STATYSTYKA - KOLOKWIUM I

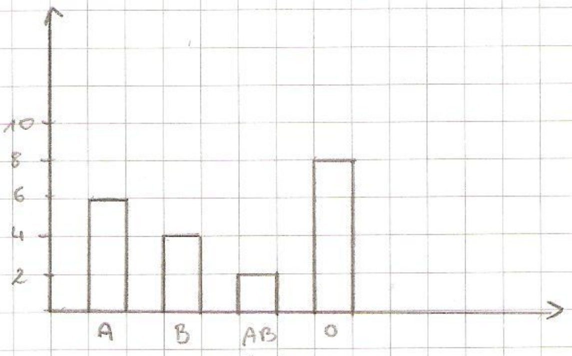
Zad. 1 Zbadano grupę krwi 20 osobom. Otrzymano wyniki: A, A, B, O, O, AB, B, O, AB, O, O, A, O, A, B, O, A, O, B, A

A - 6

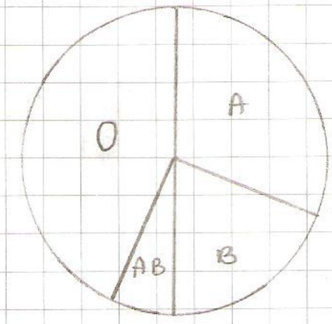
B - 4

AB - 2

O - 8



rozkład



Zad. 2 Zanotowano czasy obsługi 11 klientów w pewnym banku w minutach: 3, 5, 3, 7, 4, 5, 6, 3, 6, 5, 25

Oblicz średni czas, medianę i modę dla próbki, narysuj wykresy i zinterpretuj

- Średnia w próbie: $\frac{x_1 + x_2 + \ldots + x_n}{n} = \overline{x}$ $\quad \frac{1}{n}\sum_{i=1}^{n} x_i$

$\overline{x} = \frac{72}{11} \approx 6{,}54$

- Mediana $\quad \overset{50\%}{\longleftrightarrow} \quad \overset{50\%}{\longleftrightarrow}$

$3, 3, 3, 4, 5, \boxed{5}, 5, 6, 6, 7, 25$ // dzieli na // procentowe części

mediana $Q_2 = 5$

$$Q_2 = \begin{cases} X\left(\frac{n+1}{2}\right) & \text{dla } n \text{ nieparzystych} \\ \frac{X\left(\frac{n}{2}\right) + X\left(\frac{n}{2}+1\right)}{2} & \text{dla } n \text{ parzystych} \end{cases}$$

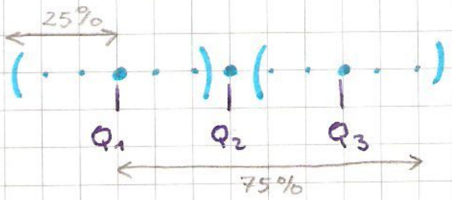
w próbie uporządkowanej

- Moda: najczęściej występująca wartość. Tutaj: próbka dwumodalna, bo są dwie mody: $M_1 = 3 \land M_2 = 5$

- Kwartyle : dolny kwartyl Q_1

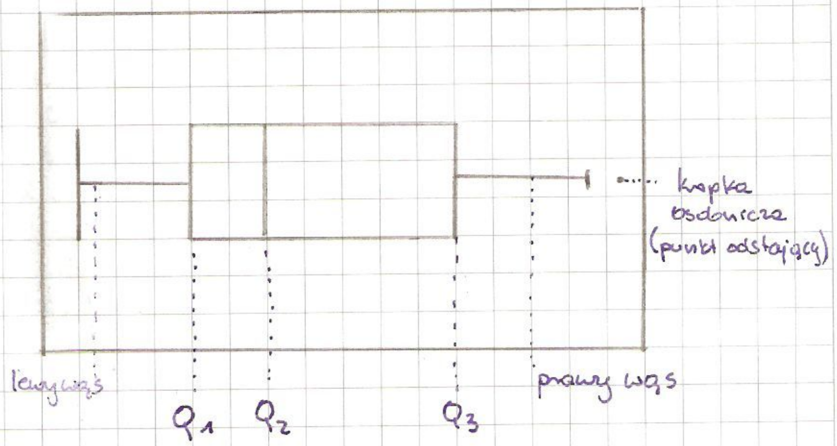
  (mediana Q_2)

  górny kwartyl Q_3



// ogólnie

- Wykres ramkowy:



Ograniczenia na wąsy: prawy wąs < $Q_3 + 1,5 \cdot IQR$ *

lewy wąs > $Q_1 - 1,5 \cdot IQR$

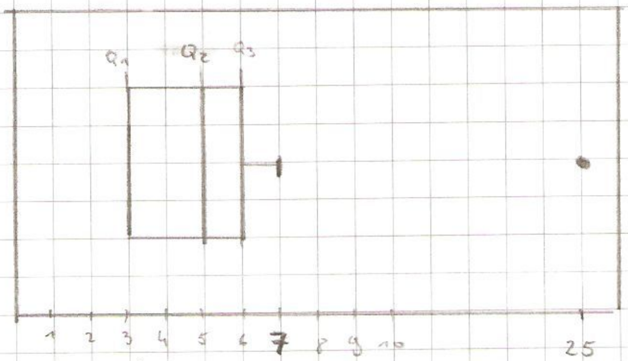
$IQR = Q_3 - Q_1$ (rozstęp międzykwartylowy)

// w zadaniu:

$Q_1 = 3$

$Q_3 = 6$

$IQR = 3$



- Nie ma lewego wąsa (najmniejsza wartość = $Q_1 = 3$)

- Prawy wąs: $6 + 1,5 \cdot 3 = 10,5$* wąs < 10,5

  musi być elementem próby - u nas największa możliwa to 7. 25 będzie wartością odstającą

Zad. 3 Ceny benzyny na 5 losowo wybranych stacjach wyniosły : 3,71 ; 3,76 ; 3,70 ; 3,69 ; 3,64 . Oblicz średnią cenę i jej wariancję

- średnia : $\overline{x} = 3{,}7$
- wariancja : $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2$
- odchylenie standardowe : $s = \sqrt{s^2}$ nie może być ujemna

$s^2 = \frac{1}{5-1} \cdot \left[(3{,}71 - 3{,}7)^2 + \ldots + (3{,}64 - 3{,}7)^2 \right] \approx 0{,}0185$

Zad. 4 Zanotowano czas rozwiązania zadania w grupie 30 uczniów: 14, 15, 25, 33, 20, 24, 15, 20, 28, 24, 25, 12, 21, 28, 30, 12, 29, 15, 22, 24, 18, 30, 20, 26, 18, 19, 22, 32, 16, 21

Narysuj histogram liczności i gęstości. Opisz kształt, uporządkuj w klasy, pakiety statystyczne

liczba pomiarów	liczba klas
30 - 60	6 - 8
60 - 100	7 - 10
100 - 200	9 - 12
200 - 500	11 - 17
500 - 1500	16 - 25

// w zadaniu

$X_{(1)} = 12$

$X_{(2)} = 33$

rozstęp = 33 - 12 = 21

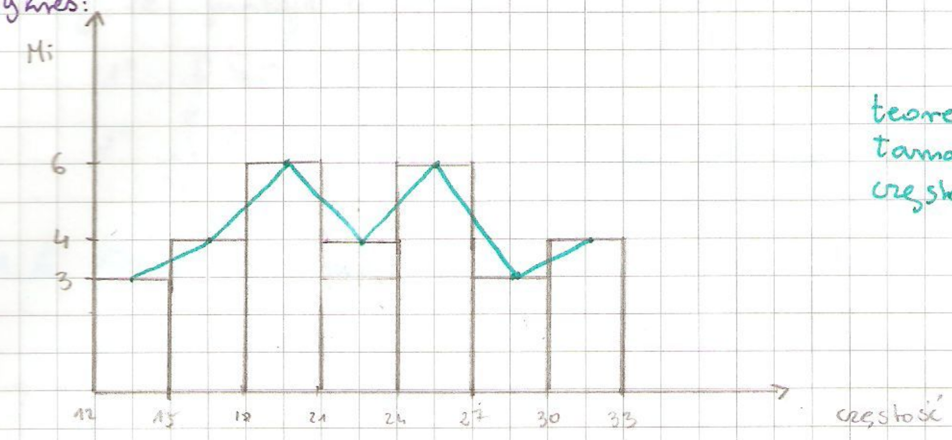
$\frac{21}{7} = 3$ szerokość klasy

tyle klas

NR	KLASA	LICZNOŚĆ	częstość
1	[12 ; 15)	3	0,1
2	[15 ; 18)	4	0,13
3	[18 ; 21)	6	0,2
4	[21 ; 24)	4	0,13
5	[24 ; 27)	6	0,2
6	[27 ; 30)	3	0,1
7	[30 ; 33]	4	0,13

$\frac{3}{30} \cdot 100\%$

· Wykres:



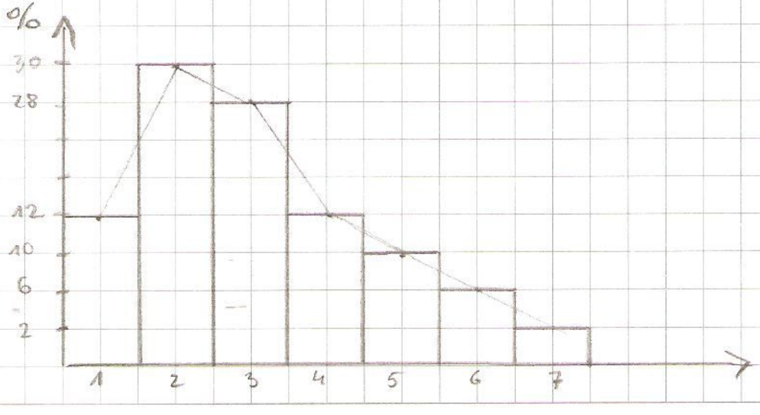
teoretyczna łamana częstości

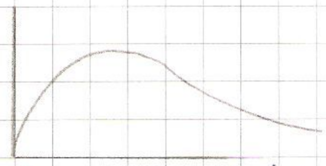
· Nie jest symetryczny (byłby, gdyby dwa ostatnie słupki były odwrotnie)

· Jest dwumodalny (ma dwa największe słupki)

Zad. 5 Miesięczne dochody w tys. zł pracowników pewnej firmy zgrupowano w tabeli

NR	DOCHÓD	LICZBA OSÓB	%
1	1,4 - 1,8	6	12%
2	1,8 - 2,2	15	30%
3	2,2 - 2,6	14	28%
4	2,6 - 3,0	6	12%
5	3,0 - 3,4	5	10%
6	3,4 - 3,8	3	6%
7	3,8 - 4,2	1	2%





Łamana gęstości
PRAWOSTRONNA SKOŚNOŚĆ

82% zarabia > 3000

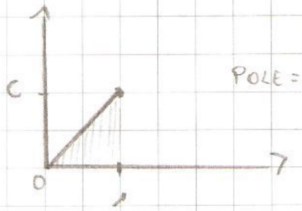
Mówimy, że funkcja f jest gęstością pewnej cechy x jeśli

1° $f \geq 0$

2° $\int_{-\infty}^{\infty} f(x)\,dx = 1$ // pole pod wykresem = 1

Zad. 6 Cecha x ma gęstość: $f(x) = \begin{cases} c \cdot x, & x \in \langle 0;1 \rangle \\ 0, & x \notin \langle 0;1 \rangle \end{cases}$

Znajdź stałą c.



$$1 = \int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 cx\,dx = c \cdot \int_0^1 x\,dx =$$

$$= c \cdot \left[\frac{x^2}{2} \right]_0^1 = c \cdot \left[\frac{1^2}{2} - \frac{0^2}{2} \right] = \frac{1}{2} c$$

$$\frac{1}{2} c = 1 \implies c = 2$$

☆ CAŁKA OZNACZONA:

$$\int_a^b f(x)\,dx = \left[F(x) \right]_a^b = F(b) - F(a) \qquad \text{gdzie } F(x) = \int f(x)\,dx$$

☆ CAŁKA NIEOZNACZONA:

$$\int 0\,dx = k$$

$$\int 1\,dx = x + k$$

$$\int x\,dx = \frac{x^2}{2} + k$$

$$\int x^n\,dx = \frac{x^{n+1}}{n+1} + k$$

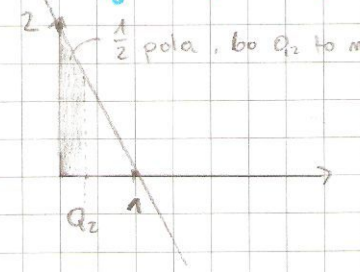
Zad. 7 Cecha x ma gęstość $f(x) = \begin{cases} 2(1-x) & \text{dla } x \in \langle 0;1) \\ 0 & \text{dla } x \notin \langle 0;1) \end{cases}$

Oblicz medianę gęstości f oraz kwantyl rzędu 0,9

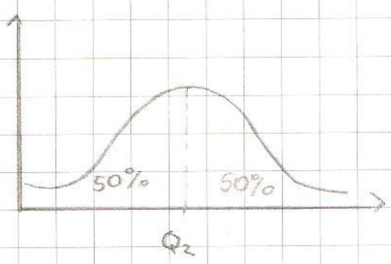
1° $2 - 2x \geq 0$

$-2x \geq 2$

$x \leq 0$



$\longrightarrow$



// pole to zbiór wystąpień

2° Jak liczyć medianę

$$Q_2 = \frac{1}{2} = \int_{-\infty}^{Q_2} f(x)\,dx$$

$$Q_3 = \frac{3}{4} = \int_{-\infty}^{Q_3} f(x)\,dx$$

$$\frac{1}{2} = \int_{-\infty}^{Q_2} f(x)\,dx = \int_0^{Q_2} (2-2x)\,dx = 2 \cdot \int_0^{Q_2} (1-x)\,dx = 2 \cdot \left(\int_0^{Q_2} 1\,dx - \int_0^{Q_2} x\,dx \right) =$$

$$= 2 \cdot \left(\left[x \right]_0^{Q_2} - \left[\frac{x^2}{2} \right]_0^{Q_2} \right) = 2 \cdot \left((Q_2 - 0) - \left(\frac{Q_2^2}{2} - \frac{0^2}{2} \right) \right) = 2 \cdot \left(Q_2 - \frac{Q_2^2}{2} \right) =$$

$$= 2Q_2 - Q_2^2 = \frac{1}{2}$$

$$-Q_2^2 + 2Q_2 - \frac{1}{2} = 0$$

$\Delta = \ldots$

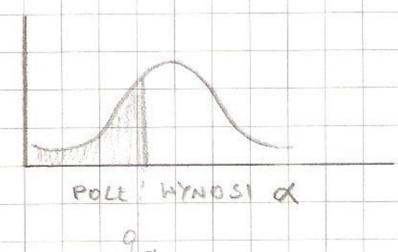
TO JEST DOBRE, Q_2 musi być < 1

$$Q_{2(1)} = \frac{-2+\sqrt{2}}{-2} = 1 - \frac{\sqrt{2}}{2} \quad !$$

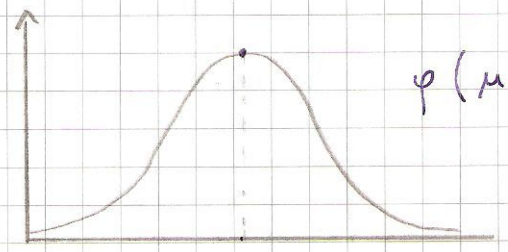
$$Q_{2(2)} = \frac{-2-\sqrt{2}}{-2} = 1 + \frac{\sqrt{2}}{2}$$

Kwantyle: kwantyl rzędu α, $\alpha \in (0;1)$

$$\alpha = \int_{-\infty}^{q_\alpha} f(x)\,dx$$



Rozkład normalny: GAUSSOWSKI $N(\mu, \sigma)$



μ

$\varphi(\mu, \sigma)$

μ - mediana / średnia

σ - odchylenie standardowe

"WYOSTRZENIE/SPŁASZCZENIE CYCKA" ^^

φ - gęstość rozkładu normalnego

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

Spośród wszystkich μ i σ najważniejsze są $\mu = 0 \wedge \sigma = 1$

Rozkład normalny standardowy $N(0,1)$

$$\Phi(x) = \int_{-\infty}^{x} \varphi_{N(\mu,\sigma)}(t)\,dt$$ - dystrybuanta rozkładu $N(0,1)$

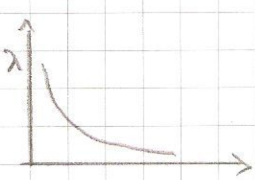
Pytania związane z postacią rozkładu:

- Czy jest symetryczny?
- Lewo- / prawostronnie skośny?
- czy jest w przybliżeniu normalny?
- czy istnieją obserwacje odstające?
- jedno- / wielomodalny?
- czy istnieją grupy danych?

- $Q_2 < \bar{x}$ prawostronna skośność $\rightarrow$

- $Q_2 > \bar{x}$ lewostronna skośność

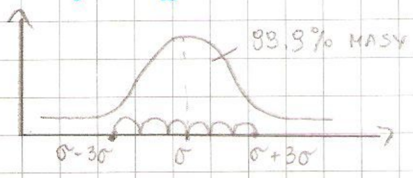
- $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0 \end{cases}$ ROZKŁAD WYKŁADNICZY



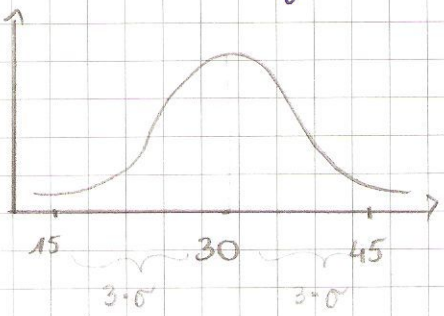
ŚREDNIA → ODCHYLENIE STD.

Zad. 8 Cecha X ma rozkład $N(\underset{\mu}{30}, \underset{\sigma}{5})$.

(a) Jaka jest przybliżona częstość obserwacji nieprzekraczających 20 000. (b) Oblicz kwantyle $q_{0,75}$ i Q_1



Twierdzenie o 3 sigmach



(a)

1. Standaryzacja całki $\displaystyle\int_{-\infty}^{\frac{20-30}{5}} \varphi(0,1)(x)\,dx = \Big[\Phi(x)\Big]_{-\infty}^{-2} =$

$$= \Phi(-2) - \overbrace{\Phi(-\infty)}^{=0} = 1 - \Phi(2) = 1 - 0{,}\underset{\text{z tablic}}{9772} = 0{,}0228 = 2{,}3\%$$

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

$\Phi(x) \approx 1$ dla $x > 3$

$\Phi(x) \approx 0$ dla $x < -3$

+ tablice !!!

(b) kwantyl rzędu 0,75

$$0{,}75 = \int_{-\infty}^{q_{0,75}} \varphi_{N(30,5)}(x)\,dx = \text{standaryzacja} = \int_{-\infty}^{\frac{q-30}{5}} \varphi_{N(0,1)}(x)\,dx =$$

$$= \Big[\Phi(x)\Big]_{-\infty}^{\frac{q-30}{5}} = \Phi\left(\frac{q-30}{5}\right) - \Phi(-\infty) = \Phi\left(\frac{q-30}{5}\right) - 0$$

$$\left.\begin{aligned} \Phi\left(\frac{q-30}{5}\right) &= 0{,}75 \\ \Phi(0{,}68) &= 0{,}75 \end{aligned}\right\} \frac{q-30}{5} = 0{,}68 \Rightarrow q = 33{,}4$$

PRAWDOPODOBIEŃSTWO

- Ω - przestrzeń zdarzeń
- A - zdarzenie
- Prawa de Morgana'a $(A \cup B)' = A' \cap B'$
  $(A \cap B)' = A' \cup B'$
- $P(\Omega) = 1$
- $P(\emptyset) = 0$
- $A \cap B = \emptyset \iff P(A \cup B) = P(A) + P(B)$

// ogólnie: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

- $P(A \cap B) = P(A) \cdot P(B)$ jeżeli zdarzenia są niezależne

Zad. 9 Doświadczenie polega na dwukrotnym rzucie kością.
Opisz przestrzeń zdarzeń elementarnych dla Ω. Niech
A - suma oczek jest nieparzysta
B - przynajmniej 1 raz wyrzucono 6
Oblicz P zdarzeń: $A \cap B$, $A \cup B$, $A \cup B'$

$$\Omega = \{(i,j),\ i,j = 1,\dots,6\}$$

$$\Omega = \{ (1,1), (1,2), (1,3), (1,4), (1,5), (1,6)$$
$$(2,1), \qquad (2,6)$$
$$(3,1), \qquad (3,6)$$
$$(4,1), \qquad (4,6)$$
$$(5,1), \qquad (5,6)$$
$$(6,1), (6,2), (6,3), (6,4), (6,5), (6,6) \}$$

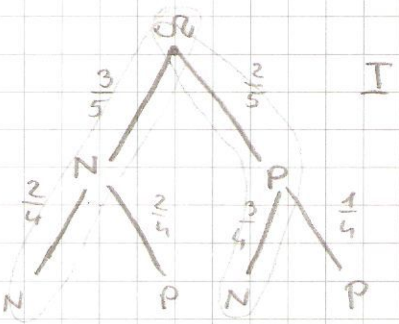
$\overline{\Omega} = 36$ $\overline{A} = 18$ $\overline{B} = 11$

$A \cap B = 6$ $P(A \cap B) = \frac{6}{36} = \frac{1}{6}$

$A \cup B = 18 + 11 - 6$ $P(A \cup B) = \frac{18+11-6}{36} = \frac{23}{36}$

$A \cup B' = 25 + 6 = 31$ $P(A \cup B') = \frac{21}{36}$

Zad. 10 Spośród cyfr 1, 2, 3, 4, 5 wylosowano jedną cyfrę, a następnie drugą z pozostałych. Oblicz P wylosowania liczby nieparzystej



a) $P(N_{I}) = \frac{3}{5}$

b) $P(N_{II}) = \frac{3}{5} \cdot \frac{2}{4} + \frac{2}{5} \cdot \frac{3}{4}$

Twierdzenie o prawdopodobieństwie całkowitym:

Z. $\Omega = \bigcup_{i=1}^{n} A_i = A_i \cap A_j = \emptyset$ dla $i \neq j$ $P(A_i) > 0$

Tw. $P(B) = P(B|A_1) \cdot P(A_1) + \ldots + P(B|A_n) \cdot P(A_n) = \sum_{i=1}^{n} P(B|A_i)$

W zadaniu: $P(N_{II}) = \underbrace{P(N_{II} | N_{I})}_{\frac{2}{4}} \cdot \underbrace{P(N_{I})}_{\frac{3}{5}} + \underbrace{P(N_{II} | P_{I})}_{\frac{3}{4}} \cdot \underbrace{P(P_{I})}_{\frac{2}{5}}$

Zad. 11 Dane są trzy klasy

A - 3 dziewczyny i 10 chłopców

B - 11 dziewczyn i 11 chłopców

C - 7 dziewcząt i 6 chłopców

a) Losujemy klasę, a następnie z klasy losujemy ucznia. Oblicz P. że wylosowano dziewczynę

b) z prawdop. proporcjonalnym do liczności klasy wybieramy klasę, a potem ucznia. Oblicz P. że wylosowano dziewczynę.

Niech: K_i - zdarzenie, że wylosowano klasę i-tą, $i = 1, 2, 3$

D - zdarzenie, że wylosowano dziewczynkę

C - " -- " -- chłopca

(a)

$$P(D) = P(D|K_1) \cdot P(K_1) + P(D|K_2) \cdot P(K_2) + P(D|K_3) \cdot P(K_3)$$

$$P(D) = \frac{3}{13} \cdot \frac{1}{3} + \frac{11}{22} \cdot \frac{1}{3} + \frac{7}{13} \cdot \frac{1}{3}$$

(b)

I KLASA	3D + 10C	$\Sigma = 13$
II KLASA	11D + 11C	$\Sigma = 22$
III KLASA	7D + 6C	$\Sigma = 13$

$$P(K_1) = \frac{13}{48}$$

$$P(K_2) = \frac{22}{48} \quad \Rightarrow \quad \frac{3}{13} \cdot \frac{13}{48} + \frac{11}{22} \cdot \frac{22}{48} + \frac{7}{13} \cdot \frac{13}{48}$$

$$P(K_3) = \frac{13}{48}$$

c) Wybieramy klasę a z klasy ucznia. Oblicz P, że losowano z klasy A pod warunkiem, że wylosowano dziewczynkę

Tw. Bayesa:

$$Z: \ \Omega = \bigcup_{i=1}^{n} A_i \quad A_i \cap A_j = \emptyset \quad i \neq j$$

$$T: \ P(A_k|B) = \frac{P(B|A_k) \cdot P(A_k)}{P(B|A_1) \cdot P(A_1) + \ldots + P(B|A_n) \cdot P(A_n)} =$$

$$= \frac{P(B|A_k) \cdot P(A_k)}{\sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)}$$

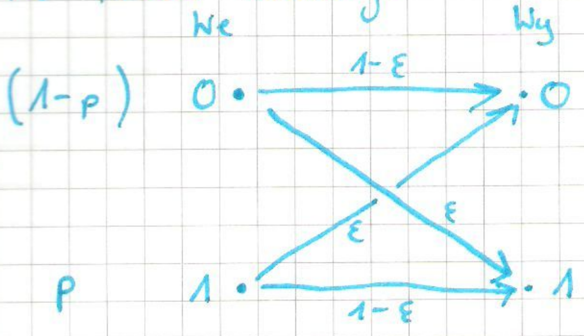
$A_k \cdot A_k$ = WAKA · WAKA!

$$c: \quad P(K_1|D) \overset{\text{Bayes}}{=} \frac{P(D|K_1) \cdot P(K_1)}{P(D|K_1) \cdot P(K_1) + P(D|K_2) \cdot P(K_2) + P(D|K_3) \cdot P(K_3)} =$$

$$= \frac{\frac{3}{13} \cdot \frac{1}{3}}{\frac{3}{13} \cdot \frac{1}{3} + \frac{11}{22} \cdot \frac{1}{3} + \frac{7}{13} \cdot \frac{1}{3}} = \ldots$$

Zad. 12 Kanał binarny:

Oblicz P, że wysłanym sygnałem było 1 jeżeli na Wy otrzymano 0



$$P'(We_1|Wy_0) = \frac{P(Wy_0|We_1)\cdot P(We_1)}{P(Wy_0|We_0)\cdot P(We_0) + P(Wy_0|We_1)\cdot P(We_1)} =$$

$$= \frac{\varepsilon \cdot p}{(1-\varepsilon)\cdot(1-p)+\varepsilon\cdot p}$$

hf & gl