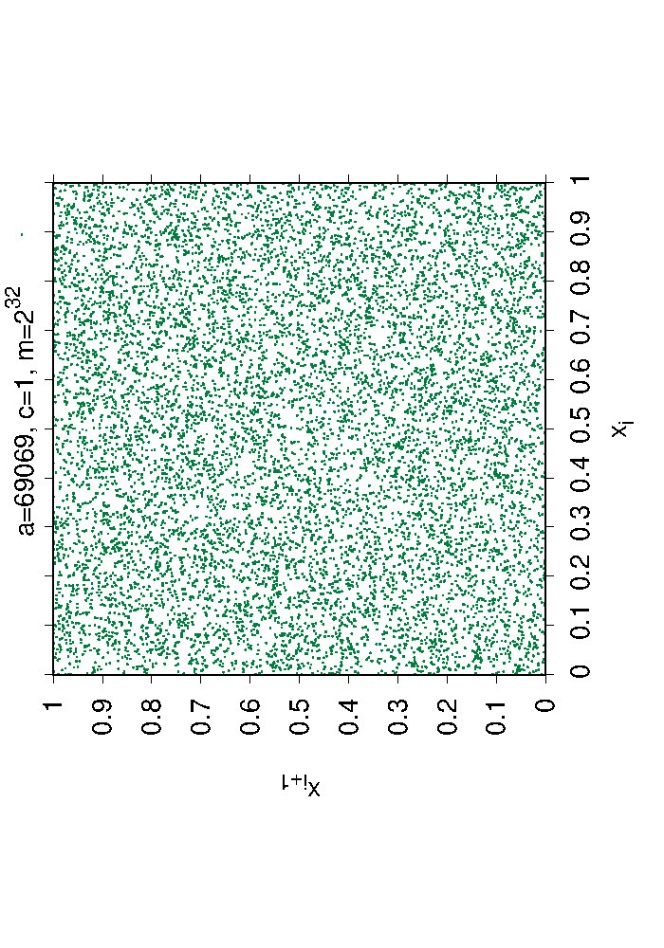
*P*(*χ*2*|ν*) = 1 *− α*˜

gdzie: *ν* = *k − r −* 1 jest liczbą stopni swobody, natomiast ˜*α* jest poziomem istotności którego nie znamy (a chcemy go poznać), korzystając z procedury bibliotecznej:

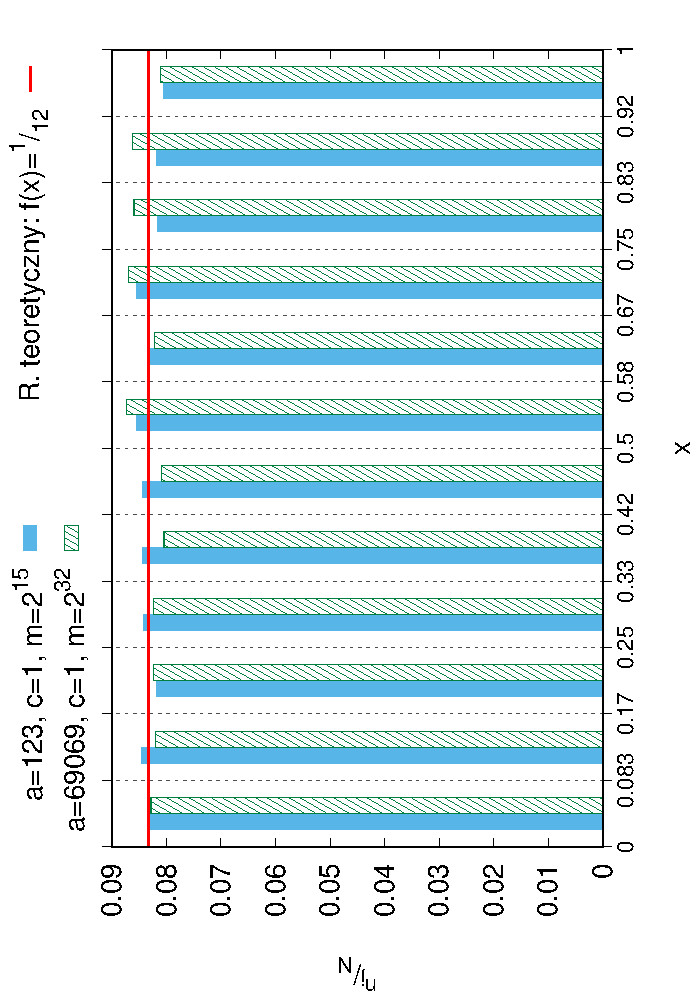


Uwaga: można tu odwrócić zagadnienie tj. zadać sobie pytanie - jaka powinna być wartość *χ*2 dla określonej wartości *α*? - i w ten sposób poszukiwać lewych granic obszarów krytycznych testu. Do poszukiwania wartości *χ*2 można użyć np. metody bisekcji.

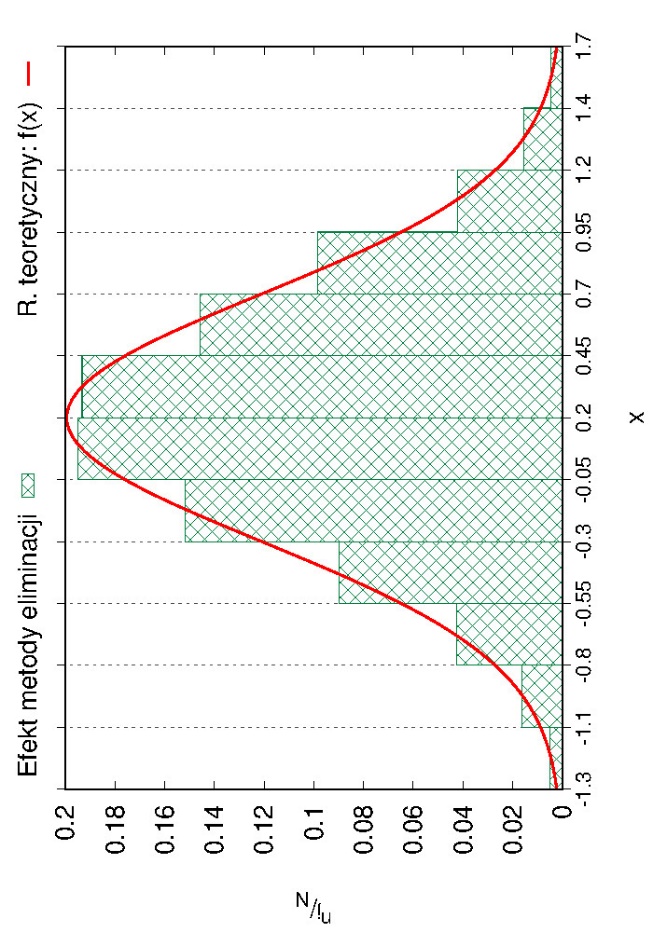
1. W sprawozdaniu proszę zamieścić histogram pokazujący wartości *ni/n* dla każdego z podprzedziałów, na tym samym rysunku proszę także zamieścić przebieg funkcji gęstości prawdopodobieństwa dla rozkładu normalnego.
2. Wyniki



Analizując oba generatory pseudolosowe drugi generator wydaje się być lepszy. Rozmieszczenie punktów na 2 wykresie wydaje się być rozmieszczeniem losowym, wykres 1 wydaje się znajdować pewien schemat (ukośne kreski).



Powyższy histogram pokazuje rozłożenie wartości rozłożonych na przedziale podzielonym na 12 przedziałów. Dla obu prób wyniki są zbliżone do wartości teoretycznej, dla obu są widoczne nie duże anomalie względem wartości teoretycznej. Wyniki są zgodne z oczekiwaniami.



Powyższy histogram przedstawia wyniki rozkładu normalnego wraz z wartościami teoretycznymi. Z histogramu można wywnioskować że pokrywa się wręcz z wartościami teoretycznymi wiec możemy powiedzieć ze algorytm został prawidłowo zaimplementowany.

1. Wnioski

Aby wygenerowanie prawdziwie losowego generatora liczb jest praktycznie nie możliwe dla komputera. Generatory które wykorzystujemy są pseudolosowe jedynie pozornie nie widzimy żadnej prawidłowości w jakiej rozkładają się wyniki.

Gdy patrzymy na losowe rozkłady punktów na wykresach w pierwszy przypadku można dojrzeć wzór, lecz w drugi przypadku miejscami punkty mają tendencje do grupowania się. To może nas utwierdzać w myśleniu, że bardzo trudne jest stworzenie prawdziwie losowego generatora.

Na histogramie z dwoma seriami widać, że oba rozkłady są teoretycznie poprawne i nie ma większej różnicy pomiędzy nimi. Dopiero gdy popatrzymy na wykresy punktowe, możemy znaleźć na pierwszym wykresie pewien schemat w jakim rozłożone są wyniki.

Test 𝝌2 potwierdza słuszność naszego algorytmu, widać to na histogramie słupkowym z krzywą w kształcie dzwonu. Gdyby zaznaczyć punktami najwyższe punkty słupków zobaczylibyśmy wtedy że punktu praktycznie pokrywają się z krzywą.