Wstęp do rachunku prawdopodobieństwa
Wykład 6: Funkcje zmiennej losowej. Zmienne losowe wielowymiarowe. Wartość oczekiwana, wariancja i kowariancja.

Definicja. *Wartością oczekiwaną zmiennej losowej dyskretnej X o wartościach (atomach) $\{x_1, x_2, \ldots\}$ nazywamy liczbę*

$$\mathbb{E}X = \sum_i x_i \mathbb{P}\left(X = x_i\right).$$

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

Twierdzenie (Prawo leniwego statystyka)**.** *Niech X będzie zmienną losową dyskretną o wartościach (atomach) $\{x_1, x_2, \ldots\}$ i $g : \mathbb{R} \to \mathbb{R}$ będzie funkcją (mierzalną), wtedy*

$$\mathbb{E}(g(X)) = \sum_i g(x_i) \mathbb{P}\left(X = x_i\right).$$

Definicja. *Wariancją zmiennej losowej X nazywamy liczbę*

$$VarX = \mathbb{E}((X - \mathbb{E}X)^2) \quad (\text{przydatny wzór: } VarX = \mathbb{E}(X^2) - (\mathbb{E}X)^2).$$

Uwaga: Istnieją zmienne losowe X, dla których $\mathbb{E}X$ nie jest określona (bo definiujący ją szereg nie jest zbieżny)! W takim przypadku Var$X$ również jest niezdefiniowana. Są również zmienna losowe X, dla których istnieje $\mathbb{E}X$, a nie istnieje Var$X$!

Funkcję (mierzalną)

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

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

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 zmienną dyskretną o atomach $\{(x_i, y_\ell) : i, \ell = 1, 2, \ldots\}$.

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

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

Wtedy $\sum_{i,\ell} p_{i,\ell} = 1$.

Definicja. Rozkłady brzegowe *zmiennych X i Y obliczamy za pomoca wzorów*

$$\mathbb{P}\left(X = x_i\right) = \sum_\ell \mathbb{P}\left(X = x_i, Y = y_\ell\right)$$

i

$$\mathbb{P}\left(Y = y_\ell\right) = \sum_i \mathbb{P}\left(X = x_i, Y = y_\ell\right).$$

Jeśli dla **wszystkich** *wartości (x_i, y_ℓ) zachodzi*

$$\mathbb{P}\left(X = x_i, Y = y_\ell\right) = \mathbb{P}\left(X = x_i\right) \mathbb{P}\left(Y = y_\ell\right),$$

to mówimy, że zmienne losowe są **niezależne**.

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

$$\mathbb{E}(g(X, Y)) = \sum_{i,\ell} g(x_i, y_\ell)\mathbb{P}\left(X = x_i, Y = y_\ell\right).$$

Zatem, np.

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

Twierdzenie. *Dla dowolnych liczb rzeczywistych* $a, b, c \in \mathbb{R}$ *i funkcji* $f, g : \mathbb{R}^2 \to \mathbb{R}$ *mamy*

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

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

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

a **współczynnik korelacji** $\rho(X, Y)$ *zdefiniowany jest wzorem*

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

Uwaga: Istnieją zmienne losowe (X, Y), dla których $\mathrm{Cov}(X, Y)$ i $\rho(X, Y)$ nie są określone (bo definiujący je szereg nie jest zbieżny)! Jesli jednak $\rho(X, Y)$ istnieje, to $|\rho(X, Y)| \leqslant 1$.

Twierdzenie. *Jeśli zmienne losowe* X *i* Y *są niezależne, to* $\mathbb{E}(XY) = \mathbb{E}X\mathbb{E}Y$, *a zatem* $Cov(X, Y) = \rho(X, Y) = 0$.

Uwaga: Twierdzenie odwrotne do powyższego nie jest prawdziwe.

1. Zadania na ćwiczenia

Zadanie 1. Zmienna losowa X posiada rozkład podany w poniższej tabeli:

k	0	1	2	3
$\mathbb{P}(X = k)$	$0,4$	$0,3$	$0,2$	$0,1$

a) Oblicz wartość oczekiwaną oraz wariancję zmiennej losowej X.
b) Znajdź rozkład zmiennej losowej $Y = |2 - X|$ i używając go, znajdź jej wartość oczekiwaną.
c) Oblicz $\mathbb{E}|2 - X|$, korzystając z "prawa leniwego statystyka".

Zadanie 2. Zmienna losowa X posiada rozkład podany w poniższej tabeli:

i	-2	-1	0	1	2	7
$\mathbb{P}(X = i)$	$0,18$	$0,16$	$0,28$	$0,23$	$0,14$	$0,01$

Znajdź rozkład zmiennej losowej $Y = \text{sgn}(X)$ i oblicz $\mathbb{E}Y$ i $\text{Var}Y$.

Zadanie 3. Zmienna losowa (X, Y) ma następujący rozkład:

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

Wyznacz rozkłady brzegowe obu zmiennych i znajdź $\rho(X, Y)$. Czy zmienne te są niezależne?

Zadanie 4. Rzucamy kostką n razy. Czy liczba wszystkich wyrzuconych jedynek X_n i liczba wszystkich wyrzuconych dwójek Y_n są zmiennymi niezależnymi?

Zadanie 5. W pierwszej urnie znajduje się 5 kul białych i 3 czarne, a w drugiej 2 białe i 2 czarne. Z pierwszej urny losujemy jedną kulę, a z drugiej również jedną kulę. Niech X będzie zmienną losową oznaczającą liczbę kul białych wśród dwóch wylosowanych kul, a Y – liczbą wylosowanych kul czarnych.

(i) Znajdź rozkład łączny (X, Y).
(ii) Czy zmienne losowe X i Y są niezależne?
(iii) Oblicz $\mathbb{E}X$ i $\mathbb{E}Y$, korzystając z rozkładów brzegowych X i Y.
(iv) Oblicz $\mathbb{E}X$, przedstawiając X w postaci sumy $X = X_1 + X_2$, gdzie X_i oznacza liczbę białych kul wyciągniętych z i-tej urny.
(v) Co można powiedzieć o rozkładzie zmiennej $Z = X + Y$?
(vi) Znajdź współczynnik korelacji $\rho(X, Y)$. Czy można to zrobić prościej, korzystając z (v) i nie używając rozkładu łącznego (X, Y)?

2. Zadania domowe

Zadanie 1. Zmienna losowa ma rozkład
$$\mathbb{P}(X=8)=\frac{1}{8} \quad \text{i} \quad \mathbb{P}(X=i)=\frac{1}{2^i}, \text{ dla } i=1,2,3.$$
Wyznacz rozkład zmiennych losowych $Y=(-2)^X$ oraz $Z=4X+1$. Oblicz wartości oczekiwane zmiennych losowych $Y=(-2)^X$ oraz $Z=4X+1$ na dwa sposoby: korzystając z rozkładu i korzystając z „prawa leniwego statystyka”.

Zadanie 2. Zmienna losowa X posiada rozkład podany w poniższej tabeli

k	-1	0	1	3
$\mathbb{P}(X=k)$	$\frac{2}{10}$	$\frac{5}{10}$	$\frac{3}{20}$	$\frac{3}{20}$

.

(a) Znajdź rozkład zmiennej losowej $Y=|2-X|$, jej wartość oczekiwaną i wariancję.
(b) Znajdź rozkład zmiennej losowej $Z=\sin(\pi X/2)$, jej wartość oczekiwaną i wariancję.

Zadanie 3. Na kole ruletki znajduje się 37 numerów: 18 czerwonych, 18 czarnych i 1 zielony. Każdy numer wypada z tym samym prawdopodobieństwem. Oblicz wartość oczekiwaną i wariancję wygranej (w żetonach) w pojedynczej grze, jeżeli grający:
(a) stawia jeden żeton na czerwone (jeśli wypadnie czerwony, wtedy gracz dostaje dodatkowo 1 żeton, jeśli nie, traci postawiony żeton);
(b) stawia jeden żeton na numer 17 (jeśli wypadnie 17, wtedy gracz dostaje dodatkowo 35 żetonów, jeśli nie, traci postawiony żeton).

Zadanie 4. Rzucamy raz symetryczną kostką. Pierwszy gracz wygrywa 2, gdy liczba oczek jest parzysta, a przegrywa 1 (tj. wygrywa -1) gdy liczba oczek jest nieparzysta. Drugi gracz wygrywa 2, gdy liczba oczek wynosi co najmniej 4, a przegrywa 1 w przeciwnym przypadku. Niech X_i oznacza wygraną i-tego gracza.
(a) Znajdź rozkład zmiennej losowej (X_1, X_2).
(b) Znajdź rozkłady brzegowe zmiennej (X_1, X_2) i sprawdź, czy X_1 i X_2 są niezależne.
(c) Oblicz współczynnik korelacji $\rho(X_1, X_2)$.
(d) Znajdź $\mathbb{E}X_1^2X_2^3$.

Zadanie 5. Z urny, w której znajdują się 2 białe i 3 czarne kule wyciągamy trzy kule. Pierwszy gracz wygrywa 6 jeśli wszystkie wyciągnięte kule są czarne i przegrywa 1 w każdym innym przypadku. Drugi gracz za każdą wyciągniętą czarną kulę wygrywa 1, a białe kule nie zwiększają, ani nie zmniejszają jego wygranej. Niech X będzie wygraną pierwszego, a Y wygraną drugiego gracza.
(a) Znajdź rozkład zmiennej losowej (X, Y).
(b) Znajdź rozkłady brzegowe zmiennej (X, Y) i sprawdź, czy X i Y są niezależne.
(c) Oblicz współczynnik korelacji $\rho(X, Y)$.