Rachunek prawdopodobieństwa i statystyka

Wykład 6

Warunkowe wartości oczekiwane

Niech będzie dany rozkład p-twa $\mathrm{P_{km}}$ dyskretnych zmiennych losowych k i m lub gęstość p-twa $f\,(x,y)$ ciągłych zmiennych losowych x i y.

Definicja: Wartością oczekiwaną dyskretnej zmiennej losowej k pod warunkiem, że zmienna losowa m przyjmuje wartość m nazywamy wielkość:

$$\mathcal{E}\left[\mathrm{k}\,|\,\mathrm{m}=m\right]\equiv\mu_{\mathrm{k}}\,(m)\equiv\sum_{k}k\,\mathrm{P}\left(\mathrm{k}=k\,|\,\mathrm{m}=m\right)=\frac{1}{\mathrm{P}_{\bullet,\mathrm{m}}}\sum_{k}k\,\mathrm{P}_{\mathrm{km}}$$

Definicja: Wartością oczekiwaną ciągłej zmiennej losowej x pod warunkiem, że zmienna losowa y przyjmuje wartość y nazywamy wielkość:

$$\mathcal{E}\left[\mathrm{x}\,|\,\mathrm{y}=y\right]\equiv\mu_{\mathrm{x}}\left(y\right)\equiv\int_{-\infty}^{+\infty}x\cdot f\left(x\,|\,y\right)dx=\frac{1}{f_2\left(y\right)}\int_{-\infty}^{+\infty}x\cdot f\left(x,y\right)dx$$

Definicja: Podobnie określamy wartość oczekiwaną dyskretnej zm. losowej m pod warunkiem, że zmienna losowa k przyjmuje wartość k:

$$\mathcal{E}\left[\mathrm{m}\,|\,\mathrm{k}=k\right]\equiv\mu_{\mathrm{m}}\,(k)\equiv\sum_{m}m\,\mathrm{P}\left(\mathrm{m}=m\,|\,\mathrm{k}=k\right)=\frac{1}{\mathrm{P}_{k,\bullet}}\sum_{m}m\,\mathrm{P}_{\mathrm{km}}$$

oraz wartość oczekiwaną ciągłej zmiennej losowej y pod warunkiem, że zmienna losowa x przyjmuje wartość x:

$$\mathcal{E}\left[\mathrm{y}\,|\,\mathrm{x}=x\right]\equiv\mu_{\mathrm{y}}\left(x\right)\equiv\int_{-\infty}^{+\infty}y\cdot f\left(y\,|\,x\right)dy=\frac{1}{f_1\left(x\right)}\int_{-\infty}^{+\infty}y\cdot f\left(x,y\right)dy$$

Warunkowe wartości oczekiwane

Twierdzenie: Niech x, y, y_1 i y_2 będą zmiennymi losowymi, $g()$ funkcją oraz c stałą.

Zachodzą następujące związki:

$$\mathcal{E}[c \mid \mathrm{x} = x] = c$$

$$\mathcal{E}[\mathrm{y}_1 + \mathrm{y}_2 \mid \mathrm{x} = x] = \mathcal{E}[\mathrm{y}_1 \mid \mathrm{x} = x] + \mathcal{E}[\mathrm{y}_2 \mid \mathrm{x} = x]$$

$$\mathcal{E}[c\,\mathrm{y} \mid \mathrm{x} = x] = c \cdot \mathcal{E}[\mathrm{y} \mid \mathrm{x} = x]$$

$$\mathcal{E}[g(\mathrm{x},\mathrm{y}) \mid \mathrm{x} = x] = \mathcal{E}[g(x,\mathrm{y}) \mid \mathrm{x} = x]$$

$$\mathcal{E}[g(\mathrm{x}) \cdot \mathrm{y} \mid \mathrm{x} = x] = g(x) \cdot \mathcal{E}[\mathrm{y} \mid \mathrm{x} = x]$$

$$\mathcal{E}[\mathrm{y} \mid \mathrm{x} = x] = \mathcal{E}[\mathrm{y}] \quad \text{jeśli zmienne losowe x i y są niezależne}$$

Twierdzenie: Niech $\mathcal{E}|\mathrm{y}| < \infty$, wówczas $\mathcal{E}[\mathcal{E}[\mathrm{y} \mid \mathrm{x} = x]] = \mathcal{E}[\mathrm{y}]$

Dowód:

$$\mathcal{E}[\mathcal{E}[\mathrm{y} \mid \mathrm{x}]] = \int_{-\infty}^{+\infty} \mathcal{E}[\mathrm{y} \mid \mathrm{x} = x]\, f_{\mathrm{x}}(x)\, dx = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} y \cdot f(y \mid x)\, dy \right) f_{\mathrm{x}}(x)\, dx =$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y \cdot f(x,y)\, dy dx = \int_{-\infty}^{+\infty} y \cdot f_{\mathrm{y}}(y)\, dy dx = \mathcal{E}[\mathrm{y}]$$

Warunkowa wariancja

Definicja: Warunkową wariancją zmiennej losowej y pod warunkiem, że zmienna losowa x przyjmuje wartość x nazywamy wielkość:

$$\mathcal{V}[\mathrm{y}\,|\,\mathrm{x}=x]=\mathcal{E}\left[\left(\mathrm{y}-\mathcal{E}[\mathrm{y}\,|\,x=x]\right)^2\,\middle|\,\mathrm{x}=x\right]$$

Twierdzenie: Niech x i y będą zmiennymi losowymi, a $g()$ funkcją, oraz niech $\mathcal{E}\left[\mathrm{y}^2\right]<\infty$ i $\mathcal{E}\left[g(\mathrm{x})^2\right]<\infty$, wówczas

$$\mathcal{E}\left[\left(\mathrm{y}-g(\mathrm{x})\right)^2\right]=\mathcal{E}\left[\mathcal{V}[\mathrm{y}\,|\,\mathrm{x}]\right]+\mathcal{E}\left[\left(\mathcal{E}[\mathrm{y}\,|\,\mathrm{x}]-g(\mathrm{x})\right)^2\right]$$

Dowód:

$$\mathcal{E}\left[\left(\mathrm{y}-g(\mathrm{x})\right)^2\right]=\mathcal{E}\left[\left(\mathrm{y}-\mathcal{E}[\mathrm{y}\,|\,\mathrm{x}]+\mathcal{E}[\mathrm{y}\,|\,\mathrm{x}]-g(\mathrm{x})\right)^2\right]=$$

$$=\mathcal{E}\left[\left(\mathrm{y}-\mathcal{E}[\mathrm{y}\,|\,\mathrm{x}]\right)^2\right]+2\mathcal{E}\left[\left(\mathrm{y}-\mathcal{E}[\mathrm{y}\,|\,\mathrm{x}]\right)\left(\mathcal{E}[\mathrm{y}\,|\,\mathrm{x}]-g(\mathrm{x})\right)\right]+\mathcal{E}\left[\left(\mathcal{E}[\mathrm{y}\,|\,\mathrm{x}]-g(\mathrm{x})\right)^2\right]=$$

$$=\mathcal{E}\left[\mathcal{E}\left[\left(\mathrm{y}-\mathcal{E}[\mathrm{y}\,|\,\mathrm{x}]\right)^2\middle|\,\mathrm{x}\right]\right]+2\mathcal{E}\left[\mathcal{E}\left[\left(\mathrm{y}-\mathcal{E}[\mathrm{y}\,|\,\mathrm{x}]\right)\left(\mathcal{E}[\mathrm{y}\,|\,\mathrm{x}]-g(\mathrm{x})\right)\middle|\,\mathrm{x}\right]\right]+\mathcal{E}\left[\left(\mathcal{E}[\mathrm{y}\,|\,\mathrm{x}]-g(\mathrm{x})\right)^2\right]=$$

$$=\mathcal{E}\left[\mathcal{V}[\mathrm{y}|\mathrm{x}]\right]+2\mathcal{E}\left[\left(\mathcal{E}[\mathrm{y}\,|\,\mathrm{x}]-g(\mathrm{x})\right)\underbrace{\mathcal{E}\left[\left(\mathrm{y}-\mathcal{E}[\mathrm{y}\,|\,\mathrm{x}]\right)\,|\,\mathrm{x}\right]}_{=0}\right]+\mathcal{E}\left[\left(\mathcal{E}[\mathrm{y}\,|\,\mathrm{x}]-g(\mathrm{x})\right)^2\right]=$$

$$=\mathcal{E}\left[\mathcal{V}[\mathrm{y}\,|\,\mathrm{x}]\right]+\mathcal{E}\left[\left(\mathcal{E}[\mathrm{y}\,|\,\mathrm{x}]-g(\mathrm{x})\right)^2\right]$$

W szczególności dla $g(\mathrm{x})=\mathcal{E}[\mathrm{y}]$ mamy: $\mathcal{V}[\mathrm{y}]=\mathcal{E}\left[\mathcal{V}[\mathrm{y}\,|\,\mathrm{x}]\right]+\mathcal{V}\left[\mathcal{E}[\mathrm{y}\,|\,\mathrm{x}]\right]$

Linie regresji I-go rodzaju

Linią regresji I-go rodzaju zmiennej losowej $\mathbf{k}$ $(\mathbf{x})$ względem zmiennej losowej $\mathbf{m}$ $(\mathbf{y})$ nazywamy zbiór punktów o współrzędnych (k,m) w przypadku zmiennych dyskretnych i (x,y) w przypadku zmiennych ciągłych, spełniających równania:

$$k = \mu_k(m) \qquad x = \mu_x(y)$$

Analogicznie, **linią regresji I-go rodzaju** zmiennej losowej $\mathbf{m}$ $(\mathbf{y})$ względem zmiennej $\mathbf{k}$ $(\mathbf{x})$ nazywamy, odpowiednio dla zmiennych dyskretnych i ciągłych, zbiory punktów (k,m) lub (x,y), spełniające równania: $m = \mu_m(k)$ $\qquad y = \mu_y(x)$

Własności:

Średnie odchylenie kwadratowe zmiennej losowej $\mathbf{x}$ od pewnej funkcji $g(y)$ jest najmniejsze, gdy ta funkcja z p-twem 1 jest równa $\mu_x(\mathbf{y})$:

$$\mathcal{E}\left[\left(\mathbf{x}-g(y)\right)^2\right] = \mathcal{E}\left[\mathcal{V}\left[\mathbf{x}|\mathbf{y}\right]\right] + \mathcal{E}\left[\left(\mu_x(y)-g(y)\right)^2\right] \geq \mathcal{E}\left[\mathcal{V}\left[\mathbf{x}|\mathbf{y}\right]\right] = \mathcal{E}\left[\left(\mathbf{x}-\mu_x(y)\right)^2\right]$$

Średnie odchylenie kwadratowe zmiennej losowej $\mathbf{y}$ od pewnej funkcji $f(x)$ jest najmniejsze, gdy ta funkcja z p-twem 1 jest równa $\mu_y(\mathbf{x})$:

$$\mathcal{E}\left[\left(\mathbf{y}-g(x)\right)^2\right] = \mathcal{E}\left[\mathcal{V}\left[\mathbf{y}|\mathbf{x}\right]\right] + \mathcal{E}\left[\left(\mu_y(x)-g(x)\right)^2\right] \geq \mathcal{E}\left[\mathcal{V}\left[\mathbf{y}|\mathbf{x}\right]\right] = \mathcal{E}\left[\left(\mathbf{y}-\mu_y(x)\right)^2\right]$$

Linie regresji II-go rodzaju

Linią regresji II-go rodzaju nazywamy zadaną krzywą $y = h(x; a,b,...)$ gdy spełnia ona warunek:

$$\mathcal{E}\left[\left(y - h(x;a,b,...)\right)^2\right] = \min(a,b,...)$$

W szczególności prostą regresji II-go rodzaju nazywamy linię prostą $y = ax+b$ spełniającą warunek:

$$\mathcal{E}\left[\left(y - a\,x - b\right)^2\right] = \mathcal{E}\left[\left(\left(y-\mu_y\right) - a\left(x-\mu_x\right) + \mu_y - a\mu_x - b\right)^2\right] =$$

$$= \sigma_y^2 + a^2\sigma_x^2 - 2a\,\mathrm{cov}[x,y] + \left(\mu_y - a\mu_x - b\right)^2 = \sigma_y^2 + a^2\sigma_x^2 - 2a\sigma_x\sigma_y\rho + \left(\mu_y - a\mu_x - b\right)^2 = \min(a,b)$$

Szukamy minimum ze względu na parametry a i b:

$$\begin{cases} \frac{\partial}{\partial a}\mathcal{E}\left[\left(y - a\,x - b\right)^2\right] = 2a\sigma_x^2 - 2\sigma_x\sigma_y\rho - 2\mu_x\left(\mu_y - a\mu_x - b\right) = 0 \\ \frac{\partial}{\partial b}\mathcal{E}\left[\left(y - a\,x - b\right)^2\right] = -2\left(\mu_y - a\mu_x - b\right) = 0 \end{cases} \Rightarrow \begin{cases} a = \rho\frac{\sigma_y}{\sigma_x} \\ b = \mu_y - a\mu_x \end{cases}$$

Prosta regresji II-go rodzaju zmiennej y względem zmiennej x ma postać:

$$y = \rho\frac{\sigma_y}{\sigma_x}x - \rho\frac{\sigma_y}{\sigma_x}\mu_x + \mu_y$$

Podobnie znajdujemy prostą regresji II-go rodzaju zmiennej x względem y:

$$x = \rho\frac{\sigma_x}{\sigma_y}y - \rho\frac{\sigma_x}{\sigma_y}\mu_y + \mu_x \qquad \Rightarrow \qquad y = \frac{1}{\rho}\frac{\sigma_y}{\sigma_x}x - \frac{1}{\rho}\frac{\sigma_y}{\sigma_x}\mu_x + \mu_y$$

Linie regresji I-go rodzaju

Przykład: Linie regresji I-go rodzaju dla dwuwymiarowego rozkładu normalnego.

$$\mathcal{N}\left(x,y;\mu_x,\mu_y,\sigma_x,\sigma_y\right)=$$

$$=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2\left(1-\rho^2\right)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2+\left(\frac{y-\mu_y}{\sigma_y}\right)^2-2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right]\right\}$$

Rozkłady brzegowe to jednowymiarowe rozkłady normalne:

$$\mathcal{N}_1\left(x;\mu_x,\sigma_x\right)=\frac{1}{\sqrt{2\pi}\sigma_x}\exp\left(-\frac{\left(x-\mu_x\right)^2}{2\sigma_x^2}\right)\qquad \mathcal{E}[x]=\mu_x\qquad \mathcal{V}[x]=\sigma_x^2$$

$$\mathcal{N}_2\left(y;\mu_y,\sigma_y\right)=\frac{1}{\sqrt{2\pi}\sigma_y}\exp\left(-\frac{\left(y-\mu_y\right)^2}{2\sigma_y^2}\right)\qquad \mathcal{E}[y]=\mu_y\qquad \mathcal{V}[y]=\sigma_y^2$$

Rozkłady warunkowe i warunkowe wartości oczekiwane:

$$g\left(y\mid x\right)=\frac{\mathcal{N}\left(x,y;\mu_x,\mu_y,\sigma_x,\sigma_y\right)}{\mathcal{N}_1\left(x;\mu_x,\sigma_x\right)}=\frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2\sigma_y^2\left(1-\rho^2\right)}\left[y-\left(\mu_y+\rho\frac{\sigma_y}{\sigma_x}\left(x-\mu_x\right)\right)\right]^2\right)$$

$$g\left(x\mid y\right)=\frac{\mathcal{N}\left(x,y;\mu_x,\mu_y,\sigma_x,\sigma_y\right)}{\mathcal{N}_2\left(y;\mu_y,\sigma_y\right)}=\frac{1}{\sqrt{2\pi}\sigma_x\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2\sigma_x^2\left(1-\rho^2\right)}\left[x-\left(\mu_x+\rho\frac{\sigma_x}{\sigma_y}\left(y-\mu_y\right)\right)\right]^2\right)$$

Linie regresji I-go rodzaju

Linie regresji I-go rodzaju to proste o równaniach:

$$y = \mathcal{E}[y \mid x] = \mu_y + \rho \frac{\sigma_y}{\sigma_x}(x - \mu_x)$$

$$x = \mathcal{E}[x \mid y] = \mu_x + \rho \frac{\sigma_x}{\sigma_y}(y - \mu_y) \qquad \Rightarrow \qquad y = \mu_y + \frac{1}{\rho}\frac{\sigma_y}{\sigma_x}(x - \mu_x)$$

Uwaga: W przypadku dwuwymiarowego rozkładu normalnego, linie regresji I-go rodzaju są jednocześnie prostymi regresji II-go rodzaju.

Linie regresji I-go rodzaju

Przykład: Chcemy umieć przewidzieć wzrost pewnego gatunku żyta na podstawie koncentracji fosforu w glebie. W tym celu badany cztery koncentracje 2, 4, 8, 16 ppm fosforu, i obserwujemy wzrost roślin przy każdej z nich począwszy od nasiona aż do momentu kwitnienia. Następnie dokonujemy pomiaru suchej masy każdej z roślin. Wyniki pomiarów przedstawia tabelka.

Roślina	Fosfor [ppm]	Masa [g]
1	2	4.1
2	2	3.8
3	2	4.0
4	2	3.9
5	4	5.2
6	4	4.9
7	4	5.0
8	4	4.8
9	8	5.7
10	8	5.9
11	8	6.0
12	8	6.2
13	16	11.7
14	16	8.9
15	16	10.1
16	16	10.3

Zmienne losowe: x - koncentrację fosforu, y - masa roślin.

$$\mu_x \equiv \mathcal{E}[x] = \frac{1}{16}(4\times 2 + 4\times 4 + 4\times 8 + 4\times 16) = 7.5$$

$$\mu_y \equiv \mathcal{E}[y] = \frac{1}{16}(4.1 + 3.8 + 4.0 + 3.9 + 5.2 + 4.9 + 5.0 + 4.8 + 5.7$$

$$+5.9 + 6.0 + 6.2 + 11.7 + 8.9 + 10.1 + 10.3) = \frac{100.5}{16} \cong 6.28$$

$$\mathcal{E}[x^2] = \frac{1360}{16} = 85 \qquad \mathcal{E}[y^2] = \frac{727.49}{16} = 45.468 \qquad \mathcal{E}[xy] = \frac{957.6}{16} = 59.85$$

$$\mathrm{cov}[x,y] = \mathcal{E}[xy] - \mathcal{E}[x]\mathcal{E}[y] = 59.85 - 7.5\times 6.28 = 12.75$$

$$\mathcal{D}[x] = \sqrt{\mathcal{E}[x^2] - (\mathcal{E}[x])^2} = \sqrt{85 - 7.5^2} = 5.36$$

$$\mathcal{D}[y] = \sqrt{\mathcal{E}[y^2] - (\mathcal{E}[y])^2} = \sqrt{45.468 - 6.28^2} = 2.46$$

$$\rho = \frac{\mathrm{cov}[x,y]}{\mathcal{D}[x]\mathcal{D}[y]} = 0.967$$

$$y = 0.44x + 2.96$$

Randomizacja – rozkład dyskretny

Przykład: Zmienna losowa $\mathbf{X}$ będąca liczbą jaj znoszonych przez kurę ma rozkład Poissona. P-two wyklucia się pisklęcia z każdego jaja wynosi p i jest niezależne od p-twa wyklucia z innych jaj. Jaki jest rozkład zmiennej losowej $\mathbf{Y}$ będącej liczbą wyklutych piskląt.

$$\mathbf{P}(\mathbf{X}=n)=\mathcal{P}_n(\lambda)=\frac{\lambda^n}{n!}e^{-\lambda} \qquad \text{gdzie} \quad n=0,1,2,\ldots$$

$$\mathbf{P}(\mathbf{Y}=j\,|\,\mathbf{X}=n)=\mathcal{B}_j(n,p)=\binom{n}{j}p^j(1-p)^{n-j} \qquad \text{gdzie} \quad j=0,1,\ldots,n$$

$$\mathbf{P}(\mathbf{Y}=j,\mathbf{X}=n)=\mathbf{P}(\mathbf{Y}=j\,|\,\mathbf{X}=n)\mathbf{P}(\mathbf{X}=n)$$

$$\mathbf{P}(\mathbf{Y}=j)=\sum_{n=j}^{\infty}\mathbf{P}(\mathbf{X}=n,\mathbf{Y}=j)=\sum_{n=j}^{\infty}\frac{\lambda^n}{n!}e^{-\lambda}\frac{n!}{j!(n-j)!}p^j(1-p)^{n-j}=$$

$$=\frac{(\lambda p)^j}{j!}e^{-\lambda}\sum_{n=j}^{\infty}\frac{[\lambda(1-p)]^{n-j}}{(n-j)!}=\frac{(\lambda p)^j}{j!}e^{-\lambda}\sum_{k=0}^{\infty}\frac{[\lambda(1-p)]^k}{k!}=$$

$$=\frac{(\lambda p)^j}{j!}e^{-\lambda}e^{\lambda(1-p)}=\frac{(\lambda p)^j}{j!}e^{-\lambda p}=\mathcal{P}_j(\lambda p)$$

Randomizacja – rozkład ciągły

Przykład: Losujemy punkt $\mathbf{X}$ z rozkładu jednostajnego na przedziale (a,b). Następnie losujemy punkt $\mathbf{Y}$ z rozkładu jednostajnego na przedziale $(\mathbf{X}$,b). Znajdź rozkład zmiennej $\mathbf{Y}$.

$$f_1(x)=\begin{cases}\dfrac{1}{b-a} & \text{dla} \quad a\le x\le b\\ 0 & \text{dla pozostalych } x\end{cases}$$

$$g_{21}(y\mid x)=\begin{cases}\dfrac{1}{b-x} & \text{dla} \quad x\le y\le b\\ 0 & \text{dla pozostalych y}\end{cases}$$

Znajdujemy łączny rozkład p-twa:

$$f(x,y)=g_{21}(y\mid x)\,f_1(x)=\begin{cases}\dfrac{1}{b-a}\cdot\dfrac{1}{b-x} & \text{dla} \quad a\le x\le y\le b\\ 0 & \text{dla pozostalych } x,y\end{cases}$$

Oraz interesujący nas rozkład brzegowy:

$$f_2(y)=\int_a^y \frac{1}{b-a}\cdot\frac{1}{b-x}dx=\frac{1}{b-a}\ln\frac{b-a}{b-y}$$

Ogólnie proces randomizacji dla rozkładów odpowiednio dyskretnych i ciągłych przebiega według schematów:

$$\mathbf{P}(\mathbf{Y}=j)=\sum_n \mathbf{P}(\mathbf{Y}=j\mid\mathbf{X}=n)\mathbf{P}(\mathbf{X}=n) \qquad f(y)=\int_{-\infty}^{\infty} g(y\mid x)h(x)dx$$

Randomizacja – rozkład mieszany

Przykład: Załóżmy, ze zmienna losowa Y przyjmuje $n>1$ wartości całkowitych z p-twami:

$$\mathrm{P}(\mathrm{Y}=i)=p_i \quad \text{przy czym} \quad p_1+p_2+\cdots+p_n=1$$

Załóżmy następnie, że dla ustalonej wartości $\mathrm{Y}=i$ zmienna losowa X ma rozkład ciągły o gęstości $\varphi_i(x)$.

P-two że zmienna X przyjmuje wartości z przedziału (a,b) natomiast zmienna $Y=i$ wynosi:

$$\mathrm{P}(a\le \mathrm{X}\le b, \mathrm{Y}=i)=p_i\int_a^b \varphi_i(x)\,dx$$

Korzystając z twierdzenia o p-twie całkowitym znajdujemy rozkład brzegowy zmiennej X:

$$\mathrm{P}(a\le \mathrm{X}\le b)=\sum_{i=1}^{n}\mathrm{P}(a\le \mathrm{X}\le b\,|\,\mathrm{Y}=i)\,P(\mathrm{Y}=i)=$$

$$=\sum_{i=1}^{n}p_i\int_a^b\varphi_i(x)\,dx=\int_a^b\sum_{i=1}^{n}p_i\varphi_i(x)\,dx$$

A więc zmienna losowa X jest zmienną ciągłą o funkcji gęstości danej przez:

$$f_X(x)=\sum_{i=1}^{n}p_i\varphi_i(x) \qquad \text{(mieszanina rozkladów)}$$