Zmienne losowe

1. Wstęp teoretyczny

Zmienne losowe – liczby charakteryzujące rezultat zjawiska losowego.
Zmienne losowe dyskretne – liczby losowe ze skończonego lub przeliczalnego zbioru wartości. Na ogół są to liczby całkowite symbolizujące rozważane zdarzenia losowe, zliczające ich krotność itp.
Zmienne losowe ciągłe – liczby rzeczywiste o losowej wartości, charakteryzujące ilościowo zjawiska losowe.

1.1. Rozkłady prawdopodobieństwa

Zdarzenia losowe odniesione do liczb losowych dotyczą wystąpienia określonych wartości zmiennych dyskretnych oraz wystąpienia wartości zmiennych ciągłych w określonych przedziałach. prawdopodobieństwa takich zdarzeń charakteryzują **rozkłady prawdopodobieństwa** zmiennych losowych.
Dystrybuantą rozkładu zmiennej losowej x nazywamy **prawdopodobieństwo** wystąpienia wartości *x* mniejszej niż argument dystrybuanty (założona wartość zmiennej) *X*:

$$F(X) = P(x < X)$$

Dystrybuanta posiada następujące cechy:

a) $F(-\infty) = 0$;
b) $F(\infty) = 1$;
c) jest funkcją lewostronnie ciągłą i niemalejącą, tzn., jeśli $X_1<X_2$ to $F(X_1) \le F(X_2)$

Dystrybuanta zmiennej losowej dyskretnej zmienia się skokowo w punktach odpowiadających kolejnym wartościom zmiennej.
Rozkład prawdopodobieństwa takich zmiennych wygodniej jest charakteryzować podając wprost prawdopodobieństwa wystąpienia poszczególnych wartości $p(x_i)$. Nazywa się to krótko **rozkładem prawdopodobieństwa zmiennych dyskretnych:**

$f(x)=\{p(x_i);\ i=1,2, ...,N\}$, gdzie N oznacza liczbę możliwych wartości zmiennej x
W przypadku zmiennych losowych **ciągłych** rozkład opisuje się tzw. funkcją **gęstości prawdopodobieństwa** $f(x)$, którą definiuje się jako pochodną dystrybuanty względem zmiennej x:

$$f(x) = \lim_{\Delta x \to o} \frac{F(x+\frac{\Delta x}{2}) - F(x-\frac{\Delta x}{2})}{\Delta x} = \frac{d}{dx}F(x)$$

Uwaga !! Funkcja gęstości prawdopodobieństwa nie jest prawdopodobieństwem, ale pozwala obliczyć *prawdopodobieństwo wystąpienia wartości X w zadanym przedziale* x_1, x_2 z wzoru:

$$P(x_1 \le X \le x_2) = \int_{x_1}^{x_2} f(x)\cdot dx$$

Wynika stąd, że: $F(X) = \int_{-\infty}^{X} f(x)\cdot dx$ oraz $\int_{-\infty}^{\infty} f(x)\cdot dx = 1$

1.2. Momenty

Rozkład prawdopodobieństwa zmiennej ciągłej charakteryzuje się przy pomocy parametrów zwanych momentami. Moment i-tego rzędu $m_i(x)$ definiuje się następująco:

$$m_i(x) = \int_{-\infty}^{\infty} x^i \cdot f(x) \cdot dx$$

Moment rzędu zerowego jest zawsze równy 1.
Moment rzędu pierwszego zmiennej X nazywa się **wartością oczekiwaną** zmiennej losowej X i oznacza się symbolem *E*(X). Wartość oczekiwana jest też nazywana **wartością przeciętną** zmiennej losowej lub **nadzieją matematyczną.**
Dla zmiennej losowej **ciągłej** wartość oczekiwaną wyraża wzór:

$$m_1(X) = E(X) \overset{def}{=} \int_{-\infty}^{\infty} X \cdot f(X) \cdot dX$$

Dla zmiennej **dyskretnej** przyjmującej wartości x_i z prawdopodobieństwem p_i wartość oczekiwaną oblicza się ze wzoru:

$$E(x) = \sum_{i=1}^{\infty} p_i \cdot x_i$$

Właściwości wartości oczekiwanej:

a) Każda ograniczona zmienna losowa ma wartość oczekiwaną.

b) Wartość oczekiwana kombinacji liniowej zmiennych losowych jest kombinacją liniową ich wartości oczekiwanych

$$E(a_1 x_1 + a_2 x_2 + ... + a_N x_N) = a_1 \cdot E(x_1) + a_2 \cdot E(x_2) + ... + a_N \cdot E(x_N)$$

c) Jeśli x_1 i x_2 są niezależnymi zmiennymi losowymi to

$$E(x_1 \cdot x_2) = E(x_1) \cdot E(x_2)$$

Zmienna losowa ciągła *x* będąca odchyłką zmiennej losowej oryginalnej X od jej wartości oczekiwanej $m_1(X)$ nazywa się zmienną losową scentrowaną:

$$x \overset{def}{=} X - m_1(X);$$

Momenty wyższego rzędu można obliczać dla oryginalnych zmiennych lub scentrowanych. Momenty dla zmiennych scentrowanych nazywa się momentami centralnymi.
Centralny moment rzędu drugiego zmiennej nazywa się **wariancją** zmiennej σ_x^2

$$\sigma_X^2 = m_2[X - E(X)] = m_2(x) = E(x^2) = \int_{-\infty}^{\infty} [X - E(X)]^2 f(X) dX = \int_{-\infty}^{\infty} x^2 f(x) dx$$

Właściwości wariancji:

a) Znając pierwszy i drugi moment oryginalnej zmiennej losowej X można obliczyć jej wariancję:

$$\sigma_X^2 = E(X^2) - E^2(X) = m_2(X) - m_1^2(X)$$

b) Jeśli x_1, x_2 ... x_N są niezależnymi zmiennymi losowymi to

$$E[(a_1 x_1 + a_2 x_2 + \mathrm{K} + a_N x_N)^2] = a_1^2 \sigma_{x1}^2 + a_2^2 \sigma_{x2}^2 + \mathrm{K} + a_N^2 \sigma_{xN}^2$$

Odchyleniem średnim (standardowym) lub **dyspersją** **σ** zmiennej losowej nazywamy pierwiastek arytmetyczny z jej wariancji**.**

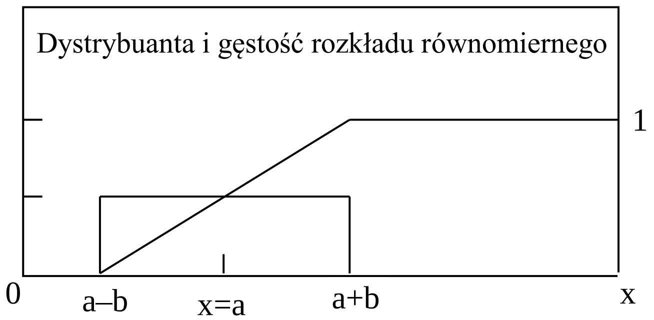
1.3. Podstawowe rozkłady prawdopodobieństwa

Rozkład jednostajny ma zmienna losowa *x*, gdy może przyjmować z tym samym prawdopodobieństwem dowolną wartość z przedziału [a – b, a + b], b > 0 i nie występuje poza tym przedziałem:

$$f(x) = \begin{cases} \dfrac{1}{2b} & gdy \quad x \in [a-b, a+b] \\ 0 & gdy \quad x \notin [a-b, a+b] \end{cases}$$

wartość oczekiwana E(x) = a

wariancja $\sigma_X^2 = \dfrac{b^2}{3}$



Dystrybuanta i gęstość rozkładu równomiernego

Rozkład równomierny przypisuje się zmiennym losowym, których wartości są naturalnie ograniczone i wynikają z oddziaływania pewnego czynnika o czysto losowym charakterze i ograniczonej „sile”. Przykładowo, taki rozkład może mieć kwota wydawana przez jednego klienta w małym sklepie spożywczym lub kiosku.

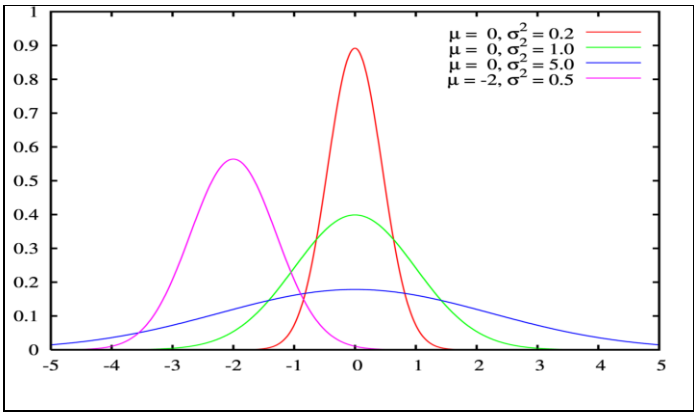
Rozkład normalny czyli **rozkład Gaussa**: dla zmiennej X o wartości oczekiwanej $m=E(X)$ i dyspersji σ (oznaczany symbolem $\boldsymbol{N(m, \sigma)}$):

$$f(X) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \exp\left(-\frac{[X - E(X)]^2}{2 \cdot \sigma^2} \right)$$

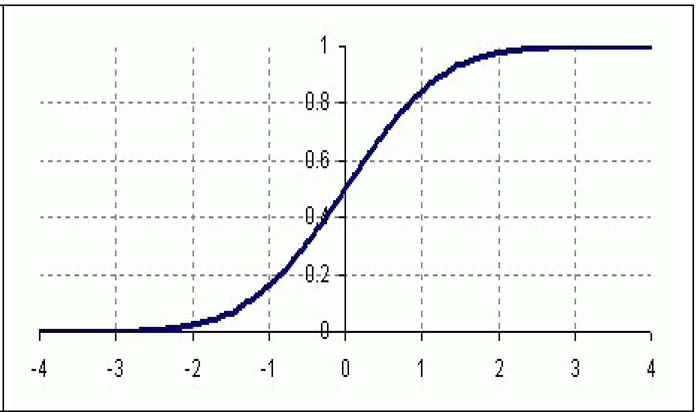
Dla zmiennej standaryzowanej x rozkład Gaussa N(0, 1) ma postać:

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2} \right)$$

i w tej postaci jest on dostępny w tablicach i generatorach liczb losowych.



gęstość rozkładu normalnego



dystrybuanta rozkładu normalnego

2. Zadania praktyczne

2.1. Generowanie sygnałów losowych o zadanym rozkładzie

a) Za pomocą pętli ***for*** i funkcji ***randn*** i ***rand*** wygeneruj sygnał losowy y_n o rozkładzie normalnym oraz sygnał losowy y_j o rozkładzie jednostajnym, o długości N = 1000 próbek:

```
clear
N=1000;
for i=1:N
    yn(i)=randn;
    yj(i)=rand;
end
```

Narysuj te sygnały:

```
subplot(2,2,1)
plot(yn),title('Sygnał o rozkładzie normalnym','FontName','MS Sans Serif')
subplot(2,2,2)
plot(yj),title('Sygnał o rozkładzie jednostajnym','FontName','MS Sans Serif')
```

b) Za pomocą funkcji ***hist*** oblicz i narysuj rozkład tych sygnałów (histogram):

```
subplot(2,2,3)
hist(yn,30),title('Rozkład normalny','FontName','MS Sans Serif')
subplot(2,2,4)
hist(yj,30),title('Rozkład jednostajny','FontName','MS Sans Serif')
```

c) Za pomocą pętli ***for*** oblicz *wartość średnią* $\bar{x}$ oraz *wariancję* σ^2 sygnałów wg wzorów:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \text{wartość średnia}$$

```
%wartość średnia
sum1=0;
sum2=0;
for i=1:N
   sum1=sum1+yn(i);
   sum2=sum2+yj(i);
end
sr_yn=sum1/N;
sr_yj=sum2/N;
disp('Wartość średnia sygnału o rozkładzie normalnym:');
disp(sr_yn)
disp('Wartość średnia sygnału o rozkładzie jednostajnym:');
disp(sr_yj)
```

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N} (x_i - \bar{x})^2 \qquad \text{wariancja}$$

```
%wariancja
sum1=0;
sum2=0;
for i=1:N
    sum1=sum1+(yn(i)-sr_yn)^2;
    sum2=sum2+(yj(i)-sr_yj)^2;
end
war_yn=sum1/(N-1);
war_yj=sum2/(N-1);
disp('Wariancja sygnału o rozkładzie normalnym:');
disp(war_yn)
disp('Wariancja sygnału o rozkładzie jednostajnym:');
disp(war_yj)
```

Uruchom powyższe programy kilka razy i obserwuj wyniki. Ponieważ za każdym razem będzie inna realizacja sygnału losowego, wyniki będą się nieznacznie różnić.

Porównaj obliczone wyniki z wartościami teoretycznymi:

	wartość średnia	wariancja
rozkład normalny	0	1
rozkład jednostajny	1/2	1/12

2.2. Zależność wartości średniej i wariancji od liczby próbek wykorzystanych do ich obliczenia

Oblicz wartość średnią i wariancję sygnałów losowych o rozkładzie normalnym i jednostajnym w funkcji liczby próbek sygnału (tzn. należy policzyć średnią i wariancję dla każdej długości sygnału – od 1 do N próbek).

```
% wartość średnia w funkcji liczby próbek sygnału
N=500;
sum1=0; sum2=0;
for i=1:N
   sum1=sum1+yn(i);
   sum2=sum2+yj(i);
   sr_yn(i)=sum1/i;
   sr_yj(i)=sum2/i;
end
%wizualizacja
figure
subplot(2,1,1)
plot(sr_yn),title('wartość średnia rozkładu normalnego','FontName','MS Sans Serif')
line([0 N],[0 0],'Color','r','LineStyle',':')
subplot(2,1,2)
plot(sr_yj),title('wartość średnia rozkładu jednostajnego','FontName','MS Sans Serif')
line([0 N],[0.5 0.5],'Color','r','LineStyle',':')
```

```
% wariancja w funkcji liczby próbek sygnału
sum1=0; sum2=0;
for i=1:N
   sum1=sum1+(yn(i)-sr_yn(i))^2;
   sum2=sum2+(yj(i)-sr_yj(i))^2;
   war_yn(i)=sum1/i;
   war_yj(i)=sum2/i;
end
%wizualizacja
figure
subplot(2,1,1)
plot(war_yn),title('wariancja rozkładu normalnego','FontName','MS Sans Serif')
line([0 N],[1 1],'Color','r','LineStyle',':')
subplot(2,1,2)
plot(war_yj),title('wariancja rozkładu jednostajnego','FontName','MS Sans
Serif')
line([0 N],[1/12 1/12],'Color','r','LineStyle',':')
```

Uruchom powyższe programy kilka razy i obserwuj wyniki.
Na podstawie wykresów oceń ile próbek sygnału jest potrzebne do wiarygodnego oszacowania wartości średniej i wariancji dla obu rozkładów.

- Zadanie do samodzielnego wykonania

Narysuj wykresy przebiegu kwadratu błędu średniej i wariancji względem wartości teoretycznych (czyli różnicy między wartościami wyliczonymi, a teoretycznymi, podniesione do kwadratu) dla obu rozkładów. Oblicz błąd średniokwadratowy MSE (ang. *Mean Square Error*), czyli sumę przebiegu kwadratu błędu, podzieloną przez liczbę próbek. Dla którego rozkładu obserwujemy najmniejszy błąd?