Wybrane rozkłady dyskretne cech statystycznych

rozkład dwumianowy oraz rozkład Poissona

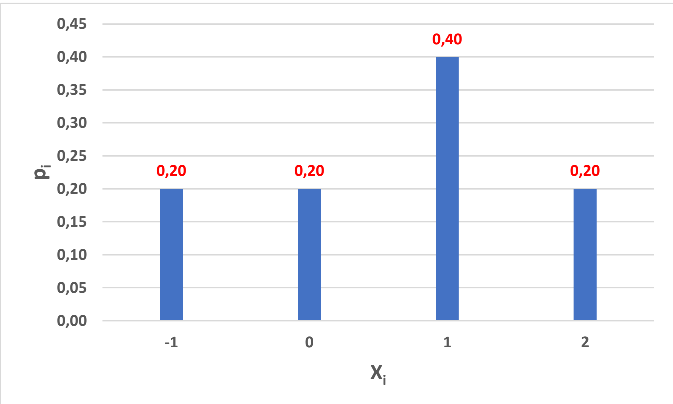
(literatura: Ostasiewicz, Rusnak, Siedlecka. Statystyka elementy teorii i zadania. Wydawnictwo AE we Wrocławiu, Wrocław 1999 (2006 – wydanie 6) – rozdział 5

Przykład 1. Funkcja rozkładu prawdopodobieństwa (x_i, p_i) dla dyskretnej zmiennej losowej X podany jest w tablicy:

(i)	1	2	3	4
Wartość zmiennej (x_i)	-1	0	1	2
Prawdopodobieństwo (p_i)	0,2	0,2	0,4	0,2

a) Narysować wykres (histogram) rozkładu prawdpopodobieństwa dla tej zmiennej losowej.
b) Wyznaczyć dystrybuantę zmiennej losowej X oraz narysować jej wykres.
c) Obliczyć wartość oczekiwaną oraz wariancję dla zmiennej losowej X.
d) Zanaleźć modalną oraz medianę zmiennej losowej X.

Ad a) Histogram rozkładu przedstawia rysunek (rys. 1)



Rys. 1. Funkcja rozkładu prawdopodobieństwa skokowej (dyskretnej) zmiennej losowej X.

gdzie: $\sum_{i=1}^{n=4} p_i = 1$

Ad b)

Dystrybuantę zmiennej losowej X wyznaczymy ze wzoru $F(x) = \sum_{(x_i<x)} p_i$.

Zatem

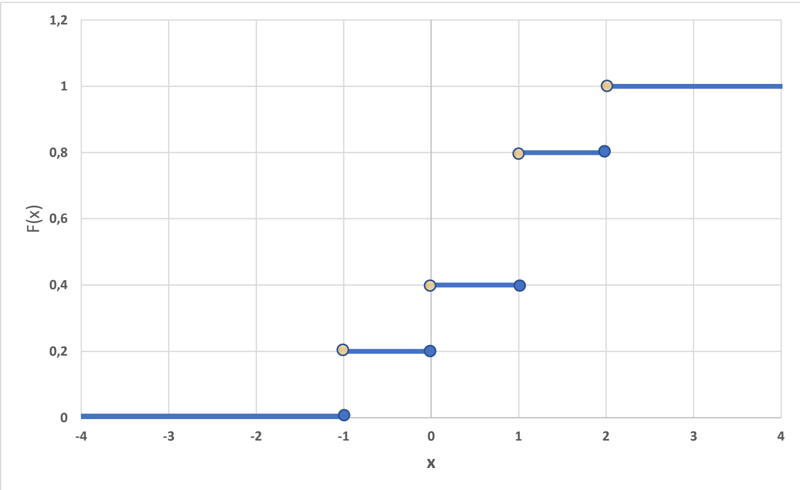
- Dla $x \leq -1$ $(x \in (-\infty, 1])$ wartość dystrybuanty wynosi $F(x) = 0$
- Dla $x \in (-1{,}0]$ wartość dystrybuanty wynosi $F(x) = p_1 = 0{,}2$.
- Dla $x \in (0{,}1]$ wartość dystrybuanty wynosi $F(x) = p_1 + p_2 = 0{,}2 + 0{,}2 = 0{,}4$.
- Dla $x \in (1{,}2]$ wartość dystrybuanty wynosi $F(x) = p_1 + p_2 + p_3 = 0{,}2 + 0{,}2 + 0{,}4 = 0{,}8$.

- Dla $x > 2$ ($x \in (2, +\infty)$) wartość dystrybuanty wynosi $F(x) = p_1 + p_1 + p_1 + p_4 = 1$.

Ostateczna funkcja dystrybuanty rozkładu badanej statystyki ma postać:

$$F(x) = \begin{cases} 0 & dla \quad x \leq -1 \\ 0{,}2 & dla \quad -1 < x \leq 0 \\ 0{,}4 & dla \quad 0 < x \leq 1 \\ 0{,}8 & dla \quad 1 < x \leq 2 \\ 1 & dla \quad x > 2 \end{cases}$$

Wykres dystrybuanty przedstawia rysunek (rys. 2).



Rys. 2. Dystrybuanta rozkładu prawdopodobieństwa dla skokowej (dyskretnej) zmiennej losowej X.

Ad c)

Wartość średnią (oczekiwaną) dla rozważanej statystyki (zmiennej losowej) dyskretnej X wyznaczamy ze wzoru: $E[x] = \bar{x} = \sum_{i=1}^{n=4} x_i * p_i$

Zatem $E[x] = -1 * 0{,}2 + 0 * 0{,}2 + 1 * 0{,}4 + 2 * 0{,}2 = 0{,}6$

Wariancję dla rozważanej statystyki (zmiennej losowej) dyskretnej X wyznaczamy ze wzoru:

$\mathrm{V}[x] = \sum_{i=1}^{n=4} (x_i - E[x])^2 * p_i$

Zatem $\mathrm{V}[x] = (-1 - 0{,}6)^2 * 0{,}2 + (0 - 0{,}6)^2 * 0{,}2 + (1 - 0{,}6)^2 * 0{,}4 + (2 - 0{,}6)^2 * 0{,}2 = 1{,}04$

Odchylenie standardowe wyniesie wtedy: $\sigma[x] = \sqrt{V[x]} = \sqrt{1{,}04} = 1{,}02$.

Wariancję możemy policzyć także z innego równoważnego wzoru:

$\mathrm{V}[x] = \sum_{i=1}^{n=4} {x_i}^2 * p_i - E[x]^2 = 1{,}4 - 0{,}6^2 = 1{,}04$ – co daje ten sam wynik.

Ad d)

Modalna dla rozkładu statystyki X istnieje i wynosi $M_o = x_3 = 1$, dla której prawdopodobieństwo realizacji jest największe (wynoszące 0,4).

Medianą rozkładu nazywamy wartość zmiennej X spełniającą łacznie układ nierówności (warunków):

$P(X \leq x) \geq \frac{1}{2}$ oraz $P(X \geq x) \geq \frac{1}{2}$.

Zatem $M_e = x_3 = 1$, gdyż: $P(X \leq 1) = \frac{8}{10} > \frac{1}{2}$ oraz $P(X \geq 1) = \frac{6}{10} > \frac{1}{2}$.

Przykład 2. Wiadomo, że 25% wszystkich szkód zgłaszanych do PZU przez firmy transportowe stanowią kolizje i zdarzenia drogowe. Jakie jest prawdopodobieństwo, że wśród zgłoszonych 10 szkód liczba kolizji będzie:

a) równa 5
b) większa niż 2
c) nie przekroczy 4
d) nie mniejsza niż 4 i nie większa niż 6 ?
e) Obliczyć wartość oczekiwaną i wariancję dla zmiennej losowej X

Zmienna losowa (statystyka) X ma rozkład dwumianowy, gdzie $X = k$ – oznacza liczbę (tzw. sukcesów) szkód bedących kolizjami drogowymi w wilokrotnym powtórzeniu pewnego zdarzenia losowego (identyfikacjoi rodzaju szkody zgłaszanej do PZU).

W naszym przypadku liczba powtórzeń doświadczenia losowego (n=10) – liczba zgłoszonych szkód. Prawdopodobieństwo sukcesu $p = 0{,}25$ (szkoda była kolizją), zaś prawdopodobieństwo zdarzenia przeciwnego (porażki) $q = 1 - p = 0{,}75$ – inny rodzaj szkody zgłoszonej do PZU.

Zatem rozkład badanej (modelowanej) zmiennej losowej o rozkładzie dwumianowym jest nastepujacy:

$X = \{x_i = k\} = \{0,1,2,\dots,10\}$, z prawdopodobieństwem $p_i = P(X = x_i = k) = B(n,p,k) = \binom{n}{k} * p^k * q^{n-k}$, $p + q = 1$.

Wyznaczmy prawdopodobieństwa rozkładu:

$$p_1 = P(X = x_1 = 0) = B(10; 0{,}25; 0) = \binom{10}{0} * 0{,}25^0 * 0{,}75^{10-0} = 1 * 1 * 0{,}75^{10} = 0{,}056.$$

$$p_2 = P(X = x_2 = 1) = B(10; 0{,}25; 1) = \binom{10}{1} * 0{,}25^1 * 0{,}75^{10-1} = 0{,}188.$$

$$p_3 = P(X = x_3 = 2) = B(10; 0{,}25; 2) = \binom{10}{2} * 0{,}25^2 * 0{,}75^{10-2} = 0{,}282.$$

$$p_4 = P(X = x_4 = 3) = B(10; 0{,}25; 3) = \binom{10}{3} * 0{,}25^3 * 0{,}75^{10-3} = 0{,}250.$$

$$p_5 = P(X = x_5 = 4) = B(10; 0{,}25; 4) = \binom{10}{4} * 0{,}25^4 * 0{,}75^{10-4} = 0{,}146.$$

$$p_6 = P(X = x_6 = 5) = B(10; 0{,}25; 5) = \binom{10}{5} * 0{,}25^5 * 0{,}75^{10-5} = 0{,}058.$$

$$p_7 = P(X = x_7 = 6) = B(10; 0{,}25; 6) = \binom{10}{6} * 0{,}25^6 * 0{,}75^{10-6} = 0{,}016.$$

$$p_8 = P(X = x_8 = 7) = B(10; 0{,}25; 7) = \binom{10}{7} * 0{,}25^7 * 0{,}75^{10-7} = 0{,}003.$$

$$p_9 = P(X = x_9 = 8) = B(10; 0{,}25; 8) = \binom{10}{8} * 0{,}25^8 * 0{,}75^{10-8} = 0{,}0004.$$

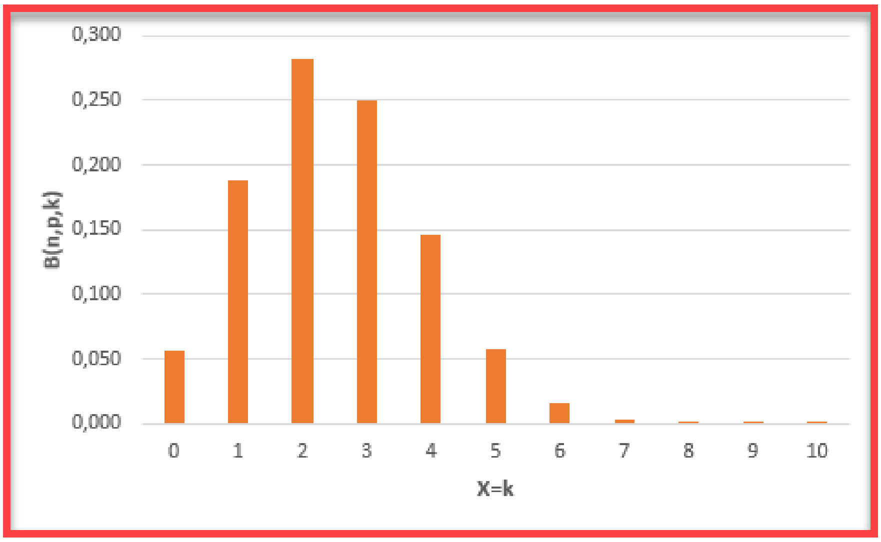
$$p_{10} = P(X = x_{10} = 9) = B(10; 0{,}25; 9) = \binom{10}{9} * 0{,}25^9 * 0{,}75^{10-9} = 0{,}00003.$$

$$p_{11} = P(X = x_{11} = 10) = B(10; 0{,}25; 10) = \binom{10}{10} * 0{,}25^{10} * 0{,}75^{10-10} = 0{,}000001.$$

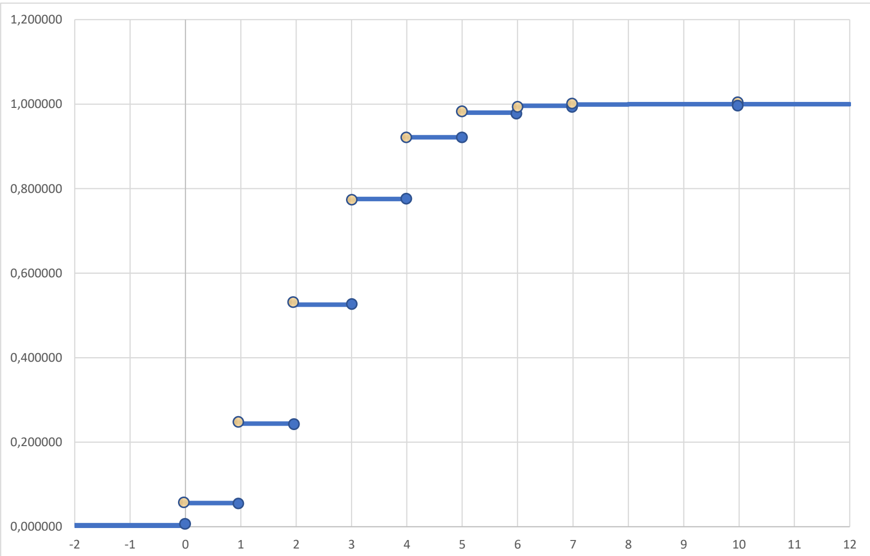
Funkcja dystrybuanty rozkładu badanej statystyki ma postać:

$$F(x) = \begin{cases} 0 & dla \quad x \leq 0 \\ 0{,}056 & dla \quad 0 < x \leq 1 \\ 0{,}244 & dla \quad 1 < x \leq 2 \\ 0{,}526 & dla \quad 2 < x \leq 3 \\ 0{,}776 & dla \quad 3 < x \leq 4 \\ 0{,}922 & dla \quad 4 < x \leq 5 \\ 0{,}980 & dla \quad 5 < x \leq 6 \\ 0{,}996 & dla \quad 6 < x \leq 7 \\ 0{,}9995 & dla \quad 7 < x \leq 8 \\ 0{,}99997 & dla \quad 8 < x \leq 9 \\ 0{,}999999 & dla \quad 9 < x \leq 10 \\ 1 & dla \quad x > 10 \end{cases}$$

Rysunek (rys. 3) przedstawia funkcję rozkładu prawdopodobieństwa, zaś rysunek (rys. 4) dystrybuantę rozkładu badanej statystyki X.



Rys. 3. Funkcja rozkładu prawdopodobieństwa analizowanej zmiennej losowej X o rozkładzie dwumianowym.



Rys. 4. Dystrybuanta rozkładu prawdopodobieństwa analizowanej zmiennej losowej X o rozkładzie dwumianowym.

Szukane prawdopodobieństwa wynoszą:

Ad a)

$p_6 = P(X = x_6 = 5) = B(10; 0{,}25; 5) = 0{,}058$ (5,8%)

Ad b)

Skrzystamy ze zdarzenia przeciwnego

$P(X > 2) = 1 - P(X \le 2) = 1 - P(X = 0) - P(X = 1) - P(X = 2) = 1 - 0{,}056 - 0{,}188 - 0{,}282 = 0{,}474$ (47%).

Ad c)

$P(X \le 4) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = 0{,}922$ (92%)

lub inaczej - skorzystamy z wartości dystrybuanty:

$P(X \le 4) = F(4) + P(X = 4) = 0{,}776 + 0{,}146 = F(5) = 0{,}922.$

Ad d)

$P(4 \le X \le 6) = P(X = 4) + P(X = 5) + P(X = 6) = 0{,}146 + 0{,}058 + 0{,}016 = 0{,}22$ (22%).

lub inaczej – oszacujemy szukane prawdopodobieństwo korzystając z wartości dystrybuanty:

$P(4 \le X \le 6) = F(6) - F(4) = 0{,}980 - 0{,}776 = 0{,}2$ (20%)

Ad e)

Wartość oczekiwana wynosi:

$$E[x] = \sum_{i=1}^{n=11} x_i * p_i = 2{,}5$$

Wariancja wynosi:

$\mathrm{V}[x] = \sum_{i=1}^{n=11} (x_i - E[x])^2 * p_i = \sum_{i=1}^{n=11} {x_i}^2 * p_i - E[x]^2 = 8{,}125 - 2{,}5^2 = 1{,}875.$

Odchylenie standardowe wynosi zatem: $\sigma[x] = \sqrt{V[x]} = \sqrt{1{,}875} = 1{,}37.$

Przykład 3. Oszacowano, że w 1999 roku 3% samochodów w Polsce miało zainstalowane katalizatory. W pewnym dniu stacja benzynowa obsłużyła 200 samochodów. Jakie jest prawdopodobieństwo, że:

a) 2 samochody miały katalizatory
b) więcej niż 2 samochody miały katalizatory
c) żaden samochód nie miał katalizatora
d) nie więcej niż 3 samochody miały katalizatory ?

Statystyka X=k – liczba obsłużonych samochodów na stacji benzynowej, które miały katalizator ma oczywiście rozkład dwumianowy Bernoulliego B(n,p,k), gdzie: n=200, p=0,03 – prawdopodobieństwo, że obsługiwany samochód ma katalizator, $q = 1 - p = 1 - 0{,}03 = 0{,}97$ – prawdopodobieństwo, że nie ma katalizatora.

Jednak przy tak dużym n=200 nie dałoby się wyznaczyć takich prawdopodobieństw bezpośrednio ze wzorów dla rozkładu dwumianowego. Wiemy jednak, że dla dużych n ($n > 20, n \to \infty$), gdy mamy zachowany stały iloczyn $n * p = m$ oraz $p < 0{,}2$ (nieduże), to rozkład dwumianowy możemy przybliżać innym rozkładem dyskretnym, a mianowicie rozkładem Poissona (z parametrem m – który jest dla niego wartością oczekiwaną - średnią oraz wariancją).

Zmienna losowa dyskretna posiada rozkład Poissona z parametrem „m”, gdy przyjmuje wartości:

$$X = k = 0,1,2,3, ...$$

z prawdopodobieństwem:

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

Dla naszego przykładu możemy przyblizać analizowany rozkład dwumianowy rozkładem Poissona z parametrem $m = n * p = 200 * 0{,}03 = 6$.

Dla granicznego rozkładu Poissona policzmy prawdopodobieństwa rozkładu dla kilku początkowych wartości „k”:

$P(X = k = 0) = \frac{6^0}{0!} * 2{,}7173^{-6}$ =0,0025

$P(X = k = 1) = \frac{6^1}{1!} * 2{,}7173^{-6}$ =0,015

$P(X = k = 2) = \frac{6^2}{2!} * 2{,}7173^{-6}$ =0,045

$P(X = k = 3) = \frac{6^3}{3!} * 2{,}7173^{-6}$ =0,089

$P(X = k = 4) = \frac{6^4}{4!} * 2{,}7173^{-6}$ =0,134

$$P(X = k = 5) = \frac{6^5}{5!} * 2{,}7173^{-6} = 0{,}161$$

$$P(X = k = 6) = \frac{6^6}{6!} * 2{,}7173^{-6} = 0{,}161$$

$$P(X = k = 7) = \frac{6^7}{7!} * 2{,}7173^{-6} = 0{,}138$$

$$P(X = k = 8) = \frac{6^8}{8!} * 2{,}7173^{-6} = 0{,}103$$

$$P(X = k = 9) = \frac{6^9}{9!} * 2{,}7173^{-6} = 0{,}069$$

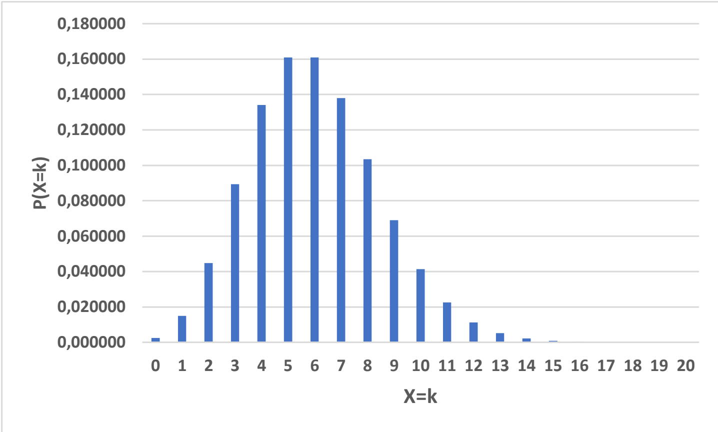
$$P(X = k = 10) = \frac{6^{10}}{10!} * 2{,}7173^{-6} = 0{,}041$$

……..

$P(X = k = 20) = \frac{6^{20}}{20!} * 2{,}7173^{-6}$ =0,000004

……...

Rysunek (rys. 5) przedstawia histogram rozkładu prawdopodobieństwa granicznego rozkładu Poissona z parametrem $m = 6$.



Rys. 5. Funkcja rozkładu prawdopodobieństwa analizowanej zmiennej losowej X o rozkładzie Poissona ($m = 6$)

Szukane prawdopodobieństwa oszacowane na podstawie granicznego rozkładu Poissona wynoszą:

Ad a)

$P(X = 2) = 0{,}045$ (4,5%).

Wartości rozkładu Poissona (dystrybuanty tego rozkładu) są stablicowane w tablicach statystycznych (np. Tablica V – na końcu książki: Ostasiewicz S. (red.), Statystyka elementy teorii i zadania, str. 384).

Podane tam są skumulowane wartości prawdopodobieństwa: $P(X \leq k) = \sum_{r=0}^{k} \frac{m^r}{r!} * e^{-m}$.

Korzystając z tablic mamy: $P(X = 2) = P(X \leq 2) - P(X \leq 1) = 0{,}062 - 0{,}017 = 0{,}045$ (4,5%)

Ad b)

Skorzystamy ze zdarzenia przeciwnego oraz tablic skumulowanego rozkładu Poissona:

$P(X > k = 2) = 1 - P(X \leq k = 2) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)] = 1 - 0{,}062 = 0{,}94$ (94%).

Ad c)

Z tablic lub ze wzoru: $P(X = k = 0) = 0{,}002$

Ad d)

Z tablic lub ze wzorów: $(X \leq k = 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 0{,}151$.

Policzmy jeszcze podstawowe parametry badanego rozkładu. Wartość oczekiwana dla analizowanego rozkładu Poissona wynosi $E[x] = m = 6$, zaś wariancja również wynosi $V[x] = m = 6$, zatem odchylenie standardowe jest równe $\sigma[x] = \sqrt{V[x]} = \sqrt{6} = 2{,}45$. Zatem średnia liczba samochodów z katalizatorem obsługiwanych na stacji benzynowej to 6 z odchyleniem $\pm 2{,}45$ samochodu.

Zadania do samodzielnego rozwiązania (Ostasiewicz S. [red.], Statystyka elementy teorii i zadania, str. 185-193): Zad. 5.5, 5.6, 5.10, 5.45