

# Rachunek prawdopodobieństwa i statystyka

Rachunek prawdopodobieństwa

Prof. UEK dr hab. Paweł Ulman

## Rachunek prawdopodobieństwa - definicje

Zdarzenie elementarne – wyniki doświadczenia losowego

Własności zdarzeń elementarnych:

1. Dane zdarzenie elementarne może zaistnieć lub nie
2. Jedno ze zdarzeń elementarnych na pewno zaistnieje
3. Zaistnienie jednego zdarzenia elementarnego wyklucza zaistnienie innego.

Przestrzeń zdarzeń elementarnych – oznaczenie  $E$  lub  $\Omega$

## Rachunek prawdopodobieństwa - definicje

Przestrzeń zdarzeń elementarnych może być:

1. Skończona
2. nieskończona, ale przeliczalna
3. nieskończona i nieprzeliczalna

Ad 1) zbiór wszystkich podzbiorów zbioru  $E$  jest mocy  $2^n$ .

Ad 2) zbiór wszystkich podzbiorów zbioru  $E$  jest mocy  $2^{\aleph_0}$ , czyli continuum.

Ad 3) zbiór wszystkich podzbiorów zbioru  $E$  jest mocy  $2^c > c$ .

## Rachunek prawdopodobieństwa - definicje

W sytuacji Ad 3) należy ograniczyć się do rozważania klasy zbiorów borelowskich ( $\sigma$  ciała – przeliczalnie addytywnym ciałem zbiorów). Jest to niepusta klasa  $Z$  podzbiorów przestrzeni zdarzeń elementarnych  $E$  spełniająca trzy warunki:

1.  $E \in Z$
2.  $A \in Z \Rightarrow \bar{A} \in Z$
3.  $A_1 \in Z, A_2 \in Z, \dots \Rightarrow (A_1 \cup A_2 \cup \dots) \in Z.$

Wybiera się więc spośród wielu możliwych ciał podzbiorów  $E$  ciało najmniejsze, które nazywamy  $\sigma$  ciałem.  $\sigma$  ciało istnieje i nie zawiera zbiorów niemierzalnych.

## Rachunek prawdopodobieństwa - definicje

Zdarzenie losowe – każdy element  $\sigma$  ciała podzbiorów przestrzeni zdarzeń elementarnych.

$E$  – zdarzenie pewne;  $\emptyset$  - zdarzenie niemożliwe;

$\bar{A}$  – zdarzenie przeciwne do zdarzenia  $A$ ;

Zdarzenia  $A$  i  $B$  są parami rozłączne, gdy  $A \cap B = \emptyset$ ;

Jeśli  $A \subset B$  to zdarzenie  $A$  pociąga za sobą zdarzenie  $B$ ;

Suma zdarzeń  $A \cup B$ ;

Równość zdarzeń  $A = B$ .

## Rachunek prawdopodobieństwa - definicje

### Definicja prawdopodobieństwa Laplace'a

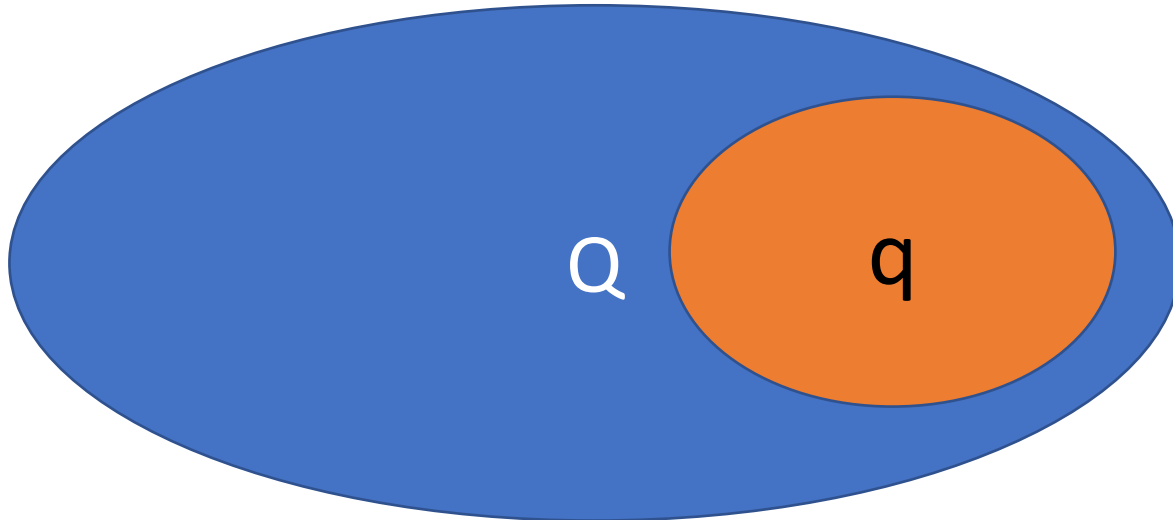
Jeśli zdarzenie  $E$  rozkłada się na  $n$  wykluczających się wzajemnie i jednakowo możliwych zdarzeń elementarnych, spośród których  $m$  sprzyja zaistnieniu interesującego nas zdarzenia  $A$ , to prawdopodobieństwo zaistnienia zdarzenia  $A$  nazywamy ułamek:

$$P(A) = \frac{m}{n}$$

## Rachunek prawdopodobieństwa - definicje

### Geometryczna definicja prawdopodobieństwa

Jeśli  $Q$  i  $q$  są to dwa zbiory w przestrzeni  $r$  wymiarowej oraz jeśli  $q \subset Q$  to prawdopodobieństwo, że dowolny punkt należący do  $Q$  będzie również należał do  $q$  równa się stosunkowi miary zbioru  $q$  do miary zbioru  $Q$ .



## Rachunek prawdopodobieństwa - definicje

### Statystyczna definicja prawdopodobieństwa

Jeśli przy wielokrotnej realizacji doświadczenia, w wyniku którego może wystąpić zdarzenie  $A$ , częstość tego zdarzenia przejawia wyraźną prawidłowość oscylując wokół pewnej nieznanej liczby  $p$  i jeśli wahania częstości przejawiają tendencję malejącą w miarę wzrostu liczby doświadczeń to liczba  $p$  nazywa się prawdopodobieństwem zdarzenia  $A$ .



## Rachunek prawdopodobieństwa - definicje

Aksjomatyczna definicja prawdopodobieństwa (Kołmogorow, 1931)

Niech  $E$  będzie przestrzenią zdarzeń elementarnych doświadczenia losowego  $D$ ,  $Z$  – jego zbiorem zdarzeń losowych.

Prawdopodobieństwem nazywamy funkcję  $P$  przyporządkowującą każdemu zdarzeniu  $A \in Z$  liczbę  $P(A)$  zgodnie z warunkami:

1.  $P(A) \geq 0$
2.  $P(E) = 1$
3. Jeśli  $A_1, A_2, \dots$  jest dowolnym ciągiem parami rozłącznych zdarzeń ze zbioru  $Z$  to  $P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$

## Rachunek prawdopodobieństwa - własności

Własności prawdopodobieństwa:

1.  $P(\emptyset) = 0$ ;
2. Jeśli  $A \subset B$  to  $P(A) \leq P(B)$ ;
3.  $P(A) \leq 1$ ;
4.  $P(A) + P(\bar{A}) = 1$ ;
5.  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Przestrzeń probabilistyczna:  $(E, Z, P)$

## Rachunek prawdopodobieństwa - własności

Prawdopodobieństwo warunkowe:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}; \quad A, B \in \mathcal{Z}, P(B) > 0.$$

Prawdopodobieństwo iloczynu zdarzeń:

$$P(A \cap B) = P(A) * P(B|A) = P(B) * P(A|B), \text{ gdy odpowiednio } P(A) \text{ i } P(B) > 0$$

$$P(A \cap B \cap C) = P(A) * P(A|B) * P(C|A \cap B), \text{ gdy } P(A \cap B) > 0.$$

## Rachunek prawdopodobieństwa - własności

Niezależność dwóch zdarzeń:

$$P(A \cap B) = P(A) * P(B).$$

Niezależność zespołowa zdarzeń:

Zdarzenia  $A_1, A_2, \dots, A_n$  są niezależne zespołowo, gdy prawdopodobieństwo łącznego zaistnienia dowolnych  $m \leq n$  różnych zdarzeń spośród nich jest równe iloczynowi prawdopodobieństwa tych zdarzeń.

$$P(A \cap B \cap C) = P(A) * P(B) * P(C)$$

oraz  $P(A \cap B) = P(A) * P(B)$  oraz  $P(A \cap C) = P(A) * P(C)$  oraz  $P(B \cap C) = P(B) * P(C)$

## Rachunek prawdopodobieństwa - własności

### Twierdzenie o prawdopodobieństwie zupełnym

Jeśli B jest dowolnym zdarzeniem, zdarzenia  $A_1, A_2, \dots, A_n$  stanowią układ zupełny zdarzeń (wykluczają się parami i wypełniają całą przestrzeń zdarzeń elementarnych) to prawdopodobieństwo zdarzenia B wyraża się wzorem:

$$P(B) = P(A_1) * P(B|A_1) + P(A_2) * P(B|A_2) + \dots + P(A_n) * P(B|A_n)$$

## Rachunek prawdopodobieństwa - własności

### Twierdzenie Bayes'a

Jeśli zdarzenie B jest dowolnym zdarzeniem o dodatnim prawdopodobieństwie, zdarzenia  $A_1, A_2, \dots, A_n$  stanowią układ zupełny zdarzeń to:

$$P(A_i|B) = \frac{P(A_i) * P(B|A_i)}{P(B)}$$

## Rachunek prawdopodobieństwa - własności

Przykład

70 kobiet i 30 mężczyzn; 35 kobiet i 10 mężczyzn uzyskało ocenę 5,0

$A_1$  – kobieta;  $A_2$  – mężczyzna;  $B$  – otrzymanie oceny 5,0

$$\begin{aligned} P(B) &= P(A_1) * P(B|A_1) + P(A_2) * P(B|A_2) = \\ &= 70/100 * 35/70 + 30/100 * 10/30 = 45/100 \end{aligned}$$

$$P(A_1|B) = \frac{P(A_1) * P(B|A_1)}{P(B)} = \frac{\frac{70}{100} * \frac{35}{70}}{45/100} = 0,78$$

## Rachunek prawdopodobieństwa – zmienna losowa

Zmienna losowa – taka wielkość, która w wyniku doświadczenia losowego przyjmuje określoną wartość, znaną po zrealizowaniu doświadczenia, ale nie dającą się przewidzieć przed realizacją tego doświadczenia.

Zmienną losową  $X$  nazywa się funkcję  $X=X(e)$  określoną na zbiorze zdarzeń elementarnych  $E$ , o wartościach ze zbioru liczb rzeczywistych taką, że dla każdej liczby rzeczywistej  $x$  zbiór  $A$  zdarzeń elementarnych  $e \in E$ , dla których  $X(e) < x$  spełnia warunek  $A \in \mathcal{Z}$ .

Zmienna dyskretna i ciągła



# Rachunek prawdopodobieństwa – zmienna losowa

## Dystrybuanta zmiennej losowej

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

Własności:

1.  $0 \leq F(x) \leq 1$
2.  $\lim_{x \rightarrow -\infty} F(x) = 0; \lim_{x \rightarrow +\infty} F(x) = 1$
3.  $F(x)$  jest funkcją niemalejącą
4.  $F(x)$  jest funkcją co najmniej lewostronnie ciągłą
5.  $P(a \leq X < b) = F(b) - F(a)$

## Rachunek prawdopodobieństwa – zmienna losowa

Rozkład prawdopodobieństwa zmiennej losowej skokowej

$$P(X = x_i) = p_i \qquad \sum_{i=1}^n p_i = 1$$

Charakterystyki rozkładu

Wartość oczekiwana -  $E(X) = \sum_{i=1}^n x_i p_i$

$$E(a) = a$$

$$E(X+a) = E(X) + a; E(X*a) = E(X) * a$$

$$E(X-E(X)) = 0$$

$$E(X+Y) = E(X) + E(Y); E(X*Y) = E(X) * E(Y), \text{ gdy } X, Y \text{ są niezależne}$$

## Rachunek prawdopodobieństwa – zmienna losowa

### Wariancja

$$D^2(X) = \sum_{i=1}^n (x_i - E(X))^2 p_i = E(X^2) - E(X)^2$$

$$D^2(a) = 0$$

$$D^2(a * X) = a^2 * D^2(X)$$

$$D^2(X+a) = D^2(X)$$

$$D^2(X+Y) = D^2(X) + D^2(Y), \text{ gdy } X, Y \text{ są niezależne}$$

# Rachunek prawdopodobieństwa – rozkłady prawdopodobieństwa

Rozkład jednopunktowy

$$P(X=x) = 1; \quad E(X)=x; \quad D^2(X) = 0$$

Rozkład dwupunktowy

$$P(X=x_1) = p_1; P(X=x_2) = p_2;$$

Rozkład zero-jedynkowy

$$P(X=0) = 1-p=q; P(X=1) = p; E(X)=p; \quad D^2(X) = p \cdot q = p \cdot (1-p)$$

Rozkład jednostajny skokowy

$$P(X=x_1) = 1/n; P(X=x_2) = 1/n; \dots, P(X=x_n) = 1/n$$

$$E(X) = \frac{1}{n} \sum_{i=1}^n x_i; \quad D^2(X) = \frac{1}{n} \sum_{i=1}^n (x_i - E(X))^2$$

# Rachunek prawdopodobieństwa – rozkłady prawdopodobieństwa

Rozkład Bernoulliego (dwumianowy);  $X \sim B(n, p)$

Założenia:

1.  $n$  niezależnych doświadczeń losowych
2. każde z tych doświadczeń może zakończyć się sukcesem lub porażką
3. prawdopodobieństwo sukcesu jest jednakowe dla każdego doświadczenia i wynosi  $p$

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

$$E(X) = np.;$$

$$D^2(X) = npq$$

## Rachunek prawdopodobieństwa – rozkłady prawdopodobieństwa

Rozkład Poissona;  $X \sim P(\lambda)$

Warunki jak w przypadku r. Bernoulliego i jeden dodatkowy

Liczba doświadczeń jest duża (zmierza do nieskończoności)

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

$$E(X) = \lambda;$$

$$D^2(X) = \lambda$$

# Rachunek prawdopodobieństwa – rozkłady prawdopodobieństwa

Rozkłady zmiennej losowej ciągłej

Gęstość prawdopodobieństwa

$$P(x_1 \leq X < x_2) = F(x_2) - F(x_1),$$

gdy  $\Delta x = x_2 - x_1 \rightarrow 0$  to  $P(x_1 \leq X < x_2) = dF(x) + r(x)$

$dF(x) = F'(x)dx$  więc

$$F'(x) = \lim_{\Delta x \rightarrow 0} \frac{F(x_2) - F(x_1)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{P(x_1 \leq X < x_2)}{\Delta x}$$

Jeśli dystrybuanta  $F(x)$  ma pochodną w punkcie  $x$  to pochodna ta nazywa się gęstością prawdopodobieństwa zmiennej losowej  $X$  w punkcie  $x$ , więc

# Rachunek prawdopodobieństwa – rozkłady prawdopodobieństwa

Rozkłady zmiennej losowej ciągłej

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

Własności:

$$f(x) \geq 0; \quad P(x_1 \leq X < x_2) = \int_{x_1}^{x_2} f(t)dt$$

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \quad P(X = x) = 0$$



# Rachunek prawdopodobieństwa – rozkłady prawdopodobieństwa

## Rozkłady zmiennej losowej ciągłej

Zmienną losową  $X$  przyjmującą wszystkie wartości z pewnego przedziału, dla której istnieje nieujemna funkcja  $f$  taka, że dystrybuantę  $F$  zmiennej losowej  $X$  można przedstawić w postaci  $F(x) = \int_{-\infty}^x f(t)dt$  nazywamy zmienną losową ciągłą, a funkcję  $f$  jej gęstością.

Zmienna losowa  $X$  jest zmienną ciągłą w danym przedziale, jeśli w tym przedziale gęstość  $f(x)$  istnieje i jest funkcją ciągłą względem  $x$  w całym przedziale z wyjątkiem co najwyżej skończonej liczby punktów.

# Rachunek prawdopodobieństwa – rozkłady prawdopodobieństwa

## Rozkłady zmiennej losowej ciągłej

Wartość oczekiwana

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

Wariancja

$$D^2(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 f(x) dx = E(X^2) - E(X)^2$$

# Rachunek prawdopodobieństwa – rozkłady prawdopodobieństwa

## Rozkłady zmiennej losowej ciągłej

### Rozkład jednostajny ciągły

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{poza tym} \end{cases} \quad F(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a < x \leq b \\ 1 & x > b \end{cases}$$

$$E(X) = \frac{a+b}{2} \quad D^2(X) = \frac{(b-a)^2}{12}$$

# Rachunek prawdopodobieństwa – rozkłady prawdopodobieństwa

Rozkłady zmiennej losowej ciągłej

Rozkład normalny  $X \sim N(\mu, \sigma)$

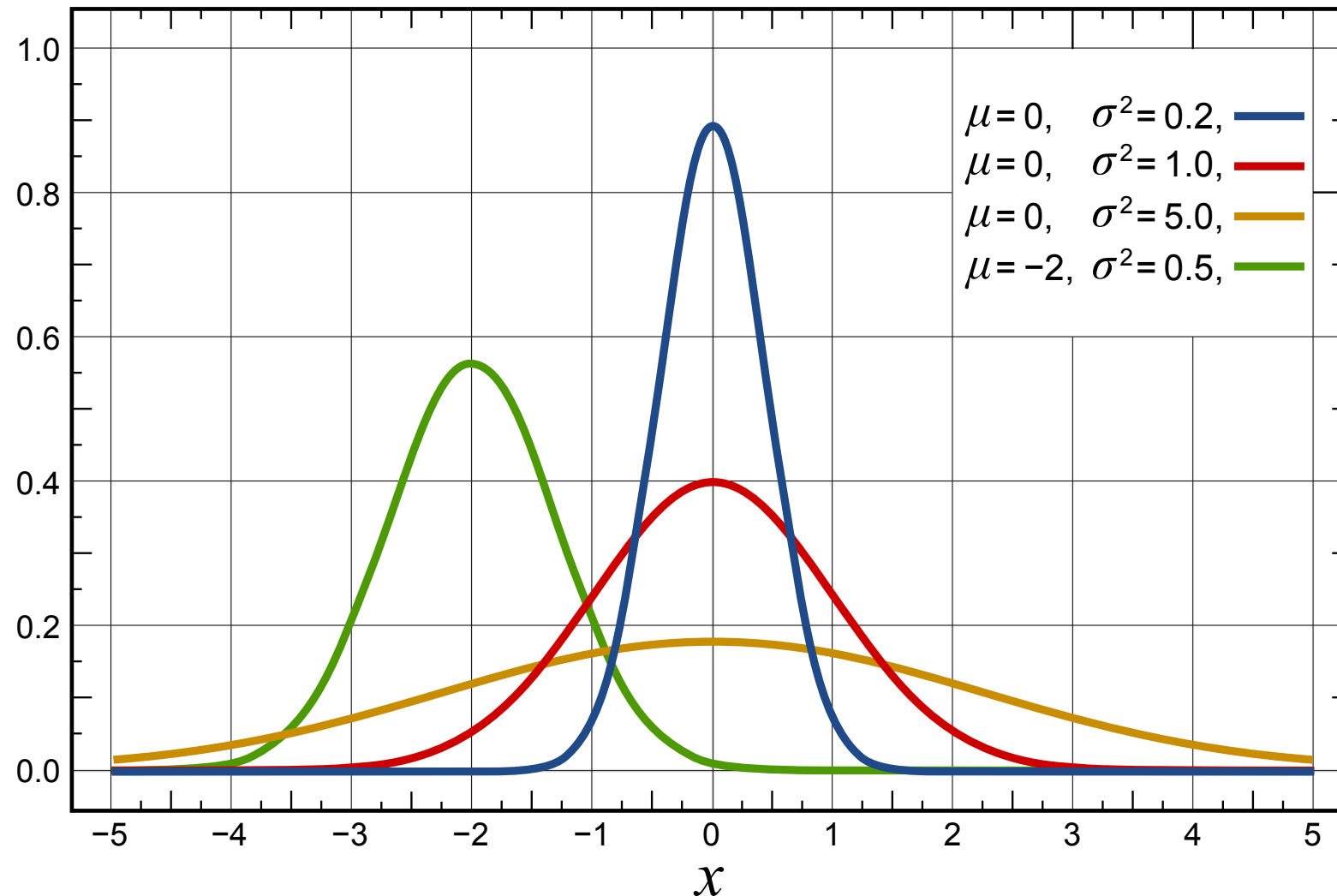
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$E(X) = Mo = Me = \mu$$

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

# Rachunek prawdopodobieństwa – rozkłady prawdopodobieństwa

## Rozkłady zmiennej losowej ciągłej



Źródło: Wikipedia

# Rachunek prawdopodobieństwa – rozkłady prawdopodobieństwa

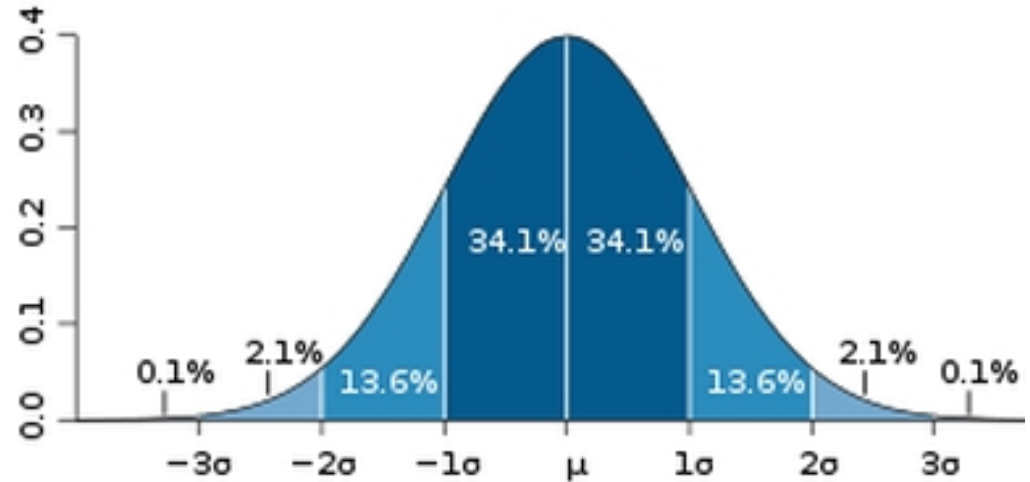
Rozkłady zmiennej losowej ciągłej

Reguła trzech sigm

$$P(\mu - \sigma < X < \mu + \sigma) = 0,6826$$

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0,9545$$

$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0,9973$$



Źródło: Wikipedia

# Rachunek prawdopodobieństwa – rozkłady prawdopodobieństwa

Rozkłady zmiennej losowej ciągłej

Standaryzacja: gdy  $X \sim N(\mu, \sigma)$

$$U = \frac{X - \mu}{\sigma}$$

więc:  $U \sim N(0, 1)$

$$P(X < x) = \Phi\left(\frac{x - \mu}{\sigma}\right) = \Phi(u)$$

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

# Rachunek prawdopodobieństwa – rozkłady prawdopodobieństwa

## Rozkłady zmiennej losowej ciągłej

Tablica 2. Dystrybuanta rozkładu normalnego  $F(u) = P(U \leq u)$  dla  $u \geq 0$

$u$	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	$u$
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359	0,0
0,1	,5398	,5438	,5478	,5517	,5557	,5596	,5636	,5675	,5714	,5753	0,1
0,2	,5793	,5832	,5861	,5910	,5948	,5987	,6026	,6064	,6103	,6141	0,2
0,3	,6179	,6217	,6255	,6293	,6331	,6368	,6406	,6443	,6480	,6517	0,3
0,4	,6554	,6591	,6628	,6664	,6700	,6736	,6772	,6808	,6844	,6879	0,4
0,5	,6915	,6950	,6985	,7019	,7054	,7088	,7123	,7157	,7190	,7224	0,5
0,6	,7257	,7291	,7324	,7357	,7389	,7422	,7454	,7486	,7517	,7549	0,6
0,7	,7580	,7611	,7642	,7673	,7703	,7734	,7764	,7794	,7823	,7852	0,7
0,8	,7881	,7910	,7939	,7967	,7995	,8023	,8051	,8078	,8106	,8133	0,8
0,9	,8159	,8186	,8212	,8238	,8264	,8289	,8315	,8340	,8365	,8389	0,9
1,0	,8413	,8438	,8461	,8485	,8508	,8531	,8554	,8577	,8599	,8621	1,0
1,1	,8643	,8665	,8686	,8708	,8729	,8749	,8770	,8790	,8810	,8830	1,1
1,2	,8849	,8869	,8888	,8907	,8925	,8944	,8962	,8980	,8997	,90147	1,2
1,3	,90320	,90490	,90658	,90824	,90988	,91149	,91309	,91466	,91621	,91774	1,3
1,4	,91924	,92073	,92220	,92354	,92507	,92647	,92785	,92922	,93056	,93189	1,4
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408	1,5
1,6	,94520	,94630	,94738	,94845	,94950	,95053	,95154	,95254	,95352	,95449	1,6
1,7	,95543	,95637	,95728	,95818	,95907	,95994	,96080	,96164	,96246	,96327	1,7
1,8	,96407	,96485	,96562	,96638	,96712	,96784	,96856	,96926	,96995	,97062	1,8
1,9	,97128	,97193	,97257	,97320	,97381	,97441	,97500	,97558	,97615	,97670	1,9
2,0	,97725	,97778	,97831	,97882	,97932	,97982	,98030	,98077	,98124	,98169	2,0



# Rachunek prawdopodobieństwa – rozkłady prawdopodobieństwa

Dwuwymiarowa i wielowymiarowa zmienna losowa

$(X, Y)$  oraz  $(X_1, X_2, \dots, X_n)$

Badanie wielowymiarowych zmiennych losowych jest interesujące z powodu zależności między zmiennymi składowymi.

Rozkłady brzegowe

Rozkłady warunkowe,

Korelacja, regresja

Wielowymiarowy rozkład normalny.

## Rachunek prawdopodobieństwa – twierdzenia graniczne

### Centralne twierdzenie graniczne Linderberg'a-Levy'ego

Jeżeli  $\{X_n\}$  jest losowym ciągiem niezależnych zmiennych o jednakowym rozkładzie, o wartości przeciętnej  $\mu$  i skończonej wariancji  $\sigma^2 > 0$  to ciąg  $(F_n)$  dystrybuant standaryzowanych średnich arytmetycznych  $\bar{X}_n$  (lub standaryzowanych sum  $\sum_{i=1}^n X_i$ )

$$Y_n = \frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma \sqrt{n}}$$

jest zbieżny do dystrybuanty  $\Phi$  rozkładu  $N(0, 1)$ .

## Rachunek prawdopodobieństwa – twierdzenia graniczne

### Twierdzenie graniczne Moivre'a-Laplace'a

Jeżeli  $\{X_n\}$  jest losowym ciągiem niezależnych zmiennych o rozkładzie dwumianowym z parametrami  $n, p$  to ciąg  $(F_n)$  dystrybuant standaryzowanych sum  $\sum_{i=1}^n X_i$

$$Y_n = \frac{\sum_{i=1}^n X_i - np}{\sqrt{npq}}$$

jest zbieżny do dystrybuanty  $\Phi$  rozkładu  $N(0, 1)$ .

## Rachunek prawdopodobieństwa – twierdzenia graniczne

### Przykład: Centralne twierdzenie graniczne Linderberg'a-Levy'ego

Inwestor uważa, że cena akcji pewnej spółki giełdowej ( $X$ ) wzrośnie jutro o pewną wartość z przedziału  $[0, 10]$ . Niech  $X$  ma rozkład jednostajny. Jakie jest prawdopodobieństwo, że wartość posiadanych przez inwestora akcji wzrośnie więcej niż o 480 zł, jeśli kupił on dzisiaj 100 akcji tej spółki.

$$E(X) = \frac{0+10}{2} = 5 \qquad D^2(X) = \frac{(10-0)^2}{12} = 0,833 \qquad D(X) = 0,913$$

$$\sum X = N(n\mu, \sigma\sqrt{n}) \qquad \sum X = N(500, 9,13)$$

$$\begin{aligned} P(\sum X > 480) &= 1 - P(\sum X < 480) = 1 - \Phi\left(\frac{480 - 500}{9,13}\right) \\ &= 1 - \Phi(-2,19) = 1 - 1 + \Phi(2,19) = 0,986 \end{aligned}$$

## Rachunek prawdopodobieństwa – twierdzenia graniczne

### Przykład: Twierdzenie graniczne Moivre'a-Laplace'a

PZU ocenia, że każdego roku 1% ubezpieczonych mężczyzn traci życie w wypadkach. Jakie jest prawdopodobieństwo, że w danym roku PZU będzie musiało wypłacić odszkodowanie więcej niż trzy razy, jeśli ubezpieczyło się od wypadków 100 mężczyzn?

$$X \sim N(np, \sqrt{npq}) \qquad X \sim N(100 * 0,01, \sqrt{100 * 0,01 * 0,99})$$
$$X \sim N(1, 0,995)$$

$$P(X > 3) = 1 - P(X \leq 3) = 1 - \Phi\left(\frac{3-1}{0,995}\right) = 1 - \Phi(2,01) = 1 - 0,97778 = 0,02222$$