

# ZMIENNE LOSOWE i ich ROZKŁADY

## (wybrane zagadnienia)

- zmienne losowe (definicja, podział, oznaczenia)
- dystrybuanta, funkcja prawdopodobieństwa, funkcja gęstości
- wybrane parametry rozkładu zmiennej losowej
- standaryzacja zmiennej losowej
- wybrane rozkłady zmiennych losowych (normalny, chi-kwadrat, t-Studenta)
- wykorzystanie tablic statystycznych (odczytywanie informacji)

**Jeżeli wartości zmiennej (cechy) są określone przez przypadek (tzn. przyjmuje ona te wartości z określonymi prawdopodobieństwami), to mówimy, że zmienna ta jest zmienną losową.**

**Zmienne losowe dzielimy na:**

- **ciągłe**; zmienna przyjmuje dowolne wartości z określonego przedziału (w szczególności cały zbiór liczb rzeczywistych)
- **skokowe (dyskretne)**; zmienna przyjmuje dowolne wartości ze zbioru przeliczalnego (np. zbiór liczb całkowitych z określonego przedziału)

**Oznaczenia (analogicznie jak przy cechach statystycznych):**

- duże litery (X, T, U, ...) - zmienna losowa
- małe litery (x, t, u, ...) - wartości zmiennej losowej

## PRZYKŁAD

Rzucamy kostką sześcienną do gry.

Liczba wyrzuconych oczek jest zmienną losową (X).

Wynik każdego rzutu jest wartością tej zmiennej (x).

Zbiór wartości zmiennej losowej jest następujący:  $x \in \{1, 2, 3, 4, 5, 6\}$ .

Zatem liczba wyrzuconych oczek jest zmienną losową skokową (dyskretną).

# **DYSTRYBUANTA zmiennej losowej X**

jest to funkcja zdefiniowana następująco

$$F(x) = P(X \leq x)$$

(czytamy: „Dystrybuanta dla konkretnej wartości zmiennej losowej tj. dla  $X=x$  ) jest równa prawdopodobieństwu tego, że zmienna losowa X będzie przyjmowała wartości nie większe niż konkretna wartość  $x$ ”.)

## **Własności dystrybuanty**

- a.  $0 \leq F(x) \leq 1$
- b.  $F(x)$  jest niemalejąca
- c.  $\lim_{x \rightarrow -\infty} F(x) = 0$  ,  $\lim_{x \rightarrow +\infty} F(x) = 1$

Zmienne losowe są opisywane za pomocą funkcji (rozkładów).

W zależności od rodzaju zmiennej są to:

1. **funkcja prawdopodobieństwa** (zmienne losowe skokowe)
2. **funkcja gęstości** (zmienne losowe ciągłe)

### **Funkcja prawdopodobieństwa zmiennej losowej skokowej**

U podstaw tej funkcji leży uporządkowany zbiór par  $(x_i, p_i)$  gdzie:

$x_i$  - wartości jakie przyjmuje zmienna losowa X

$p_i$  - prawdopodobieństwa z jakimi przyjmuje ona wartości  $x_i$

Funkcja:  $P(X = x_i) = p_i \quad i=1, 2, \dots, n \quad \sum_{i=1}^n p_i = 1$

Dystrybuanta:  $F(x) = \sum_{x_i \leq x} p_i$

## Funkcja gęstości zmiennej losowej ciągłej

Jest to funkcja  $f(x)$  określona na zbiorze liczb rzeczywistych i spełniająca następujące warunki:

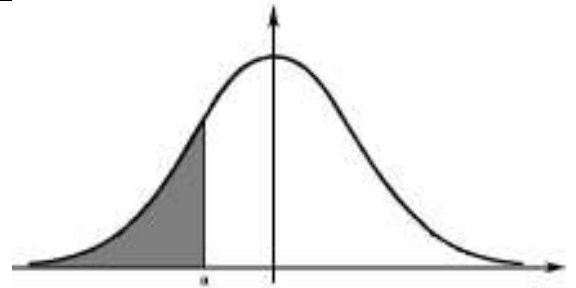
1.  $f(x) \geq 0$                       jest określona nieujemnie

2.  $\int_{-\infty}^{+\infty} f(x) dx = 1$                       pole powierzchni pomiędzy wykresem a osią  $Ox$   
jest równe jedności

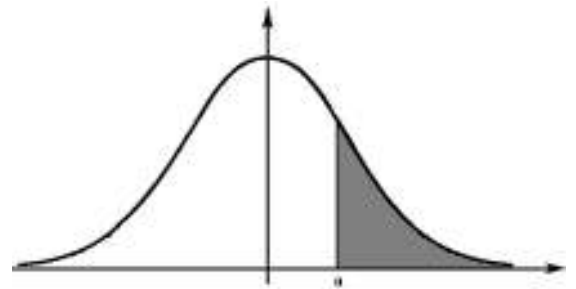
Dystrybuanta: 
$$F(x) = \int_{-\infty}^x f(t) dt$$

### Własności dystrybuanty (cd.)

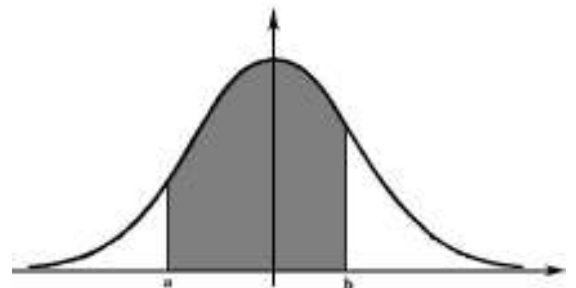
a.  $P(X \leq a) = F(a)$



b.  $P(X \geq a) = 1 - F(a)$



c.  $P(a \leq X \leq b) = F(b) - F(a)$



## Charakterystyki liczbowe rozkładu (wybrane) (parametry rozkładu)

Wartość oczekiwana  $E(X)$  (nadzieja matematyczna)

$$E(X) = m$$

Wartość  $m$  jest to taka wartość zmiennej losowej  $X$ , wokół której skupiają się wyniki wielokrotnych realizacji tej zmiennej. Innymi słowy, oczekuje się (ma się nadzieję), że wielokrotne realizacje zmiennej losowej  $X$  będą skupiały się wokół liczby  $m$ .

Wartość oczekiwana należy do miar położenia. Można ją wyliczyć jako:

$$E(X) = \sum_{i=1}^n p_i x_i \quad - \text{ dla zmiennych losowych skokowych}$$

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx \quad - \text{ dla zmiennych losowych ciągłych}$$

Wariancja  $V(X)$

$$V(X) = \sigma^2$$

Wariancja należy do miar rozproszenia. Można ją wyliczyć jako:

$$V(X) = E[X - E(X)]^2 = E(X^2) - E(X)^2$$

Odchylenie standardowe  $\sigma$

$$\sqrt{V(X)} = \sigma$$

Mediana  $M_e$

Jest to taka wartość zmiennej losowej  $X$ , dla której dystrybuanta wynosi 1/2:

$$F(M_e) = 1/2$$

Podobnie można definiować wiele pozostałych charakterystyk (modalna, kurtoza, współczynnik asymetrii, itp.).

# STANDARYZACJA zmiennej losowej

Dana jest zmienna losowa  $X$  o dowolnym rozkładzie z parametrami:

wartość oczekiwana  $E(X) = m$

odchylenie standardowe  $\sqrt{V(X)} = \sigma$

Zabieg standaryzacji polega na utworzeniu nowej zmiennej losowej  $T$  wg następującego wzoru:

$$T = \frac{X - m}{\sigma}$$

Nowa zmienna losowa  $T$  będzie miała ten sam typ rozkładu co zmienna losowa  $X$ .

Parametry rozkładu nowej zmiennej  $T$  będą zawsze następujące:

wartość oczekiwana  $E(T) = 0$

odchylenie standardowe  $\sqrt{V(T)} = 1$

# Wybrane ROZKŁADY zmiennych losowych i ich TABLICE

## Rozkład NORMALNY

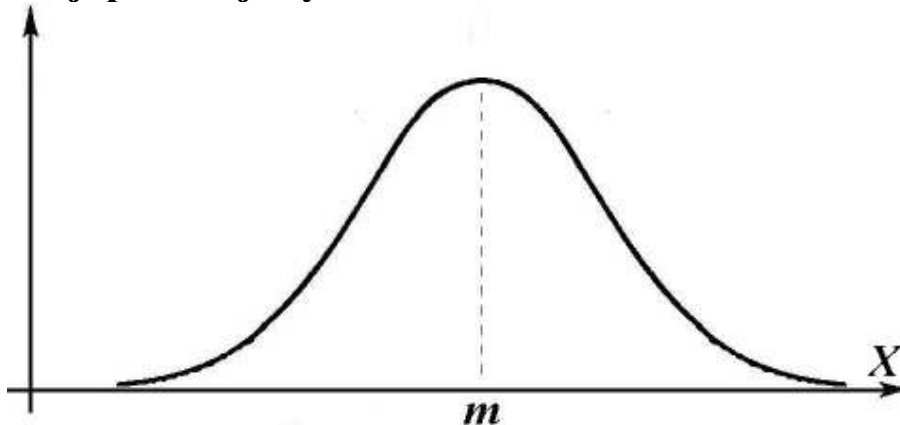
Zmienna losowa  $X$  (ciągła) ma rozkład normalny z parametrami  $m$  i  $\sigma$  jeśli jej funkcja gęstości jest określona wzorem:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad \text{dla } -\infty < x < +\infty$$

Będziemy to zapisywali krótko:

$$X : N(m; \sigma)$$

Wykres tej funkcji pokazuje rysunek:



Rozkład zmiennej losowej  $X$  ma następujące parametry:

wartość oczekiwana  $E(X) = m$

odchylenie standardowe  $\sqrt{V(X)} = \sigma$

modalna (dominanta)  $M_o = m$

mediana  $M_e = m$

Rozkład normalny jest rozkładem symetrycznym ponieważ:

$$E(X) = M_e = M_o = m$$

W celu odczytywania prawdopodobieństw dla zmiennych losowych o rozkładzie normalnym korzysta się z tablic dystrybuanty dla standaryzowanej zmiennej losowej U o rozkładzie normalnym. (w podręczniku [2] jest ona oznaczana przez T).

Każde pytanie o prawdopodobieństwa związane ze zmienną losową  $X$  musi być zawsze poprzedzone standaryzacją:

$$X : N(m; \sigma) \qquad U = \frac{X - m}{\sigma} \qquad U : N(0;1)$$

## PRZYKŁAD 2

Z badań producenta opon wynika, że ich „żywołność” (mierzona przebiegiem) ma rozkład normalny z wartością oczekiwaną 60 [tys. km] i odchyleniem standardowym 8 [tys. km].

Oblicz prawdopodobieństwo tego, że zakupiona opona będzie miała „żywołność” pomiędzy 50, a 80 [tys. km].

Oznaczając „żywołność” opony przez  $X$  możemy zapisać krótko:

$$X: N(60 ; 8).$$

Pytanie kupującego można zapisać również krótko:

$$P(50 < X < 80) = ?.$$

Aby precyzyjnie odpowiedzieć na to pytanie musimy kompleksowo:

1. skorzystać z własności (c) dystrybuanty (s. 3),
2. dokonać standaryzacji wybranych wartości zmiennej losowej  $X$  (standaryzować graniczne „żywołności” opony) oraz
3. odczytać z tablic wartości dystrybuanty dla granicznych „żywołności” poddanych standaryzacji.

$$\begin{aligned} P(50 < X < 80) &= F(80) - F(50) = \\ &= F\left(\frac{80-60}{8}\right) - F\left(\frac{50-60}{8}\right) = \Phi(2,50) - \Phi(-1,25) = \\ &= 0,99389 - 0,10565 = 0,88824 \end{aligned}$$

Zatem szansa, że zakupiona opona będzie miała „żywołność” pomiędzy 50, a 80 [tys. km] wynosi w przybliżeniu 89%.

**Wartość  $\Phi(2,50)$  odczytano na stronie [2] tablic dystrybucyj standardyzowanego rozkładu normalnego  $N(0;1)$**

$u$	<b>0,00</b>	0,01	0,02	0,03	0,04	0,05
0,00	0,50000	0,50399	0,50798	0,51197	0,51595	0,51994
0,10	0,53983	0,54380	0,54778	0,55172	0,55567	0,55962
...	...	...	...	...	...	...
2,40	0,99180	0,99202	0,99224	0,99245	0,99266	0,99286
<b>2,50</b>	<b>0,99389</b>	0,99396	0,99413	0,99430	0,99446	0,99461
2,60	0,99534	0,99547	0,99560	0,99573	0,99586	0,99598

**Wartość  $\Phi(-1,25)$  odczytano na stronie [1] tablic dystrybucyj standardyzowanego rozkładu normalnego  $N(0;1)$**

$u$	0,00	-0,01	...	-0,04	<b>-0,05</b>	-0,06
0,00	0,50000	0,49601		0,48405	0,48006	0,47608
-0,10	0,46017	0,45620		0,44433	0,44038	0,43644
-0,20	0,42064	0,41684		0,40516	0,40129	0,39743
-0,30	0,38209	0,37828		0,36693	0,36317	0,35942
-0,40	0,34458	0,34090		0,32997	0,32635	0,32276
-0,50	0,30844	0,30503		0,29460	0,29116	0,28774
-0,60	0,27425	0,27093		0,26109	0,25785	0,25463
-0,70	0,24196	0,23885		0,22965	0,22663	0,22363
-0,80	0,21185	0,20897		0,20045	0,19766	0,19489
-0,90	0,18406	0,18141		0,17361	0,17106	0,16853
-1,00	0,15865	0,15625		0,14917	0,14686	0,14457
-1,10	0,13567	0,13350		0,12714	0,12507	0,12302
<b>-1,20</b>	0,11507	0,11314		0,10749	<b>0,10565</b>	0,10383
-1,30	0,09680	0,09510		0,09012	0,08851	0,08691
-1,40	0,08076	0,07927		0,07493	0,07353	0,07214
-1,50	0,06671	0,06552		0,06178	0,06057	0,05938



Ważną umiejętnością, wykorzystywaną do wnioskowania na podstawie próby (patrz wykłady 9 i 10), jest „odwrotne” odczytywanie w tablicach dystrybucyjności  $N(0;1)$ .

„Odwrotność” odczytu polega na tym, że zadajemy sobie pewne prawdopodobieństwo i odczytujemy jakiej wartości zmiennej normalnej ono odpowiada.

### PRZYKŁAD 3

Jaka jest wartość zmiennej losowej  $U:N(0;1)$ , która spełnia warunek

$$P( U < ? ) \approx 0,05$$

Jest to równoważne znalezieniu takiej liczby „?”, która spełnia warunek

$$\Phi( ? ) \approx 0,05$$

W tablicy dystrybucyjności rozkładu normalnego  $N(0;1)$  na stronie [1] (w jej części wewnętrznej) wyszukujemy wartość najbliższą liczbie 0,05 .

Na brzegach tablicy (*boczek i główka tablicy*) odczytujemy poszukiwaną wartość zmiennej  $U$ .

<i>u</i>	0,00	...	-0,03	-0,04	-0,05	-0,06
0,00	0,50000		0,48803	0,48405	0,48006	0,47608
-0,10	0,46017		0,44828	0,44433	0,44038	0,43644
-0,20	0,42064		0,40905	0,40516	0,40129	0,39743
-0,30	0,38209		0,37070	0,36693	0,36317	0,35942
-1,40	0,08076		0,07636	0,07493	0,07353	0,07214
-1,50	0,06671		0,06301	0,06178	0,06057	0,05938
-1,60	0,05480		0,05155	0,05050	0,04947	0,04846
-1,70	0,04457		0,04181	0,04093	0,04006	0,03920

Poszukiwaną wartością zmiennej losowej  $U$  jest liczba -1,64 .

Spełnia ona warunek  $P( U < -1,64 ) \approx 0,05$  .

Zauważmy jednocześnie, że liczba przeciwna do niej (1,64) spełnia warunek  $P( U < 1,64 ) = \Phi( 1,64 ) \approx 0,95$  .

Oznacza to, że na mocy własności (b) ze s. 3 spełnia ona również warunek  $P( U > 1,64 ) \approx 0,05$

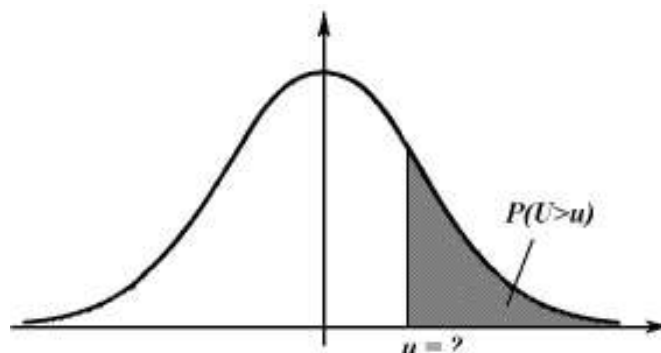
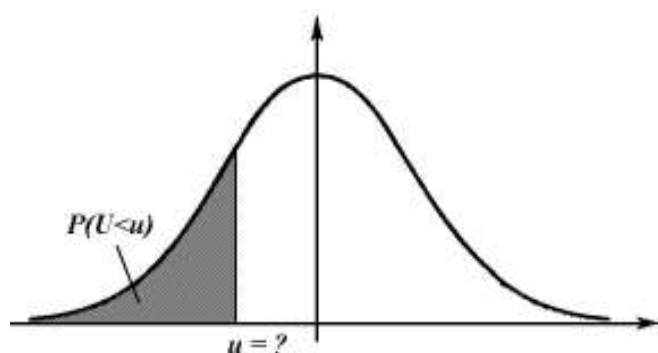
ponieważ  $P( U > 1,64 ) = 1 - P( U < 1,64 ) = 1 - \Phi( 1,64 ) \approx 1 - 0,95 \approx 0,05$  .

# Tabela przydatnych odczytów z tablic dystrybuanty rozkładu normalnego $N(0;1)$

W wykładach 9 i 10 (estymacja, weryfikacja hipotez statystycznych) często znajdzie potrzeba odczytania tzw. wartości krytycznych.

**Dobrym ćwiczeniem w celu sprawnego posługiwania się tablicami dystrybuanty rozkładu  $N(0;1)$  (przy wnioskowaniu statystycznym na podstawie informacji z próby) będzie uzupełnienie poniższej tabeli.**

dane	szukane		dane	dane do odczytu	szukane
$P(U < u)$	$u$		$P(U > u)$	$P(U > u) = 1 - P(U < u)$	$u$
0,005			0,005		
0,010			0,010		
0,020			0,020		
0,025			0,025		
0,050	-1,64		0,050	0,950	1,64
0,100			0,100		
0,200			0,200		



# Rozkład chi-kwadrat (rozkład $\chi^2$ )

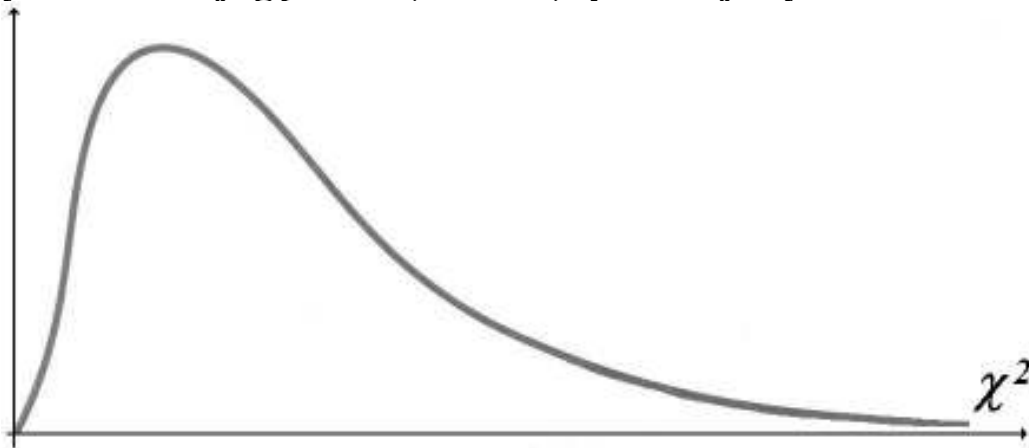
Danych jest  $k$  ciągłych zmiennych losowych o rozkładzie normalnym z wartością oczekiwaną 0 i odchyleniem standardowym 1, tj. każda zmienna  $X_i : N(0;1)$  ( $i=1,2, \dots, k$ ).

Zdefiniujemy nową zmienną losową o nazwie chi-kwadrat ( $\chi^2$ ):

$$\chi^2 = X_1^2 + X_2^2 + \dots + X_k^2$$

Rozkład tak zdefiniowanej zmiennej  $\chi^2$  nazywamy rozkładem zmiennej losowej chi-kwadrat o  $k$  stopniach swobody.

Typowy wykres funkcji gęstości (dla  $k>2$ ) pokazuje rysunek:



Rozkład zmiennej losowej  $\chi^2$  o  $k$  stopniach swobody ma następujące parametry:

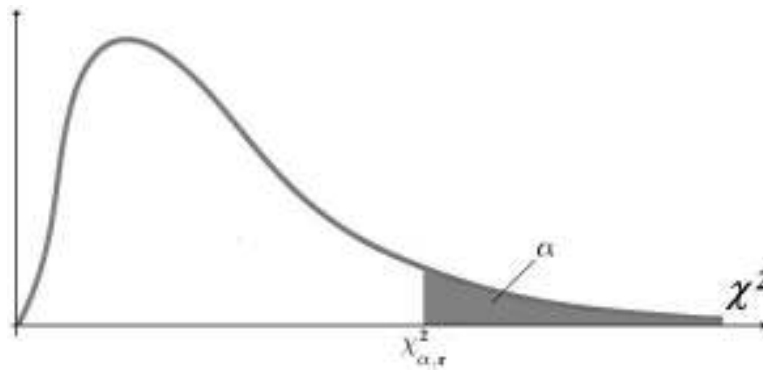
wartość oczekiwana  $E(\chi^2) = k$

odchylenie standardowe  $\sqrt{V(\chi^2)} = \sqrt{2k}$

Rozkład zmiennej losowej  $\chi^2$  o  $k$  stopniach swobody jest rozkładem pomocniczym używanym we wnioskowaniu statystycznym.

Tablice zmiennej losowej  $\chi^2$  o  $k$  stopniach swobody zostały opracowane tak, że podają przy założonym prawdopodobieństwie ( $\alpha$ ) taką wartość (oznaczymy ją  $\chi^2_{\alpha, k}$ ) zmiennej losowej  $\chi^2$ , dla której :

$$P(\chi^2 > \chi^2_{\alpha, k}) = \alpha$$



**UWAGA !!!** Stopnie swobody są oznaczone w tablicach przez ***r***.

#### PRZYKŁAD 4

Jaka jest wartość zmiennej losowej  $\chi^2$  o 5 stopniach swobody, która spełnia warunek  $P(\chi^2 > \chi^2_{0,05;5}) = 0,05$  ?

<b>r</b>	<b><math>\alpha</math></b>				
	<b>0,99</b>	...	<b>0,10</b>	<b>0,05</b>	<b>0,02</b>
<b>1</b>	0,0 <sup>3</sup> 157	...	2,706	3,841	5,412
<b>2</b>	0,0201	...	4,605	5,991	7,824
<b>3</b>	0,115	...	6,251	7,815	9,837
<b>4</b>	0,297	...	7,779	9,488	11,668
<b>5</b>	0,554	...	9,236	<b>11,070</b>	13,388
<b>6</b>	0,872	...	10,645	12,592	15,033

Poszukiwaną wartością zmiennej losowej  $\chi^2$  o 5 stopniach swobody jest liczba  $\chi^2_{0,05;5} = 11,07$ .

Spełnia ona warunek  $P(\chi^2 > 11,07) = 0,05$

# Rozkład t - Studenta

Dana są dwie zmienne losowe:

1. zmienna losowa  $U:N(0;1)$  oraz

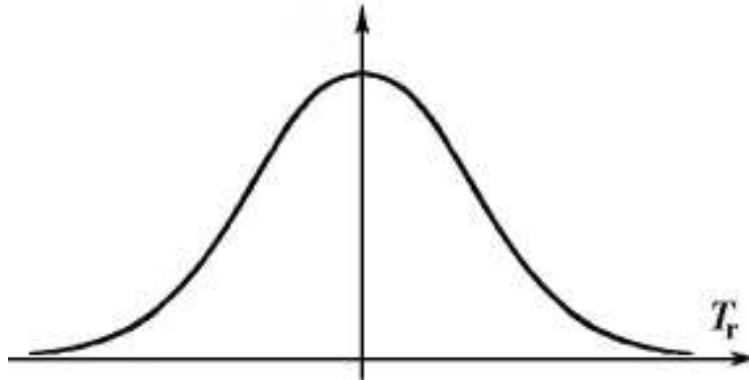
2. zmienna losowa  $\chi^2$  o  $k$  stopniach swobody

Zdefiniujemy nową zmienną losową postaci :

$$T_k = \frac{U}{\sqrt{\chi^2}} \sqrt{k}$$

Rozkład tak zdefiniowanej zmiennej  $T_k$  nazywamy rozkładem zmiennej losowej t-Studenta o  $k$  stopniach swobody.

Wykres funkcji gęstości pokazuje rysunek:



Rozkład zmiennej losowej t-Studenta o  $k$  stopniach swobody ma następujące parametry:

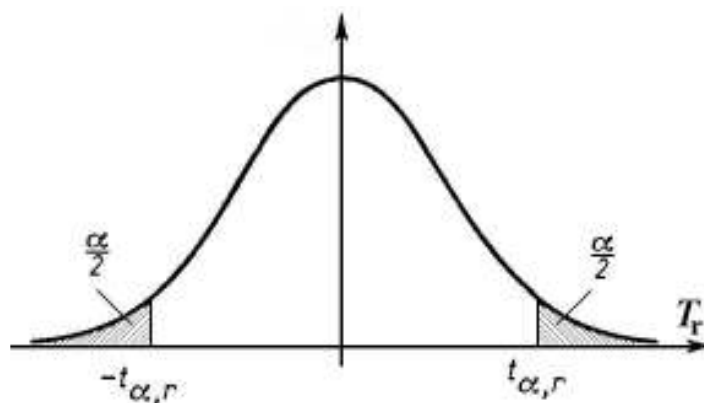
wartość oczekiwana  $E(T_k) = 0$

odchylenie standardowe  $\sqrt{V(T_k)} = \sqrt{\frac{k}{k-2}}$

Rozkład zmiennej losowej t-Studenta  $T_k$  o  $k$  stopniach swobody jest rozkładem pomocniczym używanym we wnioskowaniu statystycznym.

Tablice zmiennej losowej t-Studenta  $T_k$  o  $k$  stopniach swobody zostały opracowane tak, że podają przy założonym prawdopodobieństwie ( $\alpha$ ) taką wartość (oznaczymy ją  $t_{\alpha, k}$ ) zmiennej losowej  $T_k$ , dla której :

$$P(|T_k| \geq t_{\alpha, k}) = \alpha$$



**UWAGA !!!** Stopnie swobody są oznaczone w tablicach przez  **$r$** .

## PRZYKŁAD 5

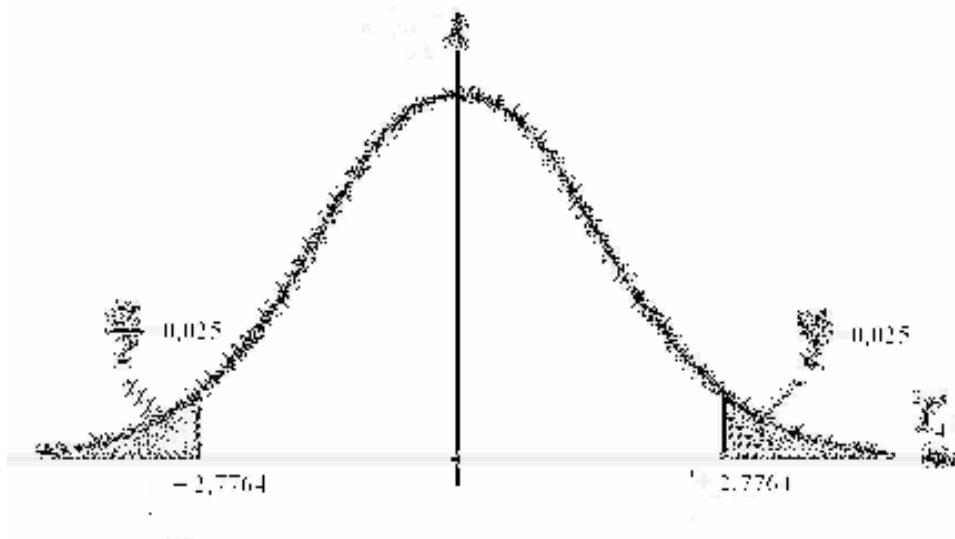
Jaka jest wartość zmiennej losowej t-Studenta o 4 stopniach swobody, która spełnia warunek  $P(|T_k| \geq t_{0,05; 4}) = 0,05$  ?

$\alpha$ <b>r</b>	0,5	...	0,1	0,05	0,02	...	$\alpha$ <b>r</b>
1	1,0000	...	6,3138	12,7062	31,8205	...	1
2	0,8165		2,9200	4,3027	6,9646		2
3	0,7649		2,3534	3,1824	4,5407		3
4	0,7407		2,1318	2,7764	3,7469		4
5	0,7267		2,0150	2,5706	3,3649		5
6	0,7176		1,9432	2,4469	3,1427		6

Jest to liczba  $t_{0,05; 4} = 2,7764$ . Spełnia ona warunek

$$P(|T_4| \geq 2,7764) = 0,05$$

Ilustruje to rysunek:



## PRZYKŁAD 6

Jaka jest wartość zmiennej losowej t-Studenta o 4 stopniach swobody, która spełnia warunek  $P(T_k \geq t_{0,05;4}) = 0,05$  ?

Aby poprawnie odczytać musimy podwoić prawdopodobieństwo  $\alpha$ .

$\alpha$ r	0,5	...	0,1	0,05	0,02	...	$\alpha$ r
1	1,0000	...	6,3138	12,7062	31,8205	...	1
2	0,8165		2,9200	4,3027	6,9646		2
3	0,7649		2,3534	3,1824	4,5407		3
4	0,7407		2,1318	2,7764	3,7469		4
5	0,7267		2,0150	2,5706	3,3649		5
6	0,7176		1,9432	2,4469	3,1427		6

Poszukiwana liczba to  $t_{0,05;4} = 2,1318$ .

Spełnia ona warunek  $P(T_4 \geq 2,1318) = 0,05$ .

## PRZYKŁAD 7

Jaka jest wartość zmiennej losowej t-Studenta o **4** stopniach swobody, która spełnia warunek  $P(T_k \leq t_{0,05;4}) = 0,05$  ?

Tutaj również aby poprawnie odczytać musimy podwoić prawdopodobieństwo  $\alpha$ .

$\alpha$ <b>r</b>	<b>0,5</b>	...	<b>0,1</b>	<b>0,05</b>	<b>0,02</b>	...	$\alpha$ <b>r</b>
<b>1</b>	1,0000	...	6,3138	12,7062	31,8205	...	<b>1</b>
<b>2</b>	0,8165		2,9200	4,3027	6,9646		<b>2</b>
<b>3</b>	0,7649		2,3534	3,1824	4,5407		<b>3</b>
<b>4</b>	0,7407		<b>2,1318</b>	2,7764	3,7469		<b>4</b>
<b>5</b>	0,7267		2,0150	2,5706	3,3649		<b>5</b>
<b>6</b>	0,7176		1,9432	2,4469	3,1427		<b>6</b>

Odczytaną z tablic liczbę należy wziąć z **przeciwnym** znakiem.

Poszukiwana liczba to  $t_{0,05;4} = -2,1318$ .

Spełnia ona warunek  $P(T_4 \leq -2,1318) = 0,05$ .