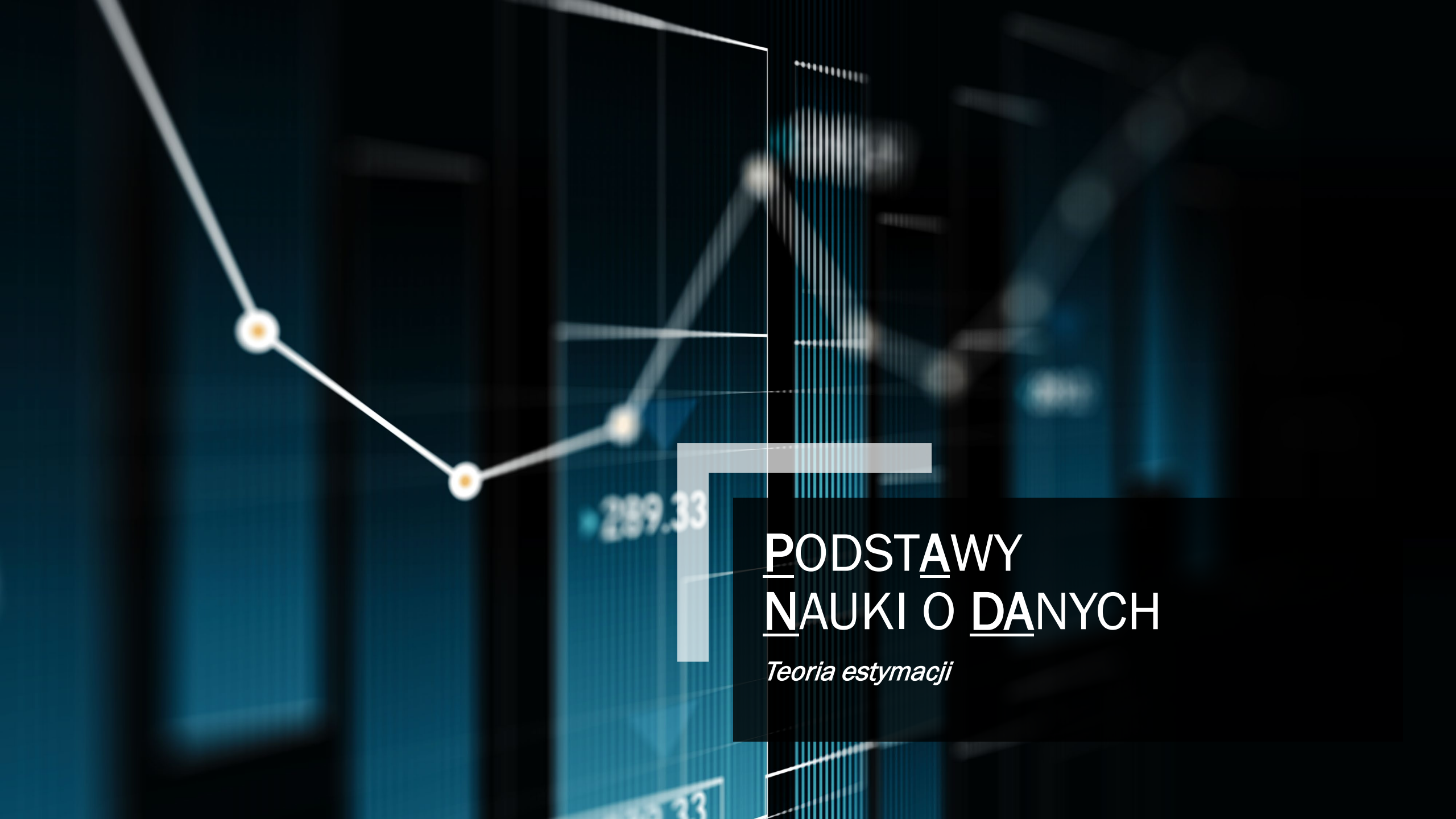
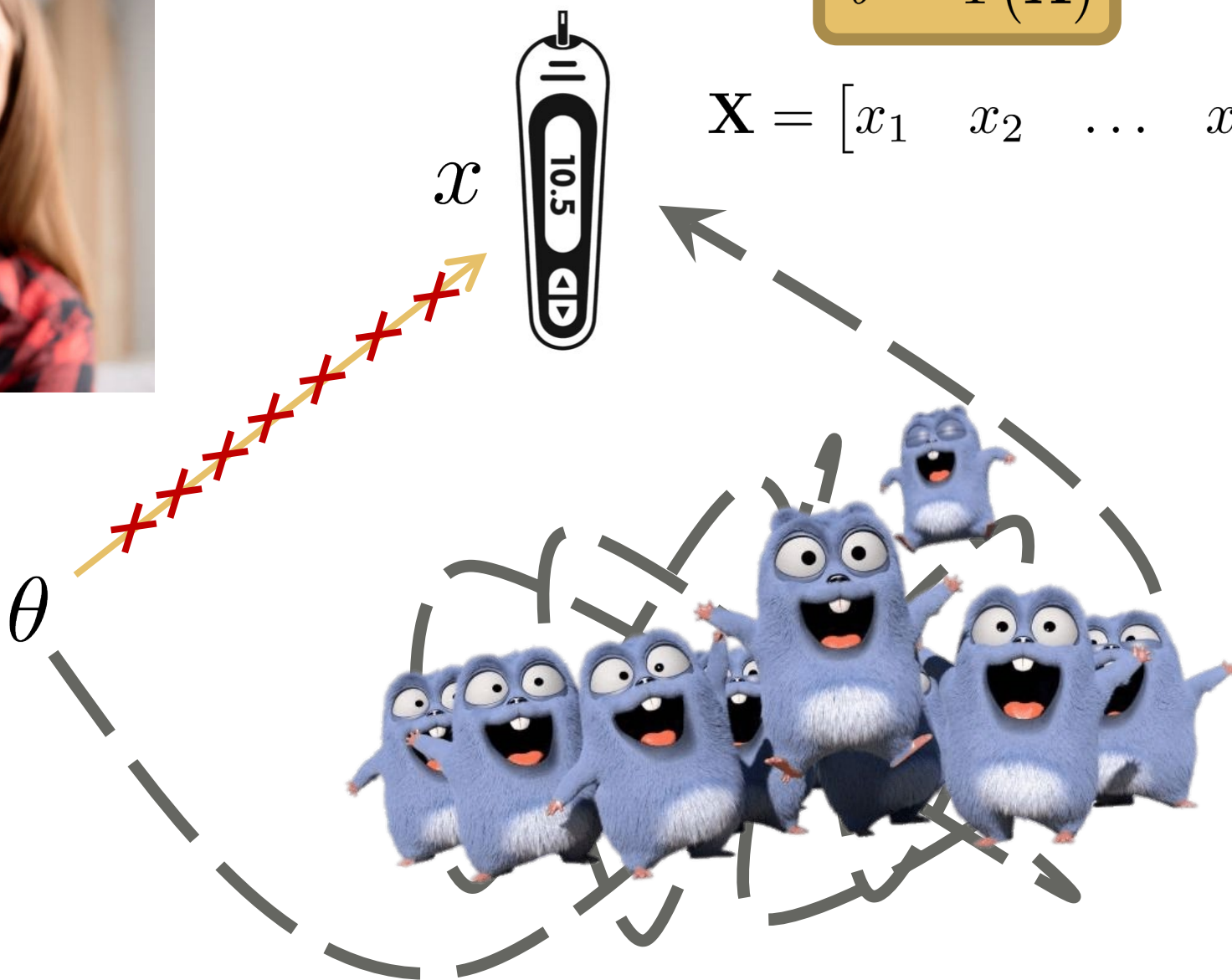
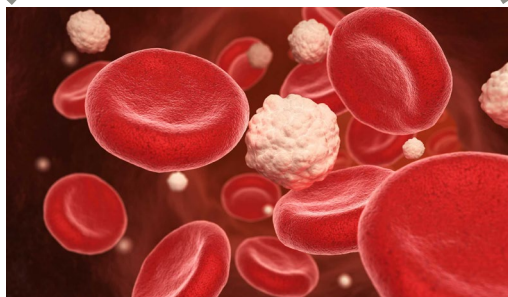
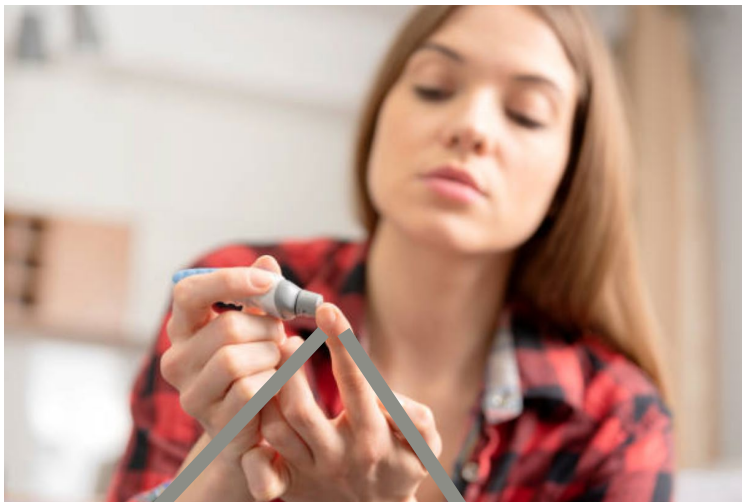


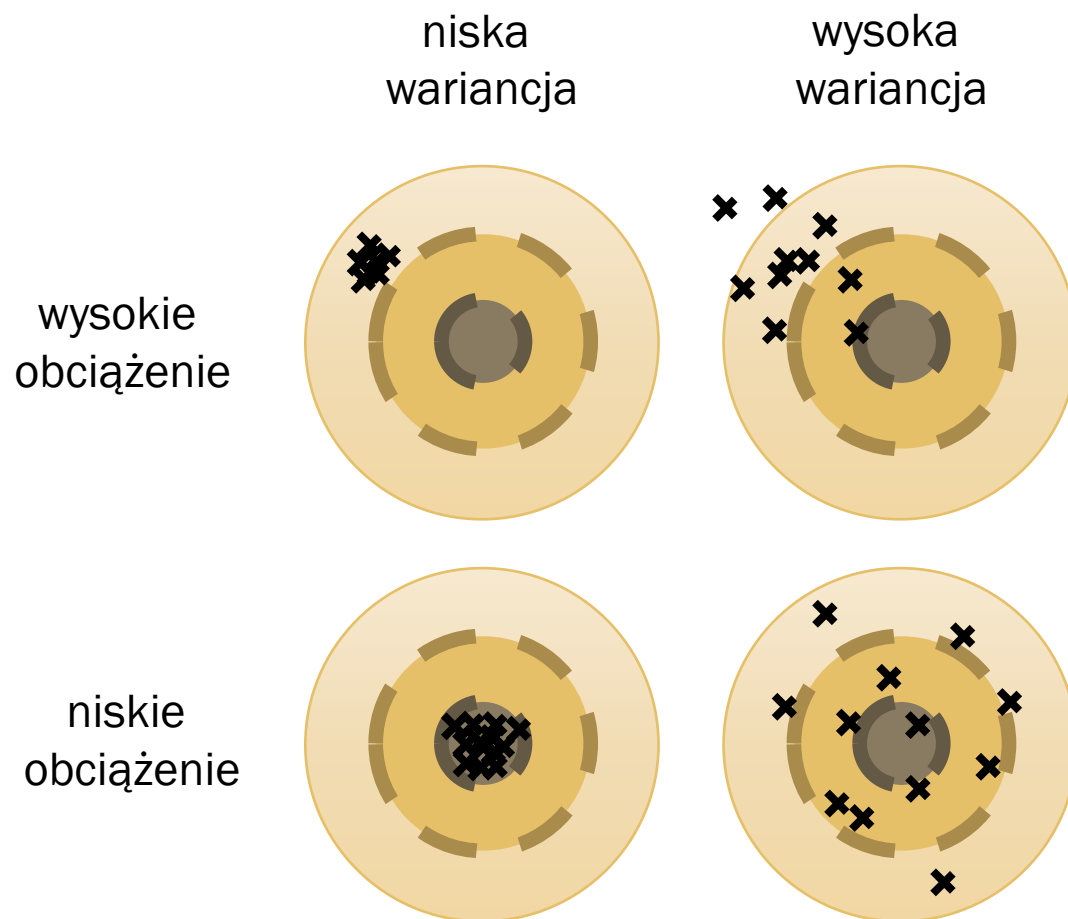


PODSTAWY NAUKI O DANYCH

Teoria estymacji



Teoria estymacji



Obciążenie (ang. *bias*)

- błąd systematyczny

Wariancja (ang. *variance*)

- błąd przypadkowy

Teoria estymacji

$$\mathbb{E}[\spadesuit + \heartsuit] = \mathbb{E}[\spadesuit] + \mathbb{E}[\heartsuit]$$

$$\mathbb{E}[\alpha \spadesuit] = \alpha \mathbb{E}[\spadesuit], \quad \alpha \in \mathbb{R}$$

$$\mathbb{E}[\text{const.}] = \text{const.}$$

Dla niezależnych $\spadesuit, \heartsuit$ zachodzi

$$\mathbb{E}[\spadesuit \cdot \heartsuit] = \mathbb{E}[\spadesuit] \cdot \mathbb{E}[\heartsuit]$$

$$\text{Var}[\spadesuit + \text{const.}] = \text{Var}[\spadesuit]$$

$$\text{Var}[\alpha \spadesuit] = \alpha^2 \text{Var}[\spadesuit], \quad \alpha \in \mathbb{R}$$

$$\text{Var}[\text{const.}] = 0$$

Dla niezależnych $\spadesuit, \heartsuit$ zachodzi

$$\text{Var}[\spadesuit + \heartsuit] = \text{Var}[\spadesuit] + \text{Var}[\heartsuit]$$

$$\hat{\theta} = \Psi \left(\{x_i\}_{i=1}^N \right)$$

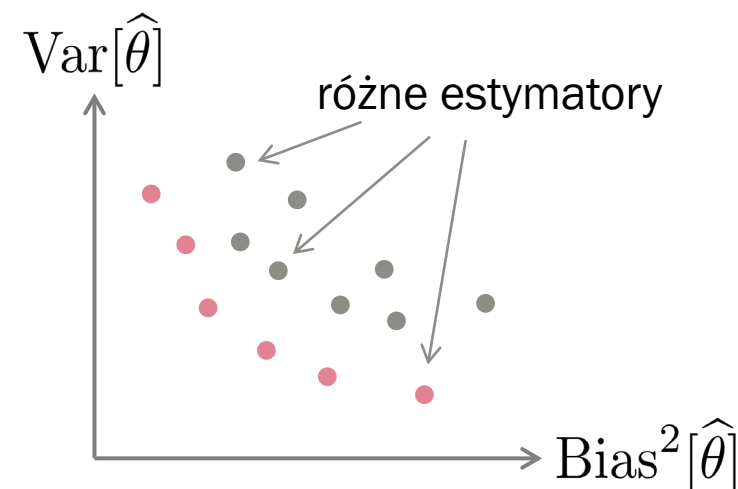
Ψ – estymator parametru

$$\text{Bias}[\hat{\theta}] \equiv \theta - \mathbb{E}[\hat{\theta}]$$

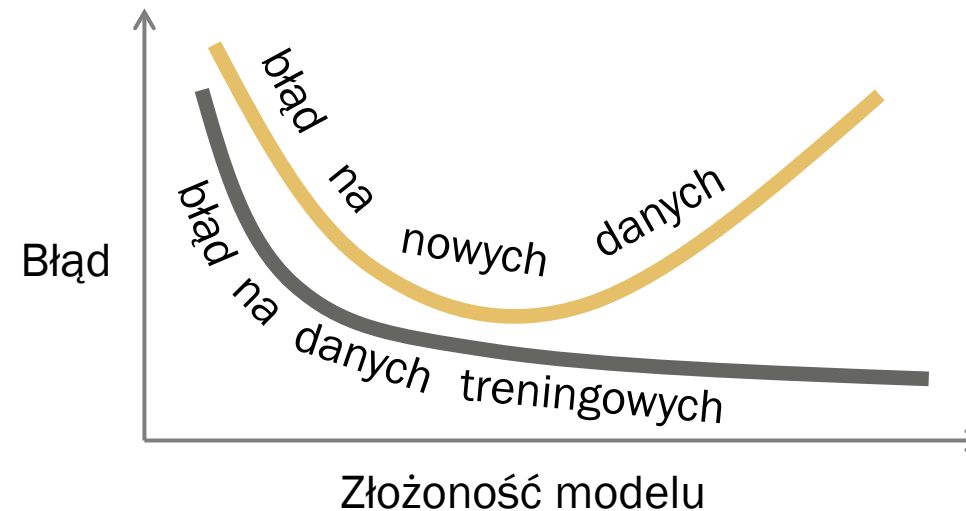
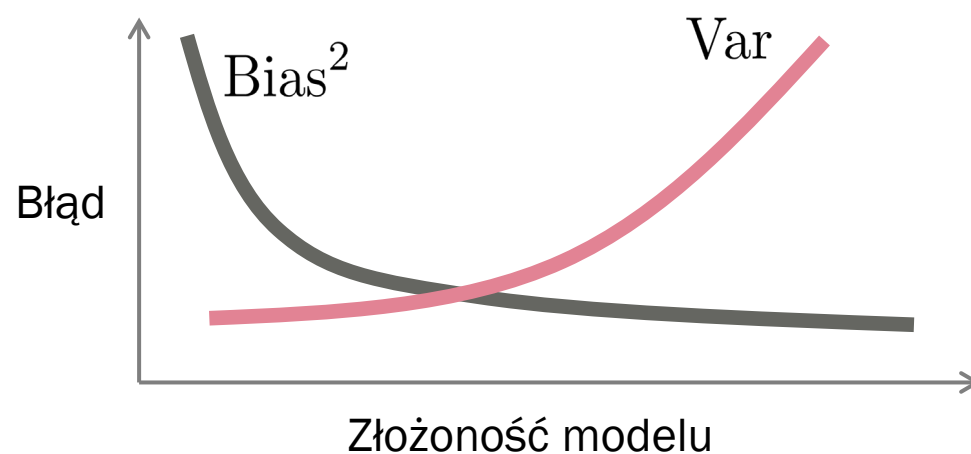
$$\begin{aligned} \text{Var}[\hat{\theta}] &\equiv \mathbb{E} \left[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2 \right] \\ &= \mathbb{E} \left[\hat{\theta}^2 \right] - \mathbb{E}^2[\hat{\theta}] \end{aligned}$$

Kompromis obciążenie–wariancja

$$\begin{aligned}\text{MSE}[\hat{\theta}] &= \mathbb{E} \left[(\theta - \hat{\theta})^2 \right] = \\ &= \mathbb{E} \left[\theta^2 - 2\theta\hat{\theta} + \hat{\theta}^2 \right] = \\ &= \mathbb{E} \left[\theta^2 \right] - \mathbb{E} \left[2\theta\hat{\theta} \right] + \mathbb{E} \left[\hat{\theta}^2 \right] = \\ &= \theta^2 - 2\theta\mathbb{E} \left[\hat{\theta} \right] + \mathbb{E} \left[\hat{\theta}^2 \right] = \\ &= \theta^2 - 2\theta\mathbb{E} \left[\hat{\theta} \right] + \mathbb{E} \left[\hat{\theta}^2 \right] + \mathbb{E}^2 \left[\hat{\theta} \right] - \mathbb{E}^2 \left[\hat{\theta} \right] = \\ &= \theta^2 - 2\theta\mathbb{E} \left[\hat{\theta} \right] + \mathbb{E}^2 \left[\hat{\theta} \right] + \mathbb{E} \left[\hat{\theta}^2 \right] - \mathbb{E}^2 \left[\hat{\theta} \right] = \\ &= (\theta - \mathbb{E} \left[\hat{\theta} \right])^2 + (\mathbb{E} \left[\hat{\theta}^2 \right] - \mathbb{E}^2 \left[\hat{\theta} \right]) = \\ &= \text{Bias}^2[\hat{\theta}] + \text{Var}[\hat{\theta}]\end{aligned}$$



Kompromis obciążenie–wariancja



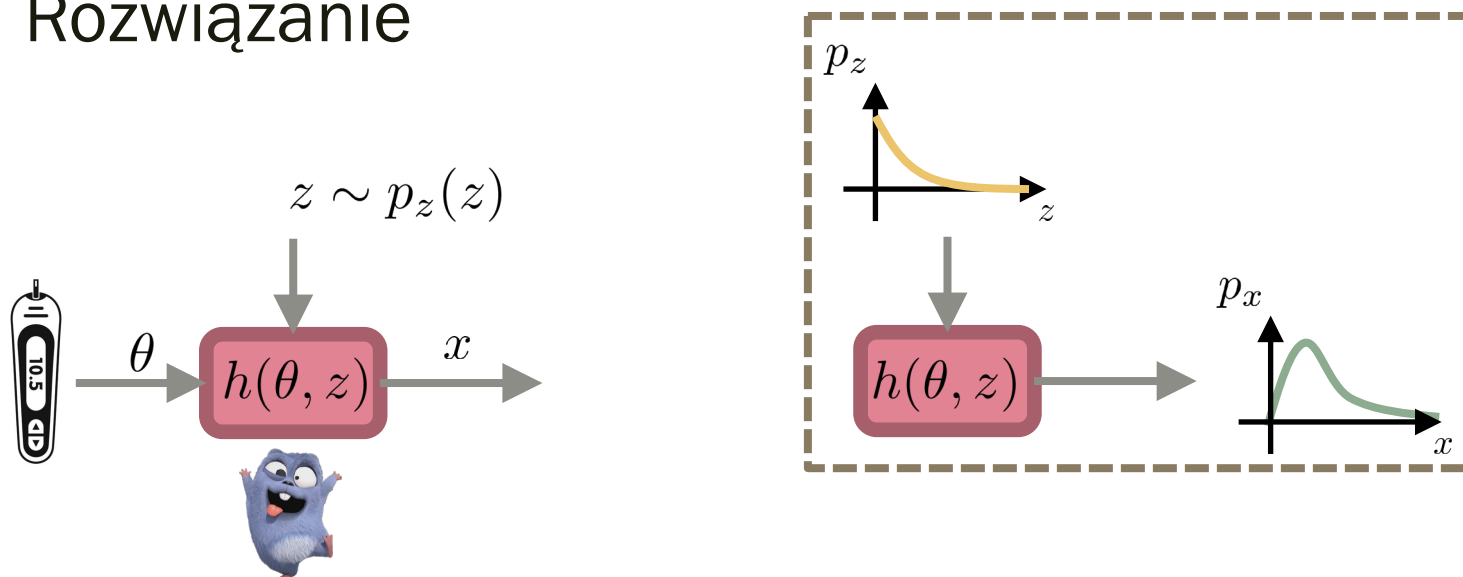
ESTYMACJA MIERZONEJ WIELKOŚCI

Metoda maksymalnej wiarygodności



Estymacja mierzonej wielkości

Rozwiązanie



$$p_x(x) = p_z(z) |\det J|$$

$$J = \frac{\partial}{\partial x} h^{-1}$$

$$L(\theta) = \prod_{i=1}^N p_x(x_i)$$

$$L(\theta) = \prod_{i=1}^N p_z(z_i) |\det J| = \prod_{i=1}^N p_z(h^{-1}(\theta, x_i)) |\det J|$$

Przykład: $x = \theta \cdot z = h(\theta, z)$

$$z = \frac{x}{\theta} = h^{-1}(\theta, x) \Rightarrow \det J = \frac{1}{\theta}$$

$$\hat{\theta}_{\text{ML}} = \Psi_{\text{ML}}(\{x_1, x_2, \dots, x_N\})$$

Przykład zadania

- Zakłócenia wymnażają się przez mierzoną wielkość i mają rozkład jednostajny na przedziale $\langle 0, 1 \rangle$

$$\text{Dane} = \left\{ \mathbf{x}_i \right\}_{i=1}^N$$

$$x = \theta \cdot z \qquad p_z(z) = \begin{cases} 1 & \text{dla } z \in \langle 0, 1 \rangle \\ 0 & \text{w przec. przyp.} \end{cases}$$

- Należy wyznaczyć wzór analityczny na wartość mierzonej wielkości tak, by maksymalizować funkcję wiarygodności

Przykład zadania

- Zakłócenia dodają się do mierzonej wielkości i mają rozkład normalny o znanych parametrach

$$\text{Dane} = \left\{ \mathbf{x}_i \right\}_{i=1}^N$$

$$x = \theta + z \quad p_z(z) = \frac{1}{\sigma_z \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{z - \mu_z}{\sigma_z} \right)^2 \right] \quad \mu_z, \sigma_z = \text{const.}$$

- Należy wyznaczyć wzór analityczny na wartość mierzonej wielkości tak, by maksymalizować funkcję wiarygodności

ESTYMACJA MIERZONEJ WIELKOŚCI

Metoda najmniejszych kwadratów



Fragment rozwiązania

$$L(\theta) = \prod_{i=1}^N p_z(z_i) |\det J| = \prod_{i=1}^N \frac{1}{\sigma_z \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{x_i - \theta - \mu_z}{\sigma_z} \right)^2 \right]$$

Założmy, że $\mu_z = 0$

$$\log L(\theta) = -N \log \sigma_z \sqrt{2\pi} - \frac{1}{2\sigma_z^2} \sum_{i=1}^N (x_i - \theta)^2$$

$$\max_{\theta} \log L(\theta) = \min_{\theta} (-\log L(\theta))$$

$$\sum_{i=1}^N (x_i - \theta)^2 = \sum_{i=1}^N z_i^2 = \text{Var}[z]$$

$$\hat{\theta}_{\text{LS}} = \arg \min_{\theta} \sum_{i=1}^N (x_i - \theta)^2$$

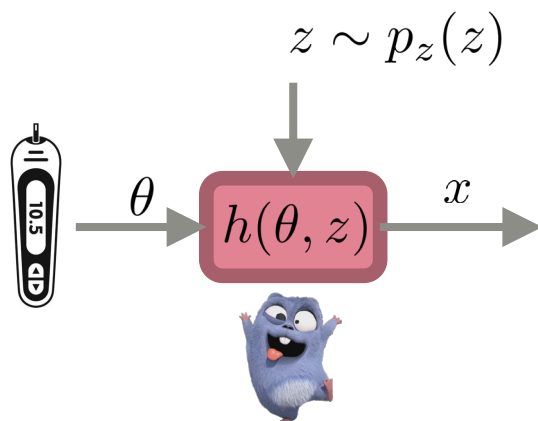
ESTYMACJA MIERZONEJ WIELKOŚCI

Metoda Bayesa

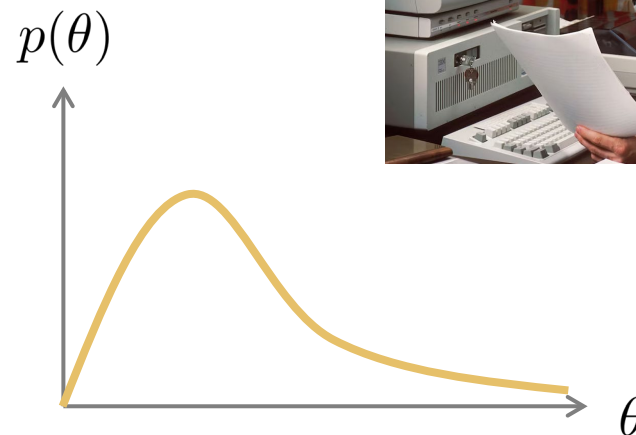


Estymacja mierzonej wielkości

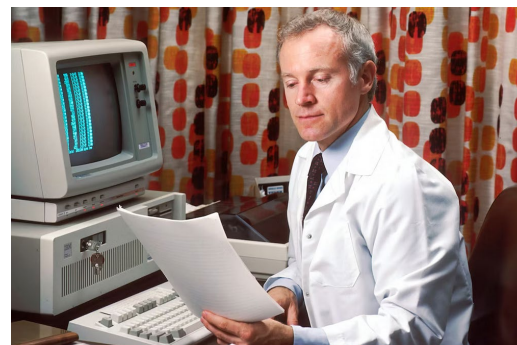
W czym problem?



Dane pomiarowe

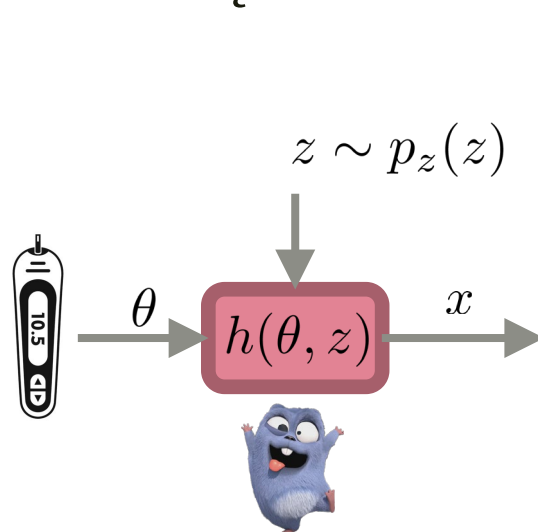


Wiedza

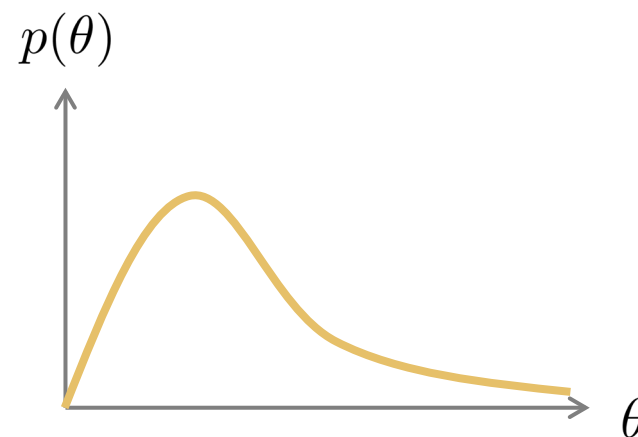


Estymacja mierzonej wielkości

Rozwiązanie



Dane pomiarowe



Wiedza

$$p(\theta | \text{Dane}) = \frac{p(\text{Dane} | \theta) \cdot p(\theta)}{\cancel{p(\text{Dane})}}$$

$$\hat{\theta}_{\text{MAP}} = \arg \max_{\theta} p(\mathbf{X} | \theta) p(\theta)$$

Przykład zadania

- Zakłócenia dodają się do mierzonej wielkości i mają rozkład normalny o znanych parametrach

$$\text{Dane} = \left\{ \mathbf{x}_i \right\}_{i=1}^N$$
$$x = \theta + z \quad p_z(z) = \frac{1}{\sigma_z \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{z - \mu_z}{\sigma_z} \right)^2 \right] \quad \mu_z, \sigma_z = \text{const.}$$

- Dodatkowo wiemy, że estymowana wielkość ma rozkład normalny:

$$p_\theta(\theta) = \frac{1}{\sigma_\theta \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{\theta - \mu_\theta}{\sigma_\theta} \right)^2 \right] \quad \mu_\theta, \sigma_\theta = \text{const.}$$

- Należy wyznaczyć estymator MAP mierzonej wielkości

Przykład zadania

- Pokaż, że estymator MAP jest średnią ważoną estymatora ML i wiedzy apriorycznej wyrażonej przez wartość oczekiwaną μ_θ .
- Jaki wzór opisuje współczynniki wagowe? Od jakich wielkości zależą wagi?
- Rozpatrz dwie skrajne sytuacje:
 - *pomiarów jest mało i są niskiej jakości (ang. unreliable data)*
 - *dostępnych jest dużo wysokiej jakości pomiarów, lecz niewiele wiemy o możliwych wartościach θ (ang. unreliable prior)*
- Jak w obu powyższych sytuacjach upraszcza się estymator MAP ?

Przykład zadania

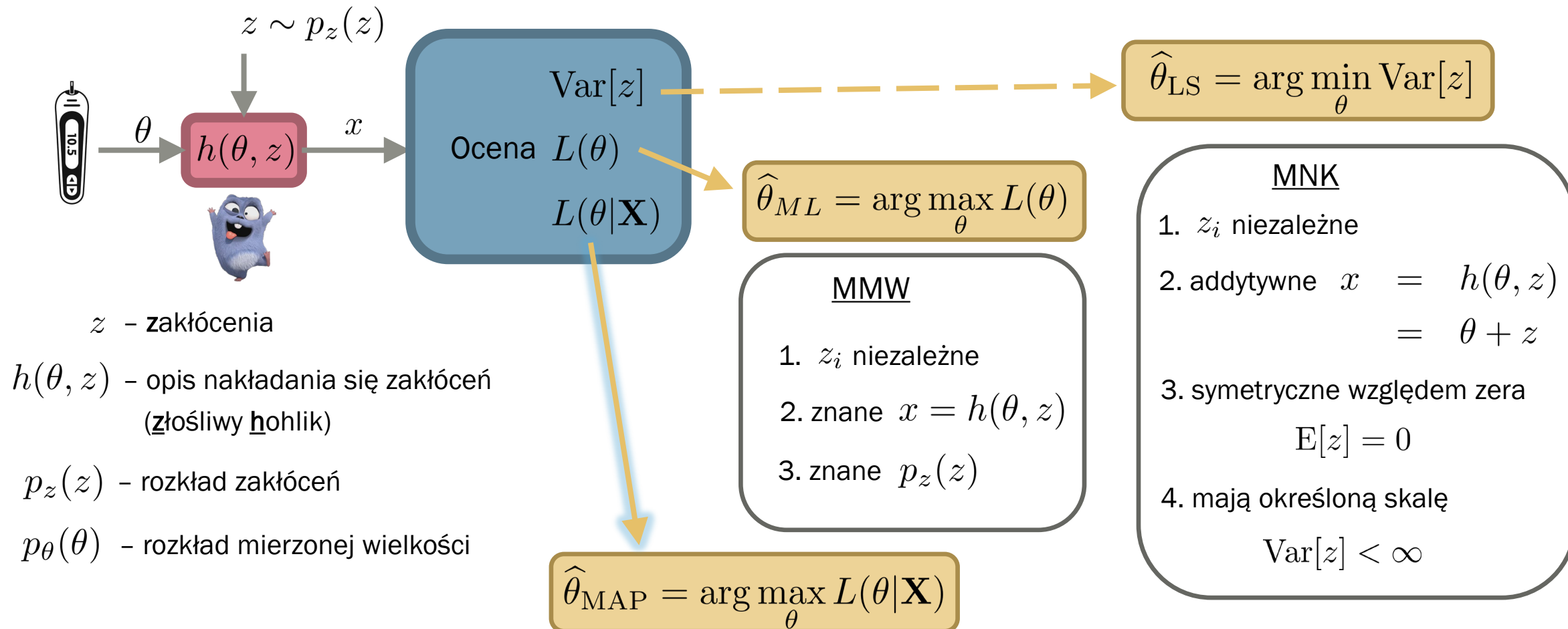
- Sprowadzimy wiedzę aprioryczną do przekonania, że wartość θ nie powinna zbytnio odbiegać od μ_θ .
- Opracuj wzór na estymator θ_K , który maksymalizuje funkcję wiarygodności z nałożoną karą za odchylenie oszacowania wartości θ od μ_θ :

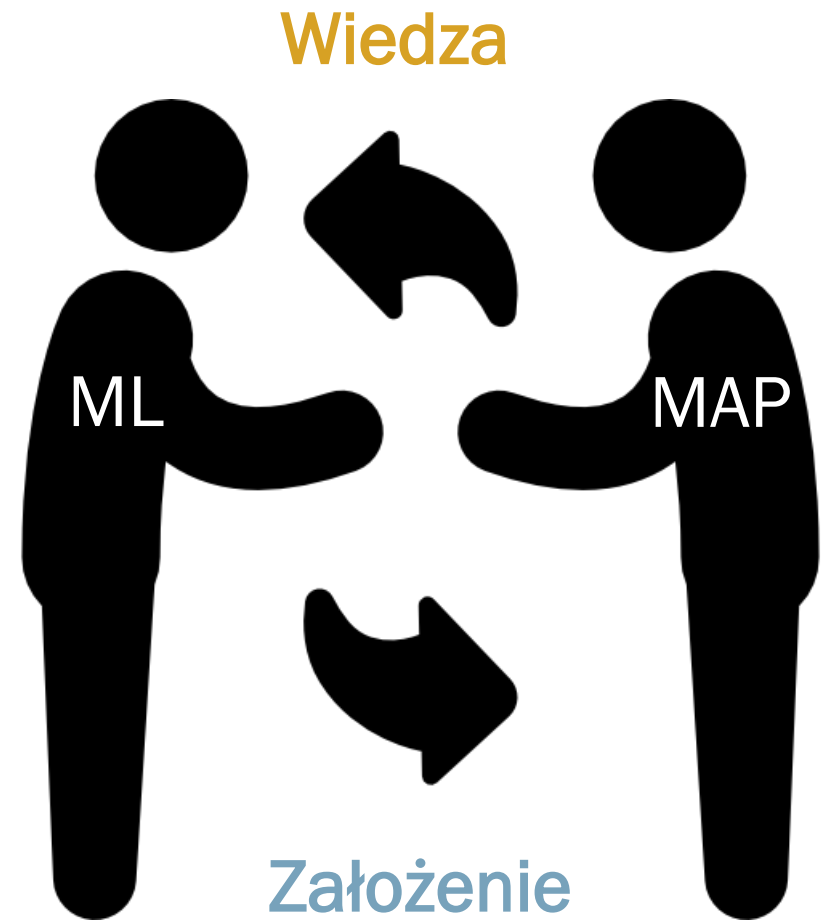
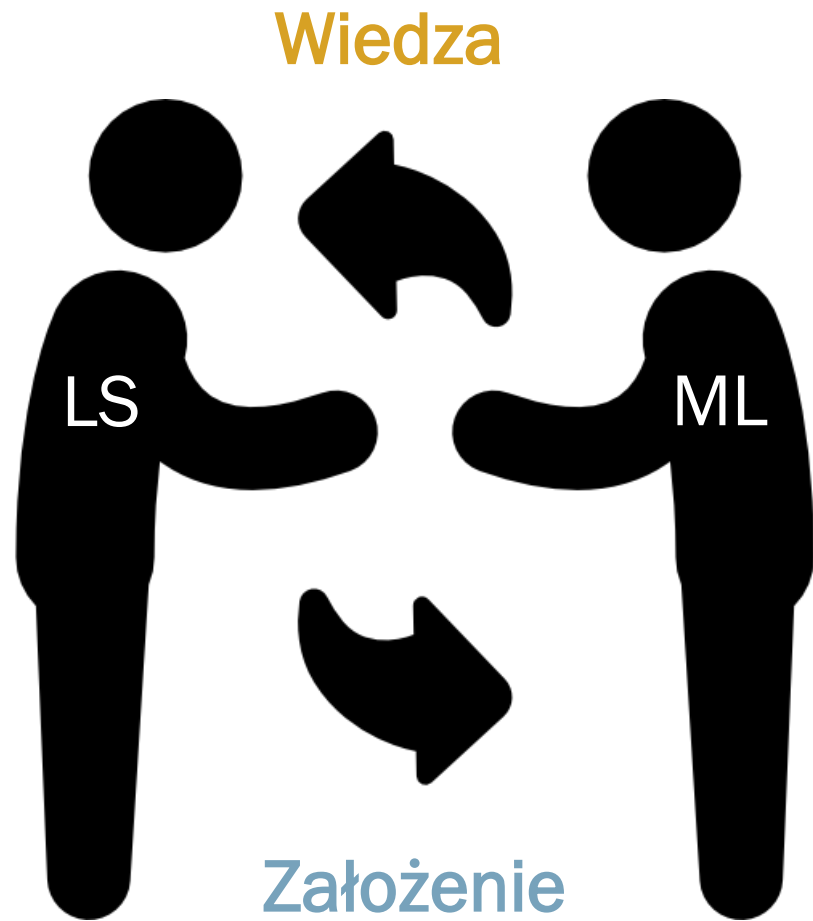
$$\theta_K = \arg \max_{\theta} \left\{ \ln p(\mathbf{X}|\theta) - \lambda(\theta - \mu_\theta)^2 \right\}, \quad \lambda > 0$$

- Zbadaj podobieństwo estymatorów θ_K i θ_{MAP} .
- Co metoda Bayesa ma wspólnego z regularyzacją?

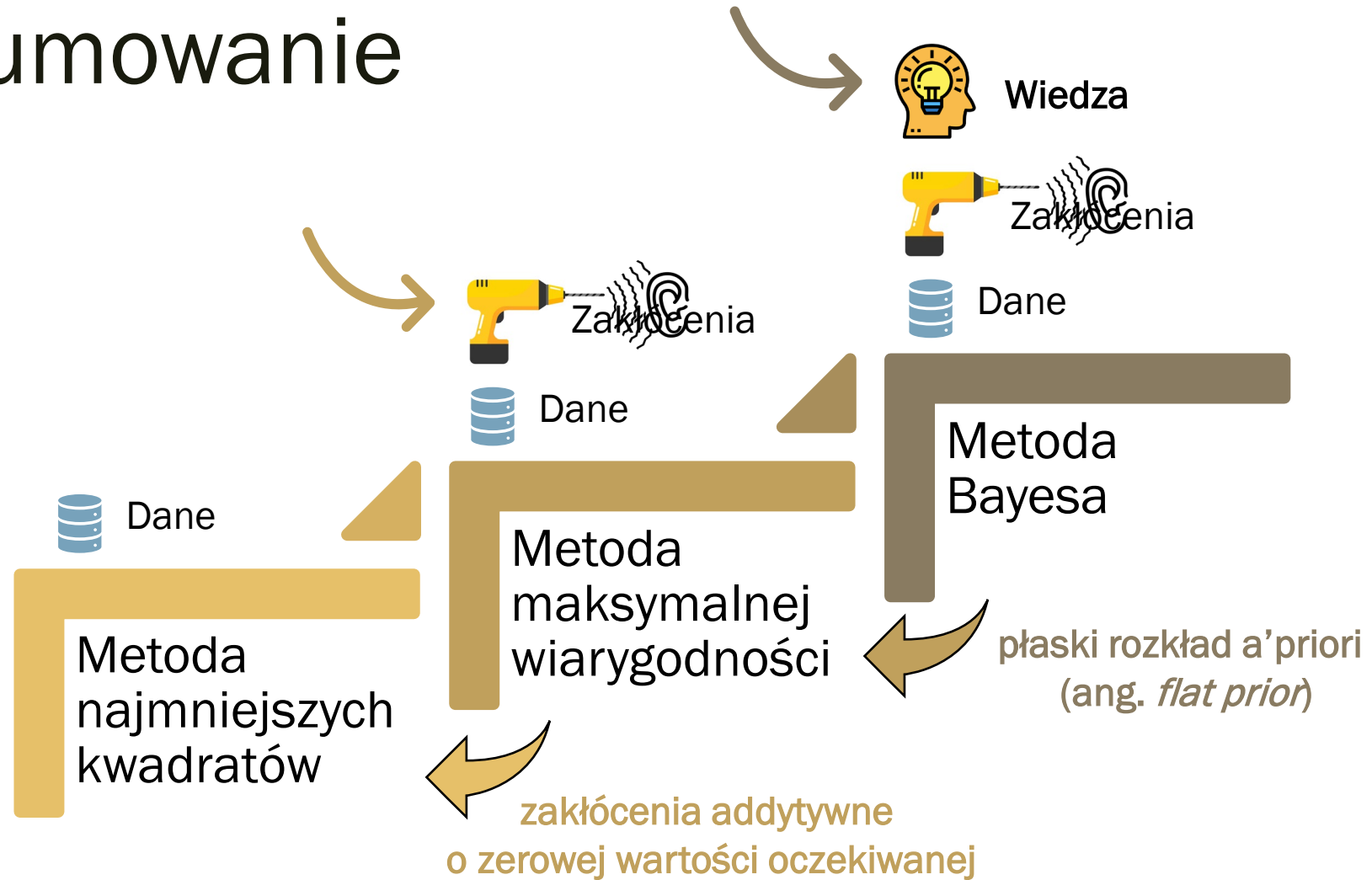
Estymacja mierzonej wielkości

Założenia stojące u podstaw metod





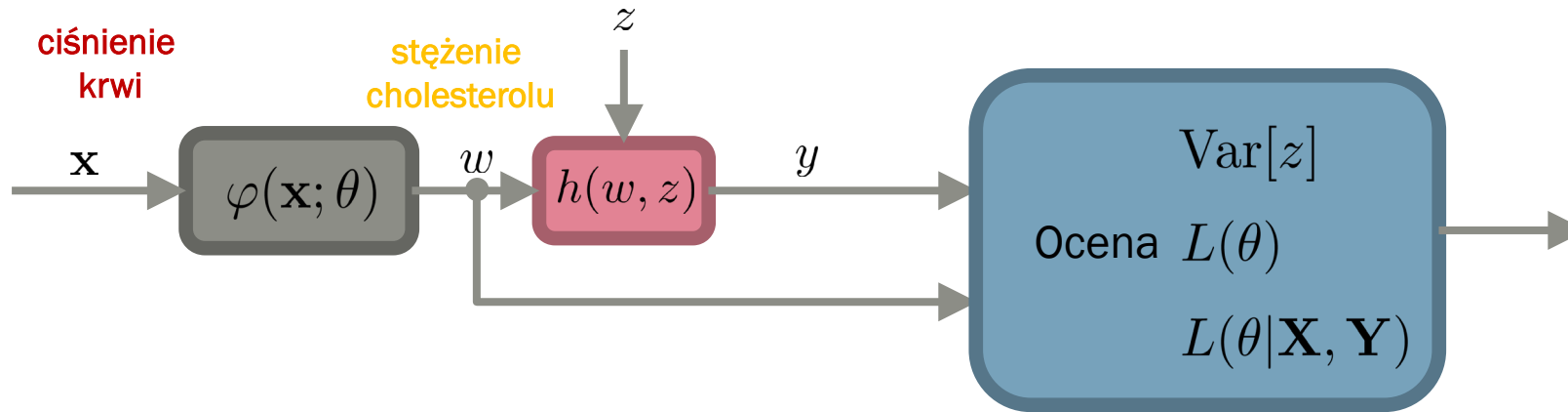
Podsumowanie



ESTYMACJA PARAMETRU MODELU



Estymacja parametrów „ubranych” w model



Zakładamy, że proces działa zgodnie z opisem $\varphi(\mathbf{x}; \theta)$

$$y = h(w, z)$$

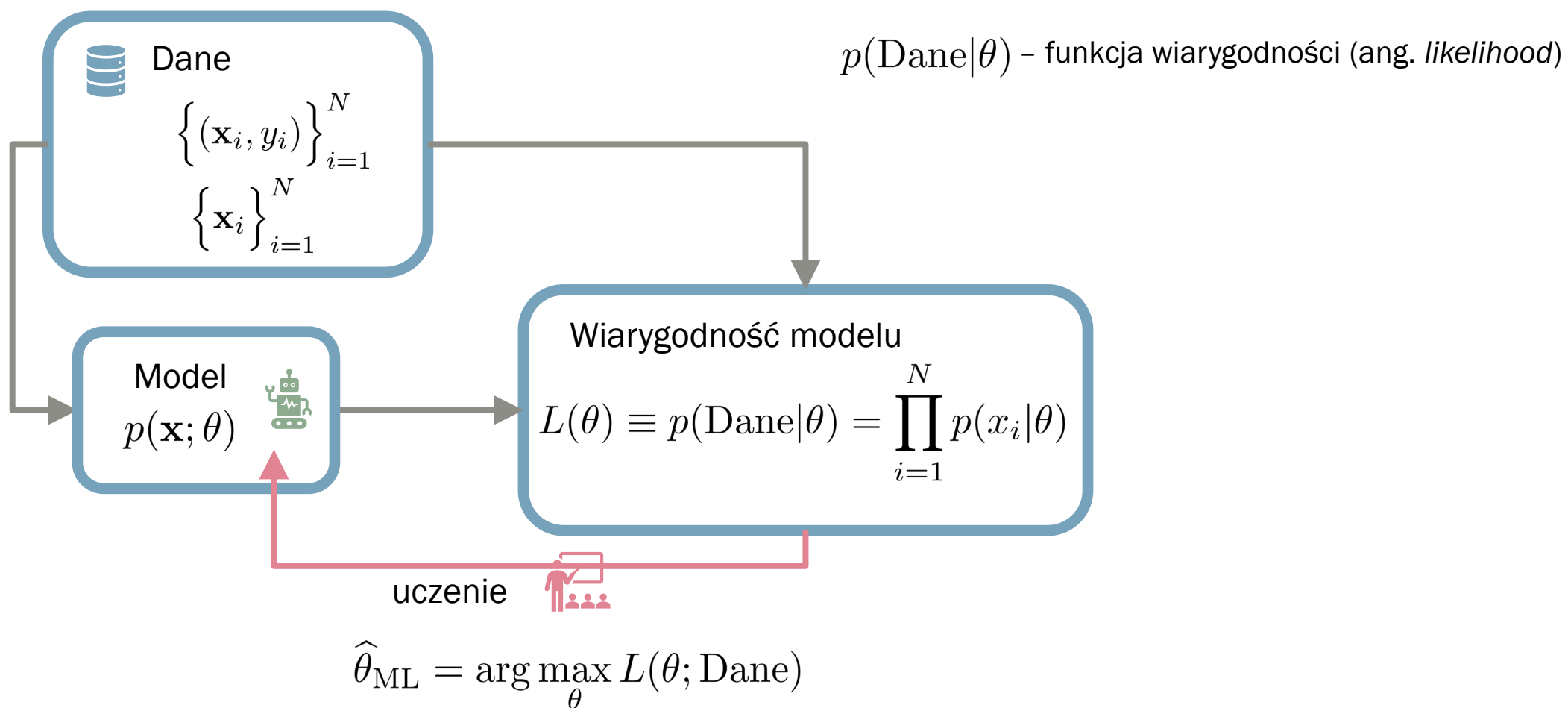
$$z = h^{-1}(w, y)$$

$$w = \varphi(\mathbf{x}; \theta)$$

$$L(\theta) = \prod_{i=1}^N p_y(y) = \prod_{i=1}^N p_z\left(h^{-1}(\varphi(\mathbf{x}_i; \theta), y_i)\right) |\det J|, \quad J = \frac{\partial}{\partial y} h^{-1}$$

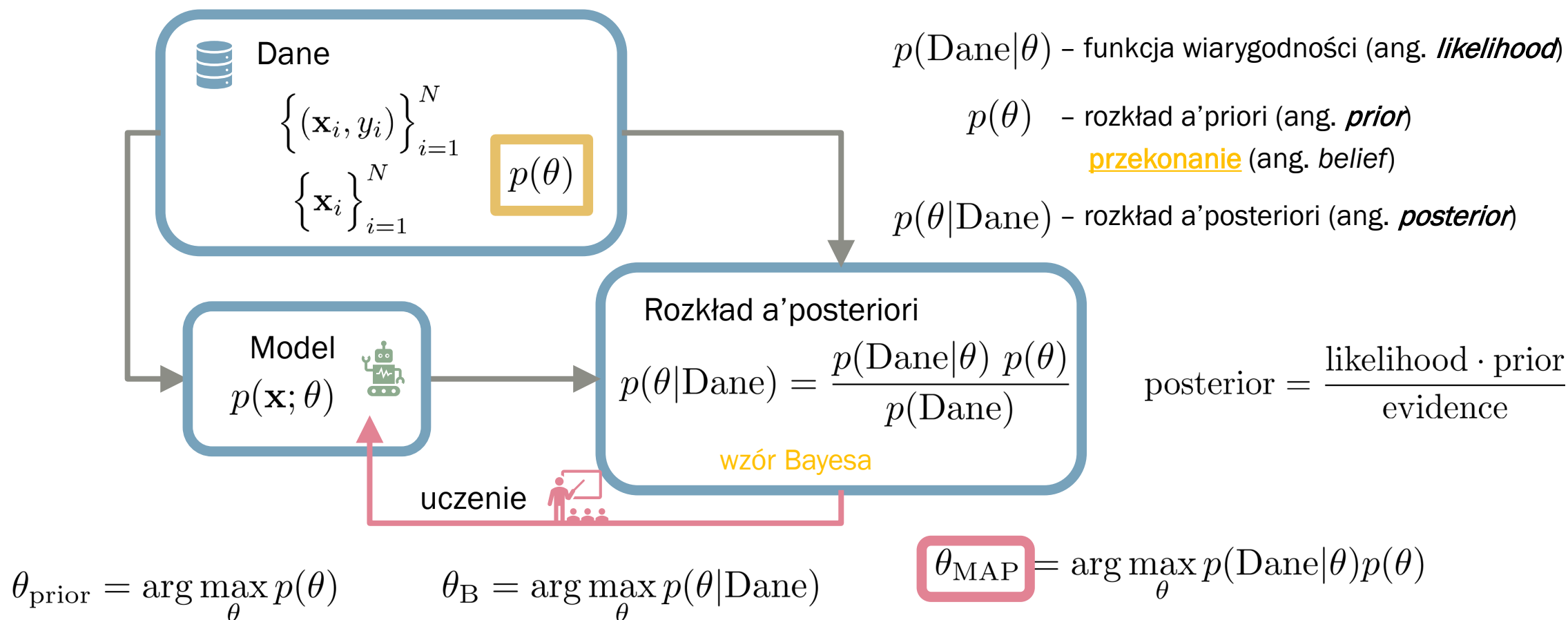
Dopasowanie modelu do danych

Metoda największej wiarygodności (ang. *Maximum Likelihood method*)



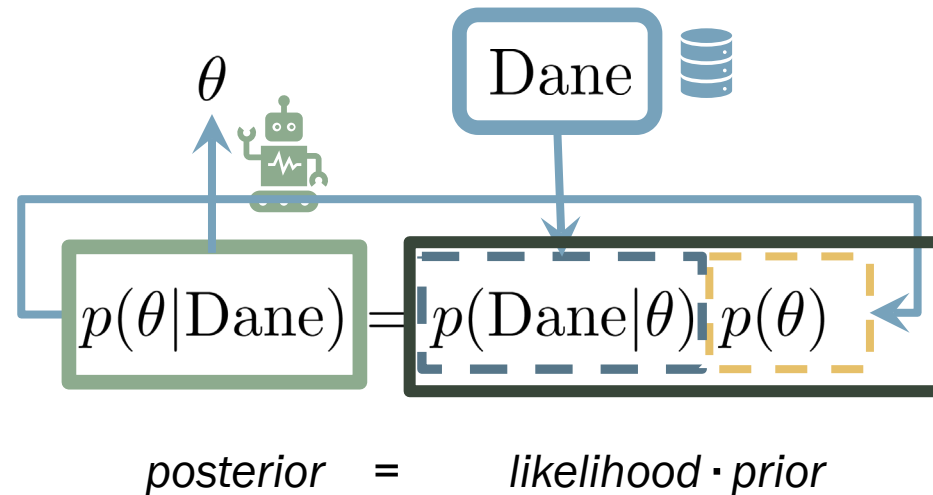
Dopasowanie modelu do danych

Metoda Bayesa (ang. *Bayesian method*)



Dopasowanie modelu do danych

Metoda Bayesa (ang. *Bayesian method*)



Przykład zadania

- Zakłócenia dodają się do wyjścia procesu liniowego i mają rozkład normalny o znanych parametrach

$$\text{Dane} = \left\{ (\mathbf{x}_i, y_i) \right\}_{i=1}^N \quad w = \theta \cdot x$$

$$y = w + z \quad p_z(z) = \frac{1}{\sigma_z \sqrt{2\pi}} \exp \left[-\frac{(z - \mu_z)^2}{2\sigma_z^2} \right] \quad \mu_z, \sigma_z = \text{const.}$$

- Należy wyznaczyć wzór analityczny na wartość parametru modelu tak, by maksymalizować funkcję wiarygodności
- Jak sprawi się MNK a jak MMW?

Przykład zadania

- Zakłócenia wymnażają się przez wyjście procesu liniowego i mają rozkład jednostajny na przedziale $\langle 0, 1 \rangle$

$$\text{Dane} = \left\{ (\mathbf{x}_i, y_i) \right\}_{i=1}^N \quad w = \theta \cdot x$$

$$y = w \cdot z \quad p_z(z) = \begin{cases} 1 & \text{dla } z \in \langle 0, 1 \rangle \\ 0 & \text{w przec. przyp.} \end{cases}$$

- Należy wyznaczyć wzór analityczny na wartość parametru modelu tak, by maksymalizować funkcję wiarygodności
- Jak sprawi się MNK a jak MMW?

Podejście bayesowskie

- Podejście bayesowskie (ang. *bayesian framework*)



$$\hat{\theta}_{\text{ML}} = \arg \max_{\theta} p(X|\theta)$$

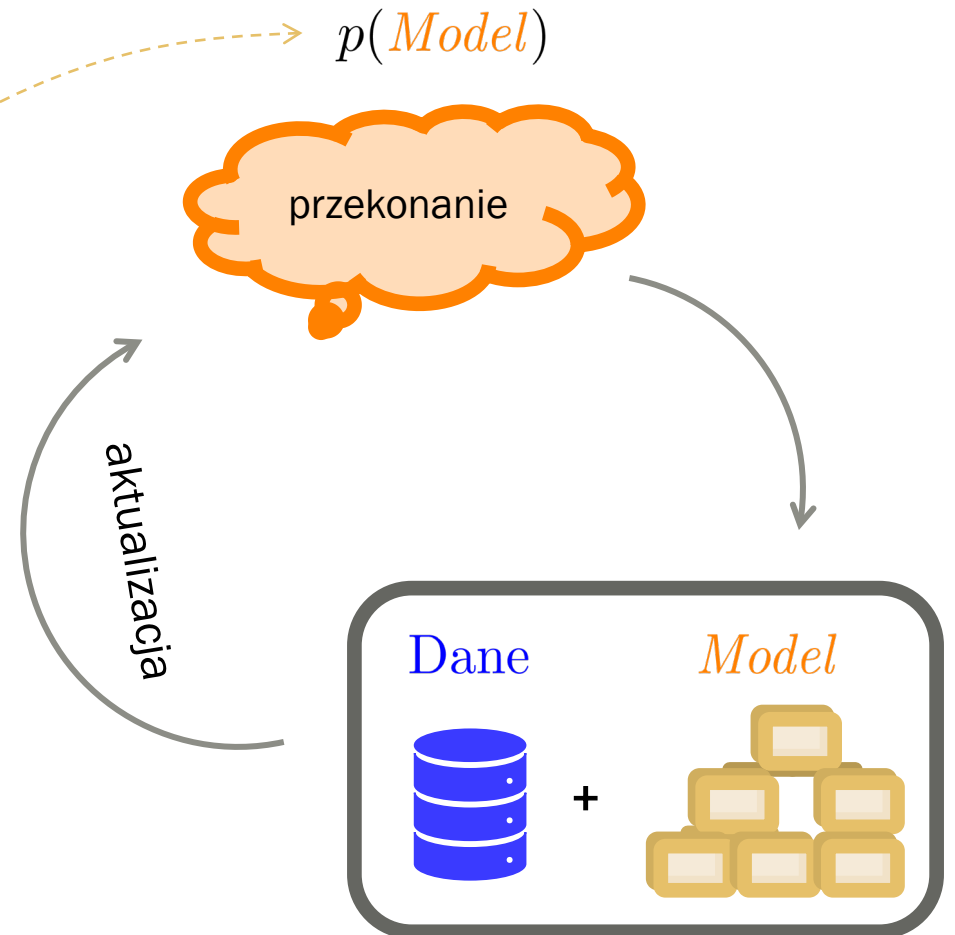
$$\begin{aligned}\hat{\theta}_{\text{B}} &= \arg \max_{\theta} p(\theta|X) \\ &= \arg \max_{\theta} \frac{p(X|\theta)p(\theta)}{p(X)}\end{aligned}$$

$$\hat{\theta}_{\text{MAP}} = \arg \max_{\theta} p(X|\theta)p(\theta)$$

Podejście bayesowskie

- Kiedy $\hat{\theta}_{\text{ML}} = \hat{\theta}_{\text{MAP}}$?
- Płaski rozkład $p(\theta)$ (ang. *flat prior*)

$$p(\textit{Model}|\textit{Dane}) = \frac{p(\textit{Dane}|\textit{Model}) \cdot p(\textit{Model})}{p(\textit{Dane})}$$



Z LOTU PTAKA

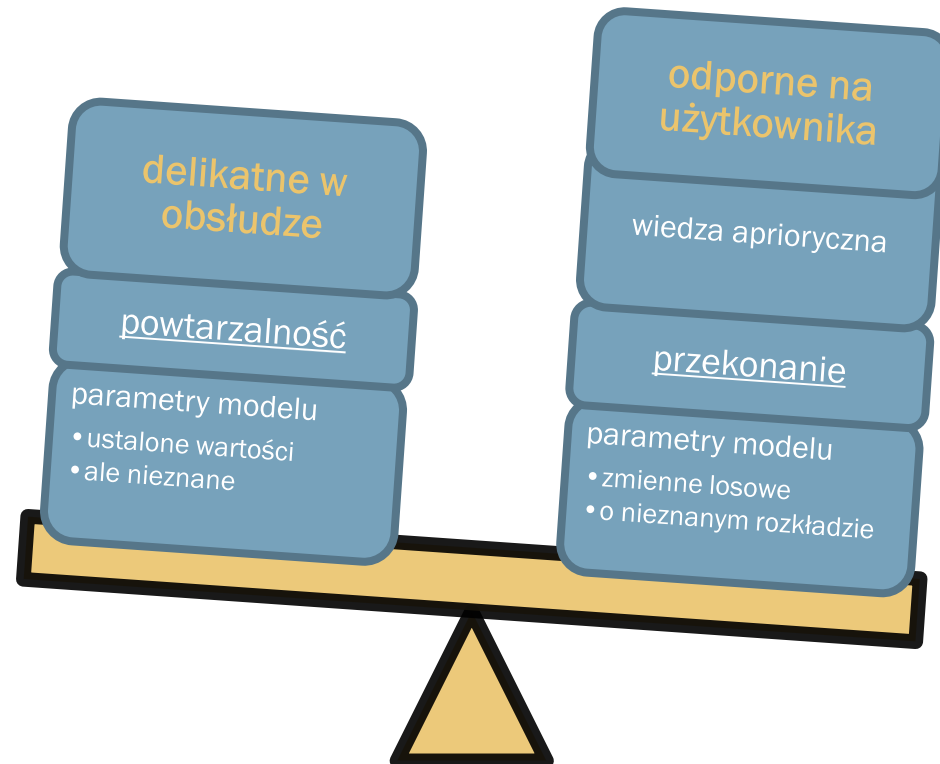


Estymacja parametrów modelu

podjęcie częstotściowe
(ang. *frequentist approach*)

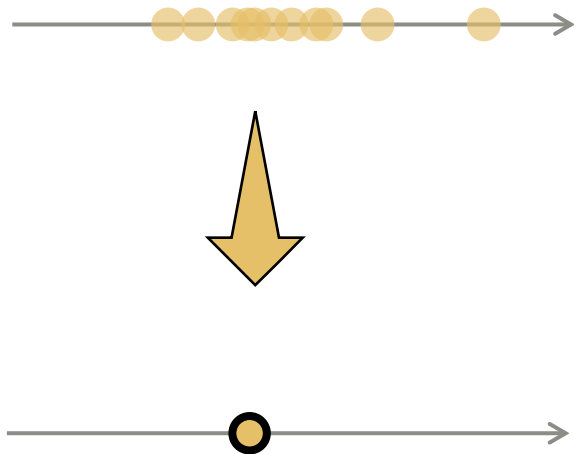
podjęcie bayesowskie
(ang. *bayesian approach*)

$p(\text{Dane} | \text{Model})$
losowe ustalony

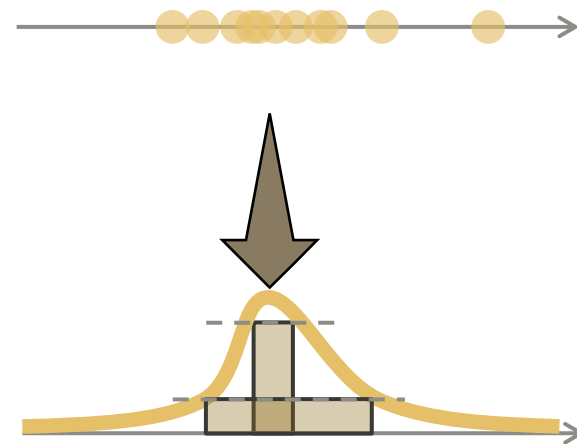


$p(\text{Model} | \text{Dane})$
losowy ustalony

Estymacja punktowa



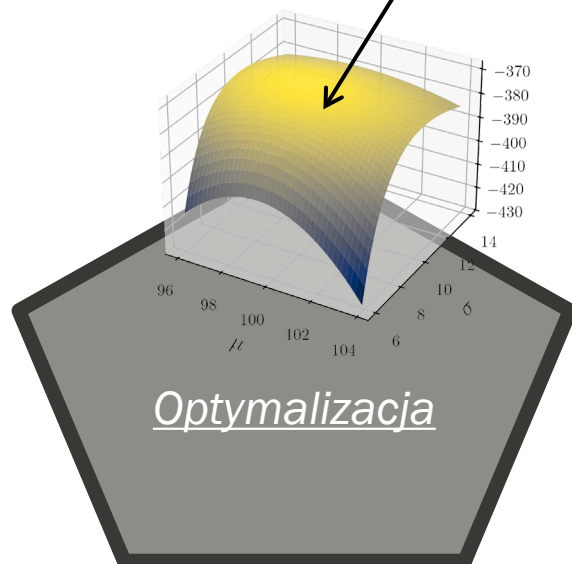
Estymacja przedziałowa



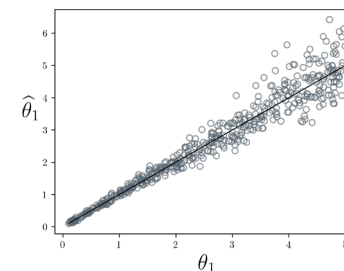
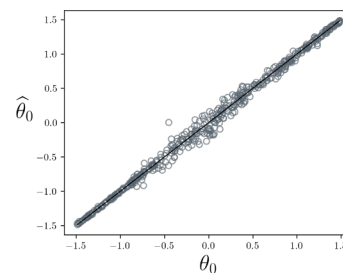
Uczenie maszynowe

jako

jako



Estymacja

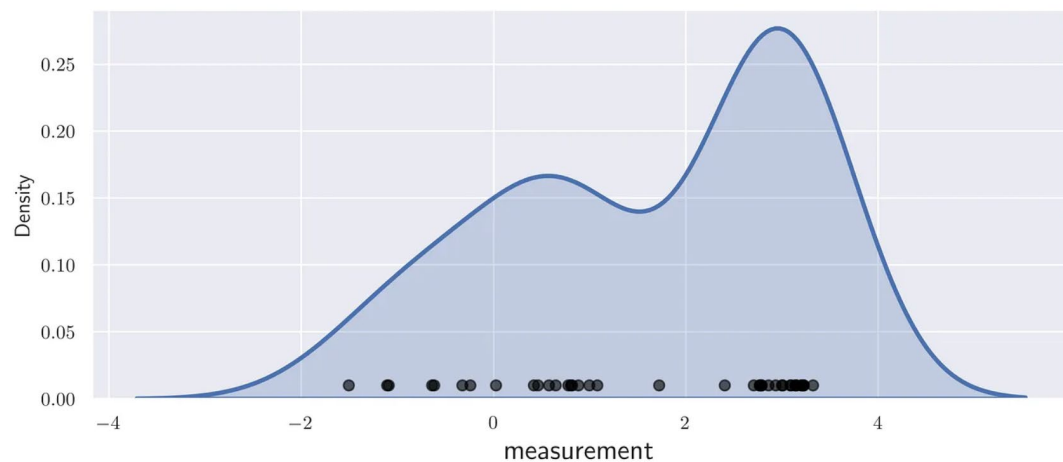


**Dopasowanie do danych
z uwzględnieniem wiedzy**

Ewaluacja modelu

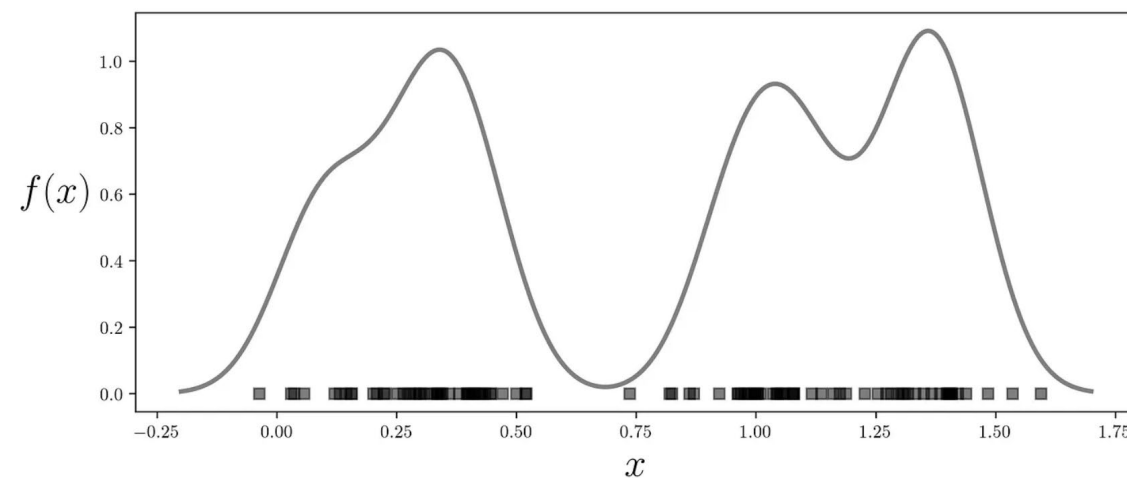
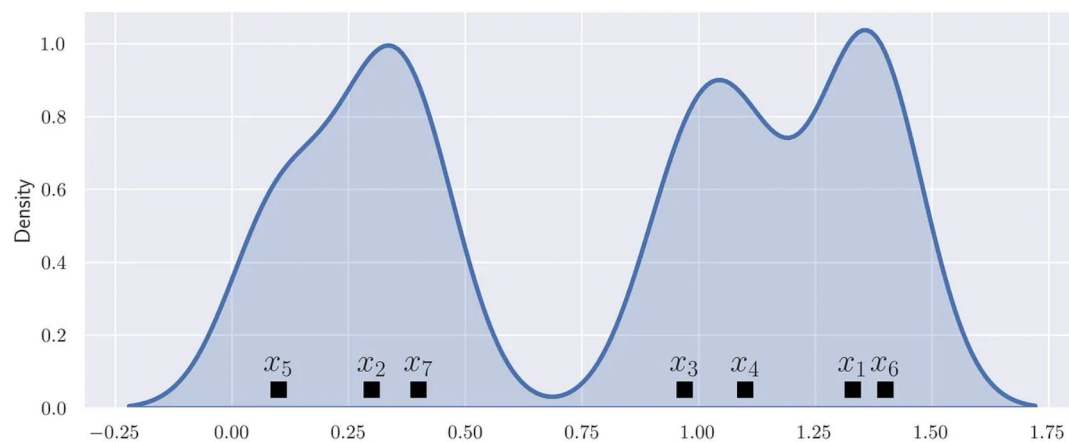
Ale można bez optymalizacji:

$$f(x) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{x - x_i}{h}\right)$$



Jądrowy estymator gęstości

syntetyczne dane





KONEC

MOVIE ENDOZA

— JÓNMY CÒNΞΙΑ —