

Zestaw 10 - Wektory losowe ciągłe

Artur Fortuna
AGH

Wektory losowe ciągłe

Jeśli mamy dwie zmienne lub więcej to ich zestawienie to **wektor losowy**.

Wektory losowe ciągłe

Jeśli mamy dwie zmienne lub więcej to ich zestawienie to **wektor losowy**. Zajmiemy się wektorami o rozkładzie ciągłym.

Wektory losowe ciągłe

Jeśli mamy dwie zmienne lub więcej to ich zestawienie to **wektor losowy**. Zajmiemy się wektorami o rozkładzie ciągłym. Zazwyczaj znamy gęstość wektora losowego.

Wektory losowe ciągłe

Jeśli mamy dwie zmienne lub więcej to ich zestawienie to **wektor losowy**. Zajmiemy się wektorami o rozkładzie ciągłym. Zazwyczaj znamy gęstość wektora losowego. W zadaniach z zestawu rozkład jest jednostajny - chodzi o to by trudność w liczeniu całek podwójnych (tak, tak, będą potrzebne) nie przesłoniła istoty zadania.

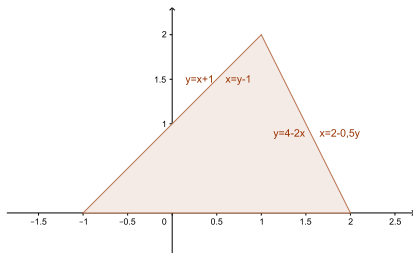
Wektory losowe ciągłe

Jeśli mamy dwie zmienne lub więcej to ich zestawienie to **wektor losowy**. Zajmiemy się wektorami o rozkładzie ciągłym. Zazwyczaj znamy gęstość wektora losowego. W zadaniach z zestawu rozkład jest jednostajny - chodzi o to by trudność w liczeniu całek podwójnych (tak, tak, będą potrzebne) nie przesłoniła istoty zadania. Ale oczywiście tak być nie musi.

Wektory losowe ciągłe

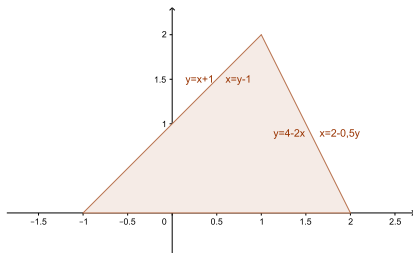
Jeśli mamy dwie zmienne lub więcej to ich zestawienie to **wektor losowy**. Zajmiemy się wektorami o rozkładzie ciągłym. Zazwyczaj znamy gęstość wektora losowego. W zadaniach z zestawu rozkład jest jednostajny - chodzi o to by trudność w liczeniu całek podwójnych (tak, tak, będą potrzebne) nie przesłoniła istoty zadania. Ale oczywiście tak być nie musi. Przyjrzymy się temu na przykładzie.

Wektory losowe ciągłe



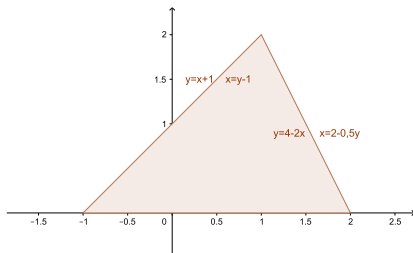
Powiedzmy, że mamy dany rozkład jednostajny skupiony w narysowanym obszarze (obszar nazwiemy D).

Wektory losowe ciągłe

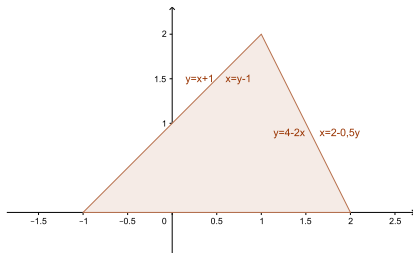


Powiedzmy, że mamy dany rozkład jednostajny skupiony w narysowanym obszarze (obszar nazwiemy D). Zaczynamy od rysunku.

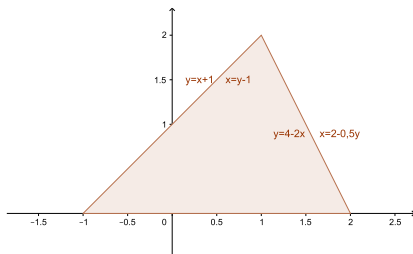
Wektory losowe ciągłe



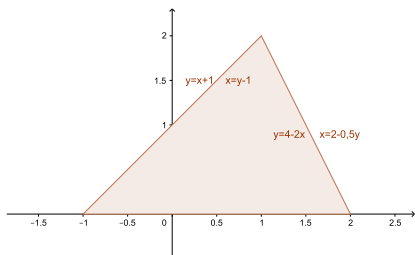
Powiedzmy, że mamy dany rozkład jednostajny skupiony w narysowanym obszarze (obszar nazwiemy D). Zaczynamy od rysunku. Podpisujemy wszystkie funkcje. Na dwa sposoby czyli zarówno $y = \dots$, jak i $x = \dots$.



Czas policzyć gęstość wektora losowego czyli $f_{X,Y}$ jeśli nie była dana wzorem.

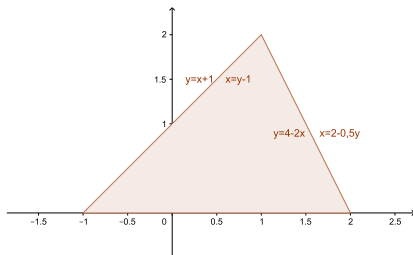


Czas policzyć gęstość wektora losowego czyli $f_{X,Y}$ jeśli nie była dana wzorem. Dla rozkładu jednostajnego zasada jest prosta. Zero poza D , a 1 przez pole $(\frac{1}{|D|})$ wewnątrz D .



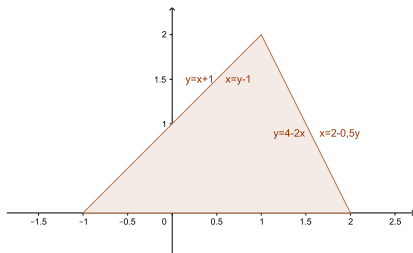
$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{3}, & (x,y) \in D \\ 0, & (x,y) \notin D \end{cases}$$

Rozkłady brzegowe



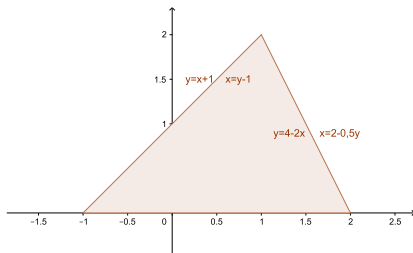
Kolejny krok to **rozkłady brzegowe**.

Rozkłady brzegowe



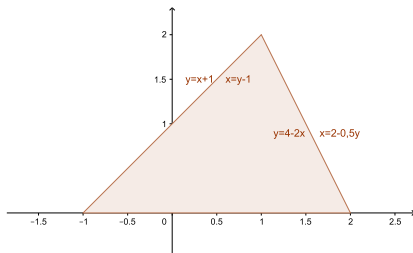
Kolejny krok to **rozkłady brzegowe**. Innymi słowy z rozkładu wspólnego wyciągamy rozkład X i Y .

Rozkłady brzegowe



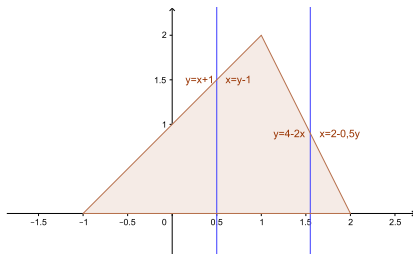
Kolejny krok to **rozkłady brzegowe**. Innymi słowy z rozkładu wspólnego wyciągamy rozkład X i Y . Będziemy liczyć gęstości X i Y czyli f_X i f_Y .

Rozkłady brzegowe



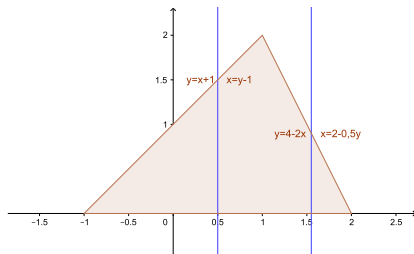
Kolejny krok to **rozkłady brzegowe**. Innymi słowy z rozkładu wspólnego wyciągamy rozkład X i Y . Będziemy liczyć gęstości X i Y czyli f_X i f_Y . Proszę sobie wyobrazić, że ściskamy płaszczyznę do odpowiedniej osi - jak teraz skupi się gęstość?

Rozkłady brzegowe



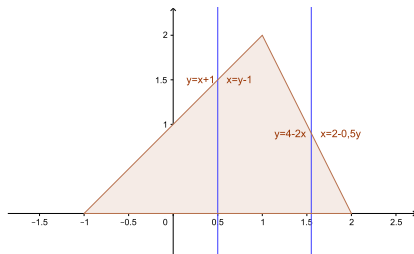
Zacniemy od X .

Rozkłady brzegowe



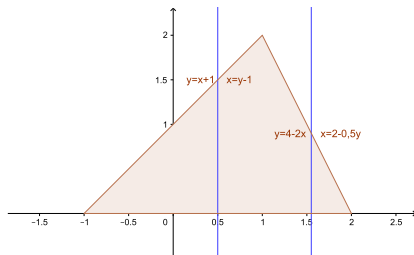
Zaczniemy od X . Rysujemy pionowe linie, wszędzie tam, gdzie ich rysunek przetnie inne funkcje ograniczające obszar.

Rozkłady brzegowe



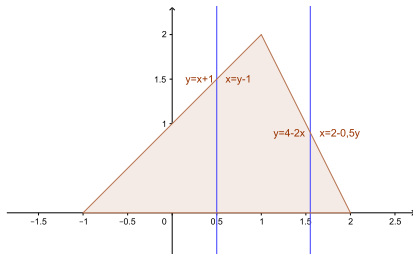
Zaczniemy od X . Rysujemy pionowe linie, wszędzie tam, gdzie ich rysunek przecina inne funkcje ograniczające obszar. A teraz całkujemy $f_{X,Y}(x,y)$ po y w granicach wyznaczonych przez przecięte funkcje.

Rozkłady brzegowe



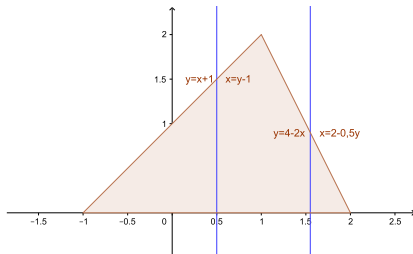
Zacniemy od X . Rysujemy pionowe linie, wszędzie tam, gdzie ich rysunek przecnie inne funkcje ograniczające obszar. A teraz całkujemy $f_{X,Y}(x,y)$ po y w granicach wyznaczonych przez przecięte funkcje. Od razu uprzedzę, to gęstość więc domknięcia przedziałów nie mają znaczenia.

Rozkłady brzegowe



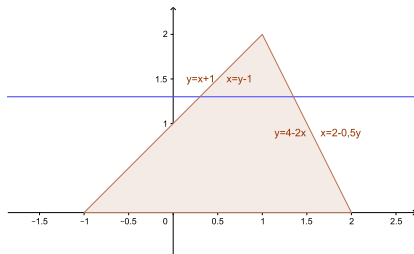
$$f_X(x) = \begin{cases} \int_0^{x+1} \frac{1}{3} dy, & x \in (-1, 1) \\ \int_0^{4-2x} \frac{1}{3} dy, & x \in (1, 2) \\ 0, & x \notin (-1, 2) \end{cases}$$

Rozkłady brzegowe



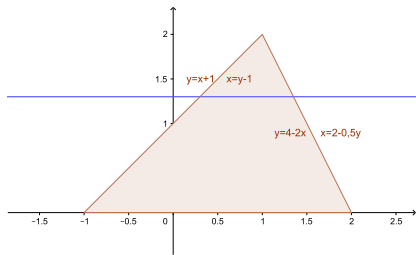
$$f_X(x) = \begin{cases} \frac{1}{3}(x + 1), & x \in (-1, 1) \\ \frac{1}{3}(4 - 2x), & x \in (1, 2) \\ 0, & x \notin (-1, 2) \end{cases}$$

Rozkłady brzegowe



To samo dla Y .

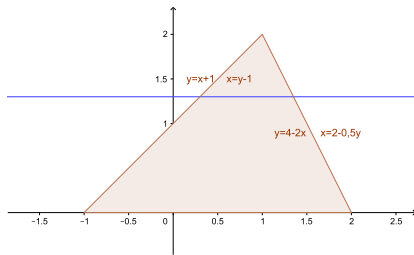
Rozkłady brzegowe



To samo dla Y .

$$f_Y(y) = \begin{cases} \int_{y-1}^{2-0,5y} \frac{1}{3} dx, & y \in (0, 2) \\ 0, & y \notin (0, 2) \end{cases}$$

Rozkłady brzegowe



To samo dla Y .

$$f_Y(y) = \begin{cases} \frac{1}{3}(3 - 1,5y), & y \in (0, 2) \\ 0, & y \notin (0, 2) \end{cases}$$

X i Y to pełnoprzeczne rozkłady, możemy zatem policzyć np. $E(X)$ i $\sigma(X)$ oraz $E(Y)$ i $\sigma(Y)$.

X i Y to pełnoprawne rozkłady, możemy zatem policzyć np. $E(X)$ i $\sigma(X)$ oraz $E(Y)$ i $\sigma(Y)$. Przydadzą się do liczenia **kowariancji**.

X i Y to pełnoprawne rozkłady, możemy zatem policzyć np. $E(X)$ i $\sigma(X)$ oraz $E(Y)$ i $\sigma(Y)$. Przydadzą się do liczenia **kowariancji**.

kowariancja

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y)$$

X i Y to pełnowymiarowe rozkłady, możemy zatem policzyć np. $E(X)$ i $\sigma(X)$ oraz $E(Y)$ i $\sigma(Y)$. Przydadzą się do liczenia **kowariancji**.

kowariancja

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y)$$

Możemy rozwinąć ten wzór by wszystko było jasne:

kowariancja

$$\text{cov}(X, Y) = \iint_{\mathbb{R}^2} xy f_{X,Y}(x, y) dx dy - \int_{\mathbb{R}} x f_X(x) dx \int_{\mathbb{R}} y f_Y(y) dy$$

Współczynnik korelacji

Znajomość kowariancji pozwala policzyć **współczynnik korelacji** zwyczajowo oznaczany grecką literą *rho*.

Współczynnik korelacji

Znajomość kowariancji pozwala policzyć **współczynnik korelacji** zwyczajowo oznaczany grecką literą *rho*.

współczynnik korelacji

$$\rho = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

Współczynnik korelacji

Znajomość kowariancji pozwala policzyć **współczynnik korelacji** zwyczajowo oznaczany grecką literą *rho*.

współczynnik korelacji

$$\rho = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

Współczynnik korelacji to liczba od -1 do 1. Mówi ona jak bardzo powiązane są zmienne.

Współczynnik korelacji

Znajomość kowariancji pozwala policzyć **współczynnik korelacji** zwyczajowo oznaczany grecką literą *rho*.

współczynnik korelacji

$$\rho = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

Współczynnik korelacji to liczba od -1 do 1. Mówi ona jak bardzo powiązane są zmienne. Wynik równy 1 oznaczałby idealną zależność - wzrost jednej zmiennej wywoływałby wzrost drugiej.

Współczynnik korelacji

Znajomość kowariancji pozwala policzyć **współczynnik korelacji** zwyczajowo oznaczany grecką literą *rho*.

współczynnik korelacji

$$\rho = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

Współczynnik korelacji to liczba od -1 do 1. Mówi ona jak bardzo powiązane są zmienne. Wynik równy 1 oznaczałby idealną zależność - wzrost jednej zmiennej wywoływałby wzrost drugiej. Wynik -1 to idealna odwrotna zależność - wzrost jednej to spadek drugiej.

Współczynnik korelacji

Znajomość kowariancji pozwala policzyć **współczynnik korelacji** zwyczajowo oznaczany grecką literą *rho*.

współczynnik korelacji

$$\rho = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

Współczynnik korelacji to liczba od -1 do 1. Mówi ona jak bardzo powiązane są zmienne. Wynik równy 1 oznaczałby idealną zależność - wzrost jednej zmiennej wywoływałby wzrost drugiej. Wynik -1 to idealna odwrotna zależność - wzrost jednej to spadek drugiej. Zero oznacza zmienne nieskorelowane - nie są ze sobą powiązane w żadną stronę.

Aby zmienne były **niezależne** musi być spełniony następujący warunek:

warunek niezależności

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$$

Aby zmienne były **niezależne** musi być spełniony następujący warunek:

warunek niezależności

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$$

Tak musi być w każdym punkcie - dla każdego x i każdego y .

Aby zmienne były **niezależne** musi być spełniony następujący warunek:

warunek niezależności

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$$

Tak musi być w każdym punkcie - dla każdego x i każdego y . To bardzo ciężki warunek, raczej należy oczekiwać, że zmienne będą zależne.

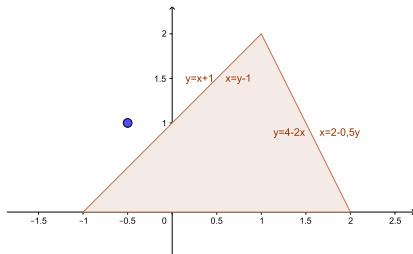
Aby zmienne były **niezależne** musi być spełniony następujący warunek:

warunek niezależności

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$$

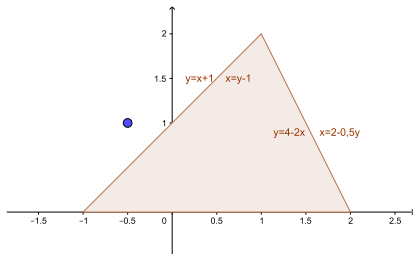
Tak musi być w każdym punkcie - dla każdego x i każdego y . To bardzo ciężki warunek, raczej należy oczekiwać, że zmienne będą zależne. Warto raczej szukać kontrprzykładu - konkretnych x i y , dla których ten warunek nie jest spełniony.

Niezależność



W naszym przykładzie interesujący jest np. punkt $(-\frac{1}{2}, 1)$ bo po lewej stronie warunku będzie zero, a po prawej nie.

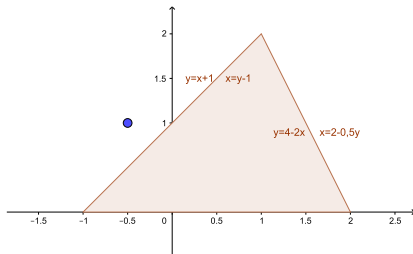
Niezależność



W naszym przykładzie interesujący jest np. punkt $(-\frac{1}{2}, 1)$ bo po lewej stronie warunku będzie zero, a po prawej nie.

$$\begin{aligned} f_{X,Y}\left(-\frac{1}{2}, 1\right) &\stackrel{?}{=} f_X\left(-\frac{1}{2}\right) \cdot f_Y(1) \\ 0 &\neq \frac{1}{3}\left(-\frac{1}{2} + 1\right) \cdot \frac{1}{3}(3 - 1, 5 \cdot 1) \end{aligned}$$

Niezależność

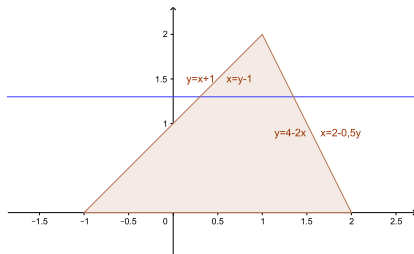


W naszym przykładzie interesujący jest np. punkt $(-\frac{1}{2}, 1)$ bo po lewej stronie warunku będzie zero, a po prawej nie.

$$\begin{aligned} f_{X,Y}\left(-\frac{1}{2}, 1\right) &\stackrel{?}{=} f_X\left(-\frac{1}{2}\right) \cdot f_Y(1) \\ 0 &\neq \frac{1}{3}\left(-\frac{1}{2} + 1\right) \cdot \frac{1}{3}(3 - 1, 5 \cdot 1) \end{aligned}$$

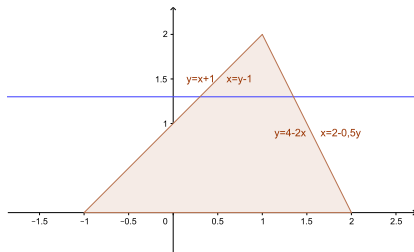
Zmienne X i Y nie są niezależne, są zależne.

Rozkłady warunkowe



Rozkłady warunkowe to odpowiedź na pytanie jak rozkłada się gęstość na danej linii.

Rozkłady warunkowe



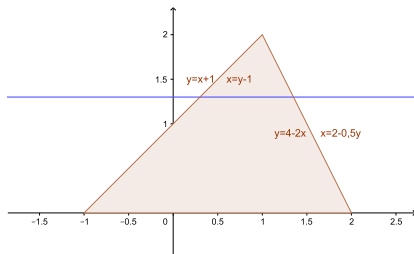
Rozkłady warunkowe to odpowiedź na pytanie jak rozkłada się gęstość na danej linii. Stawiamy warunek i liczymy rozkład z prostych wzorów.

rozkłady warunkowe

$$f(x|Y=y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

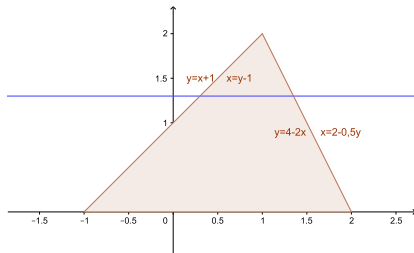
$$f(y|X=x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

Rozkłady warunkowe



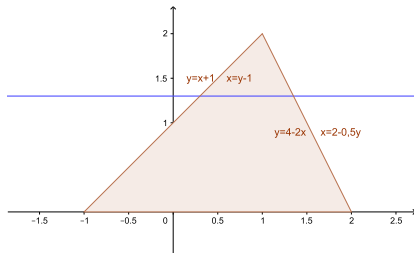
Trudniejsza sprawa to przedziały.

Rozkłady warunkowe



Trudniejsza sprawa to przedziały. Dla jakich x ów na tej linii wynik będzie niezerowy?

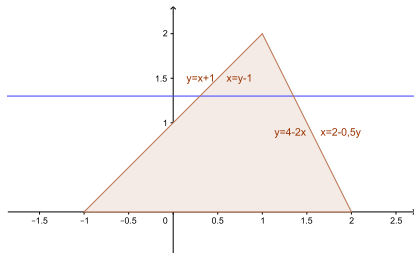
Rozkłady warunkowe



Trudniejsza sprawa to przedziały. Dla jakich x ów na tej linii wynik będzie niezerowy?

$$x \in (y - 1, 2 - 0,5y)$$

Rozkłady warunkowe

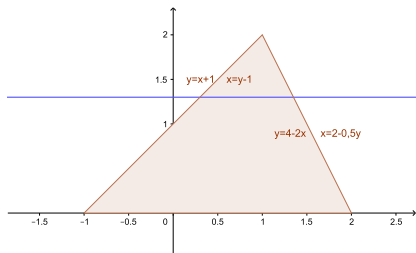


Trudniejsza sprawa to przedziały. Dla jakich x ów na tej linii wynik będzie niezerowy?

$$x \in (y - 1, 2 - 0,5y)$$

Liczymy.

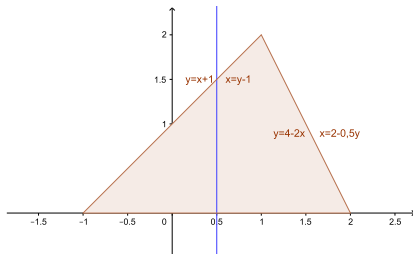
Rozkłady warunkowe



dla $y \in (0, 2)$

$$f(x|Y=y) = \begin{cases} \frac{\frac{1}{3}}{\frac{1}{3}(3-1,5y)}, & x \in (y-1, 2-0,5y) \\ 0, & x \notin (y-1, 2-0,5y) \end{cases}$$

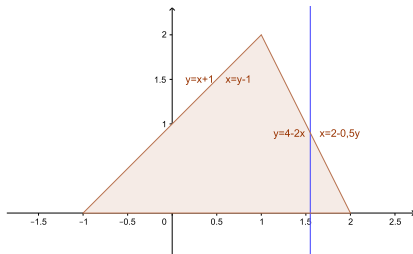
Rozkłady warunkowe



dla $x \in (-1, 1)$

$$f(y|X=x) = \begin{cases} \frac{\frac{1}{3}}{\frac{1}{3}(x+1)}, & y \in (0, x+1) \\ 0, & y \notin (0, x+1) \end{cases}$$

Rozkłady warunkowe



dla $x \in (1, 2)$

$$f(y|X=x) = \begin{cases} \frac{\frac{1}{3}}{4-2x}, & y \in (0, 4-2x) \\ 0, & y \notin (0, 4-2x) \end{cases}$$

$f(x|Y = y)$ i $f(y|X = x)$ to pełnoprawne rozkłady.

Warunkowa wartość oczekiwana

$f(x|Y = y)$ i $f(y|X = x)$ to pełnoprawne rozkłady. $f(x|Y = y)$ to funkcja zmiennej x , a $f(y|X = x)$ to funkcja zmiennej y .

Warunkowa wartość oczekiwana

$f(x|Y = y)$ i $f(y|X = x)$ to pełnoprawne rozkłady. $f(x|Y = y)$ to funkcja zmiennej x , a $f(y|X = x)$ to funkcja zmiennej y . To co jest po $|$ to tylko parametr.

Warunkowa wartość oczekiwana

$f(x|Y = y)$ i $f(y|X = x)$ to pełnoprawne rozkłady. $f(x|Y = y)$ to funkcja zmiennej x , a $f(y|X = x)$ to funkcja zmiennej y . To co jest po $|$ to tylko parametr. Skoro to rozkłady to możemy policzyć dla nich np. wartość oczekiwaną.

Warunkowa wartość oczekiwana

$f(x|Y = y)$ i $f(y|X = x)$ to pełnoprawne rozkłady. $f(x|Y = y)$ to funkcja zmiennej x , a $f(y|X = x)$ to funkcja zmiennej y . To co jest po $|$ to tylko parametr. Skoro to rozkłady to możemy policzyć dla nich np. wartość oczekiwaną. Nosi ona nazwę **warunkowa wartość oczekiwana**.

Warunkowa wartość oczekiwana

$f(x|Y = y)$ i $f(y|X = x)$ to pełnoprawne rozkłady. $f(x|Y = y)$ to funkcja zmiennej x , a $f(y|X = x)$ to funkcja zmiennej y . To co jest po $|$ to tylko parametr. Skoro to rozkłady to możemy policzyć dla nich np. wartość oczekiwaną. Nosi ona nazwę **warunkowa wartość oczekiwana**. Liczymy tak jak zawsze - całka ze zmiennej razy gęstość.

Warunkowa wartość oczekiwana

dla $y \in (0, 2)$

$$f(x|Y=y) = \begin{cases} \frac{1}{(3-1,5y)}, & x \in (y-1, 2-0,5y) \\ 0, & x \notin (y-1, 2-0,5y) \end{cases}$$

$$E(X|Y=y) = \int_{y-1}^{2-0,5y} x \frac{1}{3-1,5y} dx$$

dla $x \in (-1, 1)$

$$f(y|X=x) = \begin{cases} \frac{1}{(x+1)}, & y \in (0, x+1) \\ 0, & y \notin (0, x+1) \end{cases}$$

$$E(Y|X=x) = \int_0^{x+1} y \frac{1}{x+1} dy$$

Warunkowa wartość oczekiwana

dla $x \in (1, 2)$

$$f(y|X=x) = \begin{cases} \frac{1}{(4-2x)}, & y \in (0, 4-2x) \\ 0, & y \notin (0, 4-2x) \end{cases}$$

$$E(Y|X=x) = \int_0^{4-2x} y \frac{1}{4-2x} dy$$

Warunkowa wartość oczekiwana

dla $y \in (0, 2)$

$$E(X|Y = y) = \int_{y-1}^{2-0,5y} x \frac{1}{3-1,5y} dx$$

dla $x \in (-1, 1)$

$$E(Y|X = x) = \int_0^{x+1} y \frac{1}{x+1} dy$$

dla $x \in (1, 2)$

$$E(Y|X = x) = \int_0^{4-2x} y \frac{1}{4-2x} dy$$

Warunkowa wartość oczekiwana

dla $y \in (0, 2)$

$$E(X|Y = y) = \int_{y-1}^{2-0,5y} x \frac{1}{3-1,5y} dx$$

dla $x \in (-1, 1)$

$$E(Y|X = x) = \int_0^{x+1} y \frac{1}{x+1} dy$$

dla $x \in (1, 2)$

$$E(Y|X = x) = \int_0^{4-2x} y \frac{1}{4-2x} dy$$

Podpowiem, że zamiast podnosić do kwadratu lepiej jest skorzystać ze wzoru skróconego mnożenia na $a^2 - b^2$.

Warunkowa wartość oczekiwana

dla $y \in (0, 2)$

$$E(X|Y = y) = \frac{1}{4}y + \frac{1}{2}$$

dla $x \in (-1, 1)$

$$E(Y|X = x) = \frac{1}{2}y + \frac{1}{2}$$

dla $x \in (1, 2)$

$$E(Y|X = x) = 2 - x$$

Linie regresji to funkcje, które możemy naszkicować, gdy już policzymy warunkowe wartości oczekiwane.

Linie regresji to funkcje, które możemy naszkicować, gdy już policzymy warunkowe wartości oczekiwane.

linia regresji X względem Y

$$y \mapsto E(X|Y = y)$$

linia regresji Y względem X

$$x \mapsto E(Y|X = x)$$

Linie regresji to funkcje, które możemy naszkicować, gdy już policzymy warunkowe wartości oczekiwane.

linia regresji X względem Y

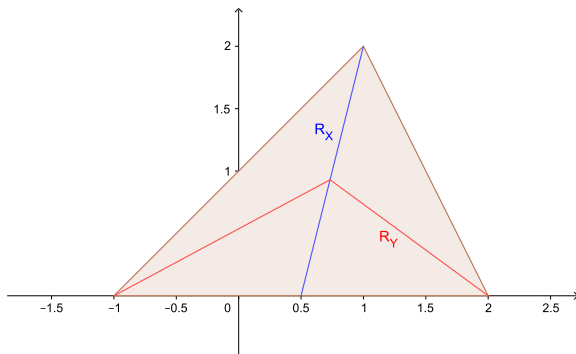
$$y \mapsto E(X|Y = y)$$

linia regresji Y względem X

$$x \mapsto E(Y|X = x)$$

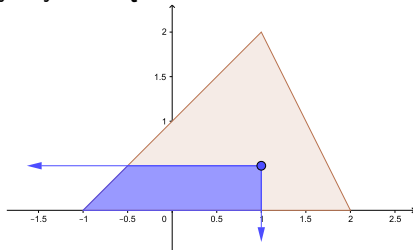
Pierwszą oznaczamy R_X , drugą R_Y .

Rozkłady warunkowe



A jak policzyć dystrybuantę?

A jak policzyć dystrybuantę?



Całka podwójna z gęstości po tej części obszaru D , która jest na lewo i w dół od wybranego punktu.

Wektory losowe dyskretne

Dla wektorów dyskretnych robimy analogicznie, tylko wszędzie tam gdzie była całka będzie suma.

Wektory losowe dyskretne

Dla wektorów dyskretnych robimy analogicznie, tylko wszędzie tam gdzie była całka będzie suma. Zobaczmy to na przykładzie.

Wektory losowe dyskretne

$Y \backslash X$	1	2	3	4
3	$\frac{3}{20}$	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{2}{20}$
5	$\frac{1}{20}$	$\frac{3}{20}$	$\frac{2}{20}$	$\frac{1}{20}$
7	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{1}{20}$

Poziomo X , pionowo Y . Dla przykładu $P(X = 2, Y = 5) = \frac{3}{20}$

Wektory losowe dyskretne

$Y \backslash X$	1	2	3	4
3	$\frac{3}{20}$	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{2}{20}$
5	$\frac{1}{20}$	$\frac{3}{20}$	$\frac{2}{20}$	$\frac{1}{20}$
7	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{1}{20}$

Poziomo X , pionowo Y . Dla przykładu $P(X = 2, Y = 5) = \frac{3}{20}$

Możemy to też zapisać krótko $p_{2,5} = \frac{3}{20}$.

Rozkłady brzegowe

Łatwo jest policzyć rozkłady brzegowe, wystarczy zsumować odpowiednio wiersze lub kolumny.

Rozkłady brzegowe

Łatwo jest policzyć rozkłady brzegowe, wystarczy zsumować odpowiednio wiersze lub kolumny.

$Y \backslash X$	1	2	3	4	$p_{\cdot y}$
3	$\frac{3}{20}$	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	$\frac{8}{20}$
5	$\frac{1}{20}$	$\frac{3}{20}$	$\frac{2}{20}$	$\frac{1}{20}$	$\frac{7}{20}$
7	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{1}{20}$	$\frac{5}{20}$
$p_{x \cdot}$	$\frac{5}{20}$	$\frac{5}{20}$	$\frac{6}{20}$	$\frac{4}{20}$	1

Rozkłady brzegowe

Łatwo jest policzyć rozkłady brzegowe, wystarczy zsumować odpowiednio wiersze lub kolumny.

$Y \backslash X$	1	2	3	4	$p_{\cdot y}$
3	$\frac{3}{20}$	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	$\frac{8}{20}$
5	$\frac{1}{20}$	$\frac{3}{20}$	$\frac{2}{20}$	$\frac{1}{20}$	$\frac{7}{20}$
7	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{1}{20}$	$\frac{5}{20}$
$p_{x \cdot}$	$\frac{5}{20}$	$\frac{5}{20}$	$\frac{6}{20}$	$\frac{4}{20}$	1

Wyjaśnień wymagają oznaczenia. Zapis p_3 byłby niejasny, bo nie wiadomo czy chodzi o X czy o Y . Dlatego używamy kropki w miejscu "usuniętej" zmiennej. Zatem $p_{3 \cdot} = P(X = 3) = \frac{6}{20}$. Zaś $p_{\cdot 3} = P(Y = 3) = \frac{8}{20}$.

Rozkłady brzegowe

$Y \backslash X$	1	2	3	4	$p_{\cdot y}$
3	$\frac{3}{20}$	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	$\frac{8}{20}$
5	$\frac{1}{20}$	$\frac{3}{20}$	$\frac{2}{20}$	$\frac{1}{20}$	$\frac{7}{20}$
7	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{1}{20}$	$\frac{5}{20}$
$p_{x \cdot}$	$\frac{5}{20}$	$\frac{5}{20}$	$\frac{6}{20}$	$\frac{4}{20}$	1

Jeśli ktoś woli formalny zapis to:

$$P_X = \frac{5}{20} \delta_1 + \frac{5}{20} \delta_2 + \frac{6}{20} \delta_3 + \frac{4}{20} \delta_4$$

$$P_Y = \frac{8}{20} \delta_3 + \frac{7}{20} \delta_5 + \frac{5}{20} \delta_7$$

X i Y to pełnowymiarowe rozkłady, możemy zatem policzyć np. $E(X)$ i $\sigma(X)$ oraz $E(Y)$ i $\sigma(Y)$.

Kowariancja i współczynnik korelacji

X i Y to pełnoprawne rozkłady, możemy zatem policzyć np. $E(X)$ i $\sigma(X)$ oraz $E(Y)$ i $\sigma(Y)$.

Wzory na kowariancję i współczynnik korelacji są takie same jak dla rozkładu ciągłego, ale z sumą zamiast całek.

Kowariancja i współczynnik korelacji

X i Y to pełnoprawne rozkłady, możemy zatem policzyć np. $E(X)$ i $\sigma(X)$ oraz $E(Y)$ i $\sigma(Y)$.

Wzory na kowariancję i współczynnik korelacji są takie same jak dla rozkładu ciągłego, ale z sumą zamiast całek. Żeby policzyć $E(XY)$ trzeba policzyć $x \cdot y \cdot p_{x,y}$ dla wszystkich możliwych par x i y i zsumować te wartości.

Aby zmienne były **niezależne** dla każdych x i y musi być spełniony następujący warunek:

warunek niezależności

$$p_{x,y} = p_{x\cdot} \cdot p_{\cdot y}$$

Aby zmienne były **niezależne** dla każdych x i y musi być spełniony następujący warunek:

warunek niezależności

$$p_{x,y} = p_x \cdot p_y$$

Ponownie: to bardzo mocny warunek więc opłaca się raczej szukać kontrprzykładu - konkretnych x i y , dla których ten warunek nie jest spełniony.

Rozkłady warunkowe

Rozkłady warunkowe liczy się bardzo prosto - wystarczy podzielić wszystkie wartości przez liczby z ostatniego wiersza lub ostatniej kolumny.

$Y \backslash X$	1	2	3	4	$p_{\cdot y}$
3	$\frac{3}{20}$	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	$\frac{8}{20}$
5	$\frac{1}{20}$	$\frac{3}{20}$	$\frac{2}{20}$	$\frac{1}{20}$	$\frac{7}{20}$
7	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{1}{20}$	$\frac{5}{20}$
$p_{x \cdot}$	$\frac{5}{20}$	$\frac{5}{20}$	$\frac{6}{20}$	$\frac{4}{20}$	1

Rozkłady warunkowe

Najpierw policzymy rozkłady, gdzie konkretna wartość Y będzie warunkiem. Dzielimy zatem wszystkie liczby z oryginalnej tabelki (czyli bez ostatniego wiersza) przez wartości z ostatniej kolumny.

$Y \backslash X$	1	2	3	4	$p_{\cdot y}$
3	$\frac{3}{20}$	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	$\frac{8}{20}$
5	$\frac{1}{20}$	$\frac{3}{20}$	$\frac{2}{20}$	$\frac{1}{20}$	$\frac{7}{20}$
7	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{1}{20}$	$\frac{5}{20}$
$p_{x \cdot}$	$\frac{5}{20}$	$\frac{5}{20}$	$\frac{6}{20}$	$\frac{4}{20}$	1

I dostajemy coś takiego:

X	1	2	3	4
$ Y = 3$	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{2}{8}$
$ Y = 5$	$\frac{1}{7}$	$\frac{3}{7}$	$\frac{2}{7}$	$\frac{1}{7}$
$ Y = 7$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{1}{5}$

Rozkłady warunkowe

Teraz niech konkretna wartość X będzie warunkiem. Dzielimy zatem wszystkie liczby z oryginalnej tabelki (czyli bez ostatniej kolumny) przez wartości z ostatniego wiersza.

$Y \backslash X$	1	2	3	4	$p_{\cdot y}$
3	$\frac{3}{20}$	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	$\frac{8}{20}$
5	$\frac{1}{20}$	$\frac{3}{20}$	$\frac{2}{20}$	$\frac{1}{20}$	$\frac{7}{20}$
7	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{1}{20}$	$\frac{5}{20}$
$p_{x \cdot}$	$\frac{5}{20}$	$\frac{5}{20}$	$\frac{6}{20}$	$\frac{4}{20}$	1

I dostajemy coś takiego:

Y	$ X = 1$	$ X = 2$	$ X = 3$	$ X = 4$
3	$\frac{3}{5}$	$\frac{1}{5}$	$\frac{2}{6}$	$\frac{2}{4}$
5	$\frac{1}{5}$	$\frac{3}{5}$	$\frac{2}{6}$	$\frac{1}{4}$
7	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{2}{6}$	$\frac{1}{4}$

Warunkowa wartość oczekiwana

X	1	2	3	4
$ Y = 3$	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{2}{8}$
$ Y = 5$	$\frac{1}{7}$	$\frac{3}{7}$	$\frac{2}{7}$	$\frac{1}{7}$
$ Y = 7$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{1}{5}$

W tej tabelce każdy wiersz to osobny rozkład. Możemy policzyć ich wartości oczekiwane, które, tak jak poprzednio noszą nazwę **warunkowe wartości oczekiwane**.

Warunkowa wartość oczekiwana

X	1	2	3	4
$ Y = 3$	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{2}{8}$
$ Y = 5$	$\frac{1}{7}$	$\frac{3}{7}$	$\frac{2}{7}$	$\frac{1}{7}$
$ Y = 7$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{1}{5}$

$$E(X|Y = 3) = 1 \cdot \frac{3}{8} + 2 \cdot \frac{1}{8} + 3 \cdot \frac{2}{8} + 4 \cdot \frac{2}{8} = \frac{19}{8}$$

$$E(X|Y = 5) = 1 \cdot \frac{1}{7} + 2 \cdot \frac{3}{7} + 3 \cdot \frac{2}{7} + 4 \cdot \frac{1}{7} = \frac{17}{7}$$

$$E(X|Y = 7) = 1 \cdot \frac{1}{5} + 2 \cdot \frac{1}{5} + 3 \cdot \frac{2}{5} + 4 \cdot \frac{1}{5} = \frac{13}{5}$$

Warunkowa wartość oczekiwana

Y	$ X = 1$	$ X = 2$	$ X = 3$	$ X = 4$
3	$\frac{3}{5}$	$\frac{1}{5}$	$\frac{2}{6}$	$\frac{2}{4}$
5	$\frac{1}{5}$	$\frac{3}{5}$	$\frac{2}{6}$	$\frac{1}{4}$
7	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{2}{6}$	$\frac{1}{4}$

W tej tabelce każda kolumna to osobny rozkład.

Warunkowa wartość oczekiwana

Y	$ X = 1$	$ X = 2$	$ X = 3$	$ X = 4$
3	$\frac{3}{5}$	$\frac{1}{5}$	$\frac{2}{6}$	$\frac{2}{4}$
5	$\frac{1}{5}$	$\frac{3}{5}$	$\frac{2}{6}$	$\frac{1}{4}$
7	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{2}{6}$	$\frac{1}{4}$

$$E(Y|X = 1) = 3 \cdot \frac{3}{5} + 5 \cdot \frac{1}{5} + 7 \cdot \frac{1}{5} = \frac{21}{5}$$

$$E(Y|X = 2) = 3 \cdot \frac{1}{5} + 5 \cdot \frac{3}{5} + 7 \cdot \frac{1}{5} = \frac{25}{5}$$

$$E(Y|X = 3) = 3 \cdot \frac{2}{6} + 5 \cdot \frac{2}{6} + 7 \cdot \frac{2}{6} = \frac{30}{6}$$

$$E(Y|X = 4) = 3 \cdot \frac{2}{4} + 5 \cdot \frac{1}{4} + 7 \cdot \frac{1}{4} = \frac{18}{4}$$

Skoro

$$E(X|Y = 3) = \frac{19}{8}$$

$$E(X|Y = 5) = \frac{17}{7}$$

$$E(X|Y = 7) = \frac{13}{5}$$

to jako linie regresji X względem Y formalnie powinniśmy narysować punkty $(\frac{19}{8}, 3)$, $(\frac{17}{7}, 5)$ i $(\frac{13}{5}, 7)$. Ale w praktyce przedłużamy je liniami i rysujemy łamane.

Tak samo dla linii regresji Y względem X .

$$E(Y|X = 1) = \frac{21}{5}$$

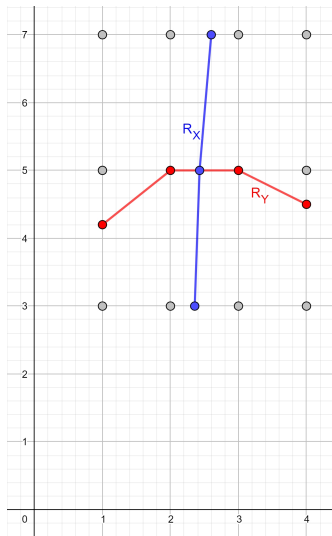
$$E(Y|X = 2) = \frac{25}{5}$$

$$E(Y|X = 3) = \frac{30}{6}$$

$$E(Y|X = 4) = \frac{18}{4}$$

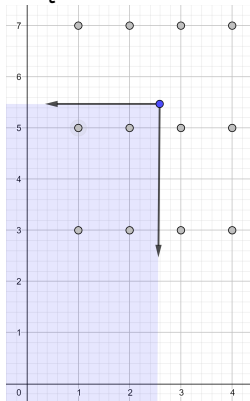
Punkty $(1, \frac{21}{5})$, $(2, \frac{25}{5})$, $(3, \frac{30}{6})$ i $(4, \frac{18}{4})$ przedłużamy liniami dołamanej.

Linie regresji



A jak policzyć dystrybuantę?

A jak policzyć dystrybuantę?



Suma z prawdopodobieństw punktów z tej części obszaru, która jest na lewo i w dół od wybranego punktu (wliczając punkty leżące równo na liniach).

Powodzenia!
Proszę pytać jeśli coś jest niejasne.