Wstęp do rachunku prawdopodobieństwa
Wykład 5: Funkcje zmiennej losowej. Zmienne losowe wielowymiarowe. Wartość oczekiwana, wariancja i kowariancja.

Funkcje zmiennej losowej

Na początek wykładu przypomnijmy, że zmienną losową nazywamy dowolną funkcję (mierzalną) $X : \Omega \to \mathbb{R}$ określoną na przestrzeni probabilistycznej $(\Omega, \mathcal{F}, \mathbb{P})$.

Przykład 1. Rozważmy grę polegającą na 10 rzutach kostką. Załóżmy, że za każdym razem, gdy wyrzucimy 6 oczek otrzymujemy 10zł, natomiast w każdym z pozostałych przypadków musimy zapłacić złotówkę. Jeśli X jest zmienną losową równą liczbie wyrzuconych szóstek w 10 rzutach kostką, to w jaki sposób możemy wyrazić „wygraną" w tej grze?

Nietrudno zauważyć, że „wygrana" równa jest

$$10 \cdot X + (-1) \cdot (10 - X) = 11 \cdot X - 10,$$

ponieważ dokładnie X razy otrzymamy 10zł, a w pozostałych przypadkach, czyli $10 - X$ razy, będziemy musieli zapłacić złotówkę. Następnie zauważmy, że wyrażenie $11 \cdot X - 10$ to nic innego jak pewna funkcja zmiennej losowej X, a dokładnie jeśli rozważymy funkcję $f : \mathbb{R} \to \mathbb{R}$ daną wzorem $f(x) = 11x - 10$, wówczas wygrana w grze wynosi $f(X)$.

Uwaga 1. Jeśli $X : \Omega \to \mathbb{R}$ jest zmienną losową, a $f : \mathbb{R} \to \mathbb{R}$ jest funkcją (mierzalną), to funkcja $f(X) : \Omega \to \mathbb{R}$ będąca złożeniem funkcji X i f **jest również zmienną losową**.

Okazuje się, że wyznaczanie wartości oczekiwanej funkcji zmiennej losowej jest stosunkowo proste i nie różni się istotnie od wyznaczania wartości oczekiwanej tej zmiennej losowej, o czym mówi kolejne twierdzenie.

Twierdzenie 1 (Prawo leniwego statystyka)**.** *Niech X będzie zmienną losową dyskretną o wartościach (atomach) w zbiorze A i $f : \mathbb{R} \to \mathbb{R}$ będzie funkcją (mierzalną), wtedy:*

$$\mathbb{E}(f(X)) = \sum_{a \in A} f(a) \cdot \mathbb{P}\left(X = a\right).$$

Przykład 2. Zauważmy, że zmienna losowa X z Przykładu 1. ma rozkład dwumianowy $Bin(10, \frac{1}{6})$. Stosując prawo leniwego statystyka, wartość oczekiwana zmiennej losowej $f(X) = 11X - 10$ możemy zatem obliczyć następująco:

$$\mathbb{E}(f(X)) = \sum_{k=0}^{10} f(k) \cdot \mathbb{P}\left(X = k\right) = \sum_{k=0}^{10} (11k - 10) \binom{10}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{10-k} = \frac{50}{6}.$$

Jedną z najczęściej rozważanych funkcji zmiennej losowej jest tzw. wariancja. Wariancja jest miarą rozproszenia zmiennej losowej. Intuicyjnie rzecz ujmując, jeśli zmienna losowa X jest dobrze skupiona wokół swojej wartości oczekiwanej $\mathbb{E}X$, czyli przyjmuje z dużym prawdopodobieństwem wartości bliskie $\mathbb{E}X$, to taka zmienna losowa będzie miała stosunkowo małą wariancję. Natomiast jeśli zmienna losowa jest rozproszona, czyli z dużym prawdopdobieństwem przyjmuje wartości dalekie od $\mathbb{E}X$, to ta zmienna losowa będzie miała stosunkowo dużą wariancję. Przy czym należy podkreślić, że w tym przypadku pojęcie bliski/daleki jest względne.

Definicja 1. Wariancją *zmiennej losowej X nazywamy liczbę*

$$VarX = \mathbb{E}((X - \mathbb{E}X)^2).$$

Uwaga 2. Przy liczeniu wariancji praktyczniejsze zastosowanie ma wzór:

$$\mathrm{Var}X = \mathbb{E}(X^2) - (\mathbb{E}X)^2.$$

Wzór ten udowodnimy później, gdy poznamy własności wartości oczekiwanej.

Proszę zwrócić uwagę, że zmienna losowa $(X - \mathbb{E}X)^2$ pojawiająca się w definicji wariancji przyjmuje jedynie wartości nieujemne, a zatem, jak nietrudno się przekonać, jej wartość oczekiwana też musi być nieujemna. Płynie stąd następujący wniosek.

Wniosek 2. Wariancja $\mathrm{Var}X$ zmiennej losowej X jest **zawsze nieujemna**. Co więcej, $\mathrm{Var}X = 0$ wtedy i tylko wtedy, gdy $X = \mathbb{E}X$, to znaczy wtedy i tylko wtedy, gdy zmienna losowa X ma **rozkład jednopunktowy** (z prawdopodobieństwem równym 1 przyjmuje jedną ustaloną wartość równą $\mathbb{E}X$).

Przykład 3. Niech X będzie zmienną losową, która zwraca 1 jeśli w pojedynczym rzucie kostką wypadnie liczba parzysta oraz -1, jeśli wypadnie liczba nieparzysta.

Przy liczeniu wariancji zazwyczaj zaczynamy od wyznaczenia wartości oczekiwanej rozpatrywanej zmiennej losowej. A zatem:

$$\mathbb{E}X = 1 \cdot \left(\frac{1}{6} + \frac{1}{6} + \frac{1}{6}\right) + (-1) \cdot \left(\frac{1}{6} + \frac{1}{6} + \frac{1}{6}\right) = 1 \cdot \frac{1}{2} + (-1) \cdot \frac{1}{2} = 0.$$

Następnie wyznaczamy wartość oczekiwaną zmiennej losowej X^2, czyli:

$$\mathbb{E}(X^2) = 1 \cdot \frac{1}{2} + (-1)^2 \cdot \frac{1}{2} = 1.$$

Teraz wystarczy podstawić powyższe wartości do wzoru i otrzymujemy:

$$\mathrm{Var}X = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = 1 - 0^2 = 1.$$

Uwaga 3. Istnieją zmienne losowe X, dla których $\mathbb{E}X$ nie jest określona, ponieważ definiujący ją szereg nie jest zbieżny. W takim przypadku $\mathrm{Var}X$ również jest niezdefiniowana. Istnieją również zmienne losowe X, dla których istnieje $\mathbb{E}X$, a nie istnieje $\mathrm{Var}X$. Dzieje się tak wtedy, gdy $\mathbb{E}(X^2)$ nie jest określona.

Zmienne losowe wielowymiarowe i ich parametry

Definicja 2. *Funkcję (mierzalną)*

$$(X, Y) : \Omega \to \mathbb{R}^2 : \omega \to (X(\omega), Y(\omega))$$

określoną na przestrzeni probabilistycznej $(\Omega, \mathcal{F}, \mathbb{P})$ *nazywamy* **dwuwymiarową zmienną losową** *lub* **dwuwymiarowym wektorem losowym**. *Współrzędne tej funkcji, X i Y, są „zwykłymi” zmiennymi losowymi określonymi na tej samej przestrzeni probabilistycznej* $(\Omega, \mathcal{F}, \mathbb{P})$.

We wszystkich definicjach poniżej ograniczymy się do przypadku, gdy zmienne losowe X i Y są zmiennymi dyskretnymi o atomach odpowiednio $\{x_1, x_2, \ldots\}$ i $\{y_1, y_2, \ldots\}$. Wtedy (X, Y) nazywamy **dwuwymiarową dyskretną zmienną losową** o atomach $\{(x_i, y_j) : i, j = 1, 2, \ldots\}$.

Definicja 3. Rozkład (łączny), *lub* **funkcję masy prawdopodobieństwa**, *dyskretnej zmiennej losowej* (X, Y) *definiujemy podając wszystkie wartości*

$$p_{i,j} = \mathbb{P}\left(X = x_i, Y = y_j\right).$$

Wtedy oczywiście:

$$\sum_{i,j} p_{i,j} = \sum_i \sum_j p_{i,j} = 1.$$

Definicja 4. **Rozkłady brzegowe** *zmiennych losowych X i Y dyskretnego wektora losowego (X, Y) obliczamy za pomocą wzorów*

$$\mathbb{P}(X = x_i) = \sum_j \mathbb{P}(X = x_i, Y = y_j)$$

oraz

$$\mathbb{P}(Y = y_j) = \sum_i \mathbb{P}(X = x_i, Y = y_j).$$

Jeśli dla **każdego** *atomu (x_i, y_j) zachodzi*

$$\mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i)\,\mathbb{P}(Y = y_j),$$

to mówimy, że zmienne losowe X i Y są **niezależne**.

Przykład 4. Wektor losowy (X, Y) ma rozkład podany w poniższej tabeli.

$X \setminus Y$	-2	0	2
-1	1/8	1/2	0
1	1/4	0	1/8

Z tabeli tej możemy na przykład odczytać, że $\mathbb{P}(X = -1, Y = 0) = 1/2$. Zmienna losowa X ma zbiór atomów $\{-1, 1\}$, a zmienna losowa Y ma zbiór atomów $\{-2, 0, 2\}$. Aby odczytać rozkłady brzegowe tych zmiennych losowych, wystarczy dla zmiennej losowej X zsumować wartości w odpowiednich wierszach, a dla zmiennej losowej Y wartości w kolumnach. Na przykład mamy:

$$\mathbb{P}(X = -1) = \mathbb{P}(X = -1, Y = -2) + \mathbb{P}(X = -1, Y = 0) + \mathbb{P}(X = -1, Y = 2) = \frac{1}{8} + \frac{1}{2} + 0 = \frac{5}{8}.$$

Otrzymujemy zatem poniższe rozkłady brzegowe:

x	-1	1
$\mathbb{P}(X = x)$	5/8	3/8

y	-2	0	2
$\mathbb{P}(Y = y)$	3/8	1/2	1/8

W szczególności widzimy, że $\mathbb{P}(X = -1) = 5/8 \neq 0$ oraz $\mathbb{P}(Y = 2) = 1/8 \neq 0$, a zatem zmienne losowe X i Y nie są niezależne, ponieważ $\mathbb{P}(X = -1, Y = 2) = 0 \neq \mathbb{P}(X = -1)\,\mathbb{P}(Y = 2)$.

Przykład 5. Niech zmienne losowe X i Y zliczają odpowiednio maksymalny i minimalny wynik w trzykrotnym rzucie standardową kostką. Chcemy wyznaczyć rozkład prawdopodobieństwa wektora losowego (X, Y).

Po pierwsze zauważmy, że zarówno zmienna losowa X, jak i zmienna losowa Y, mogą przyjmować tylko wartości w zbiorze $\{1, 2, \ldots, 6\}$. Następnie możemy zauważyć, że dla każdego zdarzenia elementarnego ω zachodzi $X(\omega) \geqslant Y(\omega)$, ponieważ maksimum nie może być mniejsze od minimum. Możemy to zapisać krócej w postaci $X \geqslant Y$. Ponadto dla $k \in \{1, 2, \ldots, 6\}$, zdarzenie $\{X = Y = k\}$ zachodzi wtedy i tylko wtedy, gdy w każdym rzucie uzyskamy ten sam wynik k, co oczywiście dzieje się z prawdopodobieństwem $(1/6)^3$. Z kolei dla $6 \geqslant k > \ell \geqslant 1$, zdarzenie $\{X = k, Y = \ell\}$ zachodzi wtedy i tylko wtedy, gdy w jednym z trzech rzutów uzyskamy wynik k, w innym wynik ℓ, a w pozostałym wynik pomiędzy ℓ i k. Otrzymujemy zatem następujący rozkład prawdopodobieństwa:

$X \setminus Y$	1	2	3	4	5	6
1	$(1/6)^3$	0	0	0	0	0
2	$6 \cdot (1/6)^3$	$(1/6)^3$	0	0	0	0
3	$12 \cdot (1/6)^3$	$6 \cdot (1/6)^3$	$(1/6)^3$	0	0	0
4	$18 \cdot (1/6)^3$	$12 \cdot (1/6)^3$	$6 \cdot (1/6)^3$	$(1/6)^3$	0	0
5	$24 \cdot (1/6)^3$	$18 \cdot (1/6)^3$	$12 \cdot (1/6)^3$	$6 \cdot (1/6)^3$	$(1/6)^3$	0
5	$30 \cdot (1/6)^3$	$24 \cdot (1/6)^3$	$18 \cdot (1/6)^3$	$12 \cdot (1/6)^3$	$6 \cdot (1/6)^3$	$(1/6)^3$

Mając daną funkcję (mierzalną) $f : \mathbb{R}^2 \to \mathbb{R}$ oraz wektor losowy (X, Y), gdzie X i Y są zmiennymi losowymi określonymi na pewnej przestrzeni probabilistycznej $(\Omega, \mathcal{F}, \mathbb{P})$, możemy rozważać nową zmienną losową $f(X, Y)$ określoną na tej samej przestrzeni probabilistycznej.

Twierdzenie 3 (Prawo leniwego statystyka)**.** *Niech (X, Y) będzie zmienną losową dyskretną o wartościach (atomach) $\{(x_i, y_j) : i, j = 1, 2, \dots\}$, a $f : \mathbb{R}^2 \to \mathbb{R}$ będzie funkcją (mierzalną), wtedy:*

$$\mathbb{E}(f(X, Y)) = \sum_{i,j} f(x_i, y_j)\mathbb{P}\left(X = x_i, Y = y_j\right).$$

Z powyższego twierdzenie wynika na przykład:

$$\mathbb{E}(XY) = \sum_{i,j} x_i y_j \mathbb{P}\left(X = x_i, Y = y_j\right).$$

Jedną z ważniejszych własności wartości oczekiwanej, z której będziemy często korzystać, jest jej liniowość.

Twierdzenie 4 (Liniowość wartości oczekiwanej)**.** *Niech X, Y będą zmiennymi losowymi określonymi na tej samej przestrzeni probabilistycznej $(\Omega, \mathcal{F}, \mathbb{P})$. Wówczas dla dowolnych liczb rzeczywistych $a, b, c \in \mathbb{R}$ i funkcji (mierzalnych) $f, g : \mathbb{R}^2 \to \mathbb{R}$ mamy:*

$$\mathbb{E}(a \cdot f(X, Y) + b \cdot g(X, Y) + c) = a \cdot \mathbb{E}f(X, Y) + b \cdot \mathbb{E}g(X, Y) + c.$$

W szczególności zachodzi:

$$\mathbb{E}(a \cdot X + b \cdot Y + c) = a \cdot \mathbb{E}X + b \cdot \mathbb{E}Y + c.$$

Dowód. W przypadku dyskretnych zmiennych losowych liniowość wynika bezpośrednio z własności sumowania szeregów. Ponadto stałą c możemy traktować jako zmienną losową o rozkładzie jednopunktowym, czyli zmienną losową stale równą c. Mamy wówczas:

$$\begin{aligned}
\mathbb{E}(a \cdot X) &= \sum_i a x_i \mathbb{P}\left(X = x_i\right) = a \sum_i x_i \mathbb{P}\left(X = x_i\right) = a \cdot \mathbb{E}X, \\
\mathbb{E}(X + Y) &= \sum_i \sum_j (x_i + y_j)\mathbb{P}\left(X = x_i, Y = y_j\right) \\
&= \sum_i \sum_j x_i \mathbb{P}\left(X = x_i, Y = y_j\right) + \sum_j \sum_i y_j \mathbb{P}\left(X = x_i, Y = y_j\right) \\
&= \sum_i x_i \sum_j \mathbb{P}\left(X = x_i, Y = y_j\right) + \sum_j y_j \sum_i \mathbb{P}\left(X = x_i, Y = y_j\right) \\
&= \sum_i x_i \mathbb{P}\left(X = x_i\right) + \sum_j y_j \mathbb{P}\left(Y = y_j\right) = \mathbb{E}X + \mathbb{E}Y.
\end{aligned}$$

Oczywiście nietrudno zauważyć, że powyższy wzór można uogólnić na dowolną sumę zmiennych losowych, a pamiętając że zarówno $f(X, Y)$ jak i $g(X, Y)$ są zmiennymi losowymi, otrzymamy interesujące nas równania. □

Przykład 6. Korzystając z liniowości wartości oczekiwanej, pamiętając że $\mathbb{E}X$ jest po prostu stałą, możemy udowodnić Uwagę 2. następująco:

$$\mathrm{Var}X = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}(X^2 - 2\mathbb{E}X \cdot X + (\mathbb{E}X)^2) = \mathbb{E}(X^2) - 2\mathbb{E}X \cdot \mathbb{E}X + (\mathbb{E}X)^2 = \mathbb{E}(X^2) - (\mathbb{E}X)^2.$$

Możemy teraz zdefiniować kolejny ważny parametr dotyczący wektorów losowych, mianowicie kowariancję. Kowariancja określa pewnego rodzaju współzależność pomiędzy dwiema zmiennymi losowymi. Jeżeli pomiędzy dwoma zmiennymi losowymi zachodzi zauważalna zależność liniowa, to kowariancja będzie przyjmowała niezerową wartość – dodatnią gdy wraz ze wzrostem wartości jednej zmiennej losowej rosną wartości drugiej, ujemną w przeciwnym przypadku.

Definicja 5. Kowariancją *zmiennej losowej* (X,Y) *nazywamy liczbę*

$$Cov(X,Y)=\mathbb{E}(X-\mathbb{E}X)(Y-\mathbb{E}Y)=\mathbb{E}(XY)-\mathbb{E}X\mathbb{E}Y,$$

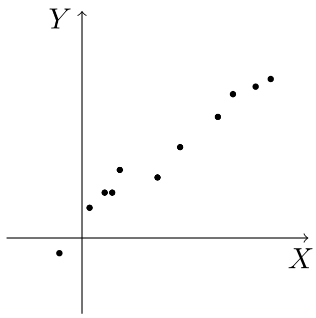
o ile wartości oczekiwane $\mathbb{E}X$, $\mathbb{E}Y$ *oraz* $\mathbb{E}(XY)$ *istnieją.* **Współczynnik korelacji** $\rho(X,Y)$ *zdefiniowany jest wzorem*

$$\rho(X,Y)=\frac{Cov(X,Y)}{\sqrt{VarX}\sqrt{VarY}},$$

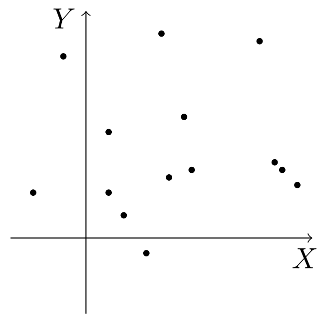
pod warunkiem, że wariancje zmiennych losowych X *i* Y *istnieją i są niezerowe.*

Uwaga 4. Istnieją zmienne losowe X i Y, dla których $\mathrm{Cov}(X,Y)$ i $\rho(X,Y)$ nie są określone (np. gdy nie jest określona wartość oczekiwana lub wariancja którejś z nich). Jeśli jednak $\rho(X,Y)$ istnieje, to $|\rho(X,Y)|\leqslant 1$. Co więcej, $|\rho(X,Y)|=1$ tylko w przypadku liniowej zależności między zmiennymi losowymi, to znaczy gdy istnieją stałe $a,b\in\mathbb{R}$ takie, że $Y=aX+b$.

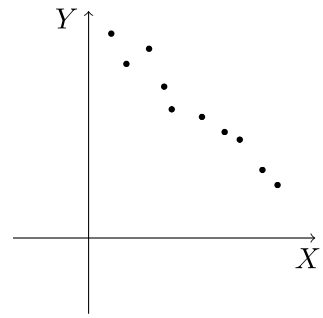
Na poniższych wykresach zaznaczono te punkty (x,y), dla których $\mathbb{P}\,(X=x,Y=y)\neq 0$.



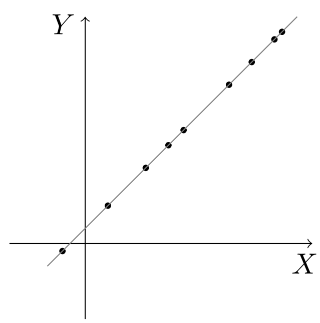
dodatnia korelacja $\rho>0$



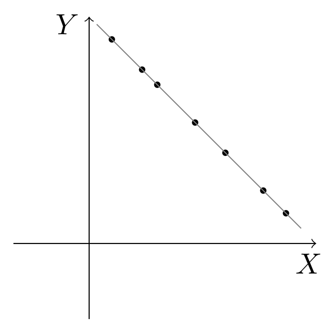
brak korelacji $\rho=0$



ujemna korelacja $\rho<0$



$\rho=1$



$\rho=-1$

Twierdzenie 5. *Jeśli zmienne losowe* X *i* Y *są niezależne, to* $\mathbb{E}(XY)=\mathbb{E}X\mathbb{E}Y$*, a zatem:*

$$Cov(X,Y)=\rho(X,Y)=0.$$

Dowód. W przypadku dyskretnych zmiennych losowych powyższa równość wynika bezpośrednio z własności sumowania szeregów.

$$\begin{aligned}\mathbb{E}(XY)&=\sum_{i,j}x_iy_j\mathbb{P}\,(X=x_i,Y=y_j)=\sum_{i,j}x_iy_j\mathbb{P}\,(X=x_i)\,\mathbb{P}\,(Y=y_j)\\&=\sum_i x_i\mathbb{P}\,(X=x_i)\sum_j y_j\mathbb{P}\,(Y=y_j)=\mathbb{E}Y\sum_i x_i\mathbb{P}\,(X=x_i)=\mathbb{E}X\mathbb{E}Y.\end{aligned}$$

□

Uwaga 5. Twierdzenie odwrotne do powyższego nie jest prawdziwe, to znaczy istnieją zmienne losowe X,Y, dla których $\mathrm{Cov}(X,Y)=0$, ale zmienne te nie są niezależne.

Przykład 7. Spójrzmy jeszcze raz na wektor losowy z Przykładu 3.

$X \setminus Y$	-2	0	2
-1	1/8	1/2	0
1	1/4	0	1/8

x	-1	1
$\mathbb{P}(X=x)$	5/8	3/8

y	-2	0	2
$\mathbb{P}(Y=y)$	3/8	1/2	1/8

Aby wyznaczyć $\mathrm{Cov}(X,Y)$ musimy policzyć najpierw wartości oczekiwane:

$$\mathbb{E}X = -1 \cdot \frac{5}{8} + 1 \cdot \frac{3}{8} = -\frac{2}{8} = -\frac{1}{4},$$
$$\mathbb{E}Y = -2 \cdot \frac{3}{8} + 2 \cdot \frac{1}{8} = -\frac{4}{8} = -\frac{1}{2},$$
$$\mathbb{E}(XY) = (-1)\cdot(-2)\cdot\frac{1}{8} + 1\cdot(-2)\cdot\frac{1}{4} + 1\cdot 2\cdot\frac{1}{8} = 0.$$

Po podstawieniu do wzoru otrzymujemy:

$$\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y = 0 - \left(-\frac{1}{4}\right)\left(-\frac{1}{2}\right) = -\frac{1}{8}.$$

Ponieważ $\mathrm{Cov}(X,Y) \neq 0$, zmienne losowe X i Y na pewno nie są niezależne. Następnie, aby wyznaczyć współczynnik korelacji, potrzebne nam są wariancje obu zmiennych losowych. W tym celu liczymy najpierw:

$$\mathbb{E}X^2 = (-1)^2 \cdot \frac{5}{8} + 1 \cdot \frac{3}{8} = 1,$$
$$\mathbb{E}Y^2 = (-2)^2 \cdot \frac{3}{8} + 2^2 \cdot \frac{1}{8} = 2,$$

skąd otrzymujemy:

$$\mathrm{Var}X = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = 1 - \left(-\frac{1}{4}\right)^2 = \frac{15}{16}\,,$$
$$\mathrm{Var}Y = \mathbb{E}(Y^2) - (\mathbb{E}Y)^2 = 2 - \left(-\frac{1}{2}\right)^2 = \frac{7}{4}\,.$$

Możemy teraz wyznaczyć współczynnik korelacji:

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{\mathrm{Var}X}\sqrt{\mathrm{Var}Y}} = \frac{-\frac{1}{8}}{\sqrt{\frac{15}{16}}\sqrt{\frac{7}{4}}} = -\frac{1}{\sqrt{105}} = -\frac{\sqrt{105}}{105}\,.$$