Rachunek prawdopodo-bieństwa i statystyka

Wykład 3

Zmienna losowa

Rozszerzenie znaczenia funkcji zmiennej rzeczywistej na przypadki, kiedy zmienna niezależna nie jest liczbą rzeczywistą:

- odległość to funkcja pary punktów,
- obwód trójkąta, to funkcja określona na zbiorze trójkątów,
- współczynnik dwumianowy, to funkcja określona dla par liczbowych (n,k), ...

Definicja: Zmienną losową X nazywamy funkcję określoną na przestrzeni zdarzeń elementarnych Ω, czyli przyporządkowanie każdemu zdarzeniu elementarnemu pewnej liczby rzeczywistej.

Przykłady:

- liczba „orłów” w rzucie monetą,
- suma oczek wyrzuconych na trzech kostkach do gry,
- liczba asów otrzymanych do ręki przy grze w brydża,
- położenie punktu na odcinku [0,1],
- masa kulek łożyskowych produkowanych w fabryce,
- energia układu fizycznego.

Dyskretna zmienna losowa

Z każdą wartością zmiennej losowej wiążemy liczbę określającą częstość jej występowania.

Przykład: Totolotek

$$n = \sum_{k=1}^{49} n_k = 18702$$

$$F_k = \frac{n_k}{n} \xrightarrow[n\to\infty]{} P_k = \frac{1}{49}$$

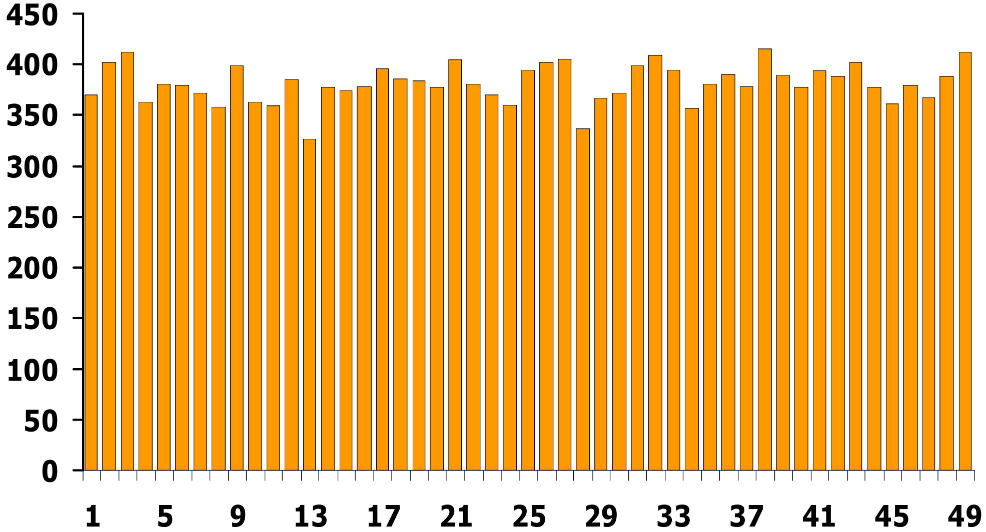
Uzyskaną w ten sposób zależność p-twa od zmiennej losowej nazywamy dyskretnym rozkładem p-twa określonym dla dyskretnej zmiennej losowej.

P_k spełniają warunek unormowania:

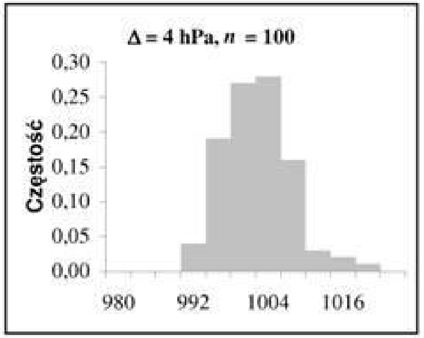
$$\sum_{k=0}^{\infty} P_k = 1 \quad \Rightarrow \quad P(i \le K \le m) = \sum_{k=i}^{m} P_k$$

Wyniki totolotka (3117 losowań 6 z 49)

1 370	2 402	3 412	4 364	5 381	6 380	7 372
8 358	9 399	10 363	11 359	12 385	13 327	14 377
15 374	16 379	17 396	18 386	19 384	20 378	21 405
22 381	23 370	24 360	25 395	26 402	27 406	28 336
29 367	30 372	31 399	32 409	33 395	34 357	35 381
36 390	37 379	38 415	39 389	40 377	41 394	42 388
43 402	44 378	45 361	46 380	47 368	48 388	49 412



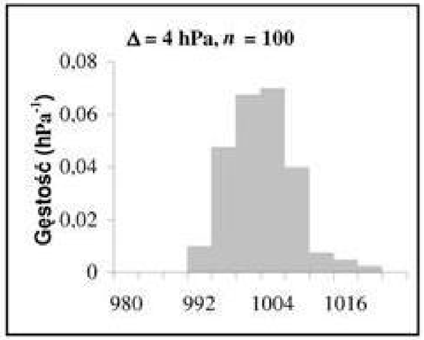
Ciągła zmienna losowa

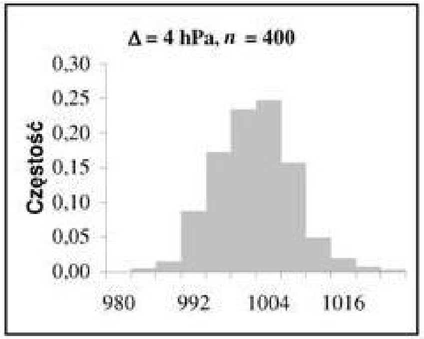


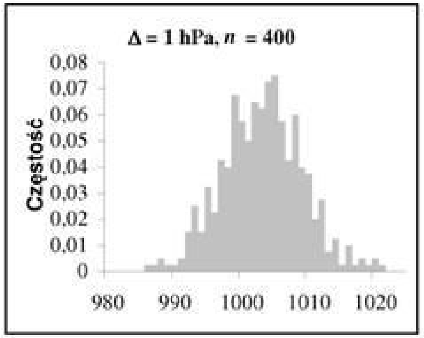
$$F_k = \frac{n_k}{n} \rightarrow f_k = \frac{F_k}{\Delta_k} = \frac{n_k}{n\Delta_k} \qquad \sum_{k=1}^{N} f_k \Delta_k = 1$$

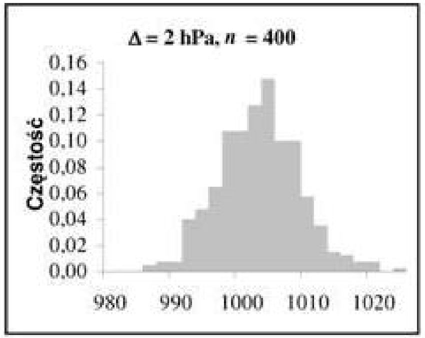
$$f_k = \frac{n_k}{n\Delta_k} \xrightarrow[n\to\infty,\ \Delta_k\to 0]{} f(x)$$

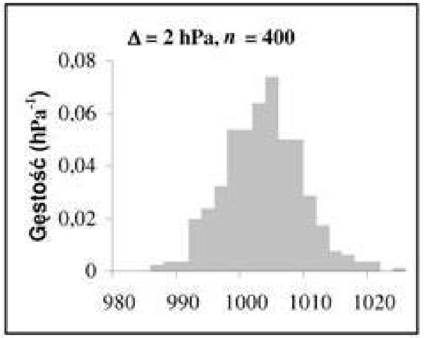
$$\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1$$

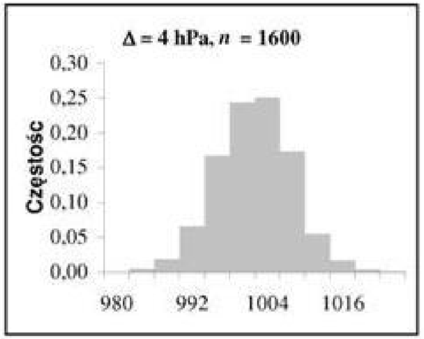


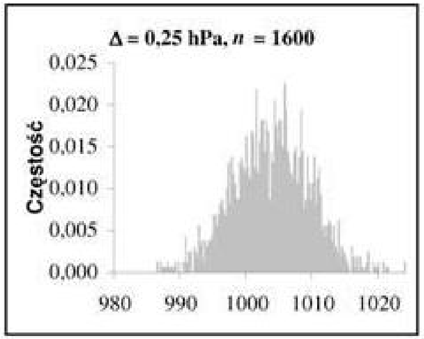


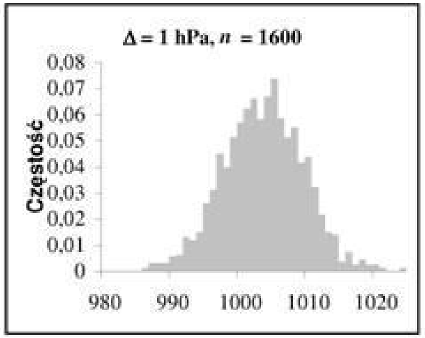


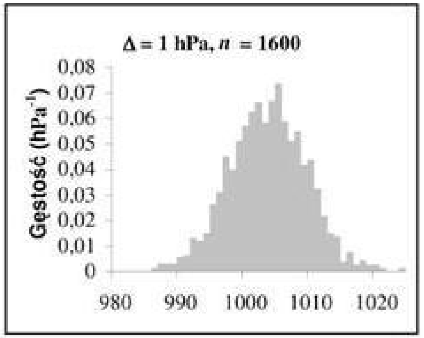












Ciągła zmienna losowa

$f(x)$ - funkcja (rozkład) gęstości prawdopodobieństwa zmiennej losowej ciągłej

Uwaga:

- dla każdego x, funkcja gęstości p-twa jest nieujemna: $f(x) \geq 0$
- funkcja gęstości p-twa jest unormowana, tzn. $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$
- wymiar funkcje gęstości p-twa jest odwrotnością wymiaru zmiennej losowej, którą opisuje.

P-two, że zmienna losowa X przyjmuje wartości z przedziału (a, b):

$$\mathrm{P}(a \leq \mathrm{X} \leq b) = \int_a^b f(x)\,\mathrm{d}x$$

Uwaga: $P(\mathrm{X} = a) = \int_a^a f(x)\,\mathrm{d}x = 0$

Przykład: Znajdź p-two, że zmienna losowa o gęstości p-twa danej przez:

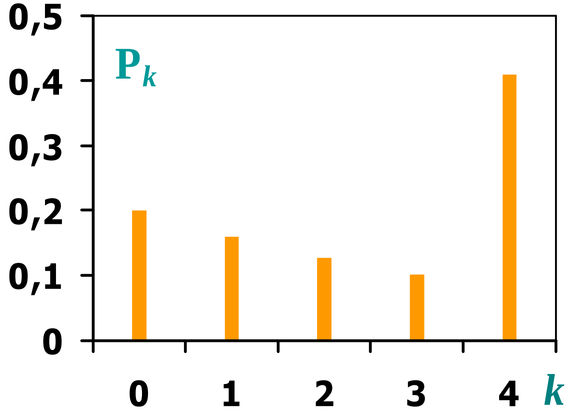
$$\mathcal{N}(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

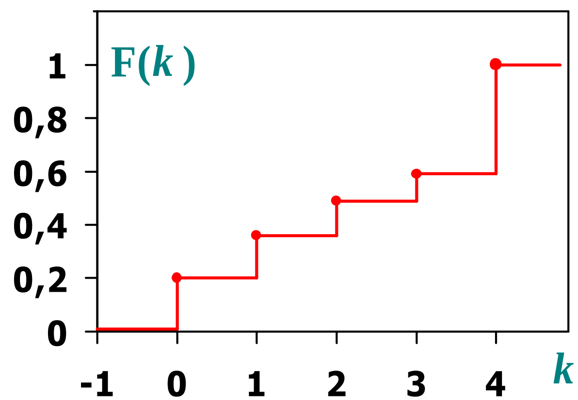
przyjmuje wartości z przedziału $(-\sigma, \sigma)$.

Dystrybuanta zmiennej losowej

Definicja: Dystrybuantą rozkładu dyskretnej zmiennej losowej K nazywamy wielkość określającą łączne p-two wszystkich możliwych wartości tej zmiennej, nie większych od pewnej zadanej wartości k:

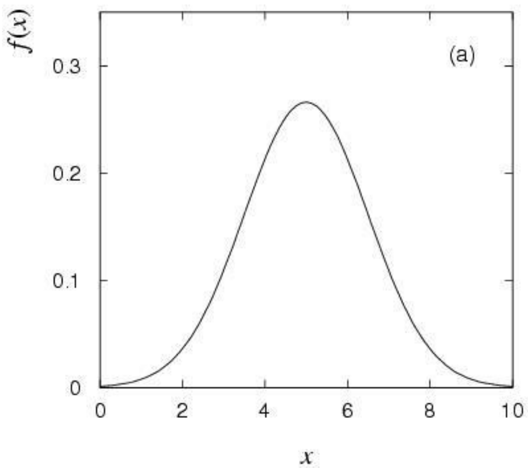
$$F(k) = P(K \le k) = \sum_{i=0}^{k} P_i$$

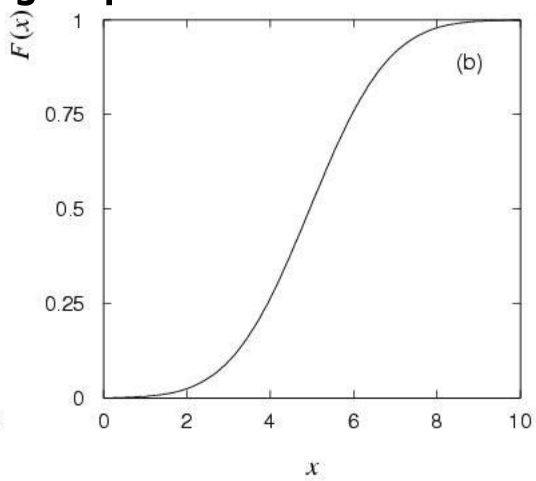




Definicja: Dystrybuantą rozkładu ciągłej zmiennej losowej X nazywamy:

$$F(x) \equiv P(X \le x) = \int_{-\infty}^{x} f(x')\,dx'$$





Dystrybuanta zmiennej losowej

Znając dystrybuantę jesteśmy w stanie określić rozkład p-twa dyskretnej zmiennej losowej: $\mathrm{P}_k = \mathrm{F}(k) - \mathrm{F}(k-1)$

... lub znaleźć funkcję gęstości p-twa ciągłej zmiennej losowej: $f(x) = \dfrac{\mathrm{d}\,\mathrm{F}(x)}{\mathrm{d}\,x}$

Uwaga: Dystrybuanta jest niemalejącą, co najmniej prawostronnie ciągłą funkcją przyjmującą wartości z przedziału [0,1].

P-two znalezienia zmiennej losowej w przedziale:

$$\mathrm{P}(a \le \mathrm{X} \le b) = \int_a^b f(x)\,\mathrm{d}\,x = \int_{-\infty}^{b} f(x)\,\mathrm{d}\,x - \int_{-\infty}^{a} f(x)\,\mathrm{d}\,x = \mathrm{F}(b) - \mathrm{F}(a)$$

$$\mathrm{P}(k_1 < \mathrm{K} \le k_2) = \sum_{k=k_1+1}^{k_2} \mathrm{P}_k = \sum_{k=1}^{k_2} \mathrm{P}_k - \sum_{k=1}^{k_1} \mathrm{P}_k = \mathrm{F}(k_2) - \mathrm{F}(k_1)$$

Definicja: Niech X będzie ciągłą zmienną losową, oraz p liczbą z przedziału [0, 1]. p-tym kwantylem (100p-tym percentylem) rozkładu zmiennej X nazywamy najmniejszą liczbę q_p taką, że:

$$\mathrm{F}(q_p) = \mathrm{P}(\mathrm{X} \le q_p) = p \quad \Rightarrow \quad q_p = \mathrm{F}^{-1}(p)$$

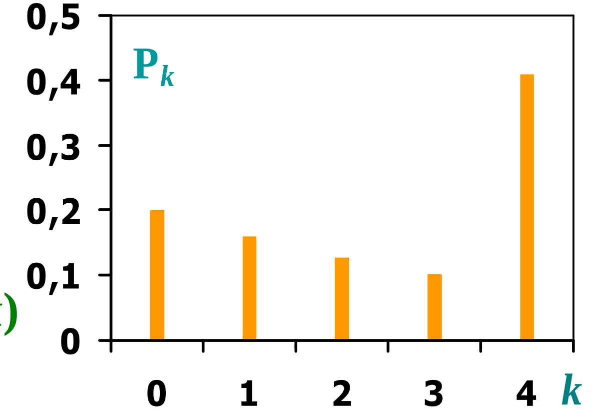
Uwaga: Kwantyl rzędu 0.5 nazywamy medianą rozkładu.

Zmienna losowa dyskretna - przykład

<u>Przykład:</u> Na drodze ruchu pociągów znajdują się w znacznej odległości od siebie 4 semafory, z których każdy (wobec znacznej odległości niezależnie od innych) zezwala na przejazd z p-twem 0.8. Niech K oznacza liczbę semaforów zezwalających na przejazd i poprzedzających pierwsze zatrzymanie lub stację docelową.

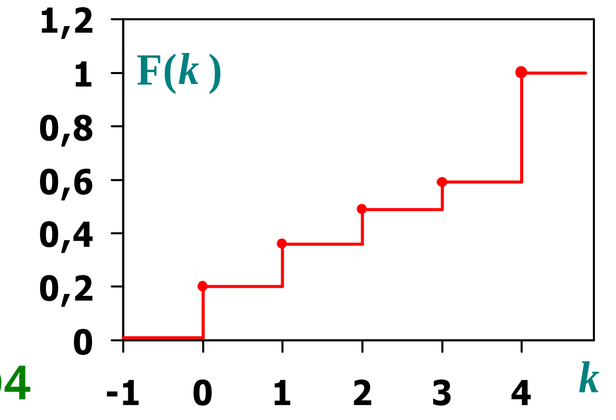
- rozkład p-twa dyskretnej zmiennej losowej K

k	0	1	2	3	4
P_k	0.2	0.16	0.128	0.1024	0.4096



- dystrybuanta rozkładu zmiennej losowej K, $F(k)=P(K<k)$

k	$(-\infty,0)$	$\langle 0,1)$	$\langle 1,2)$	$\langle 2,3)$	$\langle 3,4)$	$\langle 4,\infty)$
$F(k)$	0	0.2	0.36	0.488	0.5904	1



- $P(k=2) = F(2)-F(2^-) = F(2)-F(2^-) = 0.128$

  $P(k>2) = F(+\infty)-F(2) = 1-0.488 = 0.512$

  $P(k\geq 2) = F(+\infty)-F(2^-) = 1-0.36 = 0.64$

  $P(1\leq k\leq 3) = F(3)-F(1^-) = 0.5904-0.2 = 0.3904$

Zmienna losowa ciągła - przykład

Przykład: Zmienna losowa X opisana jest dystrybuantą:

$$F(x) = \frac{1}{\pi}\left(\operatorname{arctg} x + \frac{\pi}{2}\right) \qquad -\infty < x < \infty$$

Znajdź funkcję gęstości p-twa.

$$f(x) = \frac{d}{dx}F(x) = \frac{d}{dx}\left(\frac{1}{\pi}\left(\operatorname{arctg} x + \frac{\pi}{2}\right)\right) = \frac{1}{\pi}\frac{1}{1+x^2}$$

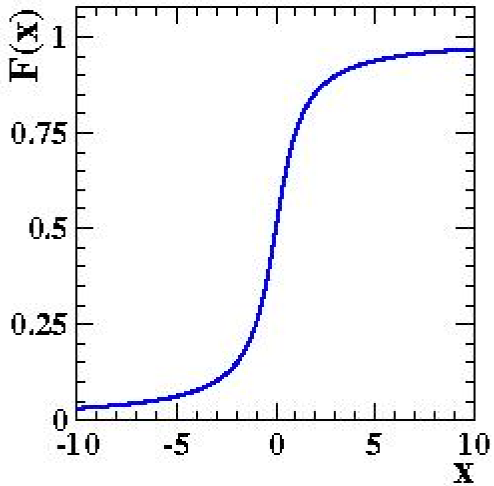
Ile wynosi p-two zdarzenia opisanego warunkiem $|X| < 1$?

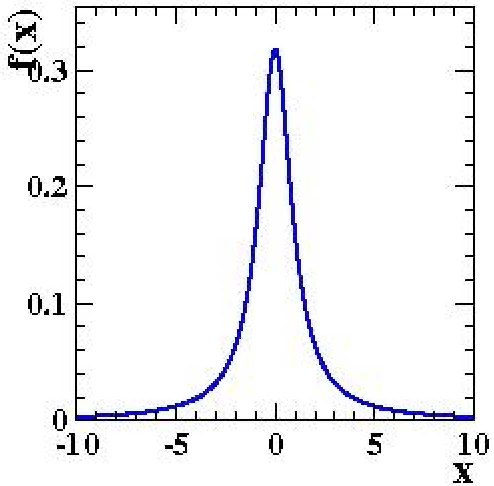
$$P(|x|<1) = P(-1<x<1) = F(1) - F(-1) = \frac{1}{\pi}\left(\frac{\pi}{4}+\frac{\pi}{2}\right) - \frac{1}{\pi}\left(-\frac{\pi}{4}+\frac{\pi}{2}\right) = 0.5$$

$$P(-1<x<1) = \int_{-1}^{1} f(x)\,dx =$$

$$= \int_{-1}^{1} \frac{1}{\pi}\frac{dx}{1+x^2} = \frac{1}{\pi}(\operatorname{arctg} x)\Big|_{-1}^{1} =$$

$$= \frac{1}{\pi}\left(\frac{\pi}{4} - \left(-\frac{\pi}{4}\right)\right) = \frac{1}{2}$$





Dwuwymiarowa zmienna losowa

Przykład: Pewną zbiorowość badamy ze względu na dwie cechy:

- ludzi ze względu na ciężar i wzrost,
- współrzędne (x,y) trafień strzałką w tarczę,
- w rzucie dwoma kostkami: sumę oczek i maksymalną liczbę oczek, ...

Definicja: Dwuwymiarową (2D) zmienna losową nazywamy parę zmiennych losowych jednowymiarowych określonych na tej samej przestrzeni zdarzeń Ω.

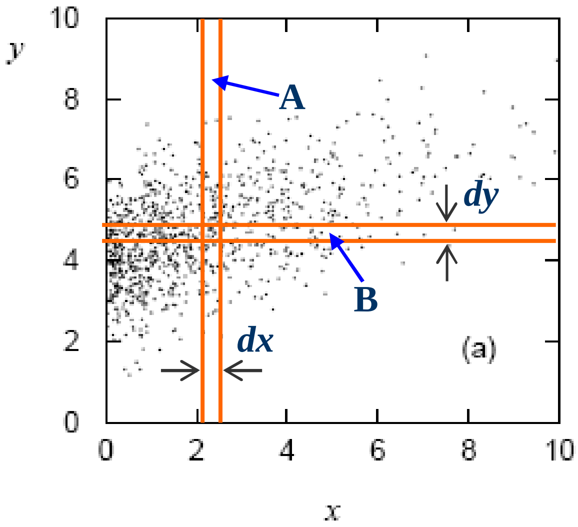
Poprzez analogię do zmiennych losowych jednowymiarowych wprowadzamy pojęcia rozkładu p-twa dwuwymiarowej dyskretnej zmiennej losowej $\mathbf{P}_{km}$ oraz funkcji gęstości p-twa dwuwymiarowej ciągłej zmiennej losowej $f(x,y)$.

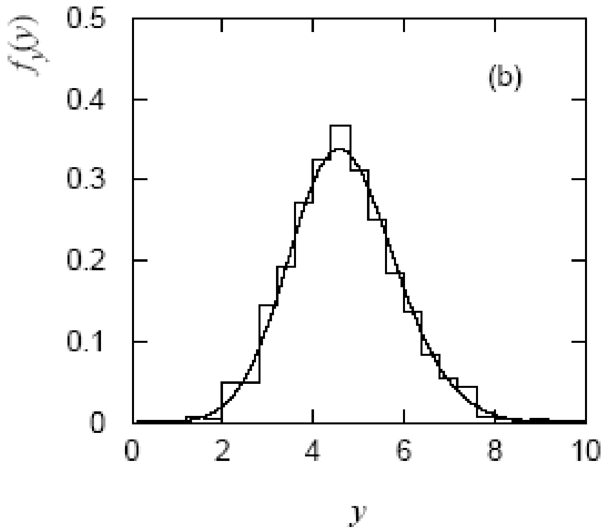
Dowolne p-twa wyliczamy ze wzorów:

$$\mathbf{P}\left(k_1 \le \mathbf{K} \le k_2, m_1 \le \mathbf{M} \le m_2\right) = \sum_{k=k_1}^{k_2} \sum_{m=m_1}^{m_2} \mathbf{P}_{km}$$

$$\mathbf{P}\left(x_1 \le \mathbf{X} \le x_2,\ y_1 \le \mathbf{Y} \le y_2\right) = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f\left(x', y'\right) \mathrm{d}\, x' \,\mathrm{d}\, y'$$

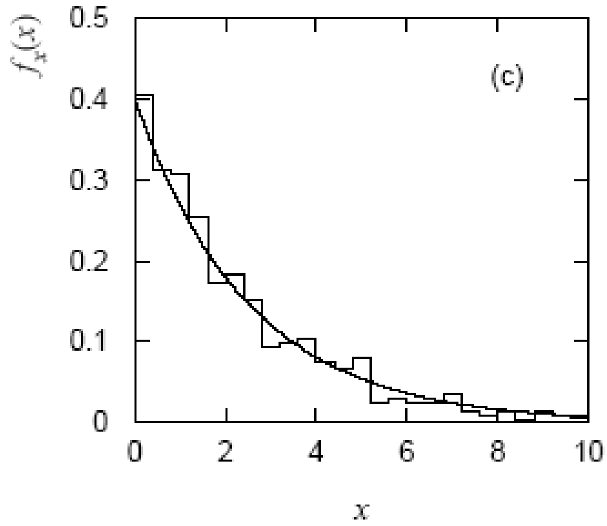
Rozkłady brzegowe





P-two łącznego zajścia zdarzeń **A** i **B**:

$$f\left(x,y\right)dxdy$$



Rozkłady brzegowe to rzuty łącznej funkcji gęstości p-twa na poszczególne osie (zmienne).

Definicja: Rozkładami brzegowymi dwuwymiarowej zmiennej losowej nazywamy:

$$\mathbf{P}_{\bullet,m} = \sum_k \mathbf{P}_{k,m} \qquad \mathbf{P}_{k,\bullet} = \sum_m \mathbf{P}_{k,m}$$

$$f_1(x) \equiv \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}\,y \qquad f_2(y) \equiv \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}\,x$$

Dystrybuanta 2D zmiennej losowej

Definicja: Dystrybuantą dwuwymiarowej zmiennej losowej nazywamy funkcję:

$$\mathrm{F}(x,y) \equiv \mathrm{P}(\mathrm{X} \le x,\ \mathrm{Y} \le y) \qquad (x,y) \in \mathcal{R}^2$$

Dystrybuanty rozkładów brzegowych zmiennej dyskretnej:

$$\mathrm{F}_1(x) \equiv \mathrm{P}(\mathrm{X} \le x) = \sum_{x_k \le x} \mathrm{P}_{k,\bullet} \qquad \mathrm{F}_2(y) \equiv \mathrm{P}(\mathrm{Y} \le y) = \sum_{y_m \le y} \mathrm{P}_{\bullet,m}$$

... i zmiennej ciągłej:

$$\mathrm{F}_1(x) \equiv \mathrm{P}(\mathrm{X} \le x) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(x',y')\,dy'dx' = \int_{-\infty}^{x} f_1(x')\,dx'$$

$$\mathrm{F}_2(y) \equiv \mathrm{P}(\mathrm{Y} \le y) = \int_{-\infty}^{y} \int_{-\infty}^{\infty} f(x',y')\,dx'dy' = \int_{-\infty}^{y} f_2(y')\,dy'$$

Pomiędzy dystrybuantą rozkładu dwuwymiarowego i dystrybuantami rozkładów brzegowych zachodzą związki:

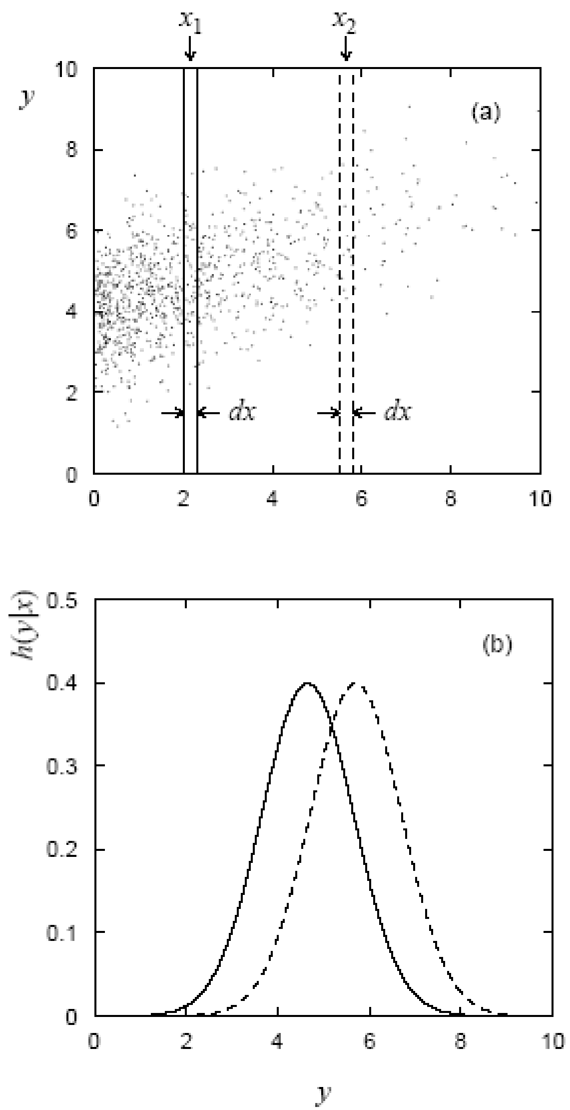
$$\mathrm{F}_1(x) \equiv \mathrm{P}(\mathrm{X} \le x) = \lim_{y \to +\infty} \mathrm{F}(x,y) \qquad \mathrm{F}_2(y) \equiv \mathrm{P}(\mathrm{Y} \le y) = \lim_{x \to +\infty} \mathrm{F}(x,y)$$

Rozkłady warunkowe

Definicja: Rozkłady warunkowe 2D zmiennej losowej:

$$\mathbf{P}(\mathbf{K}\,|\,m)=\frac{\mathbf{P}_{k,m}}{\mathbf{P}_{\bullet,m}} \qquad \mathbf{P}(\mathbf{M}\,|\,k)=\frac{\mathbf{P}_{k,m}}{\mathbf{P}_{k,\bullet}}$$

$$g(x\,|\,y)\equiv\frac{f(x,y)}{f_2(y)} \qquad h(y\,|\,x)\equiv\frac{f(x,y)}{f_1(x)}$$



Tw. Bayes'a dla zmiennych ciągłych:

$$g(x\,|\,y)=\frac{h(y\,|\,x)\,f_x(x)}{f_y(y)}$$

Analog tw. o p-twie całkowitym dla zmiennych ciągłych:

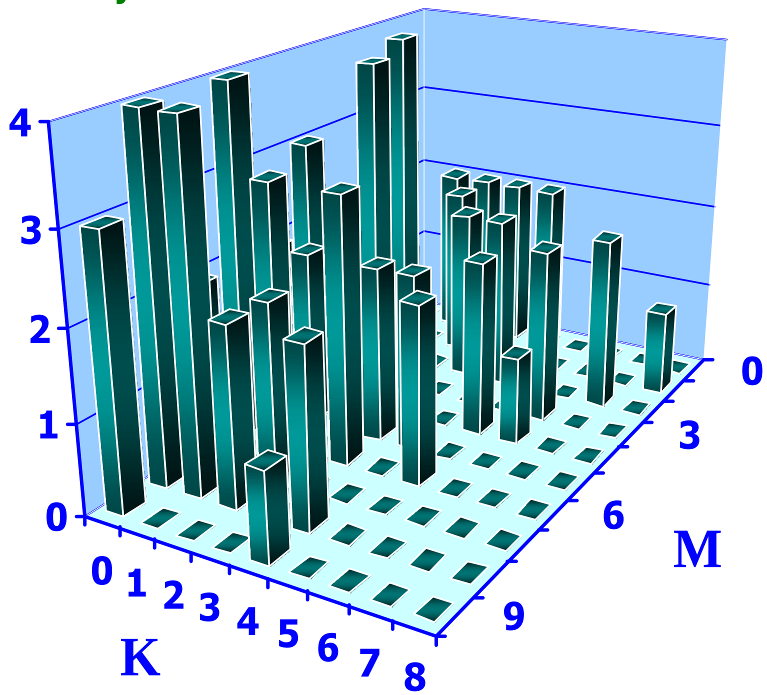
$$f_x(x)=\int_{-\infty}^{\infty} g(x\,|\,y)\,f_y(y)\,dy$$

$$f_y(y)=\int_{-\infty}^{\infty} h(y\,|\,x)\,f_x(x)\,dx$$

Rozkłady brzegowe – zm. dyskretna

<u>Przykład:</u> Dwie niezależne dyskretne zmienne losowe I oraz J mają rozkład jednostajny na zbiorze liczb całkowitych 1, 2, ..., 9. Konstruujemy iloczyn tych zmiennych i zapisujemy w postaci $10K+M$. Znajdź rozkład zmiennych losowych M i N.

K \ M	0	1	2	3	4	5	6	7	8	9	$n_{k,\bullet}$
0	0	1	2	2	3	2	4	2	4	3	23
1	2	0	4	0	2	2	3	0	4	0	17
2	2	2	0	0	4	1	0	2	2	0	13
3	2	0	2	0	0	2	3	0	0	0	9
4	2	0	2	0	0	2	0	0	2	1	9
5	0	0	0	0	2	0	2	0	0	0	4
6	0	0	0	2	1	0	0	0	0	0	3
7	0	0	2	0	0	0	0	0	0	0	2
8	0	1	0	0	0	0	0	0	0	0	1
$n_{\bullet,m}$	8	4	12	4	12	9	12	4	12	4	81



Rozkład dwuwymiarowy i rozkłady brzegowe:

$$P_{k,m} = \frac{n_{k,m}}{81} \qquad P_{k,\bullet} = \frac{1}{81}\sum_{m=0}^{9} n_{k,m} \qquad P_{\bullet,m} = \frac{1}{81}\sum_{k=0}^{8} n_{k,m}$$

Rozkłady warunkowe:

$$P(K \mid m=5) = \frac{P_{k,5}}{P_{\bullet,5}} \qquad P(M \mid k=6) = \frac{P_{6,m}}{P_{6,\bullet}}$$

Rozkłady brzegowe – zm. dyskretna

<u>Przykład:</u> (cd.) Dystrybuantę dwuwymiarowej dyskretnej zmiennej losowej (**K**, **M**) wyznaczamy z tabeli przedstawiającej rozkład p-twa $\mathbf{F}(x, y)=\mathbf{P}(\mathbf{X}\le x; \mathbf{Y}\le y)$:

K \ M	$(-\infty,0)$	$\langle 0,1)$	$\langle 1,2)$	$\langle 2,3)$	$\langle 3,4)$	$\langle 4,5)$	$\langle 5,6)$	$\langle 6,7)$	$\langle 7,8)$	$\langle 8,9)$	$\langle 9,\infty)$
$(-\infty,0)$	0	0	0	0	0	0	0	0	0	0	0
$\langle 0,1)$	0	0	1/81	3/81	5/81	8/81	10/81	14/81	16/81	20/81	23/81
$\langle 1,2)$	0	2/81	3/81	9/81	11/81	16/81	20/81	27/81	29/81	37/81	40/81
$\langle 2,3)$	0	4/81	7/81	13/81	15/81	24/81	29/81	36/81	40/81	50/81	53/81
$\langle 3,4)$	0	6/81	9/81	17/81	19/81	28/81	35/81	45/81	49/81	59/81	62/81
$\langle 4,5)$	0	8/81	11/81	21/81	23/81	32/81	41/81	51/81	55/81	67/81	71/81
$\langle 5,6)$	0	8/81	11/81	21/81	23/81	34/81	43/81	55/81	59/81	71/81	75/81
$\langle 6,7)$	0	8/81	11/81	21/81	25/81	37/81	46/81	58/81	62/81	74/81	78/81
$\langle 7,8)$	0	8/81	11/81	23/81	27/81	39/81	48/81	60/81	64/81	76/81	80/81
$\langle 8,\infty)$	0	8/81	12/81	24/81	28/81	40/81	49/81	61/81	65/81	77/81	1

Znając dystrybuantę obliczamy p-two:

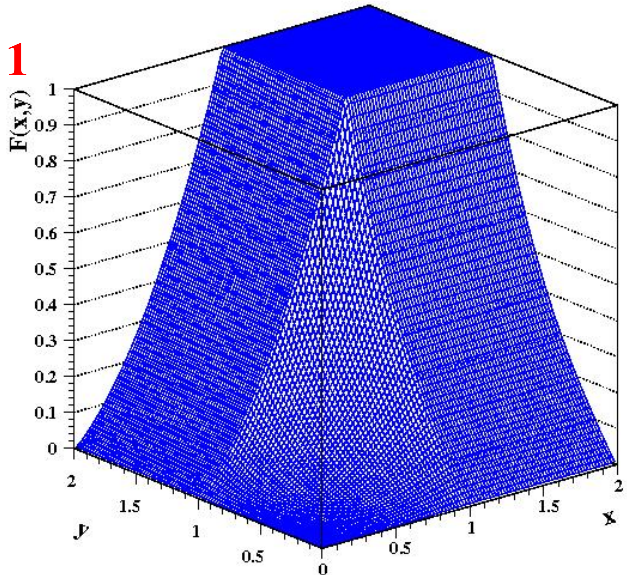
$$\mathbf{P}\left(k_1 < \mathbf{K} \le k_2;\ m_1 < \mathbf{M} \le m_2\right) = \mathbf{F}\left(k_2, m_2\right) - \mathbf{F}\left(k_2, m_1\right) - \mathbf{F}\left(k_1, m_2\right) + \mathbf{F}\left(k_1, m_1\right)$$

Dystrybuanta – 2D zmienna ciągła

Przykład: Znaleźć dystrybuantę dwuwymiarowej zmiennej losowej (X,Y) o funkcji gęstości $f(x,y) = 6/5\cdot(x + y^2)$ określonej na kwadracie $0 \le x, y \le 1$.

$$F(x,y) = P(X \le x;\ Y \le y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(x,y)dxdy = \frac{6}{5}\int_{-\infty}^{x}\int_{-\infty}^{y}(x+y^2)dxdy =$$

$$= \begin{cases} 0 & x<0 \vee y<0 \\ \frac{6}{5}\int_0^x\int_0^y (x+y^2)dxdy = \frac{6}{5}\left(\frac{1}{2}x^2y + \frac{1}{3}xy^3\right) & 0\le x<1 \wedge 0\le y<1 \\ \frac{6}{5}\int_0^x\int_0^1 (x+y^2)dxdy = \frac{6}{5}\left(\frac{1}{2}x^2 + \frac{1}{3}x\right) & 0\le x<1 \wedge 1\le y \\ \frac{6}{5}\int_0^1\int_0^y (x+y^2)dxdy = \frac{6}{5}\left(\frac{1}{2}y + \frac{1}{3}y^3\right) & 1\le x \wedge 0\le y<1 \\ \frac{6}{5}\int_0^1\int_0^1 (x+y^2)dxdy = 1 & 1\le x \wedge 1\le y \end{cases}$$

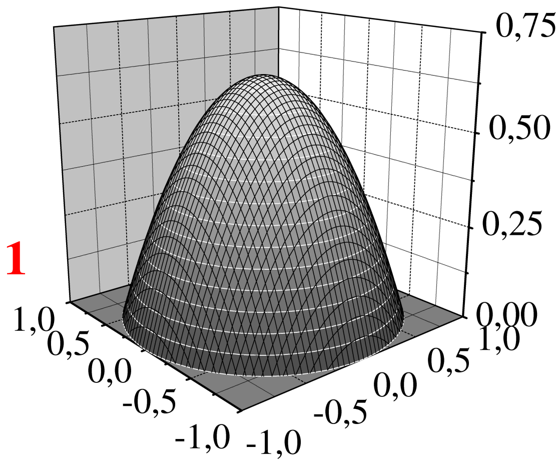


Uwaga: Znając dystrybuantę można wyznaczyć łączną funkcję gęstości p-twa:

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$$

Rozkłady brzegowe i warunkowe

<u>Przykład:</u> Rozważmy funkcję gęstości p-twa $f(x,y) \propto R^2 - x^2 - y^2$ określoną na obszarze $x^2+y^2 \leq R^2$.



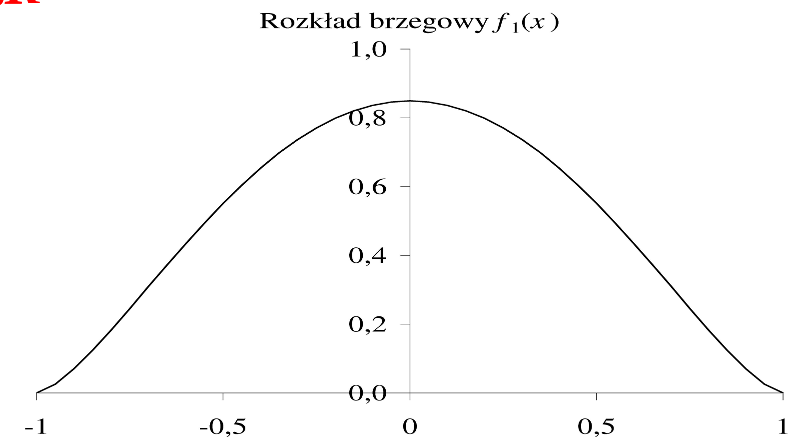
Normalizacja: $\boxed{A = \frac{2}{\pi R^4}}$

$$A \iint\limits_{x^2+y^2 \leq R^2} (R^2 - x^2 - y^2)\,dx\,dy = A\int_0^{2\pi} d\varphi \int_0^R (R^2 - r^2) r\,dr = \frac{1}{2}\pi A R^4 = 1$$

Rozkład brzegowy zmiennej X:

$$f_1(x) = \frac{2}{\pi R^4} \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} (R^2 - x^2 - y^2)\,dy = \frac{8}{3\pi R^4}(R^2 - x^2)^{\frac{3}{2}}$$

Rozkład warunkowy zmiennej Y:



$$h(y\,|\,x) = \frac{f(x,y)}{f_1(x)} = \frac{3}{4}\frac{R^2 - x^2 - y^2}{(R^2 - x^2)^{3/2}}$$