Metody probabilistyczne

8. Wielowymiarowe zmienne losowe II

Wojciech Kotłowski

Instytut Informatyki PP
http://www.cs.put.poznan.pl/wkotlowski/

28.11.2017

Funkcja zmiennych losowych

Twierdzenie

Niech $Z = f(\boldsymbol{X})$ będzie funkcją wektora dyskretnych zmiennych losowych $\boldsymbol{X} = (X_1, \ldots, X_n)$. Zachodzi:

$$EZ = \sum_{\boldsymbol{x}} f(\boldsymbol{x}) P(\boldsymbol{X} = \boldsymbol{x})$$

Funkcja zmiennych losowych

Twierdzenie

Niech $Z = f(\boldsymbol{X})$ będzie funkcją wektora dyskretnych zmiennych losowych $\boldsymbol{X} = (X_1, \ldots, X_n)$. Zachodzi:

$$EZ \;=\; \sum_{\boldsymbol{x}} f(\boldsymbol{x}) P(\boldsymbol{X} = \boldsymbol{x})$$

Dowód: Analogicznie do przypadku z jedną zmienną losową. Ponieważ

$$P(Z = z) = \sum_{\boldsymbol{x}\colon\, f(\boldsymbol{x}) = z} P(\boldsymbol{X} = \boldsymbol{x}),$$

Funkcja zmiennych losowych

Twierdzenie

Niech $Z = f(\boldsymbol{X})$ będzie funkcją wektora dyskretnych zmiennych losowych $\boldsymbol{X} = (X_1, \ldots, X_n)$. Zachodzi:

$$EZ = \sum_{\boldsymbol{x}} f(\boldsymbol{x}) P(\boldsymbol{X} = \boldsymbol{x})$$

Dowód: Analogicznie do przypadku z jedną zmienną losową. Ponieważ

$$P(Z = z) = \sum_{\boldsymbol{x}\colon f(\boldsymbol{x}) = z} P(\boldsymbol{X} = \boldsymbol{x}), \qquad \text{to:}$$

$$EZ = \sum_{z} z\, P(Z = z)$$

Funkcja zmiennych losowych

Twierdzenie

Niech $Z = f(\boldsymbol{X})$ będzie funkcją wektora dyskretnych zmiennych losowych $\boldsymbol{X} = (X_1, \ldots, X_n)$. Zachodzi:

$$EZ \;=\; \sum_{\boldsymbol{x}} f(\boldsymbol{x}) P(\boldsymbol{X} = \boldsymbol{x})$$

Dowód: Analogicznie do przypadku z jedną zmienną losową. Ponieważ

$$P(Z = z) = \sum_{\boldsymbol{x}\colon\, f(\boldsymbol{x}) = z} P(\boldsymbol{X} = \boldsymbol{x}), \qquad \text{to:}$$

$$EZ \;=\; \sum_z z\, P(Z = z) \;=\; \sum_z z \Big(\sum_{\boldsymbol{x}\colon\, f(\boldsymbol{x}) = z} P(\boldsymbol{X} = \boldsymbol{x}) \Big)$$

Funkcja zmiennych losowych

Twierdzenie

Niech $Z = f(\boldsymbol{X})$ będzie funkcją wektora dyskretnych zmiennych losowych $\boldsymbol{X} = (X_1, \ldots, X_n)$. Zachodzi:

$$EZ \;=\; \sum_{\boldsymbol{x}} f(\boldsymbol{x}) P(\boldsymbol{X} = \boldsymbol{x})$$

Dowód: Analogicznie do przypadku z jedną zmienną losową. Ponieważ

$$P(Z = z) = \sum_{\boldsymbol{x}\colon f(\boldsymbol{x})=z} P(\boldsymbol{X} = \boldsymbol{x}), \qquad \text{to:}$$

$$\begin{aligned} EZ &= \sum_z z\, P(Z = z) = \sum_z z \Big(\sum_{\boldsymbol{x}\colon f(\boldsymbol{x})=z} P(\boldsymbol{X} = \boldsymbol{x}) \Big) \\ &= \sum_z \sum_{\boldsymbol{x}\colon f(\boldsymbol{x})=z} f(\boldsymbol{x}) P(\boldsymbol{X} = \boldsymbol{x}) \end{aligned}$$

Funkcja zmiennych losowych

Twierdzenie

Niech $Z = f(\boldsymbol{X})$ będzie funkcją wektora dyskretnych zmiennych losowych $\boldsymbol{X} = (X_1, \ldots, X_n)$. Zachodzi:

$$EZ = \sum_{\boldsymbol{x}} f(\boldsymbol{x}) P(\boldsymbol{X} = \boldsymbol{x})$$

Dowód: Analogicznie do przypadku z jedną zmienną losową. Ponieważ

$$P(Z = z) = \sum_{\boldsymbol{x}\colon f(\boldsymbol{x})=z} P(\boldsymbol{X} = \boldsymbol{x}), \qquad \text{to:}$$

$$\begin{aligned} EZ &= \sum_{z} z\, P(Z = z) = \sum_{z} z \Big(\sum_{\boldsymbol{x}\colon f(\boldsymbol{x})=z} P(\boldsymbol{X} = \boldsymbol{x}) \Big) \\ &= \sum_{z} \sum_{\boldsymbol{x}\colon f(\boldsymbol{x})=z} f(\boldsymbol{x}) P(\boldsymbol{X} = \boldsymbol{x}) = \sum_{\boldsymbol{x}} f(\boldsymbol{x}) P(\boldsymbol{X} = \boldsymbol{x}) \end{aligned}$$

Funkcja zmiennych losowych

Twierdzenie

Niech $Z = f(\boldsymbol{X})$ będzie funkcją wektora dyskretnych zmiennych losowych $\boldsymbol{X} = (X_1, \ldots, X_n)$. Zachodzi:

$$EZ = \sum_{\boldsymbol{x}} f(\boldsymbol{x}) P(\boldsymbol{X} = \boldsymbol{x})$$

Uwaga: W szczególności dla $Z = f(X, Y)$, gdzie $X \in \mathcal{X}$, $Y \in \mathcal{Y}$:

$$EZ = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} f(x, y)\, P(X = x, Y = y)$$

Addytywność wartości oczekiwanej

Dla dowolnych zmiennych losowych X i Y zachodzi:

$$E(X+Y) = EX + EY$$

Addytywność wartości oczekiwanej

Dla dowolnych zmiennych losowych X i Y zachodzi:

$$E(X+Y) \;=\; EX + EY$$

Dowód: Weźmy $Z = f(X, Y) = X + Y$.

Addytywność wartości oczekiwanej

Dla dowolnych zmiennych losowych X i Y zachodzi:

$$E(X+Y) \;=\; EX+EY$$

Dowód: Weźmy $Z = f(X,Y) = X+Y$.

$$E(X+Y) \;\;=\;\; EZ \;=\; \sum_x \sum_y f(x,y)\, P(X=x, Y=y)$$

Addytywność wartości oczekiwanej

Dla dowolnych zmiennych losowych X i Y zachodzi:

$$E(X+Y) = EX + EY$$

Dowód: Weźmy $Z = f(X, Y) = X + Y$.

$$\begin{aligned} E(X+Y) &= EZ = \sum_x \sum_y f(x,y)\, P(X=x, Y=y) \\ &= \sum_x \sum_y (x+y)\, P(X=x, Y=y) \end{aligned}$$

Addytywność wartości oczekiwanej

Dla dowolnych zmiennych losowych X i Y zachodzi:

$$E(X+Y) \; = \; EX + EY$$

Dowód: Weźmy $Z = f(X,Y) = X + Y$.

$$\begin{aligned}
E(X+Y) \;&=\; EZ \;=\; \sum_x \sum_y f(x,y)\, P(X=x, Y=y) \\
&=\; \sum_x \sum_y (x+y)\, P(X=x, Y=y) \\
&=\; \sum_x x \sum_y \; P(X=x, Y=y) + \sum_y y \sum_x \; P(X=x, Y=y)
\end{aligned}$$

Addytywność wartości oczekiwanej

Dla dowolnych zmiennych losowych X i Y zachodzi:

$$E(X+Y) \;=\; EX + EY$$

Dowód: Weźmy $Z = f(X,Y) = X + Y$.

$$\begin{aligned}
E(X+Y) &= EZ = \sum_x \sum_y f(x,y)\, P(X=x, Y=y) \\
&= \sum_x \sum_y (x+y)\, P(X=x, Y=y) \\
&= \sum_x x \underbrace{\sum_y P(X=x, Y=y)}_{P(X=x)} + \sum_y y \underbrace{\sum_x P(X=x, Y=y)}_{P(Y=y)} \\
&= \sum_x x P(X=x) \;+\; \sum_y y P(Y=y) \;=\; EX + EY
\end{aligned}$$

Addytywność wartości oczekiwanej

Dla dowolnych zmiennych losowych $X_1, X_2, \ldots, X_n$ zachodzi:

$$E(X_1 + X_2 + \ldots + X_n) \; = \; EX_1 + EX_2 + \ldots + EX_n$$

Addytywność wartości oczekiwanej

Dla dowolnych zmiennych losowych $X_1, X_2, \ldots, X_n$ zachodzi:

$$E(X_1 + X_2 + \ldots + X_n) \ = \ EX_1 + EX_2 + \ldots + EX_n$$

Dowód: Przez indukcję po n

Addytywność wartości oczekiwanej

Dla dowolnych zmiennych losowych $X_1, X_2, \ldots, X_n$ zachodzi:

$$E(X_1 + X_2 + \ldots + X_n) \ = \ EX_1 + EX_2 + \ldots + EX_n$$

Dowód: Przez indukcję po n

Zadanie 1

Udowodnij to twierdzenie

Przykład

Rzucamy n razy kostką. Oblicz wartość oczekiwaną sumy oczek.

Przykład

Rzucamy n razy kostką. Oblicz wartość oczekiwaną sumy oczek.

X_i – wynik rzutu na i-tej kostce ($i = 1, \ldots, n$)

$Y = X_1 + \ldots + X_n$ – sumaryczny wynik n rzutów

Przykład

Rzucamy n razy kostką. Oblicz wartość oczekiwaną sumy oczek.

X_i – wynik rzutu na i-tej kostce ($i = 1, \ldots, n$)
$Y = X_1 + \ldots + X_n$ – sumaryczny wynik n rzutów

$$EY \;=\; EX_1 + \ldots EX_n \;=\; 3.5 \cdot n$$

Przykład

Oblicz średnią liczbę par osób mających urodziny tego samego dnia w grupie n osób.

Przykład

Oblicz średnią liczbę par osób mających urodziny tego samego dnia w grupie n osób.

Dla danej pary (i, j) (zakładamy $i < j$), niech $X_{i,j} \in \{0, 1\}$ określa, czy para ma urodziny tego samego dnia ($X_{i,j} = 1$).

Przykład

Oblicz średnią liczbę par osób mających urodziny tego samego dnia w grupie n osób.

Dla danej pary (i, j) (zakładamy $i < j$), niech $X_{i,j} \in \{0, 1\}$ określa, czy para ma urodziny tego samego dnia ($X_{i,j} = 1$).

$$P(X_{i,j} = 1) =$$

Przykład

Oblicz średnią liczbę par osób mających urodziny tego samego dnia w grupie n osób.

Dla danej pary (i, j) (zakładamy $i < j$), niech $X_{i,j} \in \{0, 1\}$ określa, czy para ma urodziny tego samego dnia ($X_{i,j} = 1$).

$$P(X_{i,j} = 1) \; = \; \frac{1}{365}, \qquad P(X_{i,j} = 0) \; = \; 1 - \frac{1}{365}$$

Przykład

Oblicz średnią liczbę par osób mających urodziny tego samego dnia w grupie n osób.

Dla danej pary (i, j) (zakładamy $i < j$), niech $X_{i,j} \in \{0, 1\}$ określa, czy para ma urodziny tego samego dnia ($X_{i,j} = 1$).

$$P(X_{i,j} = 1) \ = \ \frac{1}{365}, \qquad P(X_{i,j} = 0) \ = \ 1 - \frac{1}{365}$$

$$EX_{i,j} \ = \ \frac{1}{365}$$

Przykład

Oblicz średnią liczbę par osób mających urodziny tego samego dnia w grupie n osób.

Dla danej pary (i,j) (zakładamy $i < j$), niech $X_{i,j} \in \{0,1\}$ określa, czy para ma urodziny tego samego dnia ($X_{i,j} = 1$).

$$P(X_{i,j} = 1) = \frac{1}{365}, \qquad P(X_{i,j} = 0) = 1 - \frac{1}{365}$$

$$EX_{i,j} = \frac{1}{365}$$

$Y = \sum_{i<j} X_{i,j}$ – liczbę par mających urodziny tego samego dnia

Przykład

Oblicz średnią liczbę par osób mających urodziny tego samego dnia w grupie n osób.

Dla danej pary (i, j) (zakładamy $i < j$), niech $X_{i,j} \in \{0, 1\}$ określa, czy para ma urodziny tego samego dnia ($X_{i,j} = 1$).

$$P(X_{i,j} = 1) \ = \ \frac{1}{365}, \qquad P(X_{i,j} = 0) \ = \ 1 - \frac{1}{365}$$

$$EX_{i,j} \ = \ \frac{1}{365}$$

$Y = \sum_{i<j} X_{i,j}$ – liczbę par mających urodziny tego samego dnia

$$EY \ = \ \sum_{i<j} EX_{i,j} \ = \ \frac{n(n-1)}{2} \cdot \frac{1}{365}$$

Przykład

Wkładamy losowo n listów do różnych adresatów do n kopert. Znaleźć średnią liczbę listów włożonych prawidłowo.

Przykład

Wkładamy losowo n listów do różnych adresatów do n kopert. Znaleźć średnią liczbę listów włożonych prawidłowo.

$X_i \in \{0, 1\}$ – zmienna określająca czy i-ty list został włożony do prawidłowej koperty ($X_i = 1$)

Przykład

Wkładamy losowo n listów do różnych adresatów do n kopert. Znaleźć średnią liczbę listów włożonych prawidłowo.

$X_i \in \{0, 1\}$ – zmienna określająca czy i-ty list został włożony do prawidłowej koperty ($X_i = 1$)

$Y = X_1 + \ldots + X_n$ – liczba listów włożonych prawidłowo

Przykład

Wkładamy losowo n listów do różnych adresatów do n kopert. Znaleźć średnią liczbę listów włożonych prawidłowo.

$X_i \in \{0, 1\}$ – zmienna określająca czy i-ty list został włożony do prawidłowej koperty ($X_i = 1$)

$Y = X_1 + \ldots + X_n$ – liczba listów włożonych prawidłowo

$$P(X_i = 1) =$$

Przykład

Wkładamy losowo n listów do różnych adresatów do n kopert. Znaleźć średnią liczbę listów włożonych prawidłowo.

$X_i \in \{0, 1\}$ – zmienna określająca czy i-ty list został włożony do prawidłowej koperty ($X_i = 1$)

$Y = X_1 + \ldots + X_n$ – liczba listów włożonych prawidłowo

$$P(X_i = 1) = \frac{1}{n}, \qquad EX_i = \frac{1}{n}$$

Przykład

Wkładamy losowo n listów do różnych adresatów do n kopert. Znaleźć średnią liczbę listów włożonych prawidłowo.

$X_i \in \{0,1\}$ – zmienna określająca czy i-ty list został włożony do prawidłowej koperty ($X_i = 1$)

$Y = X_1 + \ldots + X_n$ – liczba listów włożonych prawidłowo

$$P(X_i = 1) = \frac{1}{n}, \qquad EX_i = \frac{1}{n}$$

$$EY = \sum_{i=1}^{n} EX_i = n \cdot \frac{1}{n} = 1$$

Przykład

W klasie 30 uczniów na każdej lekcji jeden losowo wybrany uczeń rozwiązuje zadanie przy tablicy. Wyznacz średnią liczbę uczniów, którzy ani razu nie pójdą do tablicy przez 30 lekcji.

Przykład

Wrzucamy losowo m kul do n urn. Wyznacz średnią liczbę pustych urn.

Przykład

Wrzucamy losowo m kul do n urn. Wyznacz średnią liczbę pustych urn.

$X_i \in \{0, 1\}$ – zmienna określająca czy i-ta urna jest pusta ($X_i = 1$)

Przykład

Wrzucamy losowo m kul do n urn. Wyznacz średnią liczbę pustych urn.

$X_i \in \{0, 1\}$ – zmienna określająca czy i-ta urna jest pusta ($X_i = 1$)

$Y = X_1 + \ldots + X_n$ – sumaryczna liczba pustych urn.

Przykład

Wrzucamy losowo m kul do n urn. Wyznacz średnią liczbę pustych urn.

$X_i \in \{0, 1\}$ – zmienna określająca czy i-ta urna jest pusta ($X_i = 1$)

$Y = X_1 + \ldots + X_n$ – sumaryczna liczba pustych urn.

$$P(X_i = 1) =$$

Przykład

Wrzucamy losowo m kul do n urn. Wyznacz średnią liczbę pustych urn.

$X_i \in \{0,1\}$ – zmienna określająca czy i-ta urna jest pusta ($X_i = 1$)

$Y = X_1 + \ldots + X_n$ – sumaryczna liczba pustych urn.

$$P(X_i = 1) = \left(1 - \frac{1}{n}\right)^m, \qquad EX_i = \left(1 - \frac{1}{n}\right)^m$$

Przykład

Wrzucamy losowo m kul do n urn. Wyznacz średnią liczbę pustych urn.

$X_i \in \{0, 1\}$ – zmienna określająca czy i-ta urna jest pusta ($X_i = 1$)

$Y = X_1 + \ldots + X_n$ – sumaryczna liczba pustych urn.

$$P(X_i = 1) \;=\; \left(1 - \frac{1}{n}\right)^m, \qquad EX_i \;=\; \left(1 - \frac{1}{n}\right)^m$$

$$EY \;=\; \sum_{i=1}^{n} \left(1 - \frac{1}{n}\right)^m \;=\; n\left(1 - \frac{1}{n}\right)^m$$

Przykład

Wrzucamy losowo m kul do n urn. Wyznacz średnią liczbę pustych urn.

$X_i \in \{0, 1\}$ – zmienna określająca czy i-ta urna jest pusta ($X_i = 1$)

$Y = X_1 + \ldots + X_n$ – sumaryczna liczba pustych urn.

$$P(X_i = 1) \;=\; \left(1 - \frac{1}{n}\right)^m, \qquad EX_i \;=\; \left(1 - \frac{1}{n}\right)^m$$

$$EY \;=\; \sum_{i=1}^{n} \left(1 - \frac{1}{n}\right)^m \;=\; n\left(1 - \frac{1}{n}\right)^m$$

W przypadku gdy $n = m$,

$$\left(1 - \frac{1}{n}\right)^n \;\simeq\; \frac{1}{e}, \qquad \text{stąd} \;\; EY \;\simeq\; \frac{n}{e} \;\simeq\; 0.37n$$

Przykład

Wrzucamy losowo m kul do n urn. Wyznacz średnią liczbę pustych urn.

$X_i \in \{0, 1\}$ – zmienna określająca czy i-ta urna jest pusta ($X_i = 1$)

$Y = X_1 + \ldots + X_n$ – sumaryczna liczba pustych urn.

$$P(X_i = 1) \; = \; \Big(1 - \frac{1}{n}\Big)^m, \qquad EX_i \; = \; \Big(1 - \frac{1}{n}\Big)^m$$

$$EY \; = \; \sum_{i=1}^{n} \Big(1 - \frac{1}{n}\Big)^m \; = \; n\Big(1 - \frac{1}{n}\Big)^m$$

W przypadku gdy $n = m$,

$$\Big(1 - \frac{1}{n}\Big)^n \; \simeq \; \frac{1}{e}, \qquad \text{stąd} \;\; EY \; \simeq \; \frac{n}{e} \; \simeq \; 0.37n$$

Ok. 37% uczniów nie pójdzie ani razu do tablicy.

Przykład

Małpa pisze na klawiaturze, za każdym razem wciskając losowo jedną z 26 liter alfabetu łacińskiego. Jaka jest oczekiwana liczba słów MALPA w ciągu $n = 1\ 000\ 000$ liter?

Przykład

Małpa pisze na klawiaturze, za każdym razem wciskając losowo jedną z 26 liter alfabetu łacińskiego. Jaka jest oczekiwana liczba słów MALPA w ciągu $n = 1\ 000\ 000$ liter?

$X_i \in \{0, 1\}$ – zmienna określająca, czy na i-tej pozycji pojawiło się słowo MALPA ($i = 1, \ldots, n-4$)

Przykład

Małpa pisze na klawiaturze, za każdym razem wciskając losowo jedną z 26 liter alfabetu łacińskiego. Jaka jest oczekiwana liczba słów MALPA w ciągu $n = 1\ 000\ 000$ liter?

$X_i \in \{0, 1\}$ – zmienna określająca, czy na i-tej pozycji pojawiło się słowo MALPA ($i = 1, \ldots, n - 4$)

$$P(X_i = 1) \ =$$

Przykład

Małpa pisze na klawiaturze, za każdym razem wciskając losowo jedną z 26 liter alfabetu łacińskiego. Jaka jest oczekiwana liczba słów MALPA w ciągu $n = 1\,000\,000$ liter?

$X_i \in \{0, 1\}$ – zmienna określająca, czy na i-tej pozycji pojawiło się słowo MALPA ($i = 1, \ldots, n-4$)

$$P(X_i = 1) \;=\; \left(\frac{1}{26}\right)^5$$

Przykład

Małpa pisze na klawiaturze, za każdym razem wciskając losowo jedną z 26 liter alfabetu łacińskiego. Jaka jest oczekiwana liczba słów MALPA w ciągu $n = 1\ 000\ 000$ liter?

$X_i \in \{0, 1\}$ – zmienna określająca, czy na i-tej pozycji pojawiło się słowo MALPA ($i = 1, \ldots, n-4$)

$$P(X_i = 1) \; = \; \left(\frac{1}{26}\right)^5$$

$Y = X_1 + \ldots + X_{n-4}$ – liczba wystąpień słowa MALPA

Przykład

Małpa pisze na klawiaturze, za każdym razem wciskając losowo jedną z 26 liter alfabetu łacińskiego. Jaka jest oczekiwana liczba słów MALPA w ciągu $n = 1\ 000\ 000$ liter?

$X_i \in \{0, 1\}$ – zmienna określająca, czy na i-tej pozycji pojawiło się słowo MALPA ($i = 1, \ldots, n-4$)

$$P(X_i = 1) \ = \ \left(\frac{1}{26}\right)^5$$

$Y = X_1 + \ldots + X_{n-4}$ – liczba wystąpień słowa MALPA

$$EY \ = \ EX_1 + \ldots + EX_{n-5} \ = \ (n-4)\left(\frac{1}{26}\right)^5 \ \simeq \ 0.08$$

Problem kolekcjonera kuponów

Rzucamy kostką, dopóki nie wypadną co najmniej raz wszystkie możliwe wyniki. Jaka jest oczekiwana liczba rzutów?

Problem kolekcjonera kuponów

Pudełko płatków śniadaniowych zawiera jeden z n różnych kuponów. Ile średnio pudełek musimy kupić, aby skompletować wszystkie kupony?

Problem kolekcjonera kuponów

Pudełko płatków śniadaniowych zawiera jeden z n różnych kuponów. Ile średnio pudełek musimy kupić, aby skompletować wszystkie kupony?

- X_1 – liczba prób do wylosowania pierwszego kuponu ($X_1 = 1$)
- X_2 – liczba dodatkowych (po wylosowaniu pierwszego kuponu) prób do wylosowania kuponu różnego od pierwszego
- ...
- X_i – liczba dodatkowych (po wylosowaniu $i - 1$ różnych kuponów) prób do wylosowania kuponu różnego od $i - 1$ pierwszych kuponów

Problem kolekcjonera kuponów

Pudełko płatków śniadaniowych zawiera jeden z n różnych kuponów. Ile średnio pudełek musimy kupić, aby skompletować wszystkie kupony?

- X_1 – liczba prób do wylosowania pierwszego kuponu ($X_1 = 1$)
- X_2 – liczba dodatkowych (po wylosowaniu pierwszego kuponu) prób do wylosowania kuponu różnego od pierwszego
- ...
- X_i – liczba dodatkowych (po wylosowaniu $i-1$ różnych kuponów) prób do wylosowania kuponu różnego od $i-1$ pierwszych kuponów

$$\underbrace{3}_{X_1}, \underbrace{3, 1}_{X_2}, \underbrace{3, 2}_{X_3}, \underbrace{1, 2, 1, 3, 5}_{X_4}, \underbrace{3, 2, 5, 1, 3, 4}_{X_5}, \ldots$$

Problem kolekcjonera kuponów

Pudełko płatków śniadaniowych zawiera jeden z n różnych kuponów. Ile średnio pudełek musimy kupić, aby skompletować wszystkie kupony?

- X_1 – liczba prób do wylosowania pierwszego kuponu ($X_1 = 1$)
- X_2 – liczba dodatkowych (po wylosowaniu pierwszego kuponu) prób do wylosowania kuponu różnego od pierwszego
- ...
- X_i – liczba dodatkowych (po wylosowaniu $i-1$ różnych kuponów) prób do wylosowania kuponu różnego od $i-1$ pierwszych kuponów

$$\underbrace{3}_{X_1}, \underbrace{3, 1}_{X_2}, \underbrace{3, 2}_{X_3}, \underbrace{1, 2, 1, 3, 5}_{X_4}, \underbrace{3, 2, 5, 1, 3, 4}_{X_5}, \ldots$$

Y – liczba prób do skompletowania wszystkich kuponów

$$Y = X_1 + X_2 + \ldots + X_n$$

Problem kolekcjonera kuponów

Jaki jest rozkład zmiennej X_i?

Problem kolekcjonera kuponów

Jaki jest rozkład zmiennej X_i?

- Skompletowaliśmy dotąd $i-1$ różnych kuponów

Problem kolekcjonera kuponów

Jaki jest rozkład zmiennej X_i?

- Skompletowaliśmy dotąd $i-1$ różnych kuponów
- Szansa wylosowania nowego (nieskompletowanego) kuponu w jednej próbie wynosi więc $p_i = \frac{\#\text{ nowych kuponów}}{\#\text{ wszystkich kuponów}} = \frac{n-(i-1)}{n}$ (nazwijmy takie zdarzenie sukcesem)

Problem kolekcjonera kuponów

Jaki jest rozkład zmiennej X_i?

- Skompletowaliśmy dotąd $i-1$ różnych kuponów
- Szansa wylosowania nowego (nieskompletowanego) kuponu w jednej próbie wynosi więc $p_i = \frac{\text{\# nowych kuponów}}{\text{\# wszystkich kuponów}} = \frac{n-(i-1)}{n}$ (nazwijmy takie zdarzenie sukcesem)
- Losujemy do uzyskania sukcesu: X_i ma rozkład $G_1\left(\frac{n-i+1}{n}\right)$

Problem kolekcjonera kuponów

Jaki jest rozkład zmiennej X_i?

- Skompletowaliśmy dotąd $i - 1$ różnych kuponów
- Szansa wylosowania nowego (nieskompletowanego) kuponu w jednej próbie wynosi więc $p_i = \frac{\#\text{ nowych kuponów}}{\#\text{ wszystkich kuponów}} = \frac{n-(i-1)}{n}$ (nazwijmy takie zdarzenie sukcesem)
- Losujemy do uzyskania sukcesu: X_i ma rozkład $G_1\left(\frac{n-i+1}{n}\right)$

$$EX_i = \frac{1}{p_i} = \frac{n}{n-i+1}$$

Problem kolekcjonera kuponów

Jaki jest rozkład zmiennej X_i?

- Skompletowaliśmy dotąd $i-1$ różnych kuponów
- Szansa wylosowania nowego (nieskompletowanego) kuponu w jednej próbie wynosi więc $p_i = \frac{\#\text{ nowych kuponów}}{\#\text{ wszystkich kuponów}} = \frac{n-(i-1)}{n}$ (nazwijmy takie zdarzenie sukcesem)
- Losujemy do uzyskania sukcesu: X_i ma rozkład $G_1\left(\frac{n-i+1}{n}\right)$

$$EX_i = \frac{1}{p_i} = \frac{n}{n-i+1}$$

$$EY = EX_1 + EX_2 + \ldots EX_n = \frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-2} + \ldots + \frac{n}{2} + \frac{n}{1}$$

Problem kolekcjonera kuponów

Jaki jest rozkład zmiennej X_i?

- Skompletowaliśmy dotąd $i-1$ różnych kuponów
- Szansa wylosowania nowego (nieskompletowanego) kuponu w jednej próbie wynosi więc $p_i = \frac{\#\text{ nowych kuponów}}{\#\text{ wszystkich kuponów}} = \frac{n-(i-1)}{n}$ (nazwijmy takie zdarzenie sukcesem)
- Losujemy do uzyskania sukcesu: X_i ma rozkład $G_1\left(\frac{n-i+1}{n}\right)$

$$EX_i = \frac{1}{p_i} = \frac{n}{n-i+1}$$

$$\begin{aligned} EY = EX_1 + EX_2 + \dots EX_n &= \frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-2} + \dots + \frac{n}{2} + \frac{n}{1} \\ &= n\underbrace{\left(1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n-1} + \frac{1}{n}\right)}_{\text{liczba harmoniczna } H_n} \end{aligned}$$

Problem kolekcjonera kuponów

Jaki jest rozkład zmiennej X_i?

- Skompletowaliśmy dotąd $i-1$ różnych kuponów
- Szansa wylosowania nowego (nieskompletowanego) kuponu w jednej próbie wynosi więc $p_i = \frac{\#\text{ nowych kuponów}}{\#\text{ wszystkich kuponów}} = \frac{n-(i-1)}{n}$ (nazwijmy takie zdarzenie sukcesem)
- Losujemy do uzyskania sukcesu: X_i ma rozkład $G_1\left(\frac{n-i+1}{n}\right)$

$$EX_i = \frac{1}{p_i} = \frac{n}{n-i+1}$$

$$\begin{aligned} EY = EX_1 + EX_2 + \ldots EX_n &= \frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-2} + \ldots + \frac{n}{2} + \frac{n}{1} \\ &= n\underbrace{\left(1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n-1} + \frac{1}{n}\right)}_{\text{liczba harmoniczna } H_n} \end{aligned}$$

Zachodzi: $\ln n \leqslant H_n \leqslant 1 + \ln n$

Problem kolekcjonera kuponów

Jaki jest rozkład zmiennej X_i?

- Skompletowaliśmy dotąd $i-1$ różnych kuponów
- Szansa wylosowania nowego (nieskompletowanego) kuponu w jednej próbie wynosi więc $p_i = \frac{\#\text{ nowych kuponów}}{\#\text{ wszystkich kuponów}} = \frac{n-(i-1)}{n}$ (nazwijmy takie zdarzenie sukcesem)
- Losujemy do uzyskania sukcesu: X_i ma rozkład $G_1\left(\frac{n-i+1}{n}\right)$

$$EX_i = \frac{1}{p_i} = \frac{n}{n-i+1}$$

$$\begin{aligned} EY = EX_1 + EX_2 + \ldots EX_n &= \frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-2} + \ldots + \frac{n}{2} + \frac{n}{1} \\ &= n\Big(\underbrace{1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n-1} + \frac{1}{n}}_{\text{liczba harmoniczna } H_n}\Big) \end{aligned}$$

Zachodzi: $\ln n \leqslant H_n \leqslant 1 + \ln n$

Stąd $EY \simeq n \ln n$

Kowariancja

Definicja

Wyrażenie:

$$C(X, Y) = E((X - EX)(Y - EY))$$

nazywamy kowariancją zmiennych losowych X i Y.

Kowariancja

Definicja

Wyrażenie:

$$C(X, Y) = E((X - EX)(Y - EY))$$

nazywamy kowariancją zmiennych losowych X i Y.

Kowariancja określa jak zmienia się jedna ze zmiennych w stosunku do drugiej.

- Jeśli X i Y są często albo obie powyżej średniej, albo obie poniżej średniej, to kowariancja jest dodatnia
- Jeśli X i Y są często po przeciwnych stronach średniej, to kowariancja jest ujemna

Kowariancja

Definicja

Wyrażenie:

$$C(X, Y) = E((X - EX)(Y - EY))$$

nazywamy kowariancją zmiennych losowych X i Y.

Kowariancja określa jak zmienia się jedna ze zmiennych w stosunku do drugiej.

- Jeśli X i Y są często albo obie powyżej średniej, albo obie poniżej średniej, to kowariancja jest dodatnia
- Jeśli X i Y są często po przeciwnych stronach średniej, to kowariancja jest ujemna

Uwaga:

$$C(X, X) = E((X - EX)^2) = D^2(X)$$
$$C(X, -X) = -D^2(X)$$

Kowariancja

Definicja

Wyrażenie:

$$C(X, Y) = E((X - EX)(Y - EY))$$

nazywamy kowariancją zmiennych losowych X i Y.

Kowariancja określa jak zmienia się jedna ze zmiennych w stosunku do drugiej.

- Jeśli X i Y są często albo obie powyżej średniej, albo obie poniżej średniej, to kowariancja jest dodatnia
- Jeśli X i Y są często po przeciwnych stronach średniej, to kowariancja jest ujemna

Uwaga:

$$\begin{aligned} C(X, X) &= E((X - EX)^2) = D^2(X) \\ C(X, -X) &= -D^2(X) \end{aligned}$$

Zmienne, dla których $C(X, Y) = 0$ nazywamy nieskorelowanymi

Wzór skróconego mnożenia dla kowariancji

Twierdzenie

$$C(X, Y) = E(XY) - (EX)(EY)$$

Wzór skróconego mnożenia dla kowariancji

Twierdzenie

$$C(X,Y) \;=\; E(XY) - (EX)(EY)$$

Dowód:

$$C(X,Y) \;=\; E((X-EX)(Y-EY))$$

Wzór skróconego mnożenia dla kowariancji

Twierdzenie

$$C(X, Y) = E(XY) - (EX)(EY)$$

Dowód:

$$\begin{aligned} C(X, Y) &= E((X - EX)(Y - EY)) \\ &= E\Big(XY - (EX)Y - X(EY) + (EX)(EY)\Big) \end{aligned}$$

Wzór skróconego mnożenia dla kowariancji

Twierdzenie

$$C(X,Y) \;=\; E(XY) - (EX)(EY)$$

Dowód:

$$\begin{aligned} C(X,Y) &= E((X-EX)(Y-EY)) \\ &= E\Big(XY - (EX)Y - X(EY) + (EX)(EY)\Big) \\ &= E(XY) - (EX)(EY) - (EX)(EY) + (EX)(EY) \end{aligned}$$

Wzór skróconego mnożenia dla kowariancji

Twierdzenie

$$C(X,Y) \;=\; E(XY) - (EX)(EY)$$

Dowód:

$$\begin{aligned}
C(X,Y) &= E((X-EX)(Y-EY)) \\
&= E\Big(XY - (EX)Y - X(EY) + (EX)(EY)\Big) \\
&= E(XY) - (EX)(EY) - (EX)(EY) + (EX)(EY) \\
&= E(XY) - (EX)(EY)
\end{aligned}$$

Kowariancja – przykład

	ω_1	ω_2	ω_3	ω_4	ω_5
$P(\{\omega\})$	0.2	0.2	0.2	0.2	0.2
$X(\omega)$	-1	-1	0	1	1
$Y(\omega)$	1	0	0	0	-1

Kowariancja – przykład

	ω_1	ω_2	ω_3	ω_4	ω_5
$P(\{\omega\})$	0.2	0.2	0.2	0.2	0.2
$X(\omega)$	-1	-1	0	1	1
$Y(\omega)$	1	0	0	0	-1
XY	-1	0	0	0	-1

Kowariancja – przykład

	ω_1	ω_2	ω_3	ω_4	ω_5
$P(\{\omega\})$	0.2	0.2	0.2	0.2	0.2
$X(\omega)$	-1	-1	0	1	1
$Y(\omega)$	1	0	0	0	-1
XY	-1	0	0	0	-1

$$EX = 0, \qquad EY = 0, \qquad E(XY) = -0.4$$

Kowariancja – przykład

	ω_1	ω_2	ω_3	ω_4	ω_5
$P(\{\omega\})$	0.2	0.2	0.2	0.2	0.2
$X(\omega)$	-1	-1	0	1	1
$Y(\omega)$	1	0	0	0	-1
XY	-1	0	0	0	-1

$$EX = 0, \qquad EY = 0, \qquad E(XY) = -0.4$$

$$C(X, Y) \; = \; E(XY) - (EX)(EY) \; = \; -0.4$$

Kowariancja – przykład

	ω_1	ω_2	ω_3	ω_4	ω_5
$P(\{\omega\})$	0.2	0.2	0.2	0.2	0.2
$X(\omega)$	0	1	2	3	4
$Y(\omega)$	-3	-2	-1	0	1

Kowariancja – przykład

	ω_1	ω_2	ω_3	ω_4	ω_5
$P(\{\omega\})$	0.2	0.2	0.2	0.2	0.2
$X(\omega)$	0	1	2	3	4
$Y(\omega)$	-3	-2	-1	0	1
XY	0	-2	-2	0	4

Kowariancja – przykład

	ω_1	ω_2	ω_3	ω_4	ω_5
$P(\{\omega\})$	0.2	0.2	0.2	0.2	0.2
$X(\omega)$	0	1	2	3	4
$Y(\omega)$	-3	-2	-1	0	1
XY	0	-2	-2	0	4

$$EX = 2, \qquad EY = -1, \qquad E(XY) = 0$$

Kowariancja – przykład

	ω_1	ω_2	ω_3	ω_4	ω_5
$P(\{\omega\})$	0.2	0.2	0.2	0.2	0.2
$X(\omega)$	0	1	2	3	4
$Y(\omega)$	-3	-2	-1	0	1
XY	0	-2	-2	0	4

$$EX = 2, \qquad EY = -1, \qquad E(XY) = 0$$

$$C(X, Y) = E(XY) - (EX)(EY) = 2$$

Wariancja sumy i różnicy zmiennych losowych

Twierdzenie

$$D^2(X \pm Y) \; = \; D^2(X) \pm 2C(X, Y) + D^2(Y)$$

Wariancja sumy i różnicy zmiennych losowych

Twierdzenie

$$D^2(X \pm Y) \; = \; D^2(X) \pm 2C(X, Y) + D^2(Y)$$

Dowód: Zdefiniujmy zmienne:

$$X' = X - EX, \qquad \text{oraz} \quad Y' = Y - EY$$

Wariancja sumy i różnicy zmiennych losowych

Twierdzenie

$$D^2(X \pm Y) \;=\; D^2(X) \pm 2C(X, Y) + D^2(Y)$$

Dowód: Zdefiniujmy zmienne:

$$X' = X - EX, \qquad \text{oraz} \quad Y' = Y - EY$$

$$D^2(X + Y) = E\Big((X + Y - E(X + Y))^2\Big)$$

Wariancja sumy i różnicy zmiennych losowych

Twierdzenie

$$D^2(X \pm Y) \;=\; D^2(X) \pm 2C(X,Y) + D^2(Y)$$

Dowód: Zdefiniujmy zmienne:

$$X' = X - EX, \qquad \text{oraz} \quad Y' = Y - EY$$

$$\begin{aligned} D^2(X+Y) &= E\Big(\big(X + Y - E(X+Y)\big)^2\Big) \\ &= E\Big((X' + Y')^2\Big) \end{aligned}$$

Wariancja sumy i różnicy zmiennych losowych

Twierdzenie

$$D^2(X \pm Y) \; = \; D^2(X) \pm 2C(X, Y) + D^2(Y)$$

Dowód: Zdefiniujmy zmienne:

$$X' = X - EX, \qquad \text{oraz} \quad Y' = Y - EY$$

$$\begin{aligned} D^2(X+Y) &= E\Big((X+Y-E(X+Y))^2\Big) \\ &= E\Big((X'+Y')^2\Big) \\ &= E(X'^2) + 2E(X'Y') + E(Y'^2) \end{aligned}$$

Wariancja sumy i różnicy zmiennych losowych

Twierdzenie

$$D^2(X \pm Y) \;=\; D^2(X) \pm 2C(X,Y) + D^2(Y)$$

Dowód: Zdefiniujmy zmienne:

$$X' = X - EX, \qquad \text{oraz} \quad Y' = Y - EY$$

$$\begin{aligned}
D^2(X+Y) &= E\Big((X+Y-E(X+Y))^2\Big) \\
&= E\Big((X'+Y')^2\Big) \\
&= E(X'^2) + 2E(X'Y') + E(Y'^2) \\
&= E((X-EX)^2) + 2E((X-EX)(Y-EY)) + E((Y-EY)^2)
\end{aligned}$$

Wariancja sumy i różnicy zmiennych losowych

Twierdzenie

$$D^2(X \pm Y) \;=\; D^2(X) \pm 2C(X,Y) + D^2(Y)$$

Dowód: Zdefiniujmy zmienne:

$$X' = X - EX, \qquad \text{oraz} \quad Y' = Y - EY$$

$$\begin{aligned}
D^2(X+Y) &= E\Big(\big(X+Y-E(X+Y)\big)^2\Big) \\
&= E\Big((X'+Y')^2\Big) \\
&= E(X'^2) + 2E(X'Y') + E(Y'^2) \\
&= \underbrace{E((X-EX)^2)}_{D^2(X)} + 2\underbrace{E((X-EX)(Y-EY))}_{C(X,Y)} + \underbrace{E((Y-EY)^2)}_{D^2(Y)}
\end{aligned}$$

Wariancja sumy i różnicy zmiennych losowych

Twierdzenie

$$D^2(X \pm Y) \;=\; D^2(X) \pm 2C(X,Y) + D^2(Y)$$

Dowód: Zdefiniujmy zmienne:

$$X' = X - EX, \qquad \text{oraz} \quad Y' = Y - EY$$

$$\begin{aligned}
D^2(X+Y) &= E\Big((X+Y-E(X+Y))^2\Big) \\
&= E\Big((X'+Y')^2\Big) \\
&= E(X'^2) + 2E(X'Y') + E(Y'^2) \\
&= \underbrace{E((X-EX)^2)}_{D^2(X)} + 2\underbrace{E((X-EX)(Y-EY))}_{C(X,Y)} + \underbrace{E((Y-EY)^2)}_{D^2(Y)}
\end{aligned}$$

Dowód dla różnicy zmiennych losowych analogiczny.

Nierówność Cauchy'ego-Schwarza

Twierdzenie

$$\left(E(XY)\right)^2 \leqslant E(X^2)E(Y^2)$$

Równość zachodzi wtedy i tylko wtedy gdy X i Y są swoimi wielokrotnościami, tzn. $Y = cX$ lub $X = cY$ dla $c \in \mathbb{R}$

Nierówność Cauchy'ego-Schwarza

Twierdzenie

$$\left(E(XY)\right)^2 \leqslant E(X^2)E(Y^2)$$

Równość zachodzi wtedy i tylko wtedy gdy X i Y są swoimi wielokrotnościami, tzn. $Y = cX$ lub $X = cY$ dla $c \in \mathbb{R}$

Zadanie 2

Udowodnij tę nierówność

Nierówność Cauchy'ego-Schwarza

Twierdzenie

$$\big(E(XY)\big)^2 \leqslant E(X^2)E(Y^2)$$

Równość zachodzi wtedy i tylko wtedy gdy X i Y są swoimi wielokrotnościami, tzn. $Y = cX$ lub $X = cY$ dla $c \in \mathbb{R}$

Zadanie 2

Udowodnij tę nierówność

Zadanie 3

Pokaż, że ta nierówność implikuje następującą nierówność:

$$\big|C(X, Y)\big| \leqslant D(X)D(Y),$$

przy czym równość zachodzi wtedy i tylko wtedy, gdy jedna ze zmiennych jest funkcją liniową drugiej, np. $Y = aX + b$

Współczynnik korelacji

Wyrażenie:

$$\rho(X, Y) = \frac{C(X, Y)}{D(X)D(Y)}$$

nazywamy współczynnikiem korelacji zmiennych losowych X i Y.

Współczynnik korelacji

Wyrażenie:

$$\rho(X, Y) = \frac{C(X, Y)}{D(X)D(Y)}$$

nazywamy **współczynnikiem korelacji** zmiennych losowych X i Y.

Ponieważ $\left|C(X, Y)\right| \leqslant D(X)D(Y)$, mamy:

$$\rho(X, Y) \in [-1, 1],$$

przy czym $\rho(X, Y) \in \{-1, 1\}$ wtedy i tylko wtedy gdy jedna zmienna jest **funkcją liniową** drugiej.

Współczynnik korelacji

Wyrażenie:

$$\rho(X, Y) = \frac{C(X, Y)}{D(X)D(Y)}$$

nazywamy współczynnikiem korelacji zmiennych losowych X i Y.

Ponieważ $|C(X, Y)| \leqslant D(X)D(Y)$, mamy:

$$\rho(X, Y) \in [-1, 1],$$

przy czym $\rho(X, Y) \in \{-1, 1\}$ wtedy i tylko wtedy gdy jedna zmienna jest funkcją liniową drugiej.

Współczynnik korelacji jest więc znormalizowaną kowariancją mierzącą siłę zależności liniowej między zmiennymi losowymi.

Niezależne zmienne losowe

Definicja

Zmienne losowe X i Y nazywamy niezależnymi jeśli:

$$P(X \in A, Y \in B) \; = \; P(X \in A)P(Y \in B)$$

Interpretacja: X i Y są niezależne, jeśli dowolne zdarzenia związane z różnymi zmiennymi są niezależne

Niezależne zmienne losowe

Definicja

Zmienne losowe X i Y nazywamy niezależnymi jeśli:

$$P(X \in A, Y \in B) \;=\; P(X \in A)P(Y \in B)$$

Interpretacja: X i Y są niezależne, jeśli dowolne zdarzenia związane z różnymi zmiennymi są niezależne

Uwaga: Jeśli $P(Y \in B) > 0$ to

$$\begin{aligned} P(X \in A | Y \in B) \;&=\; \frac{P(X \in A, Y \in B)}{P(Y \in B)} \;=\; \frac{P(X \in A)P(Y \in B)}{P(Y \in B)} \\ &=\; P(X \in A) \end{aligned}$$

Podobnie, jeśli $P(X \in A) > 0$ to $P(Y \in B | X \in A) = P(Y \in B)$

Niezależne zmienne losowe

Definicja

Zmienne losowe X i Y nazywamy niezależnymi jeśli:

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

Interpretacja: X i Y są niezależne, jeśli dowolne zdarzenia związane z różnymi zmiennymi są niezależne

Uwaga: Jeśli $P(Y \in B) > 0$ to

$$\begin{aligned} P(X \in A | Y \in B) &= \frac{P(X \in A, Y \in B)}{P(Y \in B)} = \frac{P(X \in A)P(Y \in B)}{P(Y \in B)} \\ &= P(X \in A) \end{aligned}$$

Podobnie, jeśli $P(X \in A) > 0$ to $P(Y \in B | X \in A) = P(Y \in B)$

Uwaga: Dla zmiennych niezależnych rozkład łączny da się odtworzyć z rozkładów brzegowych!

Niezależne zmienne losowe

Twierdzenie

Dyskretne zmienne X i Y są niezależne wtedy i tylko wtedy gdy:

$$P(X = x, Y = y) \; = \; P(X = x)P(Y = y) \qquad \text{dla dowolnych} \; x, y$$

Niezależne zmienne losowe

Twierdzenie

Dyskretne zmienne X i Y są niezależne wtedy i tylko wtedy gdy:

$$P(X = x, Y = y) \;=\; P(X = x)P(Y = y) \qquad \text{dla dowolnych } \; x, y$$

Dowód:

$\Longrightarrow$ Ponieważ X i Y są niezależne, korzystamy z def. niezależności wybierając $A = \{x\}$ i $B = \{y\}$

Niezależne zmienne losowe

Twierdzenie

Dyskretne zmienne X i Y są niezależne wtedy i tylko wtedy gdy:

$$P(X = x, Y = y) = P(X = x)P(Y = y) \qquad \text{dla dowolnych } x, y$$

Dowód:

$\Longrightarrow$ Ponieważ X i Y są niezależne, korzystamy z def. niezależności wybierając $A = \{x\}$ i $B = \{y\}$

$\Longleftarrow$ Dla dowolnych A, B mamy:

$$P(X \in A, Y \in B) = \sum_{x \in A} \sum_{y \in B} P(X = x, Y = y)$$

Niezależne zmienne losowe

Twierdzenie

Dyskretne zmienne X i Y są niezależne wtedy i tylko wtedy gdy:

$$P(X = x, Y = y) = P(X = x)P(Y = y) \qquad \text{dla dowolnych } x, y$$

Dowód:

$\Longrightarrow$ Ponieważ X i Y są niezależne, korzystamy z def. niezależności wybierając $A = \{x\}$ i $B = \{y\}$

$\Longleftarrow$ Dla dowolnych A, B mamy:

$$\begin{aligned} P(X \in A, Y \in B) &= \sum_{x \in A} \sum_{y \in B} P(X = x, Y = y) \\ &= \sum_{x \in A} \sum_{y \in B} P(X = x)P(Y = y) \end{aligned}$$

Niezależne zmienne losowe

Twierdzenie

Dyskretne zmienne X i Y są niezależne wtedy i tylko wtedy gdy:

$$P(X = x, Y = y) = P(X = x)P(Y = y) \qquad \text{dla dowolnych } x, y$$

Dowód:

$\Longrightarrow$ Ponieważ X i Y są niezależne, korzystamy z def. niezależności wybierając $A = \{x\}$ i $B = \{y\}$

$\Longleftarrow$ Dla dowolnych A, B mamy:

$$\begin{aligned}
P(X \in A, Y \in B) &= \sum_{x \in A} \sum_{y \in B} P(X = x, Y = y) \\
&= \sum_{x \in A} \sum_{y \in B} P(X = x)P(Y = y) \\
&= \Big(\sum_{x \in A} P(X = x)\Big)\Big(\sum_{y \in B} P(Y = y)\Big) \\
&= P(X \in A)P(Y \in B)
\end{aligned}$$

Przykłady

- X – wynik na pierwszej kostce, Y – wynik na drugiej kostce
  Dla dowolnych $i, j \in \{1, 2, 3, 4, 5, 6\}$:

$$P(X = i, Y = j) \;=\; \frac{1}{36} \;=\; \frac{1}{6} \cdot \frac{1}{6} \;=\; P(X = i)P(Y = j)$$

Przykłady

- X – wynik na pierwszej kostce, Y – wynik na drugiej kostce
  Dla dowolnych $i, j \in \{1, 2, 3, 4, 5, 6\}$:
  $$P(X = i, Y = j) \;=\; \frac{1}{36} \;=\; \frac{1}{6} \cdot \frac{1}{6} \;=\; P(X = i)P(Y = j)$$
- Losujemy kartę z talii. X – wartość karty (liczona od 1 do 13), Y – kolor karty (numerowany od 1 do 4)
  Dla dowolnych $i \in \{1, \ldots, 13\}$, $j \in \{1, 2, 3, 4\}$:
  $$P(X = i, Y = j) \;=\; \frac{1}{52} \;=\; \frac{1}{13} \cdot \frac{1}{4} \;=\; P(X = i)P(Y = j)$$

Niezależne zmienne losowe

Definicja

Zmienne losowe $X_1, \ldots, X_n$ nazywamy niezależnymi jeśli:

$$P(X_1 \in A_1, \ldots, X_n \in A_n) \;=\; P(X_1 \in A_1) \cdot \ldots \cdot P(X_n \in A_n)$$

Dyskretne zmienne losowe są niezależne wtedy i tylko wtedy gdy dla dowolnego $\boldsymbol{x} = (x_1, \ldots, x_n)$:

$$P(X_1 = x_1, \ldots, X_n = x_n) \;=\; P(X_1 = x_1) \cdot \ldots \cdot P(X_n = x_n)$$

Przykład

Rozważmy schemat n prób Bernoulliego z prawd. sukcesu p

Przykład

Rozważmy schemat n prób Bernoulliego z prawd. sukcesu p

$X_i \in \{0, 1\}$ – wynik losowania w i-tej próbie $(i = 1, \ldots, n)$

Przykład

Rozważmy schemat n prób Bernoulliego z prawd. sukcesu p

$X_i \in \{0, 1\}$ – wynik losowania w i-tej próbie ($i = 1, \ldots, n$)

Zmienne $X_1, \ldots, X_n$ są niezależne, np:

$$P(X_1 = 1, X_2 = 1, X_3 = 0) \;=\; p^2(1-p) \;=\; P(X_1 = 1)P(X_2 = 1)P(X_3 = 0)$$

Niezależne zmienne losowe

Twierdzenie

Jeśli X i Y są niezależne, to również niezależne są $f(X)$ i $g(Y)$ dla dowolnych funkcji f i g.

Niezależne zmienne losowe

Twierdzenie

Jeśli X i Y są niezależne, to również niezależne są $f(X)$ i $g(Y)$ dla dowolnych funkcji f i g.

Dowód:

$$P(f(X) \in A, g(Y) \in B) = P(X \in f^{-1}(A), Y \in g^{-1}(B))$$

Niezależne zmienne losowe

Twierdzenie

Jeśli X i Y są niezależne, to również niezależne są $f(X)$ i $g(Y)$ dla dowolnych funkcji f i g.

Dowód:

$$\begin{aligned} P(f(X) \in A, g(Y) \in B) &= P(X \in f^{-1}(A), Y \in g^{-1}(B)) \\ &\overset{(*)}{=} P(X \in f^{-1}(A)) \cdot P(Y \in g^{-1}(B)) \end{aligned}$$

gdzie w $(*)$ wykorzystaliśmy niezależność zmiennych X i Y.

Niezależne zmienne losowe

Twierdzenie

Jeśli X i Y są niezależne, to również niezależne są $f(X)$ i $g(Y)$ dla dowolnych funkcji f i g.

Dowód:

$$\begin{aligned} P(f(X) \in A, g(Y) \in B) &= P(X \in f^{-1}(A), Y \in g^{-1}(B)) \\ &\stackrel{(*)}{=} P(X \in f^{-1}(A)) \cdot P(Y \in g^{-1}(B)) \\ &= P(f(X) \in A) \cdot P(g(Y) \in B), \end{aligned}$$

gdzie w $(*)$ wykorzystaliśmy niezależność zmiennych X i Y.

Niezależne zmienne losowe

Twierdzenie

Jeśli $X_1, \ldots, X_n$ są niezależne, to zmienne:

$$\begin{aligned} Y_1 &= f_1(X_1, \ldots, X_{i_1}) \\ Y_2 &= f_2(X_{i_1+1}, \ldots, X_{i_2}) \\ &\ldots \\ Y_m &= f_m(X_{i_{m-1}+1}, \ldots, X_m) \end{aligned}$$

są niezależne.

Czyli dowolne funkcje rozłącznych podzbiorów zmiennych niezależnych są niezależne

Niezależne zmienne losowe

Twierdzenie

Jeśli $X_1, \ldots, X_n$ są niezależne, to zmienne:

$$\begin{aligned} Y_1 &= f_1(X_1, \ldots, X_{i_1}) \\ Y_2 &= f_2(X_{i_1+1}, \ldots, X_{i_2}) \\ &\ldots \\ Y_m &= f_m(X_{i_{m-1}+1}, \ldots, X_m) \end{aligned}$$

są niezależne.

Czyli dowolne funkcje rozłącznych podzbiorów zmiennych niezależnych są niezależne

Przykład: Jeśli X_1, X_2, X_3, X_4, X_5 są niezależne, to niezależne są również:

$$Y_1 = f(X_1, X_2, X_3), \qquad Y_2 = g(X_4, X_5)$$

Wartość oczekiwana iloczynu zmiennych niezależnych

Jeśli zmienne losowe X i Y są niezależne to zachodzi:

$$E(XY) = (EX)(EY)$$

Wartość oczekiwana iloczynu zmiennych niezależnych

Jeśli zmienne losowe X i Y są niezależne to zachodzi:

$$E(XY) \ = \ (EX)(EY)$$

Dowód: Weźmy $Z = f(X, Y) = XY$.

Z niezależności: $P(X = x, Y = y) = P(X = x)P(Y = y)$

Wartość oczekiwana iloczynu zmiennych niezależnych

Jeśli zmienne losowe X i Y są niezależne to zachodzi:

$$E(XY) \ = \ (EX)(EY)$$

Dowód: Weźmy $Z = f(X, Y) = XY$.

Z niezależności: $P(X = x, Y = y) = P(X = x)P(Y = y)$

$$E(XY) \ = \ EZ \ = \ \sum_x \sum_y f(x, y)\, P(X = x, Y = y)$$

Wartość oczekiwana iloczynu zmiennych niezależnych

Jeśli zmienne losowe X i Y są niezależne to zachodzi:

$$E(XY) \;=\; (EX)(EY)$$

Dowód: Weźmy $Z = f(X, Y) = XY$.

Z niezależności: $P(X = x, Y = y) = P(X = x)P(Y = y)$

$$\begin{aligned} E(XY) &= EZ = \sum_x \sum_y f(x, y)\, P(X = x, Y = y) \\ &= \sum_x \sum_y xy P(X = x)P(Y = y) \end{aligned}$$

Wartość oczekiwana iloczynu zmiennych niezależnych

Jeśli zmienne losowe X i Y są niezależne to zachodzi:

$$E(XY) = (EX)(EY)$$

Dowód: Weźmy $Z = f(X, Y) = XY$.

Z niezależności: $P(X = x, Y = y) = P(X = x)P(Y = y)$

$$\begin{aligned} E(XY) &= EZ = \sum_x \sum_y f(x, y)\, P(X = x, Y = y) \\ &= \sum_x \sum_y xy P(X = x)P(Y = y) \\ &= \Big(\sum_x x P(X = x) \Big) \Big(\sum_y y P(Y = y) \Big) \end{aligned}$$

Wartość oczekiwana iloczynu zmiennych niezależnych

Jeśli zmienne losowe X i Y są niezależne to zachodzi:

$$E(XY) = (EX)(EY)$$

Dowód: Weźmy $Z = f(X, Y) = XY$.

Z niezależności: $P(X = x, Y = y) = P(X = x)P(Y = y)$

$$\begin{aligned}
E(XY) &= EZ = \sum_x \sum_y f(x, y)\, P(X = x, Y = y) \\
&= \sum_x \sum_y xy P(X = x)P(Y = y) \\
&= \Big(\sum_x x P(X = x) \Big) \Big(\sum_y y P(Y = y) \Big) \\
&= (EX)(EY)
\end{aligned}$$

Wartość oczekiwana iloczynu zmiennych niezależnych

Jeśli zmienne losowe $X_1, X_2, \ldots, X_n$ są niezależne to zachodzi:

$$E(X_1 \cdot X_2 \cdot \ldots \cdot X_n) \;=\; (EX_1) \cdot (EX_2) \cdot \ldots \cdot (EX_n)$$

Dowód przez indukcję po n.

Wartość oczekiwana iloczynu zmiennych niezależnych

Jeśli zmienne losowe $X_1, X_2, \ldots, X_n$ są niezależne to zachodzi:

$$E(X_1 \cdot X_2 \cdot \ldots \cdot X_n) \;=\; (EX_1) \cdot (EX_2) \cdot \ldots \cdot (EX_n)$$

Dowód przez indukcję po n.

Zadanie 4

Udowodnij to twierdzenie

Kowariancja zmiennych niezależnych

Jeśli zmienne losowe X i Y są niezależne to zachodzi:

$$C(X, Y) = 0$$

Kowariancja zmiennych niezależnych

Jeśli zmienne losowe X i Y są niezależne to zachodzi:

$$C(X, Y) = 0$$

Dowód: Ze wzoru skróconego mnożenia na kowariancję:

$$C(X, Y) = E(XY) - (EX)(EY)$$

Kowariancja zmiennych niezależnych

Jeśli zmienne losowe X i Y są niezależne to zachodzi:

$$C(X, Y) = 0$$

Dowód: Ze wzoru skróconego mnożenia na kowariancję:

$$C(X, Y) = \underbrace{E(XY)}_{=(EX)(EY)} - (EX)(EY)$$

Kowariancja zmiennych niezależnych

Jeśli zmienne losowe X i Y są niezależne to zachodzi:

$$C(X, Y) = 0$$

Dowód: Ze wzoru skróconego mnożenia na kowariancję:

$$C(X, Y) = \underbrace{E(XY)}_{=(EX)(EY)} - (EX)(EY) = 0$$

Kowariancja zmiennych niezależnych

Jeśli zmienne losowe X i Y są niezależne to zachodzi:

$$C(X,Y) = 0$$

Dowód: Ze wzoru skróconego mnożenia na kowariancję:

$$C(X,Y) = \underbrace{E(XY)}_{=(EX)(EY)} - (EX)(EY) = 0$$

Uwaga: Z powyższego wynika, że:

$$\text{zmienne niezależne} \implies \text{zmienne nieskorelowane}$$

W ogólności nie zachodzi to w drugą stronę:

$$\text{zmienne nieskorelowane} \;\not\Longrightarrow\; \text{zmienne niezależne}$$

Wariancja sumy i różnicy zmiennych niezależnych

Jeśli zmienne losowe X i Y są niezależne to zachodzi:

$$D^2(X \pm Y) \ = \ D^2(X) + D^2(Y)$$

Wariancja sumy i różnicy zmiennych niezależnych

Jeśli zmienne losowe X i Y są niezależne to zachodzi:

$$D^2(X \pm Y) \; = \; D^2(X) + D^2(Y)$$

Dowód: Udowodniliśmy, że:

$$D^2(X \pm Y) \; = \; D^2(X) \pm 2C(X, Y) + D^2(Y)$$

Wariancja sumy i różnicy zmiennych niezależnych

Jeśli zmienne losowe X i Y są niezależne to zachodzi:

$$D^2(X \pm Y) \; = \; D^2(X) + D^2(Y)$$

Dowód: Udowodniliśmy, że:

$$D^2(X \pm Y) \; = \; D^2(X) \pm 2C(X, Y) + D^2(Y)$$

Ponieważ X i Y są niezależne mamy $C(X, Y) = 0$, co kończy dowód.

Wariancja sumy zmiennych niezależnych

Jeśli zmienne losowe $X_1, X_2, \ldots, X_n$ są niezależne to zachodzi:

$$D^2(X_1 + X_2 + \ldots + X_n) \ = \ D^2(X_1) + D^2(X_2) + \ldots + D^2(X_n)$$

Dowód przez indukcję po n.

Wariancja sumy zmiennych niezależnych

Jeśli zmienne losowe $X_1, X_2, \ldots, X_n$ są niezależne to zachodzi:

$$D^2(X_1 + X_2 + \ldots + X_n) \;=\; D^2(X_1) + D^2(X_2) + \ldots + D^2(X_n)$$

Dowód przez indukcję po n.

Zadanie 5

Udowodnij to twierdzenie

Zadanie

Zadanie 6

Pokaż, że jeśli X i Y są niezależne, to:

$$D^2(aX + bY) \ = \ a^2 D^2(X) + b^2 D^2(Y)$$

dla dowolnych $a, b \in \mathbb{R}$.

Podsumowanie

- Dla dowolnych zmiennych losowych:

$$
\begin{aligned}
E(X+Y) &= EX + EY \\
D^2(X \pm Y) &= D^2(X) \pm 2C(X,Y) + D^2(Y)
\end{aligned}
$$

- Dla niezależnych zmiennych losowych:

$$
\begin{aligned}
E(XY) &= (EX)(EY) \\
C(X,Y) &= 0 \\
D^2(X \pm Y) &= D^2(X) + D^2(Y)
\end{aligned}
$$

Rozkład sumy zmiennych niezależnych

Załóżmy, że X i Y są niezależne. Jaki rozkład ma $Z = X + Y$?

Rozkład sumy zmiennych niezależnych

Załóżmy, że X i Y są niezależne. Jaki rozkład ma $Z = X + Y$?

$$P(Z = z) = \sum_{x,y\,:\,x+y=z} P(X = x, Y = y)$$

Rozkład sumy zmiennych niezależnych

Załóżmy, że X i Y są niezależne. Jaki rozkład ma $Z = X + Y$?

$$
\begin{aligned}
P(Z = z) &= \sum_{x,y\,:\; x+y=z} P(X = x, Y = y) \\
&= \sum_{x,y\,:\; x+y=z} P(X = x)P(Y = y)
\end{aligned}
$$

Rozkład sumy zmiennych niezależnych

Załóżmy, że X i Y są niezależne. Jaki rozkład ma $Z = X + Y$?

$$\begin{aligned} P(Z = z) &= \sum_{x,y\,:\,x+y=z} P(X = x, Y = y) \\ &= \sum_{x,y\,:\,x+y=z} P(X = x) P(Y = y) \end{aligned}$$

Przykład: X – wynik na 1. kostce, Y – wynik na 2. kostce

$$P(Z = z) = \sum_{i,j\,:\,i+j=z} \underbrace{P(X = i)}_{=1/6} \underbrace{P(Y = j)}_{=1/6}$$

Rozkład sumy zmiennych niezależnych

Załóżmy, że X i Y są niezależne. Jaki rozkład ma $Z = X + Y$?

$$
\begin{aligned}
P(Z = z) &= \sum_{x,y\colon x+y=z} P(X = x, Y = y) \\
&= \sum_{x,y\colon x+y=z} P(X = x)P(Y = y)
\end{aligned}
$$

Przykład: X – wynik na 1. kostce, Y – wynik na 2. kostce

$$
\begin{aligned}
P(Z = z) &= \sum_{i,j\colon i+j=z} \underbrace{P(X = i)}_{=1/6}\underbrace{P(Y = j)}_{=1/6} \\
&= \frac{1}{36}\big|\{i,j\colon i+j=z\}\big|
\end{aligned}
$$

Rozkład sumy zmiennych niezależnych

Załóżmy, że X i Y są niezależne. Jaki rozkład ma $Z = X + Y$?

$$
\begin{aligned}
P(Z = z) &= \sum_{x,y\colon x+y=z} P(X = x, Y = y) \\
&= \sum_{x,y\colon x+y=z} P(X = x)P(Y = y)
\end{aligned}
$$

Przykład: X – wynik na 1. kostce, Y – wynik na 2. kostce

$$
\begin{aligned}
P(Z = z) &= \sum_{i,j\colon i+j=z} \underbrace{P(X = i)}_{=1/6}\underbrace{P(Y = j)}_{=1/6} \\
&= \frac{1}{36}\left|\{i,j\colon i + j = z\}\right|
\end{aligned}
$$

$$
P(Z = 2) = \frac{1}{36}, \quad P(Z = 3) = \frac{2}{36}, \quad P(Z = 4) = \frac{3}{36}, \ldots
$$

Zadanie

Zadanie 7

Pokaż, że jeśli X i Y są niezależne i X ma rozkład $\mathrm{Pois}(\lambda_1)$, a Y ma rozkład $\mathrm{Pois}(\lambda_2)$, to $Z = X + Y$ ma rozkład $\mathrm{Pois}(\lambda_1 + \lambda_2)$.

Rozkład dwumianowy

Rozważmy schemat n prób Bernoulliego z prawd. sukcesu p

Rozkład dwumianowy

Rozważmy schemat n prób Bernoulliego z prawd. sukcesu p

- $X_i \in \{0, 1\}$ – wynik losowania w i-tej próbie ($i = 1, \ldots, n$)

Rozkład dwumianowy

Rozważmy schemat n prób Bernoulliego z prawd. sukcesu p

- $X_i \in \{0, 1\}$ – wynik losowania w i-tej próbie ($i = 1, \ldots, n$)
- Zmienne X_i mają rozkład dwupunktowy $B(p)$ i są niezależne

Rozkład dwumianowy

Rozważmy schemat n prób Bernoulliego z prawd. sukcesu p

- $X_i \in \{0, 1\}$ – wynik losowania w i-tej próbie ($i = 1, \ldots, n$)
- Zmienne X_i mają rozkład dwupunktowy $B(p)$ i są niezależne
- $Y = X_1 + \ldots + X_n$ – zmienna określająca liczbę sukcesów w n próbach

Rozkład dwumianowy

Rozważmy schemat n prób Bernoulliego z prawd. sukcesu p

- $X_i \in \{0, 1\}$ – wynik losowania w i-tej próbie ($i = 1, \ldots, n$)
- Zmienne X_i mają rozkład dwupunktowy $B(p)$ i są niezależne
- $Y = X_1 + \ldots + X_n$ – zmienna określająca liczbę sukcesów w n próbach
- Y ma rozkład dwumianowy $B(n, p)$

Rozkład dwumianowy

Rozważmy schemat n prób Bernoulliego z prawd. sukcesu p

- $X_i \in \{0, 1\}$ – wynik losowania w i-tej próbie ($i = 1, \ldots, n$)
- Zmienne X_i mają rozkład dwupunktowy $B(p)$ i są niezależne
- $Y = X_1 + \ldots + X_n$ – zmienna określająca liczbę sukcesów w n próbach
- Y ma rozkład dwumianowy $B(n, p)$

$$EX_i = p, \qquad D^2(X_i) = p(1-p)$$

Rozkład dwumianowy

Rozważmy schemat n prób Bernoulliego z prawd. sukcesu p

- $X_i \in \{0, 1\}$ – wynik losowania w i-tej próbie ($i = 1, \ldots, n$)
- Zmienne X_i mają rozkład dwupunktowy $B(p)$ i są niezależne
- $Y = X_1 + \ldots + X_n$ – zmienna określająca liczbę sukcesów w n próbach
- Y ma rozkład dwumianowy $B(n, p)$

$$EX_i = p, \qquad D^2(X_i) = p(1-p)$$

$$\begin{aligned} EY &= EX_1 + EX_2 + \ldots + EX_n = np \\ D^2(Y) &= D^2(X_1) + D^2(X_2) + \ldots + D^2(X_n) = np(1-p) \end{aligned}$$

Otrzymaliśmy bezboleśnie wzory na wartość oczekiwaną i wariancję zmiennej o rozkładzie dwumianowym

Przykład

Rzucamy n razy kostką. Oblicz wariancję sumy oczek.

Przykład

Rzucamy n razy kostką. Oblicz wariancję sumy oczek.

X_i – wynik rzutu na i-tej kostce $(i = 1, \ldots, n)$

$X_1, \ldots, X_n$ są niezależne

$Y = X_1 + \ldots + X_n$ – sumaryczny wynik n rzutów

Przykład

Rzucamy n razy kostką. Oblicz wariancję sumy oczek.

X_i – wynik rzutu na i-tej kostce ($i = 1, \ldots, n$)

$X_1, \ldots, X_n$ są niezależne

$Y = X_1 + \ldots + X_n$ – sumaryczny wynik n rzutów

$$D^2(Y) \;=\; D^2(X_1) + \ldots + D^2(X_n) \;=\; nD^2(X_1) \;=\; n \cdot \frac{35}{12}$$