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.
  $S_X=\{0,1,2,\ldots,n\}$

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

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

1

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)

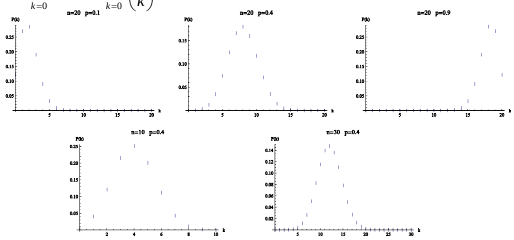
$$P_X(k)=\binom{n}{k}P(A)=\binom{n}{k}p^k(1-p)^{n-k}$$

2

Rozkład dwumienny

- Rozkład ten jest poprawnie unormowany:

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



3

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.

4

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

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

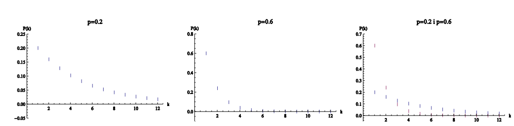
  gdzie A oznacza uzyskanie sukcesu w pojedynczej próbie.
- Normalizacja:

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

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

5

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

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

6

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

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

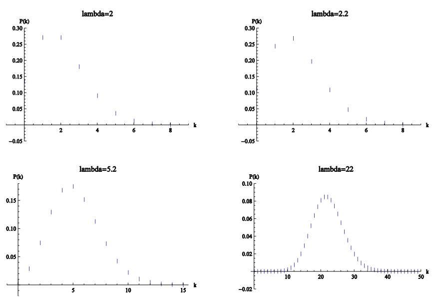
  Wartość oczekiwana w rozkładzie Poissona $E(X)=\lambda$

  Wariancja w rozkładzie Poissona $var(X)=\lambda$

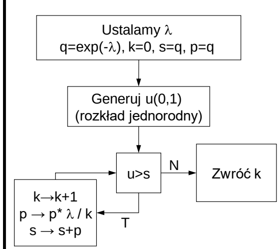
  Zastosowania: liczba ludzi pochodzących do kasy w supermarkecie w określonym przedziale czasu (np.2 minut), liczba wejść na stronę internetową w określonym przedziale czasu, liczba rozpadów jader atomowych, liczba mutacji DNA, liczba wypadków lotniczych w ciągu roku, liczba sprzedanych sztuk drogiego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

Rozkład Poissona

Rozkład Poissona dla wybranych wartości parametru λ



Rozkład Poissona – realizacja numeryczna



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

$dla\ x < 0 \quad F_X(x) = 0$

$dla\ 0 \le x < 1 \quad F_X(x) = P(k=0) = e^{-\lambda}$

$dla\ 1 \le x < 2 \quad F_X(x) = P(k=0) + P(k=1) = e^{-\lambda} + \lambda e^{-\lambda}$

$dla\ 2 \le x < 3 \quad F_X(x) = e^{-\lambda} + \lambda e^{-\lambda} + \frac{1}{2}\lambda^2 e^{-\lambda}$

$dla\ 3 \le x < 4 \quad F_X(x) = e^{-\lambda} + \lambda e^{-\lambda} + \frac{1}{2}\lambda^2 e^{-\lambda} + \frac{1}{6}\lambda^3 e^{-\lambda}$

$dla\ k = 0 \quad q = e^{-\lambda} \quad p = e^{-\lambda} \quad s = e^{-\lambda}$

$dla\ k = 1 \quad p = \frac{p\lambda}{k} = \lambda e^{-\lambda} \quad s = e^{-\lambda} + \lambda e^{-\lambda} \equiv F_X(1)$

$dla\ k = 2 \quad p = \frac{1}{2}\lambda^2 e^{-\lambda} \quad s = e^{-\lambda} + \lambda e^{-\lambda} + \frac{1}{2}\lambda^2 e^{-\lambda} \equiv F_X(2)$

$dla\ k = 3 \quad p = \frac{1}{2}\lambda^2 e^{-\lambda}\frac{\lambda}{3} = \frac{1}{6}\lambda^3 e^{-\lambda} \quad s = F_X(2) + \frac{1}{6}\lambda^3 e^{-\lambda} \equiv F_X(3)$