Inne parametry - momenty

- **Momentem rzędu k względem początku układu współrzędnych** dla zmiennej losowej X nazywamy

$$\boxed{\tilde{\mu}_k \equiv E(X^k) \qquad k = 0,1,2,\dots}$$

czyli

$$\tilde{\mu}_0 = 1 \qquad \tilde{\mu}_1 = E(X)$$

- **Momentem centralnym rzędu k** dla zmiennej losowej X nazywamy

$$\boxed{\mu_k \equiv E((X - E(X))^k) \qquad k = 0,1,2,\dots}$$

- czyli

$$\mu_0 = 1 \qquad \mu_1 = 0 \qquad \mu_2 = var(X)$$

Inne parametry – skośność i kurtoza

- **Skośność** - miara asymetrii

$$\boxed{\beta_1 \equiv \frac{\mu_3}{(\sigma(X))^3}}$$

- **Kurtoza** - miara skupienia (spłaszczenia)

$$\boxed{\beta_2 \equiv \frac{\mu_4}{(\sigma(X))^4} - 3}$$

Przykład: rozkład prostokątny i rozkłady trójkątne

Transformacje zmiennych losowych

Syt. X – zmienna losowa
g(X) – funkcja o wartościach rzeczywistych
Y=g(X) też jest zmienną losową

Pytanie: jak wygląda rozkład prawdopodobieństwa (funkcja gęstości prawdopodobieństwa) zmiennej losowej Y

- Zmienna X jest dyskretna
  Wtedy również Y jest dyskretne. Niech x_i dla *i=1,…,j* są takie, że $g(x_i)=y_k$. Wtedy

$$P_Y(y_k) = \sum_{i=1}^{j} P_X(x_i)$$

Przykład:

$S_X=\{-2,0,2\}$ P(X=-2)=1/4 P(X=0)=1/4 P(X=2)=1/2

$Y=X^2$

$S_Y=\{0,4\}$ P(Y=0)=P(X=0)=1/4

P(Y=4)=P(X=-2)+P(X=2)=1/4+1/2=3/4

Transformacje zmiennych losowych

- Zmienna X jest ciągła, g(x) jest ciągłe
  Wtedy również Y jest ciągłe.
- I sposób (przez dystrybuantę)
  Szukamy $F_Y(y)$ a z niej liczymy $f_Y(y)$
- II sposób

$$\boxed{f_Y(y) = f_X(x)\cdot\left|\frac{dx}{dy}\right|}$$

Ogólny przypadek – sumujemy po przedziałach monotoniczności g(X)

$$\boxed{f_Y(y) = \sum_k f_{X,(k)}(x)\cdot\left|\frac{dx_{(k)}}{dy}\right|}$$

Przykłady: $Y=X^2$, $Y=X^4$

Dyskretne rozkłady prawdopodobieństwa

- **Rozkład zero-jedynkowy (rozkład Bernoulliego)** o parametrze p

Zmienna losowa to ilość sukcesów w jednokrotnym powtórzeniu eksperymentu w którym możliwe są tylko dwa wyniki

$S_X=\{0,1\}$

Rozkład prawdopodobieństwa

$$P(x) = \begin{cases} 1-p & dla\ x=0 \\ p & dla\ x=1 \end{cases}$$

zatem łącznie $P(X=x)=p^x(1-p)^{1-x}$ dla x=0,1

Dystrybuanta

$$F_X(x) = \begin{cases} 0 & dla\ x<0 \\ 1-p & dla\ 0 \le x < 1 \\ 1 & dla\ x \ge 1 \end{cases}$$

Dyskretne rozkłady prawdopodobieństwa

- **Rozkład dwupunktowy** o parametrze p
  Zmienna losowa przyjmuje dwie dowolne wartości, jedną z prawdopodobieństwem p, drugą z prawdopdobieństwem 1-p.

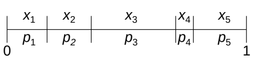
- **Rozkład wielopunktowy** o parametrach p_i, *i*=1,2,...,n
  przy czym $p_1+p_2+\ldots+p_n=1$.
  Zmienna losowa przyjmuje skończoną (n) ilość dyskretnych wartości.

- **Rozkład wielopunktowy** o parametrach p_i, *i*=1,2,...
  przy czym $p_1+p_2+\ldots=1$.
  Zmienna losowa przyjmuje nieskończoną ilość dyskretnych wartości

Dyskretne rozkłady prawdopodobieństwa – realizacja numeryczna

- 1. Odwracanie dystrybuanty (przykład w przyszłości)
- 2. Syt. Dysponujemy generatorem liczb pseudolosowych z przedziału (0,1)
- Dla rozkładu ze skończoną (=n) liczbą wartości zmiennej losowej X. Przedział (0,1) dzielimy na n przedziałów o długości p_i :

$$(0,p_1]\cup(p_1,p_1+p_2]\cup\ldots\cup\left(\sum_{k=1}^{n-1}p_k,1\right)$$



Losujemy liczbę Y $\in$(0,1) i znajdujemy dla niej przedział: $Y\in\left(\sum_{k=1}^{j-1}p_k,p_j\right]$

Jako wylosowaną wartość zmiennej losowej X przyjmujemy x_j.

Dyskretne rozkłady prawdopodobieństwa – realizacja numeryczna

- Syt. Dysponujemy generatorem liczb pseudolosowych z przedziału (0,1)
- Dla rozkładu z nieskończoną liczbą wartości zmiennej losowej X. Wybieramy ε (bardzo małe), p_i (jak największe) i n_{max}:

$$\sum_{k=1}^{n_{max}}p_k=1-\varepsilon \qquad (\varepsilon>0,\varepsilon\approx 0)$$

Przedział (0,1-ε) dzielimy na n_{max} przedziałów o długości p_i :

$$(0,p_1]\cup(p_1,p_1+p_2]\cup\ldots\cup\left(\sum_{k=1}^{n_{max}-1}p_k,1-\varepsilon\right)$$

Losujemy liczbę Y $\in$(0,1).
Jeżeli Y $\in$(0,1-ε) to znajdujemy dla niej przedział: $Y\in\left(\sum_{k=1}^{j-1}p_k,p_j\right]$

i jako wylosowaną wartość zmiennej losowej X przyjmujemy x_j.
Jeżeli Y $\in$[1-ε,1) to dodajemy przedziały powyżej n_{max} tak długo, aż znajdziemy przedział do którego należy Y.

Próba Bernoulliego i rozkład dwumienny

- **Próba Bernoulliego** to sekwencja powtórzeń n razy **tego samego** eksperymentu losowego, w wyniku którego możemy uzyskać jeden spośród dwóch wyników, zwanych jako „sukces" i „porażka".
  Muszą być spełnione dwa warunki:
  1. Powtórzenia są niezależne
  2. Prawdopodobieństwo sukcesu jest takie samo we wszystkich powtórzeniach.

  Przykład: n-krotny rzut monetą.
- **Rozkład dwumienny** (**dwumianowy**) (zwany w Polsce **rozkładem Bernoulliego**)
  Zmienna losowa X oznacza liczbę osiągniętych sukcesów w n-elementowej próbie Bernoulliego.
  S_X={0,1,2,…,n}

  Rozkład prawdopodobieństwa zmiennej X, przyjmującej wartości k$\in S_X$ jest postaci

$$\boxed{P_X(k)=\binom{n}{k}p^k(1-p)^{n-k}}$$

Rozkład dwumienny

Uzasadnienie:
Niech zdarzenie S oznacza osiągnięcie sukcesu w pojedynczym eksperymencie.
Niech zdarzenie P oznacza osiągnięcie porażki w pojedynczym eksperymencie.
Wynik próby (zdarzenie A) opisujemy jako n-elementowy ciąg wyników pojedynczych eksperymentów
np. A={S,P,S,S,…,P,P,S}
Prawdopodobieństwo uzyskania pojedynczego ciągu w którym wystąpiło k sukcesów to P(A)=$p^k(1\text{-}p)^{n\text{-}k}$ (niezależność prób)

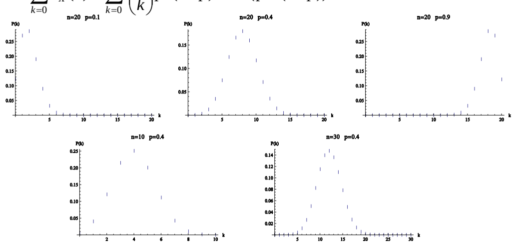
Ilość ciągów sprzyjających wystąpieniu k sukcesów to $\binom{n}{k}$ (kombinacja bez powtórzeń)
Zatem (rozłączność ciągów)

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

Rozkład dwumienny

- Rozkład ten jest poprawnie unormowany:

$$\sum_{k=0}^{n}P_X(k)=\sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k}=(p+(1-p))^n=1^n=1$$



Rozkład dwumienny

- Wartość oczekiwana w rozkładzie dwumianowym *E(X)=np*.
- Wariancja w rozkładzie dwumianowym *var(X)=np(1-p)*
- Dystrybuanta $F_X(x)=\sum_{k=0}^{\mathrm{int}(x)}P_X(k)$ (dystrybuanta jest określona na zbiorze ciągłym).
- Moda – zależy od n i p, mogą to być dwie wartości k (jest tak dla (n+1)p całkowitego, wtedy moda to k=(n+1)p i k=(n+1)p-1).
- **Rozkład wielomianowy** – wiele możliwych wyników pojedynczego eksperymentu, zachodzących z dowolnymi prawdopodobieństwami. Zachowujemy stałość prawdopodobieństw w kolejnych próbach i niezależność prób.

Rozkład geometryczny

- Zmienną losową jest ilość prób potrzebna do uzyskania sukcesu w nieskończenie długiej próbie Bernoulliego.
  Zbiór wartości S_X={1,2,3,…}
  Rozkład prawdopodobieństwa zmiennej X, przyjmującej wartości $k \in S_X$ jest postaci

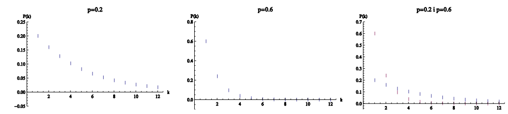
$$P_X(k) \equiv P(X=k) = P(\{\underbrace{\bar{A}\bar{A}\bar{A}...\bar{A}}_{k-1\ razy}A\}) = (1-p)^{k-1}p \equiv q^{k-1}p$$

  gdzie A oznacza uzyskanie sukcesu w pojedynczej próbie.
- Normalizacja:

$$\sum_{k=1}^{\infty} P_X(k) = \sum_{k=1}^{\infty} q^{k-1}p = p\sum_{k=1}^{\infty} q^{k-1} = p\underbrace{(1+q+q^2+\ldots)}_{|q|<1} = \frac{p}{1-q} = \frac{p}{p} = 1$$

- Wartość oczekiwana w rozkładzie geometrycznym E(X)=p^{-1}
- Wariancja w rozkładzie geometrycznym var(X)=$(1-p)p^{-2}$

Rozkład geometryczny



- Dystrybuanta (dla uproszczenia zapisu tylko w argumentach całkowitych)

$$F_X(n) = \sum_{k=1}^{n} P_X(k) = p\underbrace{(1+q+q^2+\ldots+q^{n-1})}_{|q|<1} = p\frac{1-q^n}{1-q} = \frac{p(1-q^n)}{p} = 1-q^n$$

  czyli

$$P(X>n) = 1-F_X(n) = 1-(1-q^n) = q^n$$

Rozkład geometryczny

- Zmienna losowa X o rozkładzie geometrycznym posiada własność „braku pamięci"

$$\forall k, j \in \{1,2,\ldots\}: \quad P_X(X>k+j \mid X>j) = P(X>k)$$

  czyli wcześniejsze wyniki nie wpływają na następne.
  Dowód:

$$P(X>k+j \mid X>j) = \frac{P(\{X>k+j\} \cap \{X>j\})}{P(X>j)} = \frac{P(X>k+j)}{P(X>j)} =$$

$$= \frac{1-F_X(k+j)}{1-F_X(j)} = \frac{1-(1-q^{k+j})}{1-(1-q^j)} = \frac{q^{k+j}}{q^j} = q^k = 1-(1-q^k) = 1-F_X(k) = P(X>k)$$

  Jest to jedyny dyskretny rozkład prawdopodobieństwa o tej własności.

- Istnieje również możliwość zdefiniowania rozkładu geometrycznego jako rozkładu zmiennej – liczby rzutów przed osiągnięciem pierwszego sukcesu. Wtedy S_X={0,1,2,…} i $P_X(k)=(1-p)^k p$.

Rozkład hipergeometryczny

- Opisuje ilość sukcesów (natrafień na element wybranego rodzaju) w wybieraniu n elementów spośród N, bez zwracania (w przeciwieństwie do rozkładu dwumianowego). Prawdopodobieństwo sukcesu w pierwszej próbie wynosi m/N, czyli m to liczba wyróżnionych elementów.

$$P_X(k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}} \qquad \begin{array}{l} m \in \{0,1,2,\ldots,N\} \\ n \in \{1,2,\ldots,N\} \\ k \in \{max(0,n+m-N),\ldots,min(m,n)\} \end{array}$$

  Dla n<m wartość oczekiwana E(X)=n*m/N

  Przykłady: totolotek, oznaczenie ilości wadliwych sztuk w całej partii wyrobów, oznaczenie ilości ryb w jeziorze (liczbę uzyskanych sukcesów k_{exp} traktuje się jako przybliżenie E(X)) i stąd $N = \frac{n \cdot m}{E(X)} = \frac{n \cdot m}{k_{exp}}$

Rozkład Poissona

- Opisuje prawdopodobieństwo zajścia k sukcesów w określonym przedziale czasu, jeżeli znamy średnią częstotliwość wystąpienia sukcesu, a czas oczekiwania na sukces podlega rozkładowi wykładniczemu.

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

  gdzie λ to parametr rozkładu, a *k*={0,1,2,…}

  Wartość oczekiwana w rozkładzie Poissona E(X)= λ
  Wariancja w rozkładzie Poissona var(X)= λ

  Zastosowania: liczba ludzi pochodzących do kasy w supermarkecie w określonym przedziale czasu (np.2 minut), liczba wejść na stronę internetową w określonym przedziale czasu, liczba rozpadów jader atomowych, liczba mutacji DNA, liczba wypadków lotniczych w ciągu roku, liczba sprzedanych sztuk drogiego towaru, (duża próbka i rzadkie zjawisko)

Rozkład Poissona

Rozkład Poissona dla wybranych wartości parametru λ