Dyskretne rozkłady prawdopodobieństwa

- **Rozkład zero-jedynkowy (rozkład Bernoullego)** 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 $P(x)=\begin{cases}1-p & dla\ x=0\\ p & dla\ x=1\end{cases}$
- **Rozkład dwupunktowy** o parametrze p
  Zmienna losowa przyjmuje dwie dowolne wartości, jedną z prawdopodobieństwem p, drugą z prawdopodobieństwem 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

1

Próba Bernoulliego (Bernoullego) 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.
  $S_X=\{0,1,2,\ldots,n\}$

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

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

2

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 $P(A)=p^k(1-p)^{n-k}$ (niezależność prób)

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

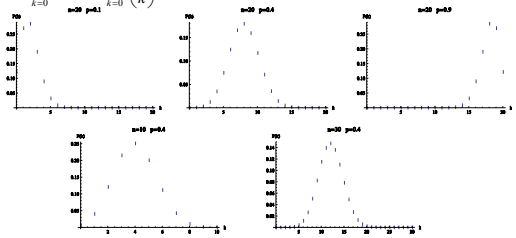
$$P_X(k)=\binom{n}{k}P(A)=\binom{n}{k}p^k(1-p)^{n-k}$$

3

Rozkład dwumienny

- Rozkład ten jest poprawnie unormowany:

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



4

Rozkład dwumianowy

- 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. Opisujemy $P(k_1,k_2,\ldots)$.

5

Rozkład geometryczny

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

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

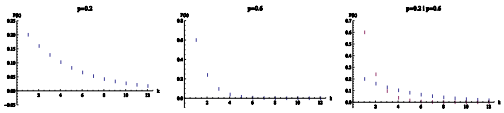
  gdzie A oznacza uzyskanie sukcesu w pojedynczej próbie.
- Normalizacja:

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

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

6

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

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

7

Rozkład geometryczny

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

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

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 $S_X=\{0,1,2,\ldots\}$ i $P_X(k)=(1-p)^k p$.

8

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 $k=\{0,1,2,\ldots\}$

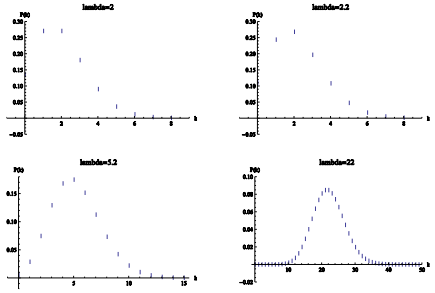
Wartość oczekiwana w rozkładzie Poissona E(X)= λ
Wariancja w rozkładzie Poissona var(X)= λ

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)

9

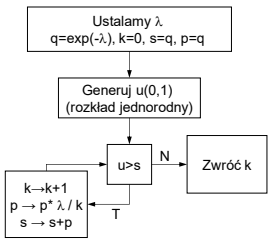
Rozkład Poissona

Rozkład Poissona dla wybranych wartości parametru λ



10

Rozkład Poissona – realizacja numeryczna



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

$$\begin{aligned}
&dla\ x<0 && F_X(x)=0\\
&dla\ 0\le x<1 && F_X(x)=P(k=0)=e^{-\lambda}\\
&dla\ 1\le x<2 && F_X(x)=P(k=0)+P(k=1)=e^{-\lambda}+\lambda e^{-\lambda}\\
&dla\ 2\le x<3 && F_X(x)=e^{-\lambda}+\lambda e^{-\lambda}+\frac{1}{2}\lambda^2 e^{-\lambda}\\
&dla\ 3\le x<4 && F_X(x)=e^{-\lambda}+\lambda e^{-\lambda}+\frac{1}{2}\lambda^2 e^{-\lambda}+\frac{1}{6}\lambda^3 e^{-\lambda}
\end{aligned}$$

$$\begin{aligned}
&dla\ k=0 && q=e^{-\lambda} && p=e^{-\lambda} && s=e^{-\lambda}\\
&dla\ k=1 && p=\frac{p\lambda}{k}=\lambda e^{-\lambda} && s=e^{-\lambda}+\lambda e^{-\lambda}=F_X(1)\\
&dla\ k=2 && p=\frac{1}{2}\lambda^2 e^{-\lambda} && s=e^{-\lambda}+\lambda e^{-\lambda}+\frac{1}{2}\lambda^2 e^{-\lambda}=F_X(2)\\
&dla\ k=3 && p=\frac{1}{2}\lambda^2 e^{-\lambda}\frac{\lambda}{3}=\frac{1}{6}\lambda^3 e^{-\lambda} && s=F_X(2)+\frac{1}{6}\lambda^3 e^{-\lambda}=F_X(3)
\end{aligned}$$

Istnieją związki pomiędzy rozkładami;

11

Przykład: rozkład Poissona jako granica rozkładu dwumianowego dla dużych n, małych p, i n·p=const= λ

Rozkład dwumianowy: $B_k^n\equiv P_n(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$ $\qquad P_X(k)=\frac{\lambda^k}{k!}e^{-\lambda}$

dla k=0

$$B_0^n=\binom{n}{0}p^0(1-p)^{n-0}=(1-p)^n=\left(1-\tfrac{\lambda}{n}\right)^n\ \rightarrow\ \lim_{n\to\infty}B_0^n=\lim_{n\to\infty}\left(1-\tfrac{\lambda}{n}\right)^n=e^{-\lambda}=P_X(X=0)$$

dla dowolnych *k* i *k+1*

$$\frac{B_{k+1}^n}{B_k^n}=\frac{\binom{n}{k+1}p^{k+1}(1-p)^{n-(k+1)}}{\binom{n}{k}p^k(1-p)^{n-k}}=\frac{n!k!(n-k)!}{(k+1)!(n-k-1)!n!}p(1-p)^{-1}=\frac{n-k}{k+1}\cdot\frac{p}{1-p}=\frac{np-pk}{(k+1)\cdot(1-p)}=$$

$$=\frac{\lambda-\frac{\lambda}{n}k}{(k+1)\cdot(1-\frac{\lambda}{n})}=\frac{\lambda(1-\frac{k}{n})}{(k+1)\cdot(1-\frac{\lambda}{n})}\ \rightarrow\ \lim_{n\to\infty}\frac{B_{k+1}^n}{B_k^n}=\lim_{n\to\infty}\frac{\lambda(1-\frac{k}{n})}{(k+1)\cdot(1-\frac{\lambda}{n})}=\frac{\lambda}{k+1}$$

stąd $\lim_{n\to\infty}\frac{B_1^n}{B_0^n}=\frac{\lim_{n\to\infty}B_1^n}{\lim_{n\to\infty}B_0^n}=\frac{\lambda}{1}=\lambda\ \rightarrow\ \lim_{n\to\infty}B_1^n=\lambda\lim_{n\to\infty}B_0^n=\lambda e^{-\lambda}=P_X(X=1)$

$$\lim_{n\to\infty}\frac{B_2^n}{B_1^n}=\frac{\lim_{n\to\infty}B_2^n}{\lim_{n\to\infty}B_1^n}=\frac{\lambda}{1+1}=\frac{\lambda}{2}\ \rightarrow\ \lim_{n\to\infty}B_2^n=\frac{\lambda}{2}\lim_{n\to\infty}B_1^n=\frac{\lambda^2 e^{-\lambda}}{2}=P_X(X=2)$$

$$\rightarrow\ \lim_{n\to\infty}B_{k+1}^n=\frac{\lambda}{k+1}\lim_{n\to\infty}B_k^n=\frac{\lambda\cdot\lambda^k e^{-\lambda}}{(k+1)\cdot k!}=\frac{\lambda^{k+1}e^{-\lambda}}{(k+1)!}=P_X(X=k+1)$$

opis „zjawisk rzadkich"

12

Zmienna losowa ciągła

- Rozkład jednorodny

$$f_X(x) = \begin{cases} \frac{1}{b-a} & dla\ x \in [a,b] \\ 0 & dla\ x \notin [a,b] \end{cases}$$

13

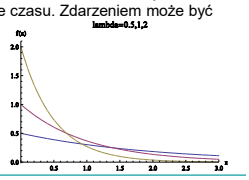
Rozkład wykładniczy (eksponencjalny)

- Jest to rozwinięcie rozkładu geometrycznego na ciągłe zmienne losowe

$$f_X(x) = \lambda\, e^{-\lambda x} \qquad x \in [0,+\infty)$$

λ – parametr rozkładu, λ >0

Gdy x interpretujemy jako czas to rozkład wykładniczy opisuje prawdopodobieństwo tego, że na zajście zdarzenia będziemy czekać czas x. Zakładamy, że zdarzenia zachodzi ze stałym prawdopodobieństwem w jednostce czasu. Zdarzeniem może być przejście układu do nowego stanu.



Zastosowania:
czas dostępu do serwera
promieniowanie kosmiczne
czas wezwania karetki pogotowia
czas obsługi klienta w banku

14

Rozkład wykładniczy -własności

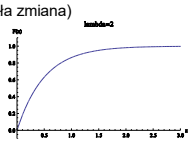
- Normalizacja

$$\int_0^{+\infty} \lambda\, e^{-\lambda x} dx = \lambda \int_0^{+\infty} e^{-\lambda x} dx = \lambda \frac{1}{(-\lambda)} e^{-\lambda x}\Big|_0^{+\infty} = (-1)\left(e^{-\lambda\infty} - e^{-\lambda 0}\right) = (-1)(0-1) = 1$$

- Dystrybuanta (prawdopodobieństwo, że zaszła zmiana)

$$F_X(x) = \int_0^x \lambda\, e^{-\lambda t} dt = (-1)\left(e^{-\lambda x} - e^{-\lambda 0}\right) = 1 - e^{-\lambda x}$$

$$\rightarrow \quad P_X(X > x) = 1 - F_X(x) = e^{-\lambda x}$$



- Posiada własność braku pamięci

$$\forall s,t > 0: \quad P_X(X > t+s \mid X > t) = P_X(X > s)$$

(dowód analog. do rozkładu geometrycznego, zastosowanie: metoda datowania węglem ^{14}C)

- Wartość oczekiwana E(X)=1/ λ var(X)=1/(λ^2)

15