Zadanie domowe 4.

zadanie 1

Najpierw zajmijmy się wyznaczeniem E(X) i var(X) wiedząc, że X ma rozkład Bin(n, 1/2).

E(X) = n * 1/2 = n/2

var(x) = n * 1/2 * (1 - 1/2) = n/4

a) Podstawiając do nierówności Markova otrzymujemy:
P(X >= 3n/5) <= n/2 * 5/(3n)
P(X >= 3n/5) <= 5/6

Z kolei jeśli spróbujemy to obliczyć przy pomocy nierówności Czebyszewa to musimy to przekształcić do postaci:
P(X - E(X) >= E(x) * 1/5) <= var(X) / (E(X) * 1/5)^2
P(X - n/2 >= n/10) <= n/4 / (n/10)^2
P(X - n/2 >= n/10) <= 25/n

b) Podstawiając do nierówności Czebyszewa:
P(|X - n/2| >= n/20) <= 100/n

Z nierówności Markova:
P(X >= 11n/20) <= n/2 * 20/(11n)
P(X >= 11n/20) <= 20/22

w celu obliczenia dokładnych wartości skorzystałem z pakietu scipy.stats

n	czebyszew	markov	prawdziwe
100	0.25	0.8333	2.84E-02
1000	0.025	0.8333	1.36E-10
10000	0.0025	0.8333	0.00E+00

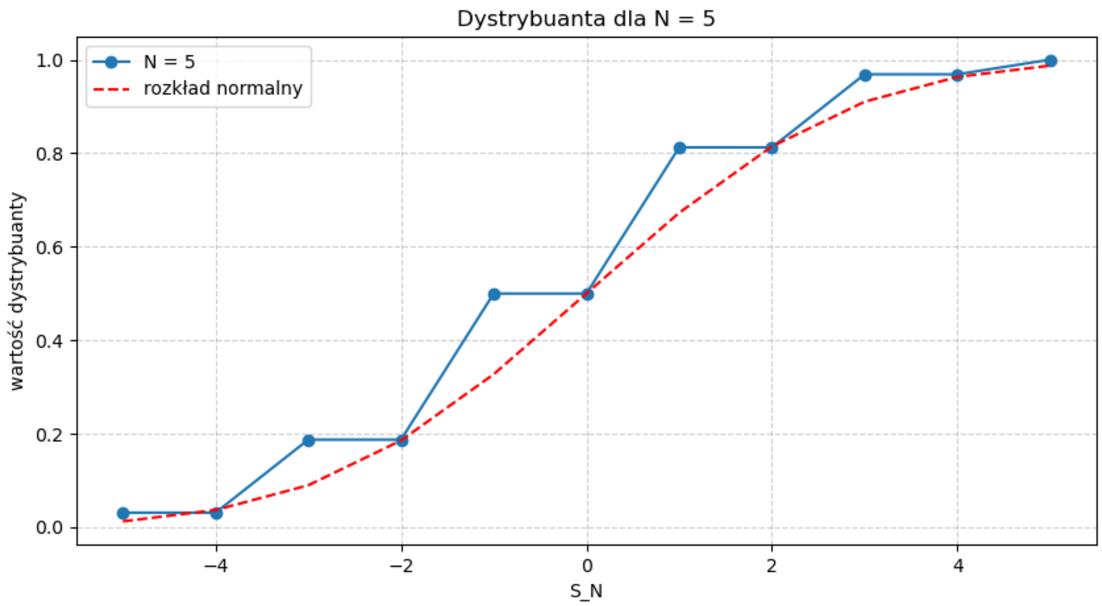
PROF

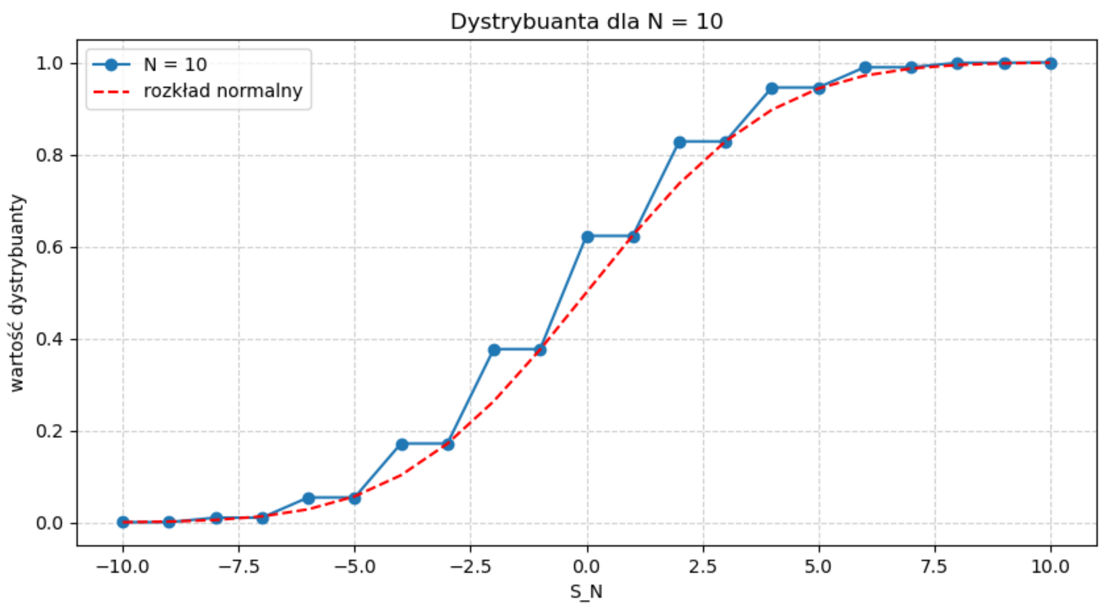
n	czebyszew	markov	prawdziwe
100	1	9.09E-01	0.368202
1000	0.1	9.09E-01	0.001731
10000	0.01	9.09E-01	0

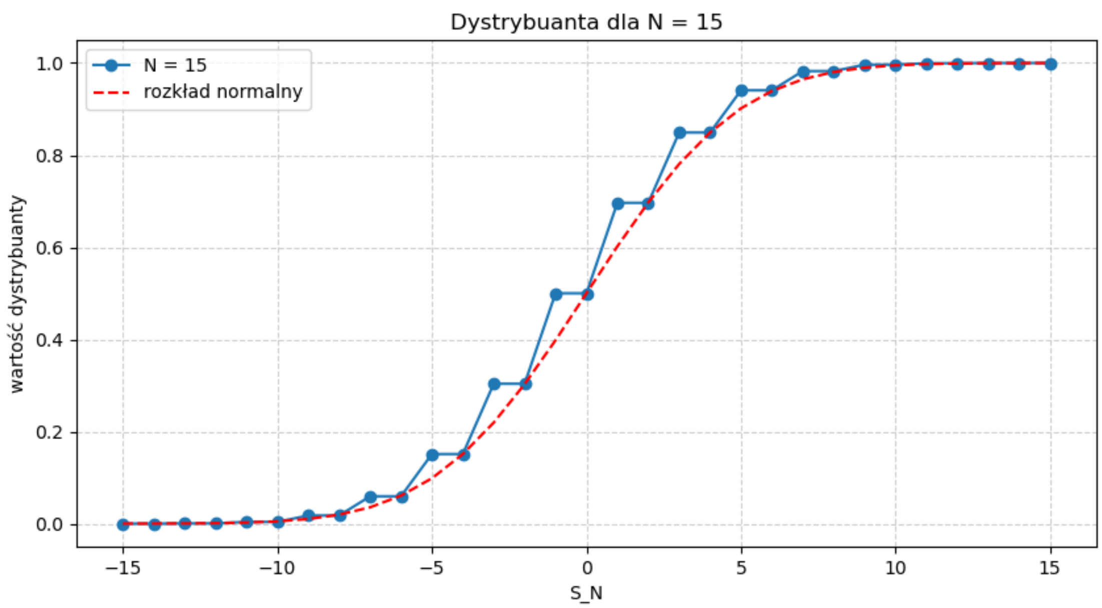
jak widać oszacowanie za pomocą nierówności Czebyszewa jest przeważnie dokładniejsze.
Ponadto im większe n tym bardziej dokładne oszacowanie Czebyszewa

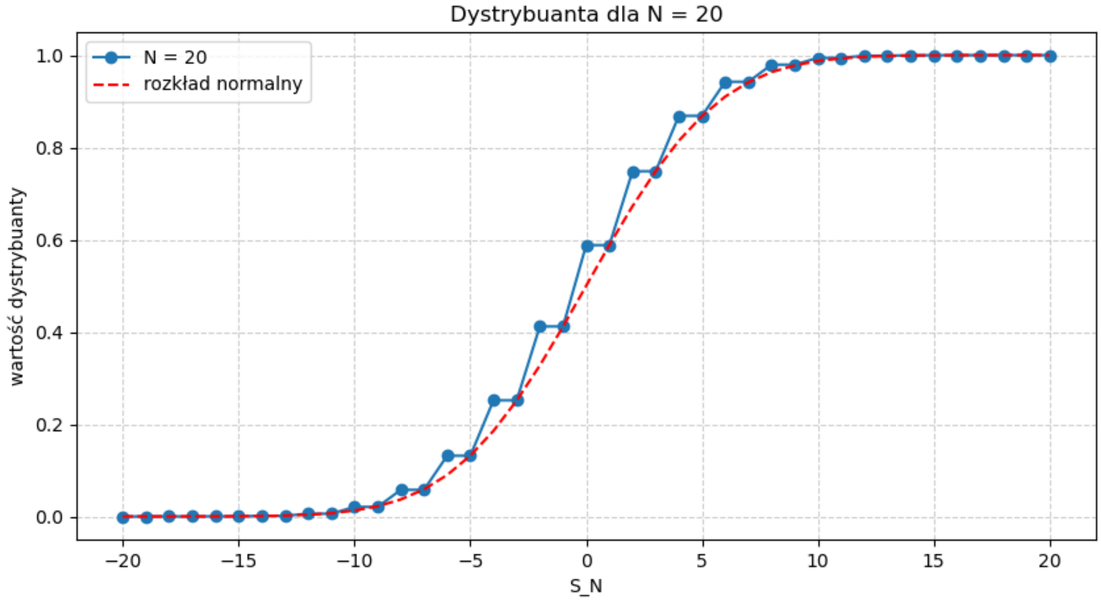
zadanie 2

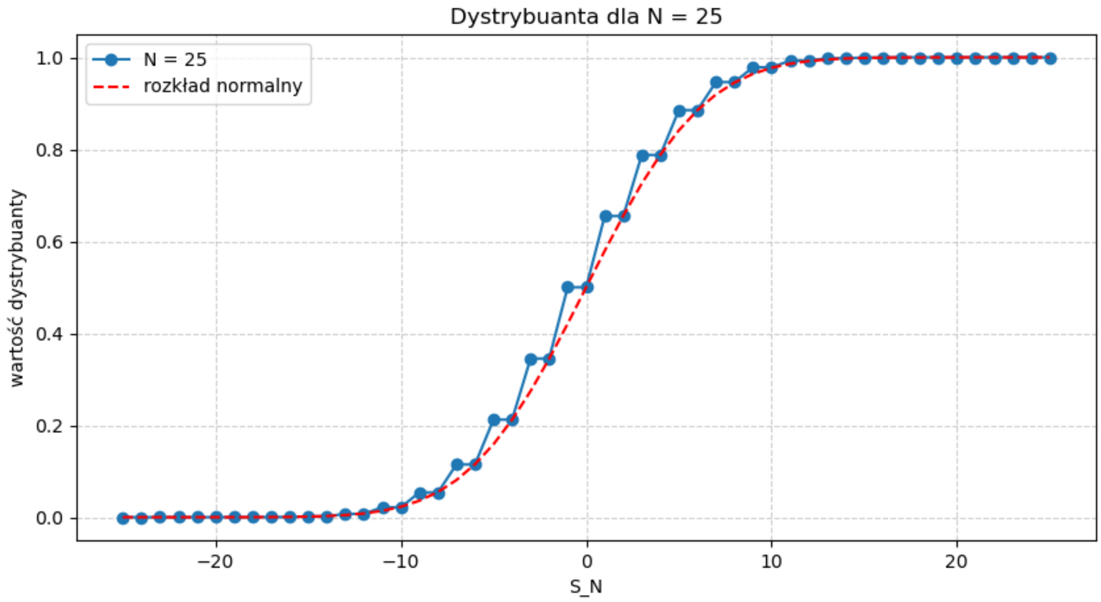
Poniżej zamieszczam wykresy przedstawiające dystrybuantę zmiennej S dla parametrów N = {5, 10, 15, 25, 30, 100} oraz dystrybuantę rozkładu normalnego, który aproksymuje rozkład S_N


Dystrybuanta dla N = 5
N = 5
rozkład normalny
wartość dystrybuanty
S_N

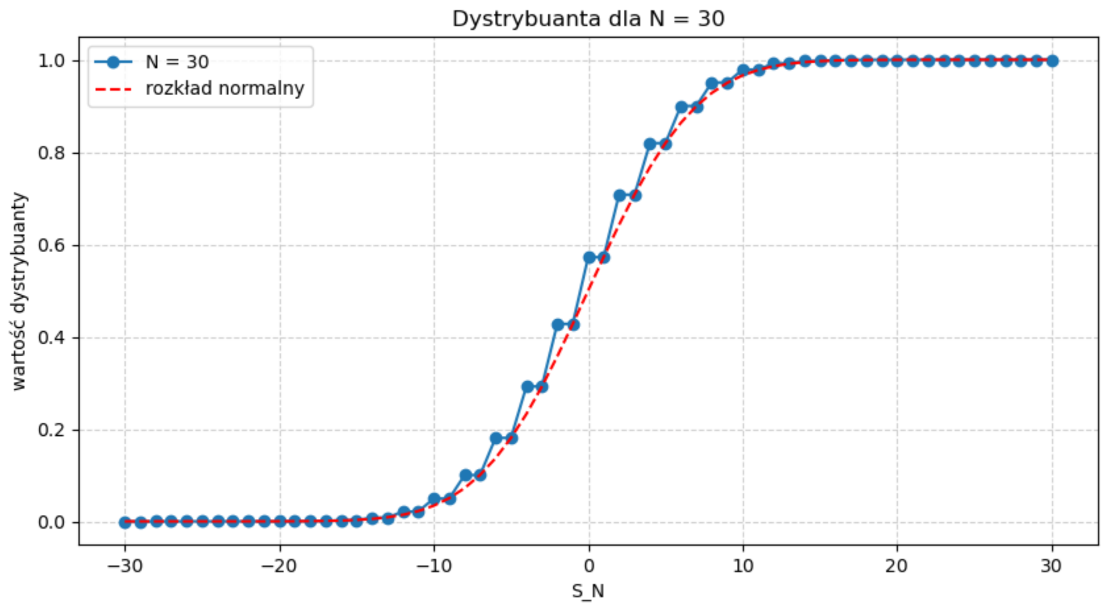

Dystrybuanta dla N = 10
N = 10
rozkład normalny
wartość dystrybuanty
S_N

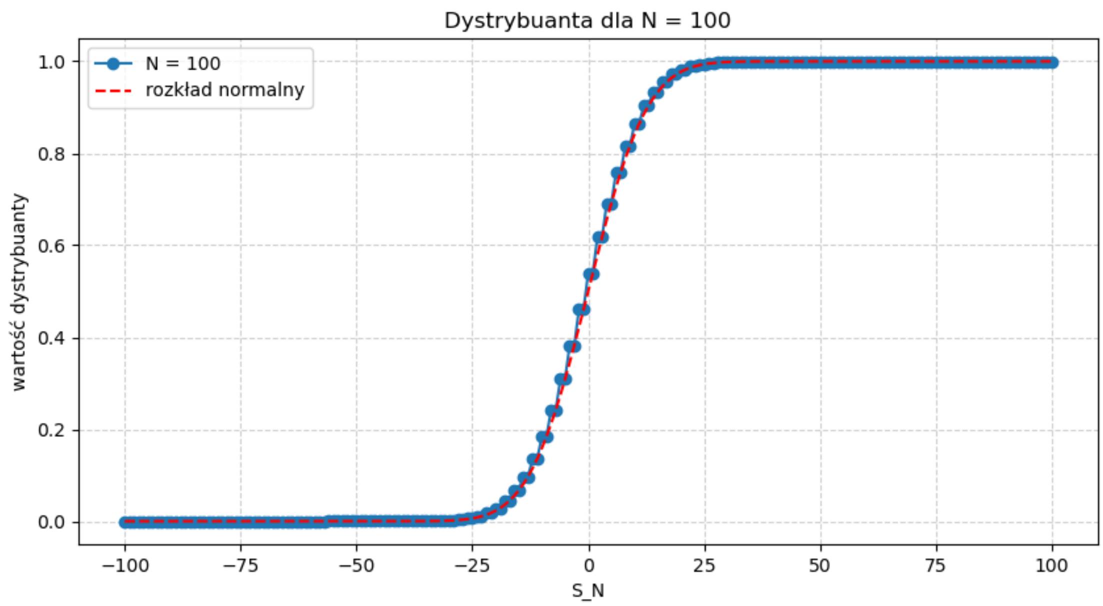

Dystrybuanta dla N = 15
N = 15
rozkład normalny
wartość dystrybuanty
S_N


Dystrybuanta dla N = 20
N = 20
rozkład normalny
wartość dystrybuanty
S_N


Dystrybuanta dla N = 25
N = 25
rozkład normalny
wartość dystrybuanty
S_N

PROF



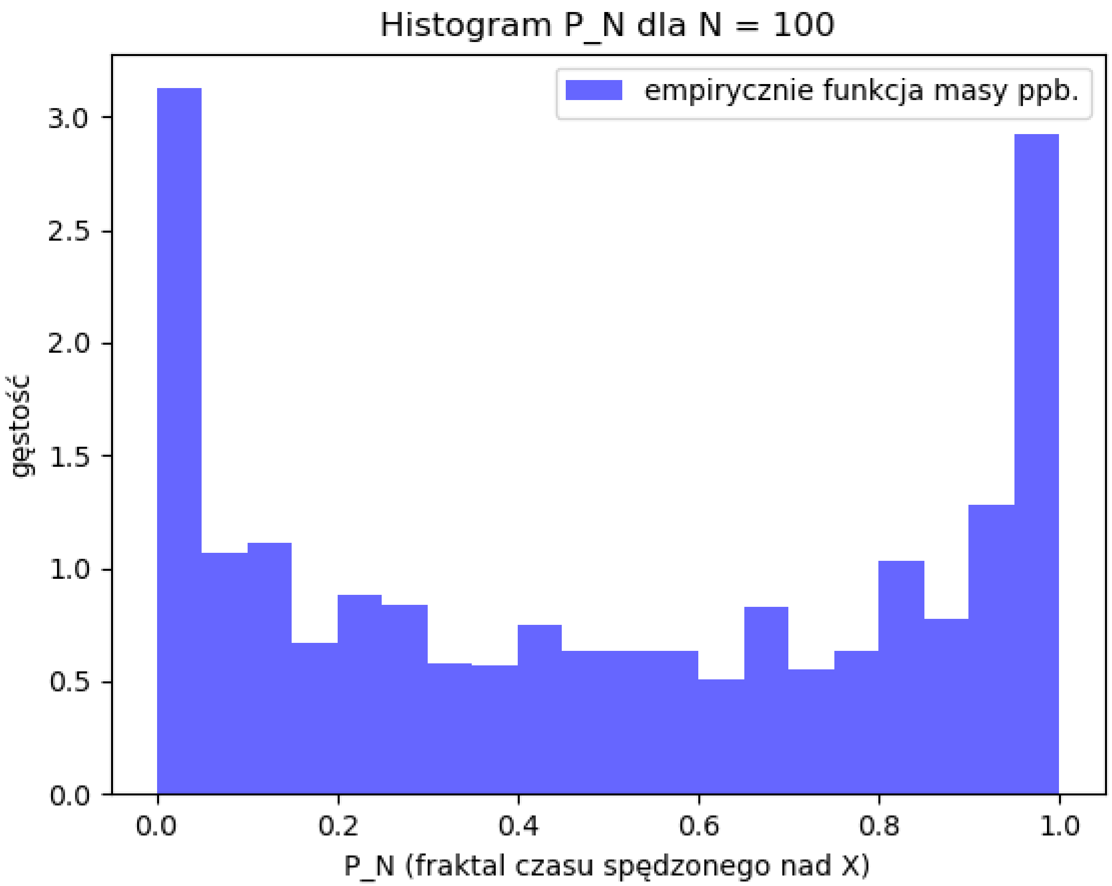


PROF

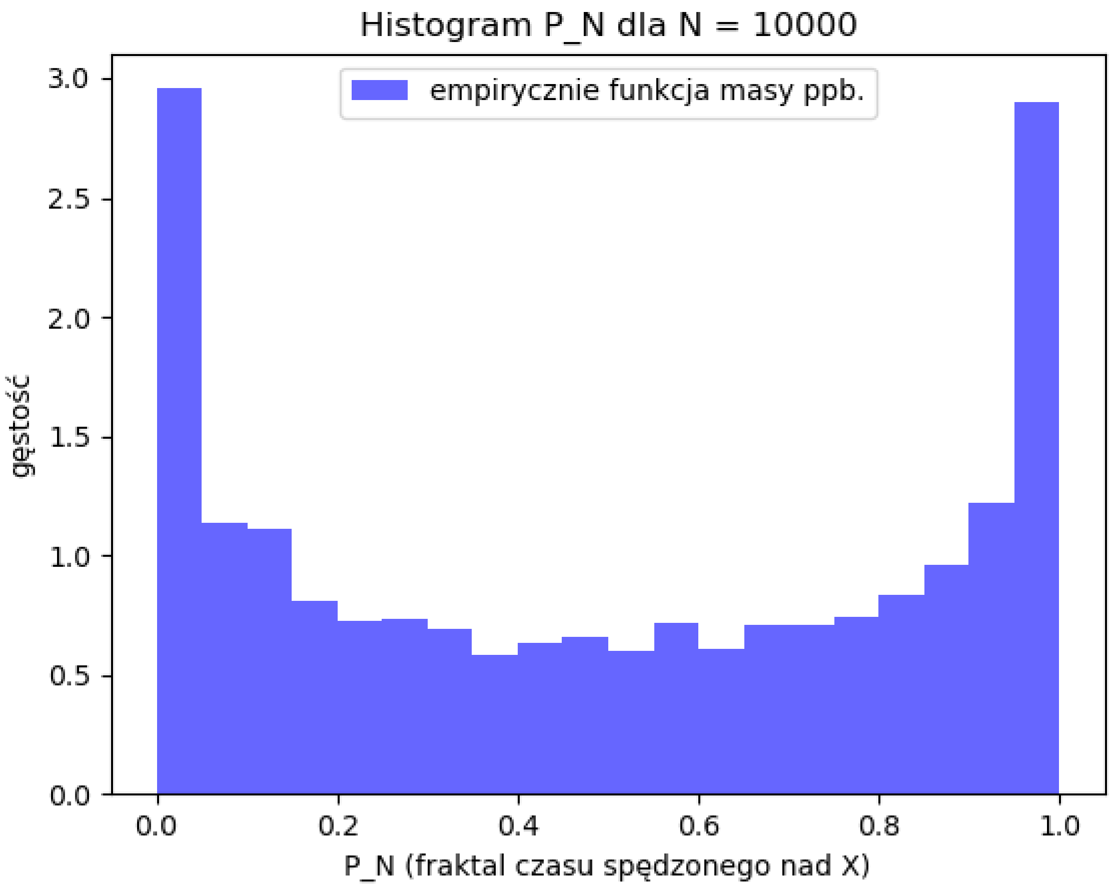
Z wykresów widać, że wraz ze wzrostem N rozkład wartości S_N coraz bardziej przypomina rozkład normalny. Wynika to z centralnego twierdzenia granicznego, które mówi, że suma (lub średnia) dużej liczby niezależnych zmiennych losowych o skończonej wariancji dąży do rozkładu normalnego, niezależnie od pierwotnego rozkładu tych zmiennych.

zadanie 3

Poniżej zamieszczam histogramy, które reprezentują oszacowane wartości funkcji masy prawdopodobieństwa zmiennej losowej P_N, dla wartosci N {100, 1000, 10000}.


Histogram P_N dla N = 100
empirycznie funkcja masy ppb.
gęstość
P_N (fraktal czasu spędzonego nad X)


Histogram P_N dla N = 1000
empirycznie funkcja masy ppb.
gęstość
P_N (fraktal czasu spędzonego nad X)



Histogramy PMF dla P_N wykazują kształt charakterystyczny dla rozkładu arcusa sinus: wartości skrajne (blisko 0 i 1) są częstsze niż wartości środkowe.
Oprócz tego widać, że podobnie jak w poprzednim zadaniu im większe N tym bardziej widać to zbliżenie do PDF arcusa sinus.
Można z tego wywnioskować, że jeżeli proces zacznie spędzać więcej czasu nad lub pod osią X to ma tendencję do utrzymania tego stanu przez dłuższy czas.