# Dyskretne zmienne losowe

dr Mariusz Grządziel

Wykład 4; 27 lutego 2012

**Definicja 1.** *Zmienną losową nazywamy dyskretną (skokową), jeśli zbiór jej wartości* $x_1, x_2, \ldots,$ *można ustawić w ciąg.*

Zmienna losowa $X$, która przyjmuje wszystkie wartości z zadanego przedziału $(a, b)$, nie jest zmienną losową dyskretną, ponieważ elementów tego przedziału nie da się ustawić w ciąg ("ponumerować").
- G. Cantor 1873— twierdzenie— wszystkich liczb rzeczywistych nie da się ustawić w ciąg.

**Rozkład dyskretnej zmiennej losowej**
Zbiór wartości dyskretnej zmiennej losowej $X$— ciąg $x_1, x_2, \ldots,$ (skończony lub nieskończony).
Rozkład zmiennej losowej dyskretnej $X$ jest określony przez nieujemne liczby $p_1, p_2, \ldots$ spełniające warunki:

$$\sum p_i = 1, \tag{1}$$

$$p_i = P(X = x_i). \tag{2}$$

**Dyskretne zmienne losowe— przykłady**
Przykłady:

- Rozkład $U$, sumy oczek w dwukrotnym rzucie kostką (patrz poprzedni wykład);
- Rozkład $Z$, gdzie $Z$ oznacza liczbę rzutów monetą, po której po raz pierwszy wypada orzeł (zdarzeniu polegającemu na tym, że orzeł wypadnie już w pierwszym rzucie, odpowiada wartość zmiennej $Z$ równa 0).

z niezależności zdarzeń:

$$P(Z = k) = \left(\frac{1}{2}\right)^{k+1}, \quad k = 0, 1, 2, \ldots.$$

Zmienna losowa $Z$ przykład dyskretnej zmiennej losowej, dla której zbiór wartości: $\{0, 1, 2, \ldots\}$ nie jest skończony.

**Rzuty osobiste— przykład**
Niech $X$- liczba trafień w wykonywanym przez koszykarza A rzucie osobistym.Niech: T odpowiada trafieniu do kosza, C odpowiada chybieniu.
Przestrzeń zdarzeń elementarnych: $\mathcal{S} = \{C, T\}$.
Niech $X$- liczba trafionych rzutów. Zmienna $X$ jest funkcją określoną na $\mathcal{S}$;

$$X(C) = 0, \quad X(T) = 1.$$

Zakładamy, że prawdopodobieństwo trafienia wynosi $0{,}9$. Rozkład zmiennej losowej $X$ można przedstawić przy pomocy tabelki:

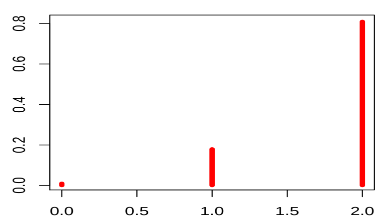
| $k$ | 0 | 1 |
|---|---|---|
| $P(X = k)$ | 0,1 | 0,9 |

**Liczba trafień $Y$ w dwóch rzutach**
Niech $Y$- liczba trafień w dwóch wykonywanych przez koszykarza A rzutach osobistych. Przyjmujemy, że prawdopodobieństwo trafienia w jednym rzucie osobistym wynosi $0{,}9$ i zdarzenie trafienia/chybienia w drugim rzucie jest niezależne od analogicznego zdarzenia w pierwszym rzucie.
Można pokazać, że:

$$P(Y = 0) = \left(\frac{1}{10}\right)^2 = 0{,}01,$$
$$P(Y = 1) = 2 \times \frac{1}{10} \times \frac{9}{10} = 0{,}18,$$
$$P(Y = 2) = \left(\frac{9}{10}\right)^2 = 0{,}81.$$

**Liczba trafień $Y$ w dwóch rzutach— c.d.**
Rozkład można przedstawić w postaci tabelki lub wykresu słupkowego:

| $k$ | 0 | 1 | 2 |
|---|---|---|---|
| $P(X = k)$ | 0,01 | 0,18 | 0,81 |




**Rozkład dwumianowy**
Symbol Newtona $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ jest równy liczbie podzbiorów k-elementowych zbioru n-elementowego ($0 \leqslant k \leqslant n$).

**Definicja 2.** *Mówimy, że zmienna losowa $X$ ma rozkład dwumianowy z parametrami $n \in \mathbb{N}$ i $0 < p < 1$, co w skrótowo zapisujemy $X \sim Bin(n,p)$ (lub $X \sim Bin(n;p)$), jeśli*

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \ldots, n.$$

**„Dziesięciokrotny rzut monetą"— przykład**
Niech $V$ oznacza liczbę orłów otrzymanych w dziesięciokrotnym rzucie monetą (zakładamy, że moneta jest „rzetelna", tj. prawdopodobieństwo otrzymania orła jest równe $\frac{1}{2}$ oraz że wyniki kolejnych rzutów są od siebie niezależne). Chcemy obliczyć prawdopodobieństwo;

$$P(V \geqslant 9).$$

Rozwiązanie $V \sim Bin(10; 0{,}5)$,

$$\begin{aligned} P(V \geqslant 9) =& P(V = 9) + P(V = 10) = \\ =& \binom{10}{9} (0{,}5)^9 (0{,}5)^1 + \binom{10}{10} (0{,}5)^{10} (0{,}5)^0 = \\ =& \frac{10}{1024} + \frac{1}{1024} = \frac{11}{1024}. \end{aligned}$$

**Pojęcie dystrybuanty rozkładu**
W obliczeniach podobnych do tych z poprzedniego przykładu użyteczne może się okazać pojęcie dystrybuanty zmiennej losowej.

**Definicja 3.** *Niech $X$ będzie dowolną zmienną losową. Dystrybuantą zmiennej losowej $X$ nazywamy funkcję $F$ określoną jako:*

$$F(x) = P(X \leqslant x).$$

**Uwaga.** W powyższej definicji nie zakładamy, że zmienna $X$ jest dyskretna.

**„Dziesięciokrotny rzut monetą"—c.d.**
$V$-liczba wyrzuconych orłów w dziesięciokrotnym rzucie monetą;

$$P(V \geqslant 9) = P(V = 9) + P(V = 10) = F_V(10) - F_V(8)$$

gdzie $F_V$ jest dystrybuantą zmiennej losowej $V$.
Obliczenia wykonane w R-rze:

```
> pbinom(10,10,0.5)- pbinom(8,10,0.5)
[1] 0.01074219
```

pbinom- pierwsza litera odpowiada „dystrybuancie", binom odpowiada rodzajowi rozkładu (ang. *binomial*- dwumianowy). Korzystając z polecenia $pbinom$ można obliczać wartości dystrybuanty rozkładu $Bin(n,p)$ dla dużych wartości $n$.

**Rozkład Poissona**

**Definicja 4.** *Mówimy, że zmienna losowa $X$ ma rozkład Poissona z parametrem $\lambda > 0$, jeśli przyjmuje ona wartości w zbiorze $\{0, 1, 2, \ldots, \}$ oraz*

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

Rozkład Poissona może być zastosowany z powodzeniem do opisu takich cech jak liczba nasion chwastów wśród nasion trawy, liczba klientów zgłaszających się dziennie do banku, liczba wypadków drogowych na placu Grunwaldzkim w danym dniu itd.

**Wartość oczekiwana zmiennej losowej dyskretnej**

**Definicja 5.** *Dla zmiennej losowej dyskretnej $X$ wartość oczekiwana, jeśli istnieje, jest liczbą określoną wzorem*

$$EX = \sum_i x_i p_i,$$

*w którym sumowanie obejmuje wszystkie wartości zmiennej $X$.*

**Uwaga** Wartość oczekiwana zmiennej losowej nazywana jest literaturze także wartością średnią zmiennej losowej. W definicji tej zakładamy, że $\sum_i x_i p_i$ jest liczbą skończoną (w przypadku, gdy liczba składników jest nieskończona, zakładamy zbieżność sumy do granicy skończonej).
Wartość oczekiwana może być interpretowana jako "środek ciężkości" układu punktów materialnych $x_1, x_2, \ldots$ o wagach $p_1, p_2, \ldots$.

**Własności wartości oczekiwanej**
Łatwo widać, że wartość oczekiwana zmiennej losowej $Y = aX + b$ jest równa

$E(aX + b) = aEX + b$ (wartość oczekiwana ma własność *liniowości*).

Jeśli wartości oczekiwane zmiennych losowych $X_1$ i $X_2$ istnieją i są równe, odpowiednio, $\mu_1$ i $\mu_2$, to

$$E(X_1 + X_2) = \mu_1 + \mu_2.$$

**Wariancja zmiennej losowej**

**Definicja 6.** *Wariancję zmiennej losowej $X$ określamy wzorem*

$$VarX = E(X - \mu)^2,$$

*gdzie $\mu = EX$.*

Wariancja jest równa wartości oczekiwanej kwadratu odchylenia wartości zmiennej losowej od swojej wartości przeciętnej.
**Uwaga:** W definicji tej nie zakładamy, że zmienna losowa $X$ jest dyskretna. Zakładamy natomiast istnienie wartości oczekiwanej $E(X - \mu)^2$.
Dla $a > 0$ mamy:

$$Var(aX + b) = a^2 Var(X).$$

Dla zmiennej dyskretnej $X$ wariancja jest równa

$$VarX = \sum_i (x_i - \mu)^2 p_i.$$

**Własności wariancji rozkładu**
Widzimy, że wariancja jest tym większa, im większa jest średnia odległość punktów $x_i$ od środka ciężkości $\mu$- wartości oczekiwanej.
Jeśli wszystkie wartości $x_i$ są sobie równe, wtedy wariancja jest równa zeru.

**Odchylenie standardowe**
Odchylenie standardowe zmiennej losowej $X$, oznaczane przez $DX$, definiujemy jako pierwiastek kwadratowy wariancji $X$. Odchylenie standardowe zmiennej losowej $X$ czesto jest też oznaczane grecką literą $\sigma$.
Można pokazać, że dla $a \geqslant 0$:

$$D(aX + b) = aDX. \tag{3}$$

**Niezależność zmiennych losowych**

**Definicja 7.** *Mówimy, że zmienne losowe $X$ i $Y$ są niezależne, jeżeli*

$$P(X \in [a, b] \wedge Y \in [c, d]) = P(X \in [a, b]) \times P(Y \in [c, d])$$

*dla dowolnych przedziałów $[a, b]$ i $[c, d]$.*

Intuicyjnie: niezależne zmienne losowe odpowiadają realizacje liczbowe niezależnych zmiennych losowych.

**Przykład**
Rozważmy jeszcze raz doświadczenie losowe polegające na wykonaniu przez zawodnika A dwóch rzutów osobistych (prawdopodobieństwo trafienia jest równe 0,9). Niech $Y_1$ oznacza wynik pierwszego rzutu (0, jeśli A chybił, 1 jeśli A trafił) a $Y_2$ wynik drugiego rzutu. Przyjęliśmy, że zdarzenie trafienia/chybienia w drugim rzucie jest niezależne od analogicznego zdarzenia w pierwszym rzucie. Przestrzeń zdarzeń elementarnych $\mathcal{S} = \{(C, C), (C, T), (T, C), (T, T)\}$

**Przykład—c.d.**
Mamy

$$\begin{aligned} P((C, C)) &= (0{,}1)^2 = 0{,}01, \\ P((C, T)) &= P((T, C)) = 0{,}1 \times 0{,}9 = 0{,}09, \\ P((T, T)) &= (0{,}9)^2 = 0{,}81, \end{aligned}$$

stąd:

$$\begin{aligned} P(Y_1 = 0) &= P(\{(C, C), (C, T)\} = 0{,}01 + 0{,}09 = 0{,}1, \\ P(Y_1 = 1) &= P(\{(T, C), (T, T)\} = 0{,}09 + 0{,}81 = 0{,}9, \\ P(Y_2 = 0) &= P(\{(C, C), (T, C)\} = 0{,}01 + 0{,}09 = 0{,}1, \\ P(Y_2 = 1) &= P(\{(C, T), (T, T)\} = 0{,}09 + 0{,}81 = 0{,}9 \end{aligned}$$

i $P((Y_1 = 0) \wedge (Y_2 = 0)) = 0{,}01 = P((Y_1 = 0)) \times P((Y_2 = 0))$ itd.; Stąd: zmienne $Y_1$ i $Y_2$ są niezależne.

**Przykład—c.d.**
Jesteśmy zainteresowani wartością oczekiwaną i wariancją zmiennej losowej $Y$.
Mamy $E(Y) = E(Y_1) + E(Y_2)$. Obliczamy $E(Y_1) = 0{,}1 \times 0 + 0{,}9 \times 1 = 0{,}9$; analogicznie $E(Y_2) = 0{,}9$; stąd $E(Y) = 2 \times 0{,}9 = 1{,}8$;
Z równości:

$$Var(Y_1) = Var(Y_2) = (0 - 0{,}9)^2 \times 0{,}1 + (1 - 0{,}9)^2 \times 0{,}9 = 0{,}09$$

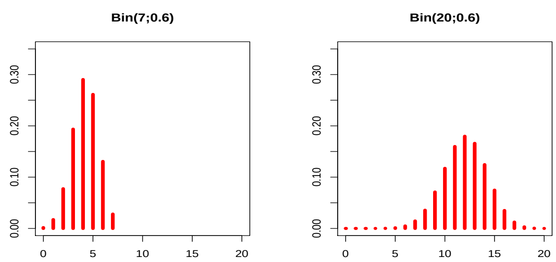
oraz z faktu, że $Y_1$ i $Y_2$ są niezależne, wynika:

$$Var(Y) = Var(Y_1) + Var(Y_2) = 2 \times 0{,}09 = 0{,}18.$$

Wariancję zmiennej losowej $Y$ można byłoby obliczyć bezpośrednio z definicji— byłoby to trochę bardziej żmudne.

**Wartości oczekiwane i wariancje rozkładów: dwumianowgo i Poissona**
Można pokazać, że wartość oczekiwana zmiennej losowej $X$, $X \sim Bin(n, p)$ wynosi $np$, a wariancja tej zmiennej jest równa $Var(Y) = np(1 - p)$. można skorzystać z przedstawienia zmiennej losowej $X$ jako sumy niezleżnych zmiennych losowych o rozkładzie $Bin(1, p)$
Można pokazać również, że dla zmienej losowej $Z$ o rozkładzie Poissona z parametrem $\lambda$ :

$$E(Z) = \lambda \text{ oraz } Var(Z) = \lambda.$$




Rysunek 1: Wykresy słupkowe przedstawiające rozkłady $Bin(7; 0{,}6)$ oraz $Bin(20; 0{,}6)$

Wariancja $X_1 \sim Bin(7; 0{,}6)$: $7 \times 0{,}6 \times 0{,}4 = 1{,}68$ Wariancja $X_2 \sim Bin(20; 0{,}6)$: $20 \times 0{,}6 \times 0{,}4 = 4{,}8$ $X_2$ jest bardziej „rozproszona”!

**Lektura uzupełniająca**
T. Bednarski, Elementy matematyki w naukach ekonomicznych. Oficyna ekonomiczna. Kraków 2004, str. 228–234.
Koronacki, J., Mielniczuk, J. Statystyka dla studentów kierunków technicznych i przyrodniczych. WNT. Warszawa 2001, podrozdział 2.2.1, str. 94–110.