KONKURS MATEMATYCZNY – KOMA 2013
SZKOŁY PONADGIMNAZJALNE – ELIMINACJE SZKOLNE
ROZKŁAD NORMALNY – KONSPEKT WYKŁADU

1) Eksperyment statystyczny i probabilistyczny
2) Co to jest rozkład prawdopodobieństwa? Przykłady rozkładów dyskretnych i ciągłych
3) Określenie rozkładu normalnego, parametry rozkładu, przykłady
4) Standardowy rozkład normalny, tablice
5) Standaryzacja innych rozkładów
6) Zastosowania rozkładu normalnego

Ad 1

- Statystyka to praktyka – opisuje wyniki doświadczeń. Probabilistyka to teoria – wyjaśnia, dlaczego wyniki praktycznych doświadczeń są właśnie takie, jak obserwujemy.
- Eksperyment statystyczny odbywa się w praktyce. Powtarzamy go i otrzymujemy zestaw wyników oraz częstości ich występowania. Możemy obliczyć wynik średni, najczęstszy, wskazać wynik środkowy, obliczyć średni rozrzut od średniego itp.
- Eksperyment probabilistyczny to opis teoretyczny. Określamy w nim teoretyczny rozkład prawdopodobieństw i obliczamy wartość oczekiwaną (najbardziej prawdopodobną) i . Dalej zajmujemy się właśnie teorią.

Ad 2

- W eksperymencie probabilistycznym każdy wynik ma określone, zawsze to samo prawdopodobieństwo zajścia. Suma tych prawdopodobieństw dla wszystkich możliwych wyników musi dać 1. Wtedy mamy do czynienia z dobrze określonym <u>rozkładem prawdopodobieństw</u>.
- Przykłady. W rzucie monetą mamy rozkład <u>jednostajny</u>, tzn. wszystkie wyniki mają takie same prawdopodobieństwa: a) $\{P(X{=}O)={}^1/_2,\ P(X{=}R)={}^1/_2\}$, b) w rzucie kostką może być rozkład niejednostajny, np. $\{P(X{=}3k)={}^1/_3,\ P(X{\neq}3k)={}^2/_3\}$.
- X nazywamy <u>zmienną losową</u>, a x – <u>wynikiem</u> doświadczenia probabilistycznego.
- <u>Dyskretny</u> rozkład prawdopodobieństw tzn. taki, w którym liczba możliwych wyników jest skończona lub przeliczalna, zapisujemy zazwyczaj w postaci tabelki, wzoru funkcji lub wykresu słupkowego.

a) skończony rozkład symetryczny

X	2	3	4	5	6	7	8
$P(X=x)$	1/16	2/16	3/16	4/16	3/16	2/16	1/16

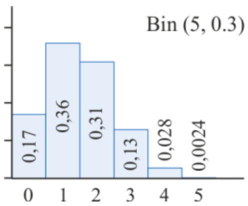
$$f(x) = P(X=x) = \begin{cases} \dfrac{x-1}{16}, & \text{for } x = 2,3,4,5 \\ \dfrac{9-x}{16}, & \text{for } x = 6,7,8 \end{cases}$$

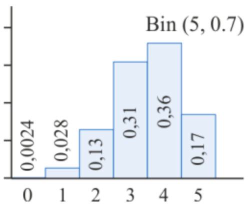
\+ naszkicować wykres
\+ sprawdzić, czy to rzeczywiście rozkład p-stwa

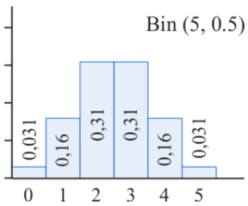
b) skończony rozkład niesymetryczny – np. dwumianowy

X – liczba sukcesów w n niezależnych próbach kończących się sukcesem lub porażką, $x = 0, 1, 2, \ldots, n$.
p – prawdopodobieństwo sukcesu

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



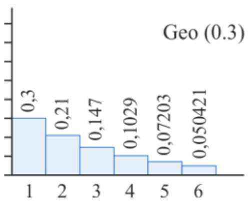


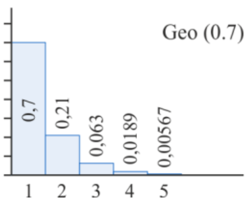


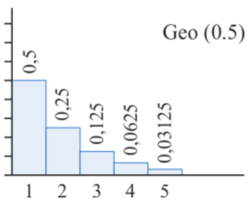
c) przeliczalny rozkład niesymetryczny – np. geometryczny

X – liczba niezależnych prób (kończących się sukcesem lub porażką) do uzyskania pierwszego sukcesu, $x = 1, 2, 3, \ldots$
p – prawdopodobieństwo sukcesu

$$f(x) = P(X=x) = (1-p)^{x-1}\, p$$

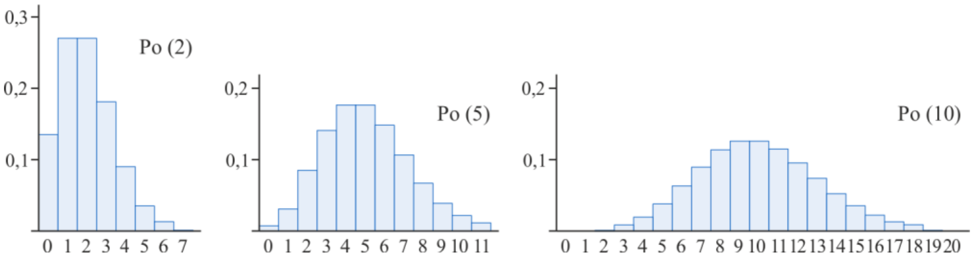






d) przeliczalny rozkład czasem zbliżony do symetrycznego – np. Poissona
Zdarzenia losowo rozmieszczone w pewnym przedziale (np. skazy na pręcie długości 1 m, przejazdy pociągu w 1 min).
X – liczba zdarzeń w danym przedziale, $x = 0, 1, 2, 3, \ldots$
λ – średnia liczba zdarzeń w danym przedziale

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



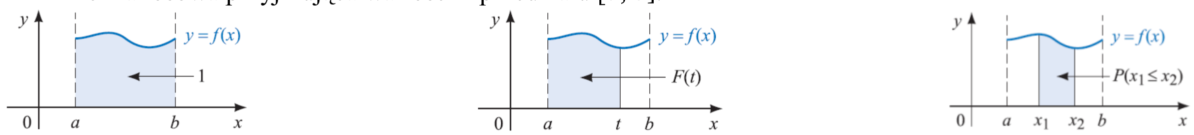
Dla dużej średniej liczby zdarzeń wykres staje się symetryczny i przypomina rozkład dwumianowy z p=1/2.

W kilku przykładach sprawdzić, czy to są rzeczywiście rozkłady prawdopodobieństw i obliczyć z rozkładów wielkości typu: $P(X\le 2)$, $P(X>3)$, $P(1\le X\le 5)$ itp.

Obliczanie wartości prawdopodobieństw dla rozkładu dwumianowego i Poissona jest dość żmudne, dlatego wartości te zebrano w tablicach dla wielu wartości n i różnych parametrów p lub λ. Często są one wbudowane do kalkulatorów lub arkuszy kalkulacyjnych.

- Rozkłady mogą być też ciągłe, tzn. takie, których wynikiem może być każda liczba rzeczywista lub liczba rzeczywista z danego przedziału. Wyników jest wtedy nieprzeliczalnie wiele i ich prawdopodobieństw też. Ale nadal muszą sumować się do jedynki. Odpada reprezentacja w formie tabelki. Pozostaje wzór i wykres. Ale jak zsumować tyle liczb? Jest to suma długości wszystkich pionowych kresek wystawionych pod wykresem funkcji, czyli pole pod tym wykresem. Z fizyki wiadomo, że mierzy je całka.

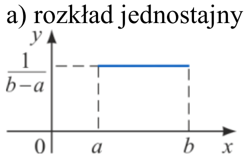
X – zmienna losowa przyjmująca wartości z przedziału $[a, b]$.



Jeśli $\int_a^b f(x)dx = 1$, to f jest rozkładem p-stwa. $P(X\le t) = \int_a^t f(x)dx$ $P(x_1 \le X \le x_2) = \int_{x_1}^{x_2} f(x)dx$

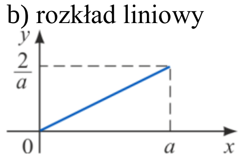
- Przykłady. W przykładach a-c zmienna przyjmuje wartość z danego przedziału, a w d) dowolną.

a) rozkład jednostajny (prostokątny)



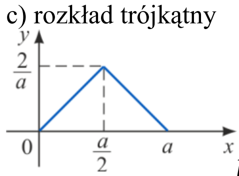
$f(x) = \dfrac{1}{b-a}$

b) rozkład liniowy



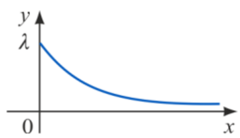
$f(x) = \dfrac{2}{a^2}x$

c) rozkład trójkątny



$$f(x) = \begin{cases} \dfrac{4}{a^2}x, \text{ dla } x \le \dfrac{a}{2} \\ \dfrac{4}{a^2}(a-x), \text{ dla } x > \dfrac{a}{2} \end{cases}$$

d) rozkład wykładniczy



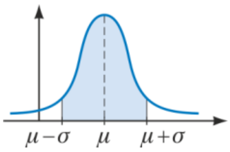
$f(x) = \lambda e^{-\lambda x}$

W przykładach a-c można sprawdzić, czy to są rzeczywiście rozkłady prawdopodobieństw. W d trzeba by umieć obliczać całki (a tego robić nie będziemy) lub posiadać tablice rozkładu wykładniczego.

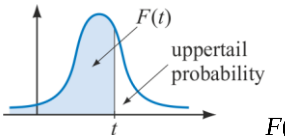
- Uwaga. W rozkładach ciągłych $f(x)$ nie oznacza już $P(X=x)$. Gdyby tak było, to np. na przedziale [0, 2] wartości z dwóch punktów dawałyby jedynkę. Prawdopodobieństwo przyjęcia przez zmienną konkretnej wartości jest równe zero (bo pole pod pojedynczym punktem jest zero).

Ad 3

- Rozkład normalny jest najważniejszym ciągłym rozkładem w statystyce, gdyż bardzo dużo wielkości, jakie mierzy się lub obserwuje w naukach przyrodniczych i społecznych, podlega temu właśnie rozkładowi (wysokość, masa, wiek, wyniki IQ, wyniki egzaminów itp.).
- Funkcja opisująca rozkład normalny jest dość skomplikowana. Jest symetryczna i ma kształt dzwonu (zwany też krzywą Gaussa), którego detale zależą od dwóch parametrów: μ - najbardziej prawdopodobna wartość, σ = mierzy skupienie wyników wokół wartości μ (ok. 60% wyników leży w odległości σ, 95% w odległości 2σ, a 98,5% w odległości 3σ. Funkcja rozkładu prawdopodobieństw jest określona na R, ale w praktyce używa się zakresu 6σ. Analiza rysunku z przykładami dla różnych parametrów.

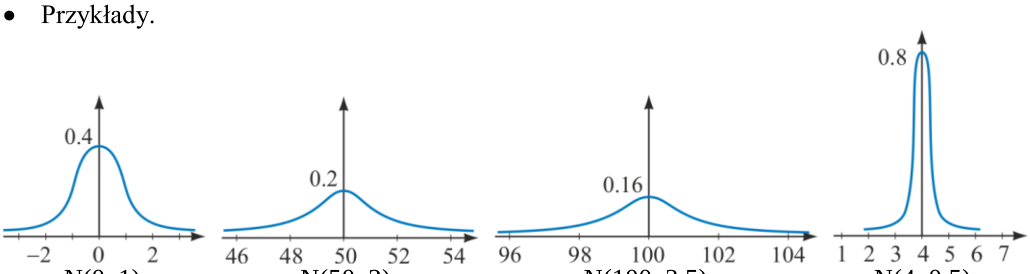


$$f(x)=P(X=t)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}$$



$$F(x)=P(X<t)=\int_{-\infty}^{t}\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}dt$$

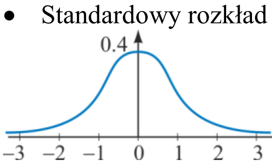
- Obliczanie całki z tej funkcji jest skomplikowane, dlatego wartości tej całki tablicuje się. Robi się to tylko dla jednej pary parametrów μ=0 i σ=1 – piszemy N(0,1). A co z innymi rozkładami normalnymi?
- Przykłady.



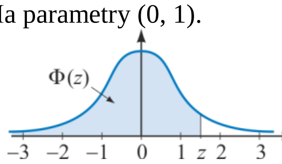
N(0, 1) N(50, 2) N(100, 2,5) N(4, 0,5)

Ad 4

- Standardowy rozkład normalny oznaczamy ϕ (małe fi). Ma parametry (0, 1).

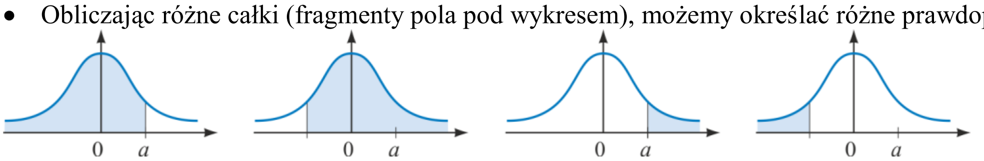


$$\phi(z)=P(Z=z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}$$



$$\Phi(z)=P(Z<z)=\int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz$$

- Obliczając różne całki (fragmenty pola pod wykresem), możemy określać różne prawdopodobieństwa.



$P(Z<a) = \Phi(a)$ $P(Z>-a) = \Phi(a)$ $P(Z>a) = 1-\Phi(a)$ $P(Z<-a) = 1-\Phi(a)$

- W tablicach standardowego rozkładu normalnego podaje się $\Phi(z)$ dla z>0. Dla z<0 korzysta się z symetrii wykresu, tzn. $\Phi(z) = 1 - \Phi(-z)$.
- Ćwiczenia w korzystaniu z tablic w obie strony, tzn. podajemy różne z i znajdujemy $P(Z<z)$, lub podajemy to prawdopodobieństwo i znajdujemy właściwe z, np. znajdź $P(Z<2{,}173)$, $P(Z>2{,}173)$, $P(Z<-2{,}173)$, $P(Z>-2{,}173)$ oraz koniunkcje i alternatywy takich warunków.

Ad 5

- Jeśli dany jest rozkład normalny z innymi parametrami, musimy go wystandaryzować (idea jest podobna jak ze wzorem na zamianę podstawy logarytmu), tzn. znaleźć taką wartość *z* (lub *P*) dla rozkładu standardowego, która odpowiada tej samej wartości *P* (lub z). Jeśli *X* ma rozkład N(μ, σ) to nowa zmienna losowa $Z=\frac{X-\mu}{\sigma}$ ma rozkład N(0, 1). Dzięki temu każdy rozważany warunek na *X* możemy przekształcić w równoważny mu warunek na *Z*.
- Przykład. a) Dla zmiennej *X* o rozkładzie N (300, 5) szukamy *P* (*X*>305).

$P(X>305) = P\left(\frac{X-300}{5} > \frac{305-300}{5}\right) = P(Z>1)$, a to już czytamy z tablic: $1-\Phi(1) = 1-0{,}8413 = 0{,}1587$.

b) Dla zmiennej X o rozkładzie N (50, √8) szukamy P (46<X<49).

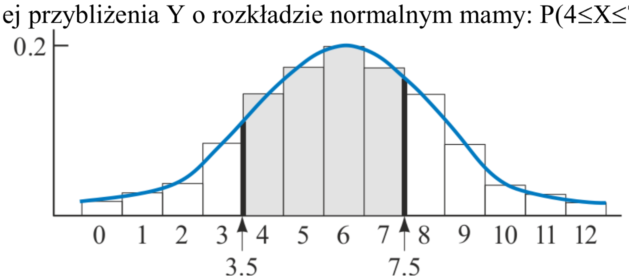
P (52<X<55)= $P\left(\frac{52-50}{\sqrt{8}} < \frac{X-50}{\sqrt{8}} < \frac{55-50}{\sqrt{8}}\right)$ = P (0,707<Z<1,768) = Φ(1,768)–Φ(0,707) = 0,9615–0,7601=0,2014.

c) Gdy znany prawdopodobieństwo, można obliczać nieznane parametry rozkładu, np. długość produkowanych w zakładzie szyn ma rozkład normalny o nieznanej średniej i odchyleniu 6 cm. Wiadomo, że 4,78% szyn ma długość większą od 82 cm. Jaka jest typowa długość szyny?

P (X>82)= $P\left(\frac{X-\mu}{6} > \frac{82-\mu}{6}\right)$= $P\left(Z > \frac{82-\mu}{6}\right)$=1- $\Phi\left(\frac{82-\mu}{6}\right)$=0,0478, czyli $\Phi\left(\frac{82-\mu}{6}\right)$=0,9522. Z tablic widzimy, że Φ(1,667)=0,9522, czyli $\frac{82-\mu}{6}$=1,667, a stąd μ=72 cm.

Ad 6

- Jeśli X ma rozkład dwumianowy Bin(n, p), to dla dużych n (>10) i p niezbyt skrajnych (niezbyt bliskich 0 lub 1) można traktować X jakby miała rozkład normalny N(np, √(npq)).
- Jeśli X ma rozkład Poissona Po(λ), to dla dużych n (>20) można traktować X jakby miała rozkład normalny N(λ, $\sqrt{\lambda}$).
- Uwaga. Jeśli rozkład dyskretny przybliżamy rozkładem ciągłym, musimy wziąć poprawkę na to, że słupki histogramu wystają po pół jednostki w każdą stronę od wartości x (tam jest bowiem środek słupka). Zatem dla zmiennej X o rozkładzie dwumianowym i jej przybliżenia Y o rozkładzie normalnym mamy: P(4≤X≤7)=P(3,5≤Y≤7,5).



UWAGI

1. Czas trwania wykładu z ćwiczeniami: 60-90 min. Czas pisania zadań 60 min.
2. Uczniowie powinni mieć kalkulator prosty.
3. Termin konkursu szkolnego:26 XI. Termin odesłania wyników: 2 XII.
4. Wyniki uporządkowane malejąco w pliku .xls pisane minuskułą. Trzy kolumny: imię/nazwisko/wynik. Nazwa szkoły tylko w nagłówku.
4. Zad. 1-11 - każdy podpunkt oceniamy zero-jedynkowo (4 punkty za zadanie). Zad. 12, 14,15, po 1 pkt za zadanie, zad. 13, 16 i 18 po 2 pkt. Za zadanie, zad. 17 – 3 pkt za zadanie. Za błędy wynikające z zastosowania symetrii każemy tylko raz (czyli można przyznać punkt za błędną odpowiedź w zadaniu z podpunktami, jeśli jest konsekwencją błędu we wcześniejszym podpunkcie).

KLUCZ

Zad. 1. T N T N
Zad. 2. T T T N
Zad. 3. T N T T
Zad. 4. T N T T
Zad. 5. 0,35; 0,68; 1; 0,12
Zad. 6. 1/3; 1/9; 5/9; 5/9
Zad. 7. 0.91621; 0,08534; 0,08534; 0.91621
Zad. 8. 0,27035; 0,27035; 0.56424; 0.56424
Zad. 9. 0,3; -0,78, 0,5; 0,28
Zad. 10. ½; ½; 0; 0,03593
Zad. 11. 2, 58, 228, 0

Zad. 12. 2/3
Zad. 13. rys odcinek łączący (0,0) z (4, 0.5); 21/64
Zad. 14. tak, 2
Zad. 15. 107,38
Zad. 16.°0,733; 0,732
Zad. 17. 0,0158; 0,8487; 0,0158
Zad. 18. 0,75804; 246315

KONKURS MATEMATYCZNY KO-MA 2012
ELIMINACJE PONADGIMNAZJALNE szkoła: ……………………………….…

Imię i nazwisko: .... klasa: ....

W zadaniach 1-4 zaznacz prawidłową odpowiedź.

Zad. 1. Czy to jest rozkład prawdopodobieństwa?

a) $\{P(X=1)=0{,}1,\ P(X=2)=0{,}6,\ P(X=3)=0{,}3\}$ TAK / NIE

b)

x	1	2	3	4	5
$P(X=x)$	0,2	0,12	0,35	0,35	0,1

TAK / NIE

c)

x	1	2	3	4	5
$P(X\le x)$	0,2	0,32	0,67	0,9	1

TAK / NIE

d) $P(X=x) = \frac{1}{2}\left(\frac{4}{5}\right)^x$, dla $x\in\mathrm{N}$ TAK / NIE

Zad. 2. *X* ma rozkład jednostajny na przedziale [0, 2]. Czy to prawda, że:

a) $P(X>2/3) = 2/3$ TAK / NIE

b) $P(0{,}5\le X\le 3) = ¾$ TAK / NIE

c) prawdopodobieństwa, że wynik leży w przedziale [0,2; 0,3] i [1,7; 1,8] są równe. TAK / NIE

d) prawdopodobieństwo, że pole kwadratu o boku *X* jest większe od 3 przekracza 0,2. TAK / NIE

Zad. 3. Czy te całki są poprawnie obliczone?

a) $\int_0^2 \frac{1}{2}\,dx=1$ TAK / NIE

b) $\int_2^7 0{,}25\,dx=1$ TAK / NIE

c) $\int_a^t \frac{1}{b-a}\,dt = \frac{t-a}{b-a}$ dla $a<t<b$ TAK / NIE

d) $\int_{-\infty}^0 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\,dz = \frac{1}{2}$ TAK / NIE

Zad. 4. *X* ma rozkład N(μ, σ). Czy to prawda, że:

a) wykres $P(X=x)$ jest symetryczny względem prostej $x=\mu$ TAK / NIE

b) żeby wystandaryzować *X* dzielimy *X* przez σ i odejmujemy μ TAK / NIE

c) maksymalna wartość $P(X=x)$ wynosi $^1/(\sigma\sqrt{(2\pi)})$ TAK / NIE

d) wykres $P(X=x)$ ma 2 punkty przegięcia TAK / NIE

W zadaniach 5-11 daj odpowiedź w każdym przykładzie.

Zad. 5. Oblicz prawdopodobieństwa, jeśli *X* spełnia:

x	1	2	3	4	5
$P(X\le x)$	0.2	0.32	0.67	0.9	1

a) $P(X=3)$ ……………………………...

b) $P(X>2)$ ……………………………...

c) $P(X<6)$ ……………………………...

d) $P(1<X<3)$ ……………………………...

Zad 6. Dla zmiennej losowej *X* przyjmującej wartości 1, 2, i 3 funkcja $F(x)= x^2/9$ określa $P(X\le x)$. Oblicz:

a) $P(X=2)$ ……………………………...

b) $P(X<2)$ ……………………………...

c) $P(X>2)$ ....

d) $P(X=3)$ ……………………………...

Zad. 7. Dla *Z* o rozkładzie N(0, 1) oblicz:

a) $P(Z<1{,}377)$ ……………………………...

b) $P(Z>1{,}377)$ ……………………………...

c) $P(Z<-1{,}377)$ ……………………………....

d) $P(Z>-1{,}377)$ ……………………………....

Zad. 8. Dla *Z* o rozkładzie N(0, 1) oblicz:

a) $P(0{,}377<Z<1{,}377)$ ……………………...

b) $P(-0{,}377<Z<1{,}377)$ ……………..........

c) $P(-1{,}377<Z<0{,}377)$ …………………....

d) $P(Z>-0{,}377$ lub $Z>1{,}377)$ ……………………...

Zad. 9. Dla *Z* o rozkładzie N(0, 1) znajdź *a*, wiedząc że:

a) $P(Z>a)=0{,}3802$ ……………………………...

b) $P(Z>a)=0{,}7818$ ……………………………...

c) $P(|Z|<a)=0{,}3802$ ……………………………...

d) $P(|Z|>a)=0{,}7818$ ……………………………...

Zad. 10. Wiedząc, że zmienna *X* ma rozkład N(300, 5), znajdź prawdopodobieństwa:

a) $P(X>300)$ …………………………

b) $P(X\leq 300)$ …………………………

c) $P(X>320)$ …………………………

d) $P(X<291)$ …………………………

Zad. 11. Janek mleczarz codziennie dostarcza mleko mieszkańcom ulicy Wspólnej. Czas, jaki mu to zabiera, ma rozkład normalny ze średnią 12 minut i odchyleniem 2 minut. Oszacuj liczbę dni w roku kiedy zajmuje to:

a) więcej niż 17 minut ……………………

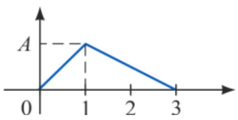
b) mniej niż 10 minut ……………………………..

c) między 9 a 13 minut ……………………

d) dokładnie 15 minut ……………………

W zadaniach 12-20 podaj krótką odpowiedź do każdego zadania.

Zad. 12. Na rysunku przedstawiono ciągły rozkład prawdopodobieństwa. Ile wynosi *A*?



……………………………………………………

Zad. 13. Naszkicuj wykres rozkładu liniowego na przedziale [0, 4] i oblicz $P(1\leq X\leq 2.5)$.

Zad. 14. Dana jest funkcja

$$f(x)=\begin{cases}\frac{1}{4}x \text{ dla } 0\leq x\leq 2\\ \frac{1}{4}(4-x) \text{ dla } 2\leq x\leq 4\\ 0 \text{ dla pozostałych } x\end{cases}.$$

Czy to jest rozkład prawdopodobieństwa? Jeśli tak, otrzymanie jakiej wartości *X* jest najbardziej prawdopodobne?

……………………………………………………

Zad. 15. Wiedząc, że *X* ma rozkład N(100, 6) i $P(X>a) = 0{,}1093$, znajdź *a*.

……………………………………………………

Zad. 16. Jakie jest prawdopodobieństwo otrzymania między 4 a 7 orłów (włącznie z 4 i 7) przy 12 rzutach monetą? Użyj:
a) rozkładu dwumianowego,
b) aproksymacji rozkładem normalnym

……………………………………………………

……………………………………………………

Zad. 17. W worku zmieszano ziarno jęczmienia i pszenicy, z czego 35% stanowi jęczmień. Pobrano próbkę 400 ziaren. Jakie jest prawdopodobieństwo, że ziaren jęczmienia będzie w niej:
a) mniej niż 120
b) między 120 a 150
c) ponad 160

……………………………………………………

……………………………………………………

……………………………………………………

Zad. 18. W regionie dolnośląskim mieszka 289.500 młodych ludzi (wiek 14-25 lat). Zbadano, że wydatki tych osób na pisma młodzieżowe w ciągu miesiąca podlegają rozkładowi normalnemu ze średnią 15,40 zł i odchyleniem standardowym 2,30 zł. Jakie jest prawdopodobieństwo, że wydatki te nie przekroczą 17 zł? Ilu młodych ludzi przeznacza na pisma młodzieżowe ponad 13 zł?

……………………………………………………

……………………………………………………