Lab 5_1 – Niezależność, warunkowa niezależność i funkcje zmiennych losowych.

Niech $(\Omega, \mathcal{F}, \Pr)$ przestrzeń probabilistyczna.

Def. Odwzorowanie $X: \Omega \to \mathbb{R}^n$ nazywamy zmienną losową, jeżeli

$$\forall A \in \mathcal{B}(\mathbb{R}^n) \ X^{-1}(A) \in \mathcal{F}$$

gdzie $\mathcal{B}(\mathbb{R}^n)$ jest rodziną zbiorów borelowskich.

Def. Rozkładem prawdopodobieństwa zmiennej losowej $X: \Omega \to \mathbb{R}^n$ nazywamy miarę μ_X na $\mathbb{R}^n$ taką, że

$$\mu_X(B) = \Pr\left(X^{-1}(B)\right), \qquad \forall B \in \mathcal{B}(\mathbb{R}^n)$$

Def. Jeżeli istnieje funkcja $f_X: \mathbb{R}^n \to \mathbb{R}$ taka, że

$$\mu_X(B) = \int_B f_X(x)\,dx, \qquad \forall B \in \mathcal{B}(\mathbb{R}^n)$$

to f_X nazywamy *gęstością zmiennej losowej* X. Zmienną losową posiadająca gęstość nazywamy *ciągłą*.

Def. Zmienne losowe $X_1, \ldots, X_k: \Omega \to \mathbb{R}$ nazywamy *niezależnymi*, jeżeli dla każdego ciągu zbiorów borelowskich $B_1, \ldots, B_k \in \mathcal{B}(\mathbb{R})$ mamy

$$\Pr(X_1 \in B_1, \ldots, X_k \in B_k) = \Pr(X_1 \in B_1), \ldots, \Pr(X_k \in B_k).$$

Tw. Niech $X_1, \ldots, X_k: \Omega \to \mathbb{R}$ zmienne losowe. Następujące warunki są równoważne:

1. Zmienne losowe są niezależne.
2. $\mu_{X_1,\ldots,X_k} = \mu_{X_1} \otimes, \ldots, \otimes \mu_{X_k}$.
3. $\forall (t_1, \ldots, t_k) \in \mathbb{R}^k$ mamy $F_{X_1,\ldots,X_k}(t_1, \ldots, t_k) = F_{X_1}(t_1), \ldots, F_{X_k}(t_k)$.

Tw. Zmienne losowe o wartościach dyskretnych, rzeczywistych $\{X_i: \Omega \to S_i \subset \mathbb{R}\}_{i=1,..,k}$ są niezależne wtedy i tylko wtedy, gdy $\forall (x_1, \ldots, x_k) \in S_1 \times, \ldots, \times S_k$ mamy

$$\Pr(X_1 = x_1, \ldots, X_k = x_k) = \Pr(X_1 = x_1), \ldots, \Pr(X_k = x_k).$$

Zmienne losowe zależne, rozkłady warunkowe.

Niech $X, Y: \Omega \to \mathbb{R}$ dwie zmienne losowe.

Niech zmienne X, Y będą dyskretne, to $\Pr\{X = x_i | Y = y_k\} = \frac{\Pr\{X=x_i, Y=y_k\}}{\Pr\{Y=y_k\}}$, $F(x|y_k) = \Pr\{X < x | Y = y_k\} = \sum_{x_i<x} \Pr\{X = x_i | Y = y_k\}$ są odpowiednio rozkładem i dystrybuantą warunkową.

Jeżeli X, Y będą zmiennymi o rozkładzie ciągłym to dystrybuantą warunkową zmiennej X pod warunkiem $Y \in B$ dla pewnego $B \in \mathcal{B}(\mathbb{R})$ nazywamy funkcję $F(x|y)$ spełniającą warunek

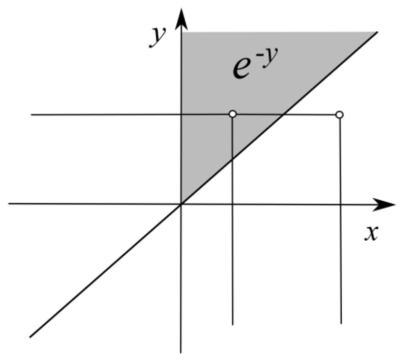
$$\forall B \in \mathcal{B}(\mathbb{R}) \ \Pr(X \leq x | Y \in B) = \int_B F(x|y) d\Pr_Y(y)$$

Jeżeli X, Y będą zmiennymi o rozkładzie ciągłym, łącznej funkcji gęstości $f(x, y)$ funkcją gęstości łącznej, $f_X(x), f_Y(y)$ gęstościami zmiennych X, Y, to

$$f(x|y) = \frac{f(x, y)}{f_Y(y)}, \qquad f(y|x) = \frac{f(x, y)}{f_X(x)}$$

będą gęstościami rozkładów warunkowych zmiennej X pod warunkiem $Y = y$, oraz zmiennej Y pod warunkiem $X = x$ odpowiednio.

Zadanie 1. Losujemy jedna kartę z talii 52 kart. Oznaczamy przez X_1 zmienną losową przyjmującą 0 dla karty nietreflowej a wartość 1 dla treflowej; przez X_2 przyjmującą wartości 5 dla asa, 4 dla króla, 3 dla damy, 0 dla pozostałych kart. Wyznacz rozkład prawdopodobieństwa zmiennej losowej $X = X_1 + X_2$.



Zadanie 2. (oparte na zadaniu 14 z Lab 5) Dla rozkładu zmiennej dwuwymiarowej danego gęstością

$$f(x, y) = \begin{cases} e^{-y} & \text{dla} \quad 0 \leq x < \infty, \ x \leq y \\ 0 & \text{dla} \quad \text{dla pozostałych } (x, y) \end{cases}$$

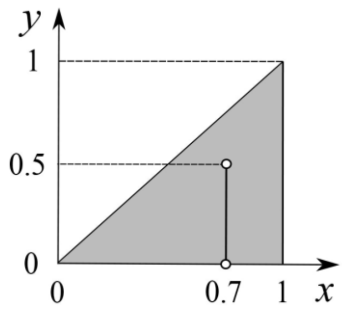
Znaleźć gęstości rozkładów warunkowych.

Zadanie 14, Lab5. Funkcja gęstości zmiennej losowej $(X, Y): \Omega \to \mathbb{R}^2$ dana jest wzorem

$$f(x,y)=\begin{cases} e^{-y} & \text{dla} \quad 0 \le x, \ \ x \le y \\ 0 & \text{dla} \quad \text{dla pozostałych } (x,y) \end{cases}$$

1) Znaleźć dystrybuantę $F(x,y)$ tego rozkładu.

2) Znaleźć gęstości i dystrybuanty brzegowe tego rozkładu.



Zadanie 3. Niech $(X,Y)\colon \Omega \to \Delta = \{(x,y) \in \mathbb{R}^2, 0 \le x, y \le 1, y \le x\}$. Rozważmy funkcje

$$\Delta \ni (\mathrm{x},\mathrm{y}) \to F(x,y) = -y^2 + 2xy$$

a) Sprawdzić, czy funkcja F może być dystrybuantą pewnej zmiennej losowej.
b) Jeżeli tak, to obliczyć funkcję gęstości $f(x,y)$ tej zmiennej.
c) Obliczyć gęstość brzegową $f_X(x)$ i gęstość warunkową $f(y|x)$ tego rozkładu.
d) Obliczyć $\Pr\{y \in [0, 0.5] | x = 0.7\}$.

Niezależność zdarzeń i zmiennych losowych

Zadanie 4. Funkcja gęstości zmiennej losowej $(X,Y)\colon \Omega \to \mathbb{R}^2$ dana jest wzorem

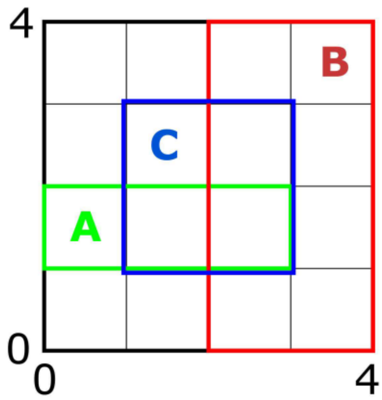
$$f(x,y)=\begin{cases} 24x^2y(1-x) & \text{dla} \quad 0 \le x \le 1, \ \ 0 \le y < 1 \\ 0 & \text{dla} \quad \text{dla pozostałych } (x,y) \end{cases}$$

Zbadać, czy zmienne X, Y są niezależne.

Warunkowa niezależność zdarzeń

Def. Zdarzenia A i B są warunkowo niezalezne wzgledem zdarzenia C w.t.w. gdy

$$\Pr(A \cap B|C) = \Pr(A|C)\Pr(B|C).$$



Zadanie 5. Rozważmy zbiór zdarzeń elementarnych $\Omega = [0,4] \times [0,4]$ oraz równomierną miarę probabilistyczną $\Pr$ na Ω. Rozważmy zdarzenia:

$A = [0,3] \times [1,2],\ B = [2,4] \times [0,4], C = [1,2] \times [1,2].$

1) Zbadaj, czy zdarzenia A, B są niezależne w $(\Omega, \mathcal{B}(\Omega), \Pr)$?
2) Zbadaj, czy zdarzenia A, B są warunkowo niezależne względem zdarzenia C?

Warunkowa niezależność zmiennych losowych

Def. Zmienne losowe $X_1, \dots, X_k, Y: \Omega \to \mathbb{R}$ nazywamy *warunkowo niezależne względem zmiennej* Y, jeżeli dla każdego ciągu zbiorów borelowskich $B_1, \dots, B_k \in \mathcal{B}(\mathbb{R})$ mamy

$$\Pr(X_1 \in B_1, \dots, X_k \in B_k | Y = y) = \Pr(X_1 \in B_1 | Y), \dots, \Pr(X_k \in B_k | Y).$$

Tw. Niech $X_1, \dots, X_k, Y: \Omega \to \mathbb{R}$ zmienne losowe. Następujące warunki są równoważne:

4. Zmienne losowe $X_1, \dots, X_k$ są warunkowo niezależne względem zmiennej Y.
5. $\forall (x_1, \dots, x_k, y) \in \mathbb{R}^{k+1}$ mamy $F_{X_1,\dots,X_k}(t_1, \dots, t_k | y) = F_{X_1}(x_1 | y), \dots, F_{X_k}(x_k | y)$.
6. Dla rozkładów dyskretnych $\forall (x_1, \dots, x_k, y) \in \mathbb{R}^{k+1}$ mamy $p(x_1, \dots, x_k | y) = p(x_1 | y), \dots, p(x_k | y)$.
7. Dla rozkładów ciągłych $\forall (x_1, \dots, x_k, y) \in \mathbb{R}^{k+1}$ mamy $f_{X_1,\dots,X_k}(t_1, \dots, t_k | y) = f_{X_1}(x_1 | y), \dots, f(x_k | y)$.

Tw. Niech $X, Y, Z: \Omega \to \mathbb{R}$ zmienne losowe, X, Y niezależne względem Z, to $\forall A \in \mathcal{B}(\mathbb{R})$, $\forall y, z \in \mathbb{R}$ mamy

$$\Pr(X \in A | Y = y, Z = z) = \Pr(X \in A | Z = z).$$

Funkcja zmiennej losowej i działania na zmiennych losowych

Niech $(\Omega, \mathcal{F}, \Pr)$ przestrzeń probabilistyczna.

Def. Odwzorowanie $X: \Omega \to \mathbb{R}^n$ nazywamy zmienną losową, jeżeli

$$\forall A \in \mathcal{B}(\mathbb{R}^n) \ X^{-1}(A) \in \mathcal{F}$$

gdzie $\mathcal{B}(\mathbb{R}^n)$ jest rodziną zbiorów borelowskich.

Def. Rozkładem prawdopodobieństwa zmiennej losowej $X: \Omega \to \mathbb{R}^n$ nazywamy miarę μ_X na $\mathbb{R}^n$ taką, że

$$\mu_X(B) = \Pr\left(X^{-1}(B)\right), \qquad \forall B \in \mathcal{B}(\mathbb{R}^n)$$

Def. Jeżeli istnieje funkcja $f_X: \mathbb{R}^n \to \mathbb{R}$ taka, że

$$\mu_X(B) = \int_B f_X(x)\,dx, \qquad \forall B \in \mathcal{B}(\mathbb{R}^n)$$

to f_X nazywamy *gęstością zmiennej losowej* X. Zmienną losową posiadająca gęstość nazywamy *ciągłą*.

Def. $g: \mathbb{R}^n \to \mathbb{R}^m$ nazywamy funkcją borelowską, jeżeli

$$\forall A \in \mathcal{B}(\mathbb{R}^m) \ g^{-1}(A) \in \mathcal{B}(\mathbb{R}^n)$$

Tw. Jeżeli $X: \Omega \to \mathbb{R}^n$ jest zmienną losową, to $g \circ X: \Omega \to \mathbb{R}^m$ jest także zmienną losową.

Główny Problem

Niech zmienna losowa Y będzie funkcją pewnej zmiennej X, tzn $Y(\omega) = g(X(\omega))$. Znając rozkład zmiennej X chcemy wyznaczyć rozkład zmiennej Y (?)

Przypadek trywialny wektora losowego

Niech $\Omega = \{e_1, \dots, e_k\}$, $X: \Omega \to \mathbb{R}^n$, wtedy wartości zmiennej X możemy kojarzyć z wektorem (tablicą) $x = (x_1, \dots, x_k)$; $x_i = X(e_i)$, $i = 1, \dots, k$. Rozkład prawdopodobieństwa zmiennej X, to wektor $p_X = (\Pr(e_1), \dots, \Pr(e_k)) \in \mathbb{R}^n$.

Niech $g: \mathbb{R}^n \to \mathbb{R}^m$ funkcja (niekoniecznie borelowska). Chcemy znaleźć rozkład prawdopodobieństwa $Y = g(X)$. Oznaczmy przez $y = (y_1, \dots, y_r) = g(x)$

Łatwo zauważyć, że $\operatorname{card}\left(Y(\Omega)\right) = \operatorname{card}\left(g\left(X(\Omega)\right)\right) = r \le k$, a równość zachodzi zawsze w przypadku, kiedy g jest injekcją. Będziemy mieli zatem

$$p_Y = (p_1, \dots, p_r); \qquad p_i = \sum_{j; g(x_j) = y_i} \Pr(e_j), \qquad i = 1, \dots, r$$

Rozwiązanie przez dystrybuanty:

$F_Y(y) = \Pr(Y < y) = \Pr(g(X) < y) = \Pr(X < g^{-1}(y)) = F_X(g^{-1}(y))$

Jeżeli dystrybuanta jest różniczkowalna, to $D(F_X(g^{-1}(y))) = f_Y(y)$.

Zadanie 6:

Znaleźć rozkład zmiennej losowej $Y: \Omega \to \mathbb{R}$ (dystrybuantę i gęstość), gdy $Y = aX + b$, gdzie $a \neq 0$, X jest zmienną losową typu ciągłego z gęstością f_X i dystrybuantą F_X.

Odwzorowania gładkie zmiennej jednowymiarowej

Niech $(\Omega, \mathcal{F}, \Pr)$ przestrzeń probabilistyczna, $X: \Omega \to \mathbb{R}$ zmienna losowa jednowymiarowa, oraz $g: \mathbb{R} \to \mathbb{R}$ pewna funkcja gładka w całej dziedzinie. Zakładając, że X ma rozkład ciągły oraz funkcję gęstości $f: \mathbb{R} \to \mathbb{R}$ będziemy starali się wyznaczyć rozkład zmiennej losowej jednowymiarowej $Y: \Omega \to \mathbb{R}$; $Y(\omega) = g\big(X(\omega)\big)\ \forall\ \omega \in \Omega$.

Tw. Jeżeli zmienna losowa $X: \Omega \to \mathbb{R}$ ma rozkład ciągły o gęstości $f_X: \mathbb{R} \to \mathbb{R}$ i $X(\Omega) \subset$ (a, b) tak, że $g|(a,b) \in C^1(a,b)$ oraz $g'|(a,b) \neq 0$, to zmienna losowa $Y = g(X)$ ma rozkład ciągły o gęstości

$$f_Y(y) = f_X\big(g^{-1}(y)\big)\, |D(g^{-1})(y)|\ \chi_{g(a,b)}(y)$$

gdzie $\chi_{g(a,b)}$ jest funkcją charakterystyczną obrazu przedziału (a,b) przez funkcję g.

Zadanie 7: Niech zmienna losowa $X: \Omega \to \mathbb{R}$ ma standaryzowany rozkład normalny (Gaussa) $N(0{,}1)$, którego gęstość dana jest wzorem $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$.

Należy wyznaczyć dystrybuantę i gęstość (jeżeli istnieje) dla zmiennych losowych $Y = e^X$ oraz $Z = X^2$.

Zadanie 8: Niech ponownie $X: \Omega \to \mathbb{R}$ ma standaryzowany rozkład normalny (Gaussa) $N(0{,}1)$. Znaleźć rozkład zmiennej $Y = aX + b, a \neq 0$.

Suma niezależnych zmiennych losowych

Zmienne dyskretne:

Niech $X, Y: \Omega \to \mathbb{Z}$ dwie niezależne zmienne niezależne losowe przyjmujące wartości całkowite oraz $\{p_i\}, \{r_k\}$; $\Pr\{X = i\} = p_i$, $\Pr\{Y = k\} = r_k$, są ich rozkładami prawdopodobieństwa. Oczywiście $\Pr\{X = i, Y = k\} = p_i r_k$. Niech $Z = X + Y$ nowa zmienna losowa dyskretna, wtedy:

$$\Pr\{Z = j\} = \Pr\{X = 0, Y = j\} + \Pr\{X = 1, Y = j - 1\} +, \dots, + \Pr\{X = j, Y = 0\}.$$

Jeżeli zmienne X, Y są niezależne, to:

$$\Pr\{Z = j\} = \sum_{i=0}^{j} p_i r_{j-i}$$

Zadanie 9:

Rzucamy dwoma kostkami do gry. Niech X będzie liczbą oczek wyrzuconej na 1 kostce, natomiast $Y = 1$ gdy liczba oczek na 2 kostce jest parzysta, $Y = 2$ w.p.p. Łatwo widać, że zmienne są niezależne. Znaleźć rozkład $Z = X + Y$.

Zmienne ciągłe:

Niech obecnie $X, Y: \Omega \to \mathbb{R}$ dwie niezależne zmienne losowe o wartościach rzeczywistych. Dystrybuanta i funkcja gęstości zmiennej losowej $X + Y$ dane są wzorami

$$F_{X+Y}(z) = \int_{-\infty}^{\infty} F_X(z - y)\, F_Y(y) dy$$

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(z - y)\, f_Y(y) dy$$

Jest to odpowiednio splot dystrybuant i splot gęstości tych zmiennych.

Jeżeli X, Y dwie niezależne zmienne losowe o funkcjach charakterystycznych $\varphi_X(t), \varphi_Y(t)$, to

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t), \forall t$$

Zadanie 10:

Obliczyć funkcję gęstości sumy ciągłej zmiennej losowej X o gęstości f_X i niezależnej zmiennej losowej Y o rozkładzie normalnym $N(0, \sigma)$.

Minimum i maksimum niezależnych zmiennych losowych

Niech $X, Y: \Omega \to \mathbb{R}$ dwie niezależne zmienne losowe i ich dystrybuanty $F_X(x), F_Y(y)$, wówczas zmienna losowa $Z = \min\{X, Y\}$ ma dystrybuantę

$$F_Z(z) = 1 - \big(1 - F_X(z)\big)(1 - F_Y(z))$$

oraz zmienna $H = \max\{X, Y\}$ ma dystrybuantę

$$F_H(z) = F_X(z)F_Y(z).$$

Jeżeli X, Y mają taki sam rozkład i dystrybuantę F, to

$$F_Z(z) = 1 - \big(1 - F(z)\big)^2, \quad F_H(z) = F(z)^2.$$

Dla niezależnych zmiennych losowych $X_1, \ldots, X_n$ o tym samym rozkładzie

$$F_Z(z) = 1 - \big(1 - F(z)\big)^n, \quad F_H(z) = F(z)^n$$

Gdzie obecnie $Z = \min\{X_1, \ldots, X_n\}$, $H = \max\{X_1, \ldots, X_n\}$.

Iloczyn i iloraz zmiennych losowych

Niech $X, Y: \Omega \to \mathbb{R}$ dwie zmienne losowe oraz $Z = X\,Y$, $U = \frac{X}{Y}$, wtedy

$$f_Z(z) = \int_{-\infty}^{+\infty} f_{X,Y}\left(x, \frac{z}{x}\right) \frac{1}{|x|}\, dx$$

gdzie $f_{X,Y}$ gęstość rozkłady łącznego zmiennych X, Y. Ponadto

$$f_U(z) = \int_{-\infty}^{+\infty} f_{X,Y}(zy, y)\, |y|\, dy.$$

Jeżeli zmienne X, Y są niezależne, to

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x)\, f_Y\left(\frac{z}{x}\right) \frac{1}{|x|}\, dx,$$

$$f_U(z) = \int_{-\infty}^{+\infty} f_X(x)\, f_Y\left(\frac{x}{z}\right) \frac{1}{|x|}\, dx.$$