

RACHUNEK PRAWDOPODOBIENSTWA

WYKŁAD 5.

PODSTAWOWE ROZKŁADY PRAWDOPODOBIENSTWA

Rozkłady skokowe

Rozkład jednopunktowy

Określamy:

$P(X = c) = 1$ gdzie c ustalona liczba.

$EX = c, \quad D^2X = 0$
(tylko ten rozkład ma zerową wariancję !!!)

Rozkład dwupunktowy (zerojedynekowy)

Niech $p \in (0, 1)$ będzie ustaloną liczbą.

Określamy:

$P(X = 0) = q, P(X = 1) = p$; gdzie $q = 1 - p$.

Umowa:

0 - porażka

1 - sukces

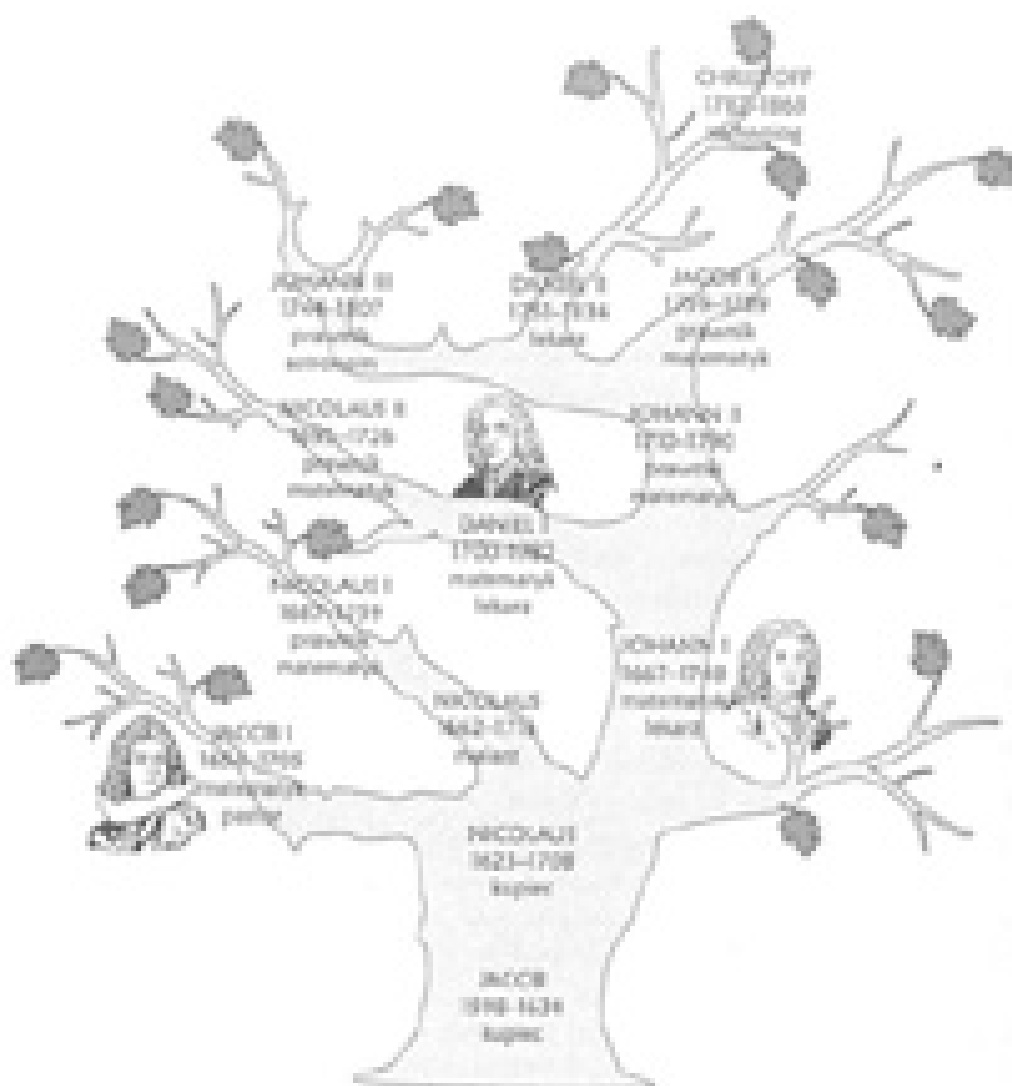
$$EX = p, \quad D^2X = pq$$

Rozkład dwumianowy

Dla danych $p \in (0, 1)$, $n \in N$ określamy funkcję prawdopodobieństwa

$$P(X = k) = \binom{n}{k} p^k q^{n-k} \quad (\text{wzór Bernoulliego})$$

gdzie $q = 1 - p$
 $k = 0, 1, 2, \dots, n$.





Jakub Bernoulli (1654 - 1705) - szwajcarski matematyk i fizyk.

Jeśli przyjmiemy, że n oznacza liczbę niezależnych doświadczeń z których każde kończy się jednym z dwóch wyników: „sukcesem” (z prawdopodobieństwem p w każdym doświadczeniu) lub „porażką” i zmienna losowa X oznacza liczbę „sukcesów” to powyższy wzór wyznacza **prawdopodobieństwo uzyskania dokładnie k sukcesów w n doświadczeniach** (próbach).

Sprawdzenie

$$\sum_{k=0}^n P(X = k) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = (p + q)^n = 1$$

$$EX = np, \quad D^2X = npq$$

Przykład

Prawdopodobieństwo uszkodzenia kserokopiarki przed upływem gwarancji wynosi 0,2. Firma zakupiła 6 kserokopiarek. Obliczyć prawdopodobieństwo, że przed upływem gwarancji 2 kserokopiarki ulegną uszkodzeniu. Jaka jest najbardziej prawdopodobna liczba uszkodzonych kserokopiarek przed upływem gwarancji.

X – liczba uszkodzonych kserokopiarek przed upływem gwarancji,

$$P(X = 2) = \binom{6}{2} 0,2^2 0,8^4 = 15 \cdot 0,04 \cdot 0,4096 = 0,24576$$

Uwaga

x_k	0	1	2	3	4	5	6
p_k	0,2621	0,3932	0,2458	0,0819	0,0154	0,0015	0,0001

Przykład

Rzucamy 4 razy kostką sześcienną. Jakie jest prawdopodobieństwo, że w co najmniej 3 rzutach liczba oczek będzie podzielna przez 3?.

Szukane prawdopodobieństwo to
 $P(X \geq 3) = P(X = 3) + P(X = 4)$,
gdzie „sukcesem” jest uzyskanie 3 lub 6
oczek, więc $p = 1/3$.

Zatem

$$P(X = 3) = \binom{4}{3} \left(\frac{1}{3}\right)^3 \left(\frac{2}{3}\right)^1 = 4 \cdot \frac{2}{81} = \frac{8}{81}$$

$$P(X = 4) = \binom{4}{4} \left(\frac{1}{3}\right)^4 \left(\frac{2}{3}\right)^0 = 1 \cdot \frac{1}{81} = \frac{1}{81}$$

$$P(X \geq 3) = P(X = 3) + P(X = 4) = \frac{8}{81} + \frac{1}{81} = \frac{1}{9}$$

Przykład

Obliczymy wartość oczekiwaną rozkładu dwumianowego.

$$\begin{aligned} EX &= \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} = \sum_{k=1}^n k \frac{n!}{k!(n-k)!} p^k q^{n-k} = \\ &= np \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} q^{n-k} = np(p+q)^{n-1} = np \end{aligned}$$

Rozkład geometryczny

X - liczba prób Bernoulliego **poprzedzających** pierwszy sukces

$$P(X = k) = pq^k$$

$$q = 1 - p$$

$$k = 0, 1, 2, \dots$$

Sprawdzenie

$$\sum_{k=0}^{\infty} P(X = k) = \sum_{k=0}^{\infty} pq^k = \frac{p}{1-q} = 1$$

$$EX = q/p; \quad D^2X = q/p^2$$

Rozkład Poissona

Dla $\lambda > 0$ określamy funkcję
prawdopodobieństwa

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad k = 0, 1, 2, \dots$$



Siméon Denis Poisson (1781–1840), francuski mechanik teoretyk, fizyk i matematyk. W matematyce zajmował się całkami oznaczonymi, równaniami różnicowymi i różniczkowymi oraz teorią prawdopodobieństwa.

Sprawdzenie

$$\begin{aligned}\sum_{k=0}^{\infty} P(X = k) &= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1\end{aligned}$$

$$EX = \lambda \quad D^2X = \lambda$$

Przykład

Obliczymy wartość oczekiwaną rozkładu Poissona.

$$EX = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

Rozkład Poissona (możliwość odczytu w tablicy) może dla dużych n (praktycznie $n \geq 30$) i małych p (praktycznie $p \leq 0,2$) przybliżać rozkład dwumianowy
(przybliżenie Poissona)

$$\binom{n}{k} p^k q^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{gdzie } \lambda = n \cdot p$$

Przykład

W pudełku jest 400 żarówek. Jakie jest prawdopodobieństwo, że wśród nich jest 5 żarówek wadliwych, jeśli wadliwość produkcji takich żarówek wynosi 0,5%?

Jaka jest najbardziej prawdopodobna liczba uszkodzonych żarówek w tym pudełku?

Zastosujemy przybliżenie Poissona,
 $\lambda = n \cdot p = 400 \cdot 0,005 = 2$.

W tablicy rozkładu Poissona (tablica I)
odczytamy, że:

$$P(X = 5) = 0,0361$$

Również w tablicy rozkładu Poissona
odczytamy, że najbardziej prawdopodobna
liczba uszkodzonych żarówek w tym pudełku
to 1 lub 2 (dla obu tych liczb
prawdopodobieństwo jest równe 0,2707).

Rozkłady ciągłe

Rozkład jednostajny

Rozkład którego gęstość jest stała w pewnym przedziale nazywamy **jednostajnym**.

Gęstość rozkładu jednostajnego w (a, b)

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a; b) \\ 0 & x \notin (a; b) \end{cases}$$

Ponieważ gęstość ta ma oś symetrii w punkcie $x = (a + b)/2$ to

$$EX = (a+b)/2$$

Pokażemy, że

$$D^2X = (b - a)^2/12$$

Przykład

Najpierw obliczymy EX^2

$$\begin{aligned} EX^2 &= \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{b-a} \frac{x^3}{3} \Big|_a^b = \\ &= \frac{1}{b-a} \left(\frac{b^3}{3} - \frac{a^3}{3} \right) = \frac{a^2 + 2ab + b^2}{3} \end{aligned}$$

Zatem

$$D^2 X = EX^2 - (EX)^2 = \frac{a^2 + 2ab + b^2}{3} - \left(\frac{a+b}{2} \right)^2 = \frac{(b-a)^2}{12}$$

Rozkład wykładniczy

Rozkład ten występuje często w zagadnieniach rozkładu czasu między zgłoszeniami (awariami) lub czasu oczekiwania na obsługę w systemach kolejkowych.

Gęstość rozkładu wykładniczego o parametrze $a > 0$ ma postać

$$f(x) = \begin{cases} ae^{-ax} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

dystribuantą tego rozkładu jest funkcja

$$F(x) = \begin{cases} 1 - e^{-ax} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

(uzasadnienie: $F'(x) = f(x)$)

Przykład

Obliczmy EX

$$EX = \int_0^{\infty} xae^{-ax} dx = \left(-xe^{-ax} - \frac{1}{a}e^{-ax} \right) \Big|_0^{\infty} = \frac{1}{a}$$

$$D^2X = \frac{1}{a^2}$$

Własność.

1) Jeśli liczba zgłoszeń w systemie kolejkowym w przedziale czasu $(t, t + T)$ ma rozkład Poissona o parametrze λT , oraz liczby zgłoszeń przychodzące w rozłącznych przedziałach czasu są niezależne to czas X między kolejnymi zgłoszeniami ma rozkład wykładniczy o parametrze $a = 1/\lambda$.

2) Dla dowolnych $t, T > 0$ mamy

$$P(X \geq t + T \mid X \geq t) = P(X \geq T)$$

(własność braku pamięci)

$$\begin{aligned} P(X \geq t + T \mid X \geq t) &= \frac{P(X \geq t + T \wedge X \geq t)}{P(X \geq t)} = \frac{P(X \geq t + T)}{P(X \geq t)} = \\ &= \frac{e^{-(t+T)a}}{e^{-ta}} = e^{-Ta} = P(X \geq T) \end{aligned}$$

Jest to jedyny rozkład ciągły o tej własności.

Rozkład normalny (Gaussa) $N(m, \sigma)$

Dla $m \in R, \sigma \in (0, +\infty)$

Określamy gęstość rozkładu

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad x \in R$$



Carl Friedrich Gauss (1777 - 1855) – niemiecki matematyk i fizyk. Jego badania związane z teorią błędów doprowadziły do odkrycia rozkładu normalnego zmiennej losowej (nazywany także rozkładem Gaussa), który jest najważniejszym rozkładem w teorii prawdopodobieństwa.



Uwaga

Jeśli X ma rozkład $N(m, \sigma)$ to zmienna losowa $Y = (X - m)/\sigma$ ma rozkład $N(0, 1)$ (takie przekształcenie nazywamy **standaryzacją**).

Wartości dystrybuanty dla argumentów
ujemnych wyznaczamy na podstawie
zależności

$$\Phi(-x) = 1 - \Phi(x)$$

Przykład

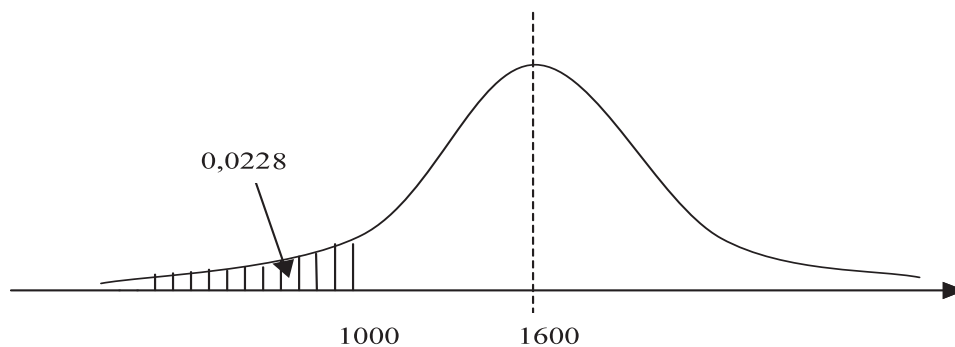
Dochód miesięczny (zł) w pewnej populacji osób ma rozkład normalny $N(1600; 300)$.

Jaki procent osób w tej populacji ma dochód miesięczny poniżej 1000 zł?

X – wysokość miesięcznego dochodu

$$P(X < 1000) = P\left(\frac{X - 1600}{300} < \frac{1000 - 1600}{300}\right) = P(Y < -2) =$$

$$= \Phi(-2) = 1 - \Phi(2) = 1 - 0,9772 = 0,0228 = 2,28\%$$



Przykład

Czas wykonania pewnego detalu (min.) jest zmienną losową o rozkładzie normalnym $N(m; \sigma)$. Wiadomo, że 80% robotników wykonuje ten detal dłużej niż 10 minut a 60% robotników dłużej niż 12 minut.

- a) wyznacz parametry rozkładu czasu wykonania detalu m i σ ,
- b) jaki odsetek robotników wykonuje ten detal w czasie krótszym niż 6 minut?

X – czas wykonania detalu.

$$P(X > 10) = 0,8 \quad \text{stąd} \quad \frac{m-10}{\sigma} = 0,84$$

$$P(X > 12) = 0,6 \quad \text{stąd} \quad \frac{m-12}{\sigma} = 0,25$$

Rozwiązując powyższy układ równań otrzymamy $m = 12,85$; $\sigma = 3,39$.

$$\begin{aligned} P(X < 6) &= P\left(\frac{X-12,85}{3,39} < \frac{6-12,85}{3,39}\right) = P(Y < -2,02) = \\ &= \Phi(-2,02) = 1 - \Phi(2,02) = 0,0217 = 2,17\% \end{aligned}$$

Prawo trzech sigm

Jeśli X ma rozkład $N(m, \sigma)$ to

$$P(m - \sigma < X < m + \sigma) = 0,683,$$

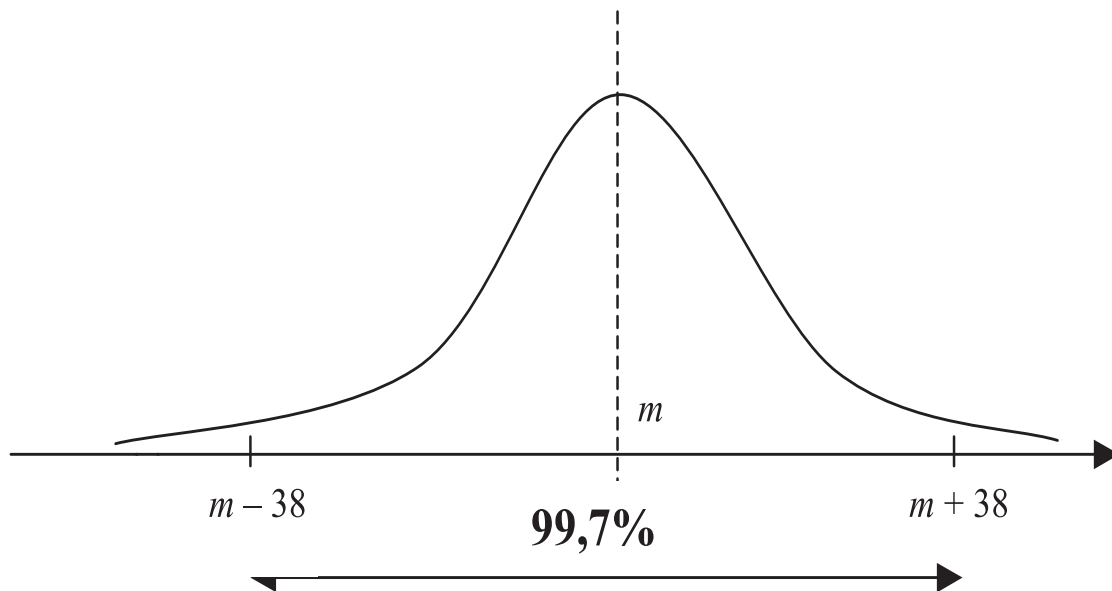
$$P(m - 2\sigma < X < m + 2\sigma) = 0,955,$$

$$P(m - 3\sigma < X < m + 3\sigma) = 0,997$$

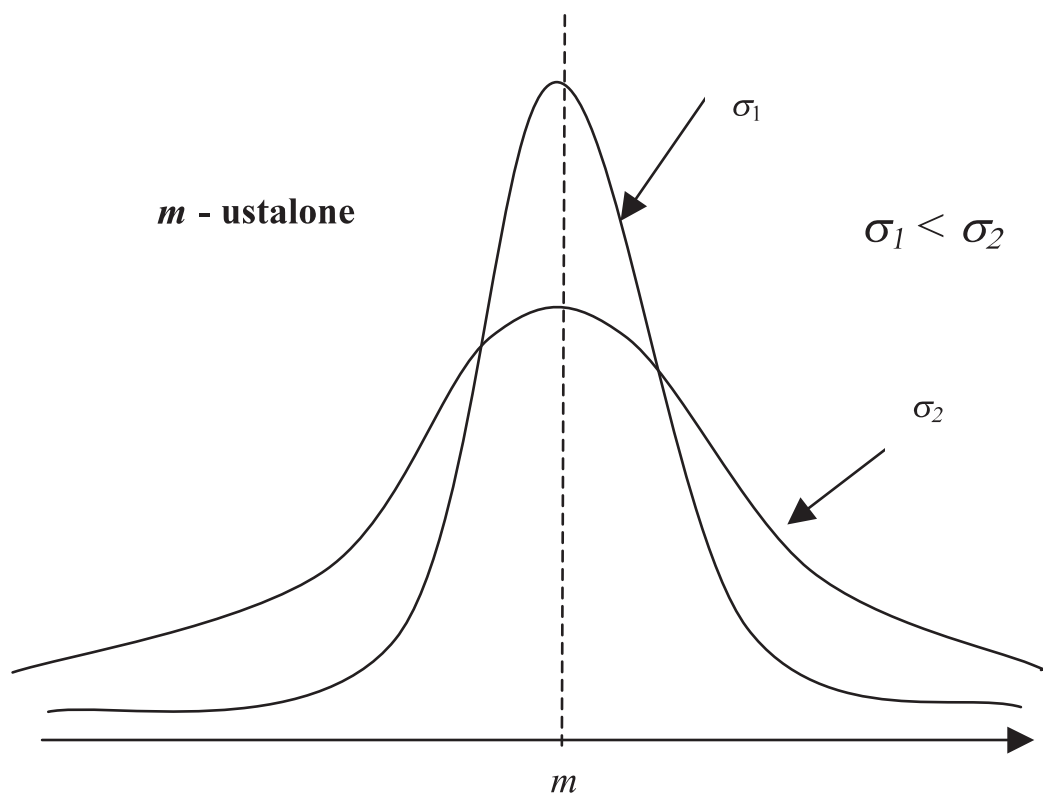
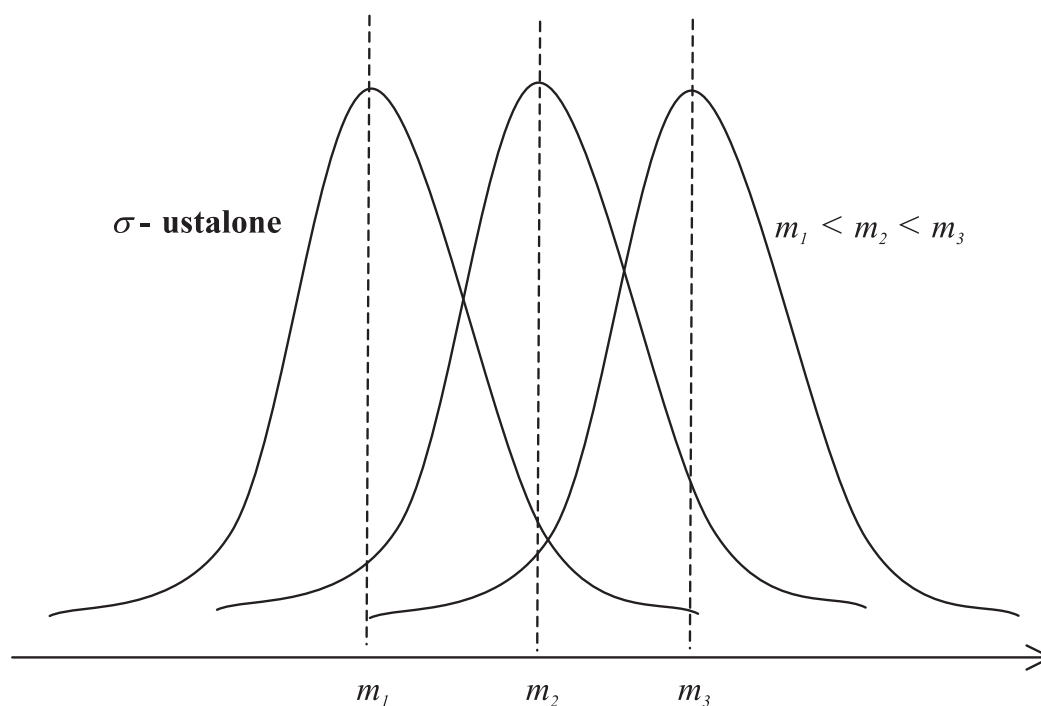
Ostatnia równość świadczy o tym, że chociaż rozkład normalny ma gęstość różną od zera na całej prostej to praktycznie niemal wszystkie realizacje skupiają się w przedziale

$$(m - 3\sigma, m + 3\sigma)$$

własność tą nazywamy **prawem trzech sigm**.



Interpretacja graficzna parametrów rozkładu $N(m, \sigma)$



Trzy rozkłady ciągłe, które mają duże znaczenie w statystyce matematycznej:

- **Rozkład chi kwadrat,**
- **Rozkład Studenta,**
- **Rozkład F – Snedecora**

Rozkłady te są tablicowane.

Rozkład chi kwadrat (χ^2) Y_n

n – liczba stopni swobody

$$Y_n = X_1^2 + \dots + X_n^2$$

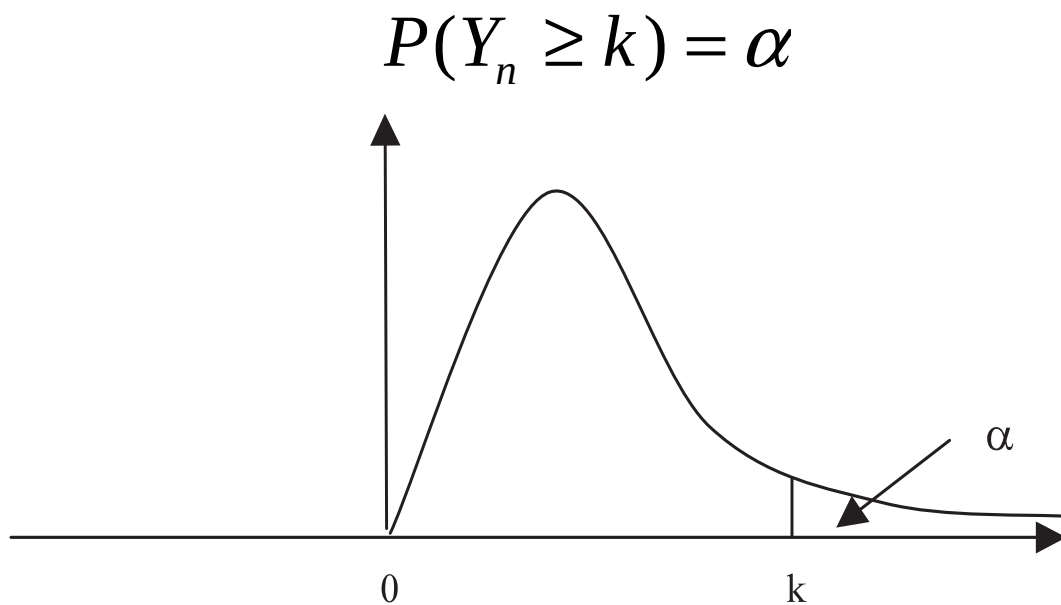
$X_1, \dots, X_n$ - niezależne, o rozkładzie $N(0, 1)$

$$EX = n;$$

$$D^2X = 2n$$

Odczyt z tablicy dla rozkładu chi kwadrat.

(podobnie interpretujemy graficznie odczyt z tablicy F – Snedecora.)



Uwaga.

1) Dla $n = 1, 2$ wykres gęstości rozkładu chi kwadrat jest inny (tylko część malejąca wykresu)

2) dla $n > 30$ stosujemy przybliżenie rozkładem normalnym.

$$\sqrt{2Y_n} \sim N(\sqrt{2n-1}; 1)$$

Rozkład Studenta T_n

n – liczba stopni swobody

$$T_n = \frac{X}{\sqrt{Y_n}}$$

X, Y_n - niezależne

X o rozkładzie $N(0, 1)$;

Y_n o rozkładzie chi kwadrat z n stopniami swobody

$$EX = 0 ; \quad \text{dla } n > 1$$

$$D^2X = n/(n-2) \quad \text{dla } n > 2$$

Uwaga. $T_n \xrightarrow{n \rightarrow \infty} N(0,1)$

Odczyt z tablicy (tablica IV) dla rozkładu Studenta.

$$P(|T_n| \geq k) = \alpha$$

