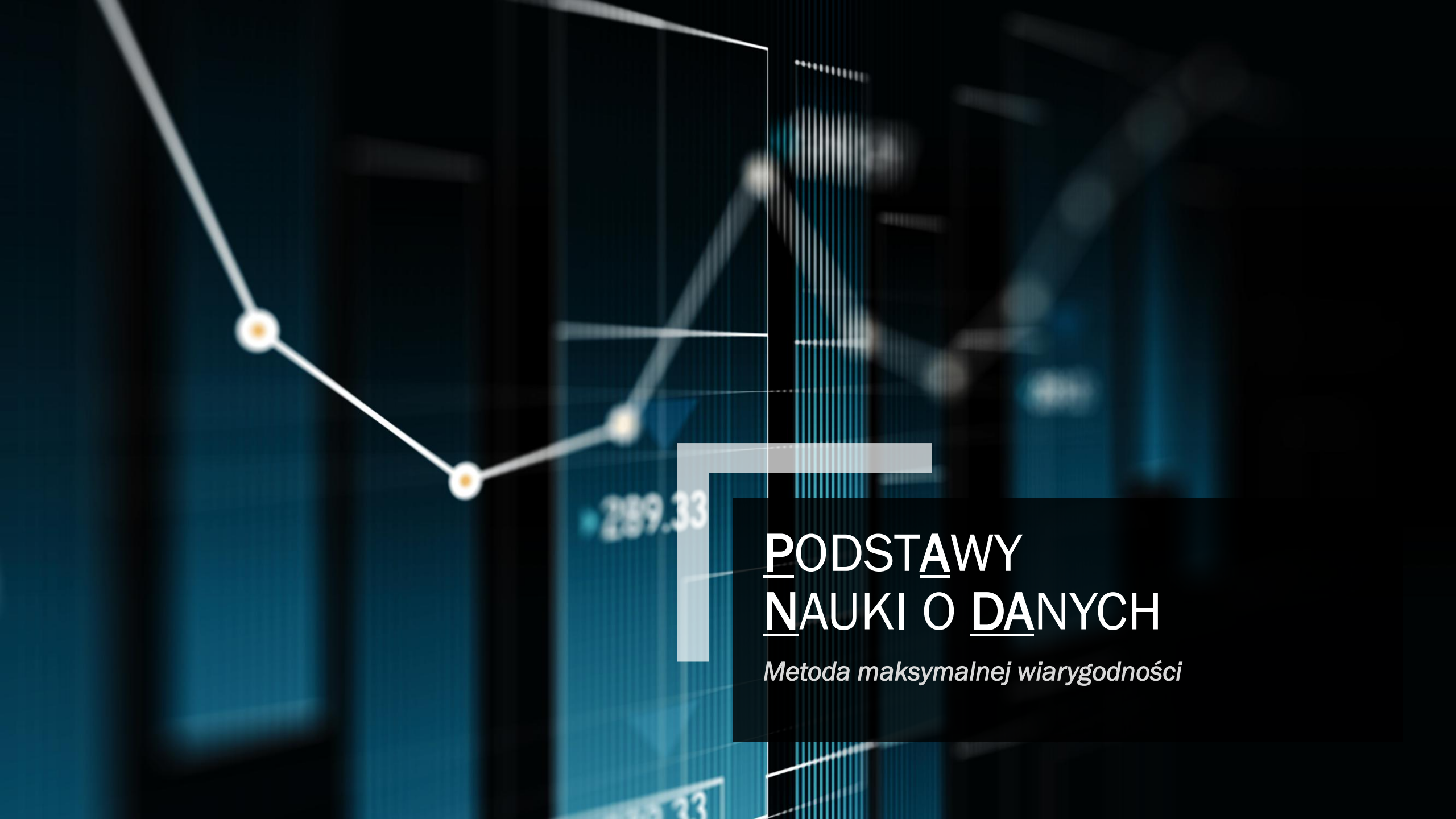




PODSTAWY NAUKI O DANYCH

Metoda maksymalnej wiarygodności

Przypominajka statystyczna

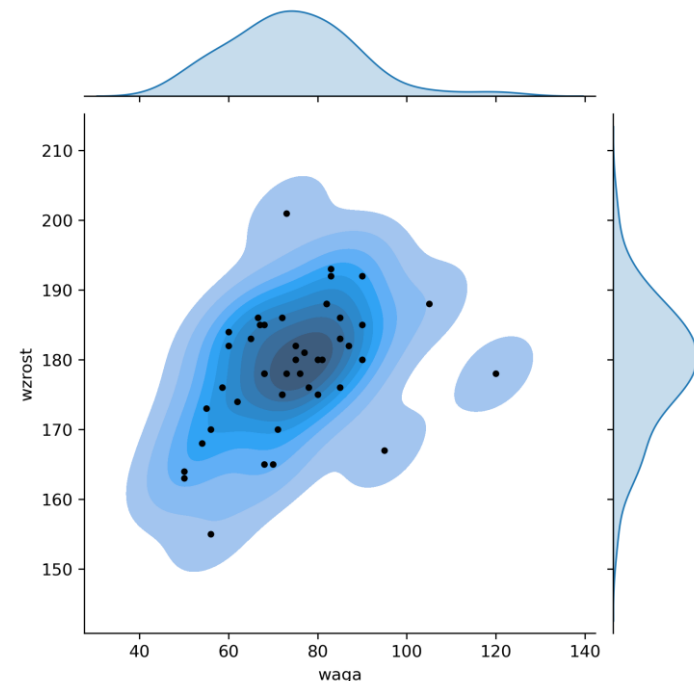
$p(X; \theta)$ – rozkład zmiennej losowej X , którego parametrem jest θ
zwróć uwagę na tę różnicę
 $p(X, \theta)$ – **rozkład łączny** dwóch zmiennych losowych X oraz θ
ang. *joint probability distribution*

$p(X|\theta)$ – **rozkład warunkowy** zmiennej losowej X pod warunkiem θ
ang. *conditional PDF*

$p(\theta|X)$ – **rozkład warunkowy** zmiennej losowej θ pod warunkiem X

PDF – Probability Density Function

PMF – Probability Mass Function



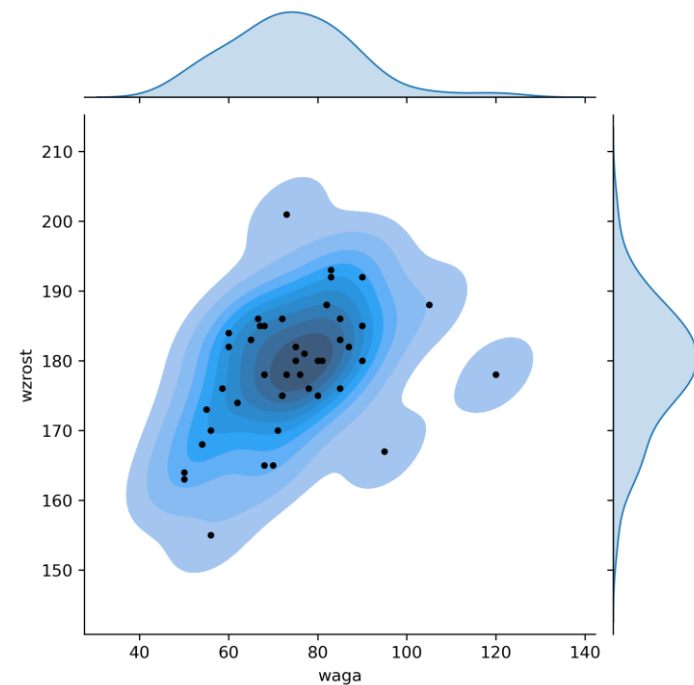
$$p(X, \theta) = p(\theta, X)$$

Przypominajka statystyczna

- Rozkład łączny – pełna informacja probabilistyczna
- Jak z rozkładu łącznego $p(X, \theta)$ wyznaczyć inne rozkłady?
- **Rozkłady brzegowe** (ang. *marginal PDFs*)

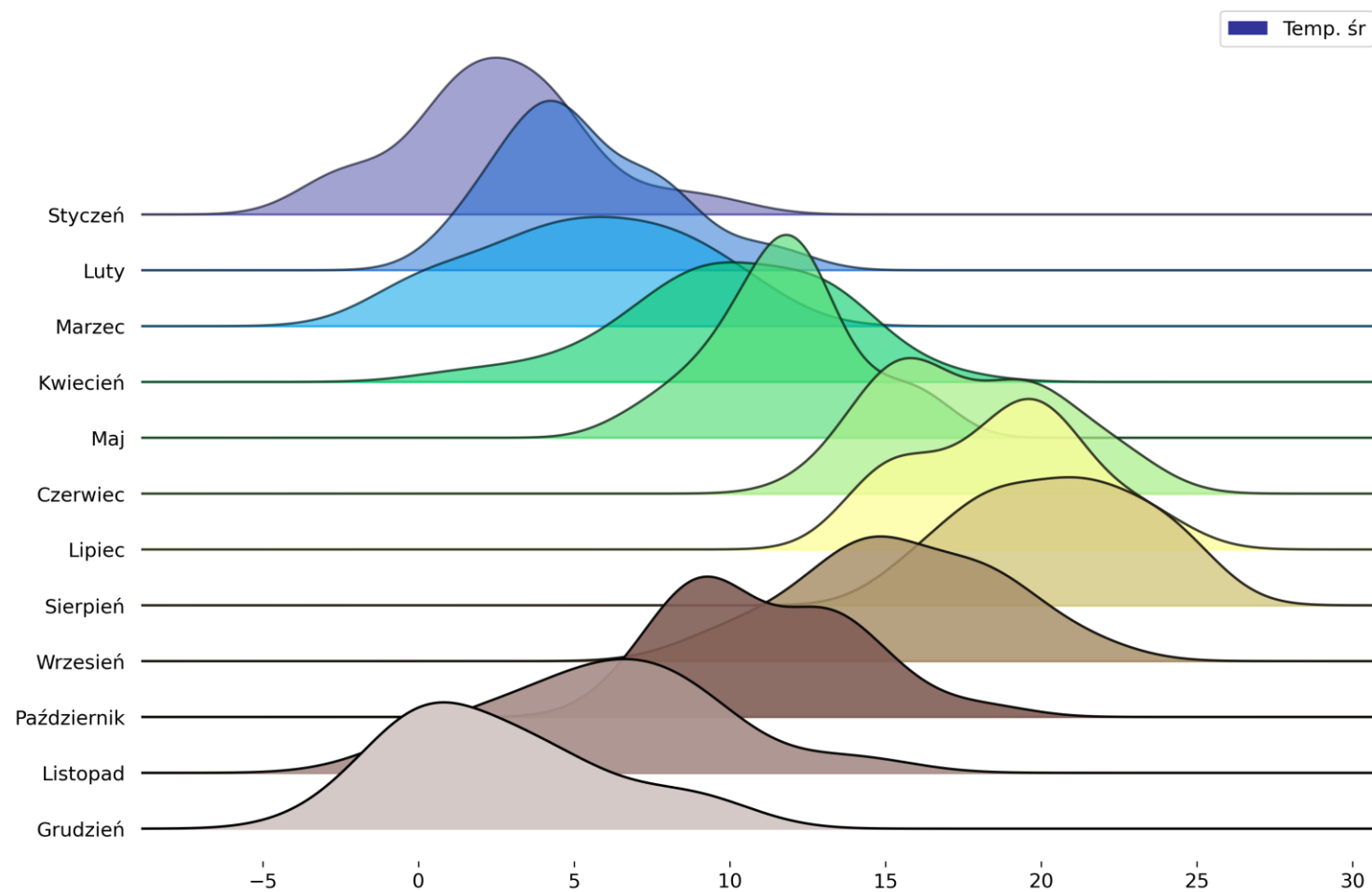
$$p(X) = \int_{D_\theta} p(X, \theta) d\theta \quad p(\theta) = \int_{D_X} p(X, \theta) dX$$

$$p(\theta|X) = \frac{p(X, \theta)}{p(X)} \quad p(X|\theta) = \frac{p(X, \theta)}{p(\theta)}$$



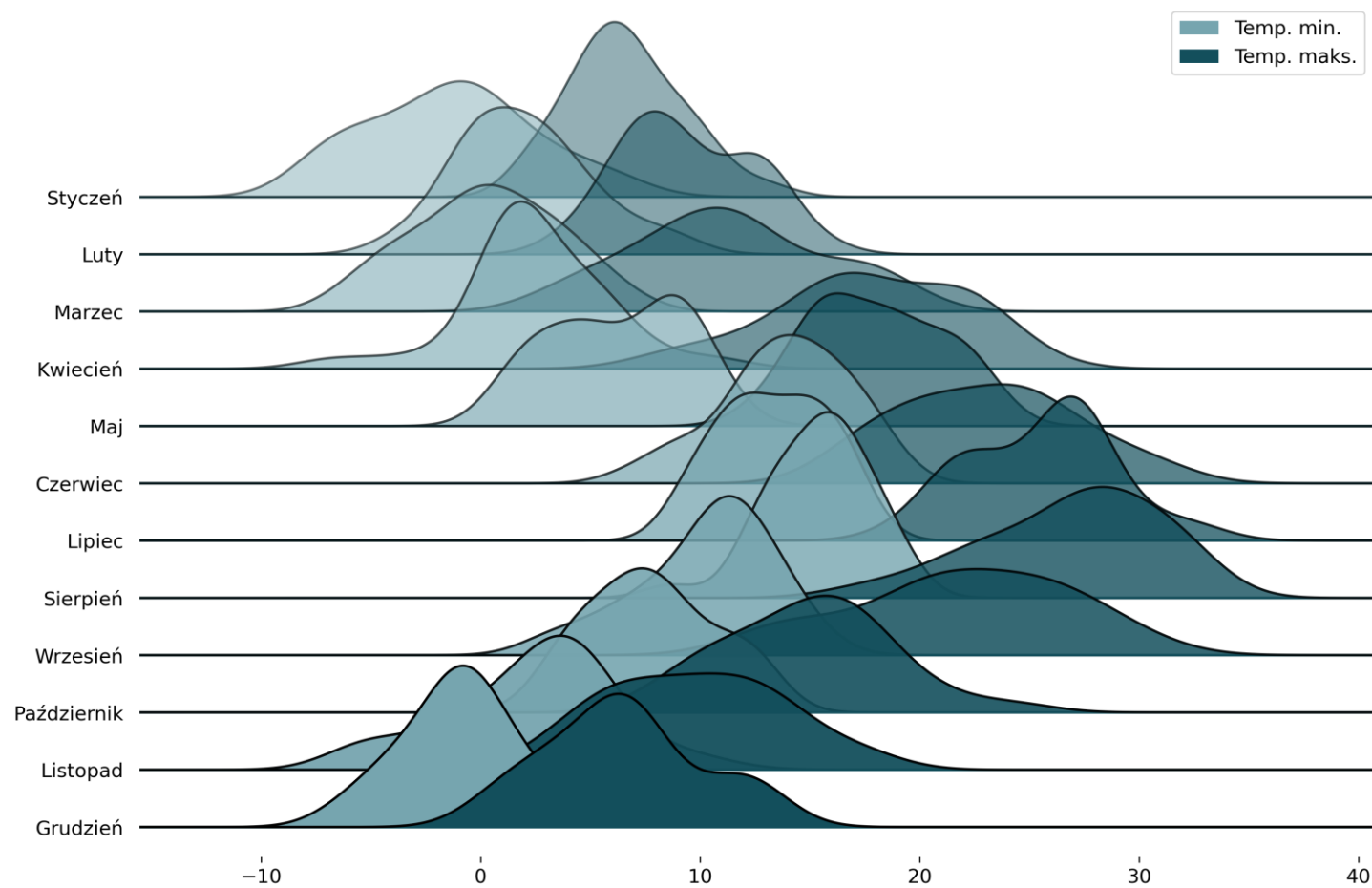
$$p(X, \theta) = p(\theta, X)$$

Przypominajka statystyczna



$$P(\text{Temp śr}|\text{miesiąc})$$

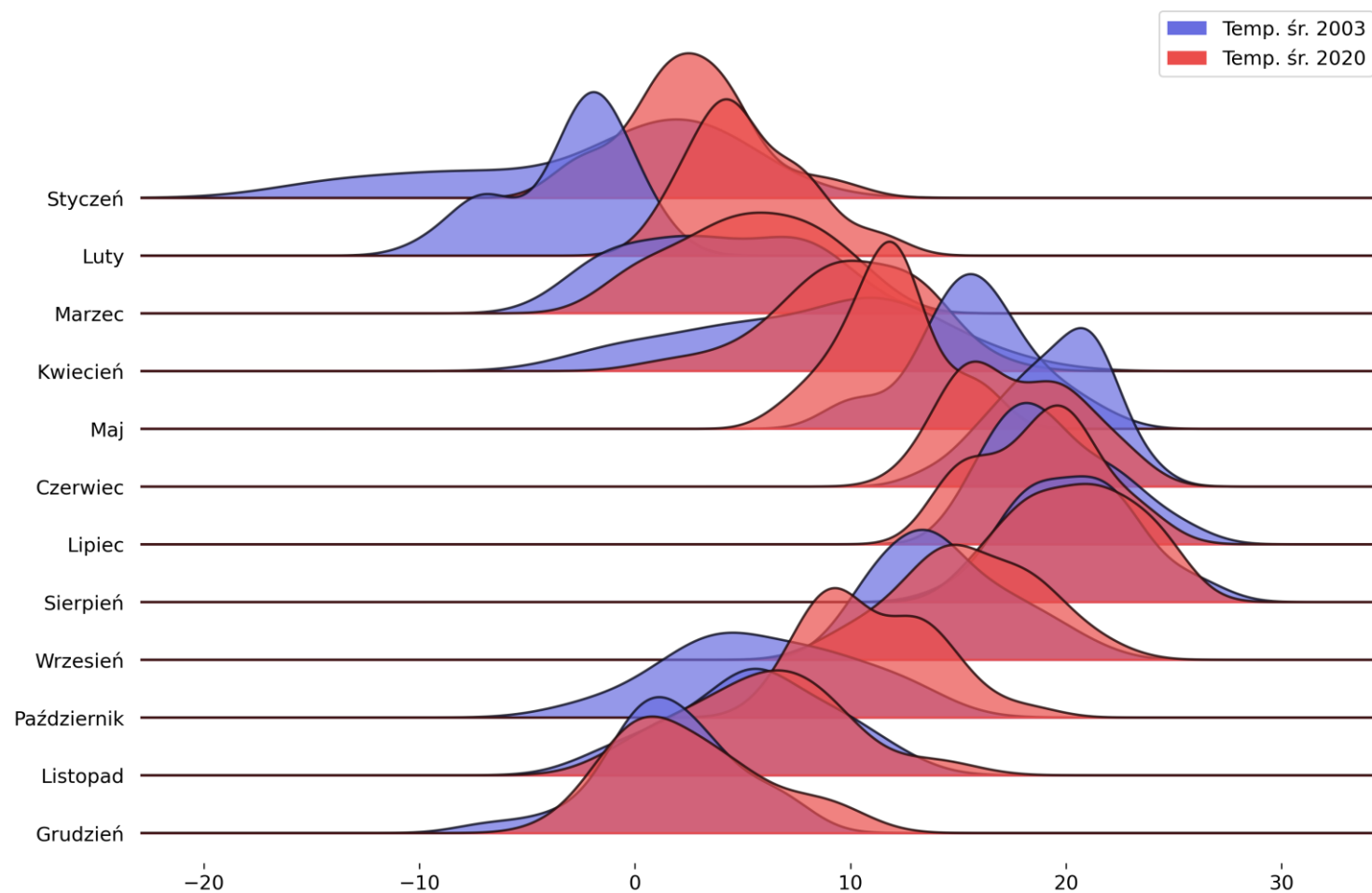
Przypominajka statystyczna



$$P(\text{Temp min}|\text{miesiac})$$

$$P(\text{Temp maks}|\text{miesiac})$$

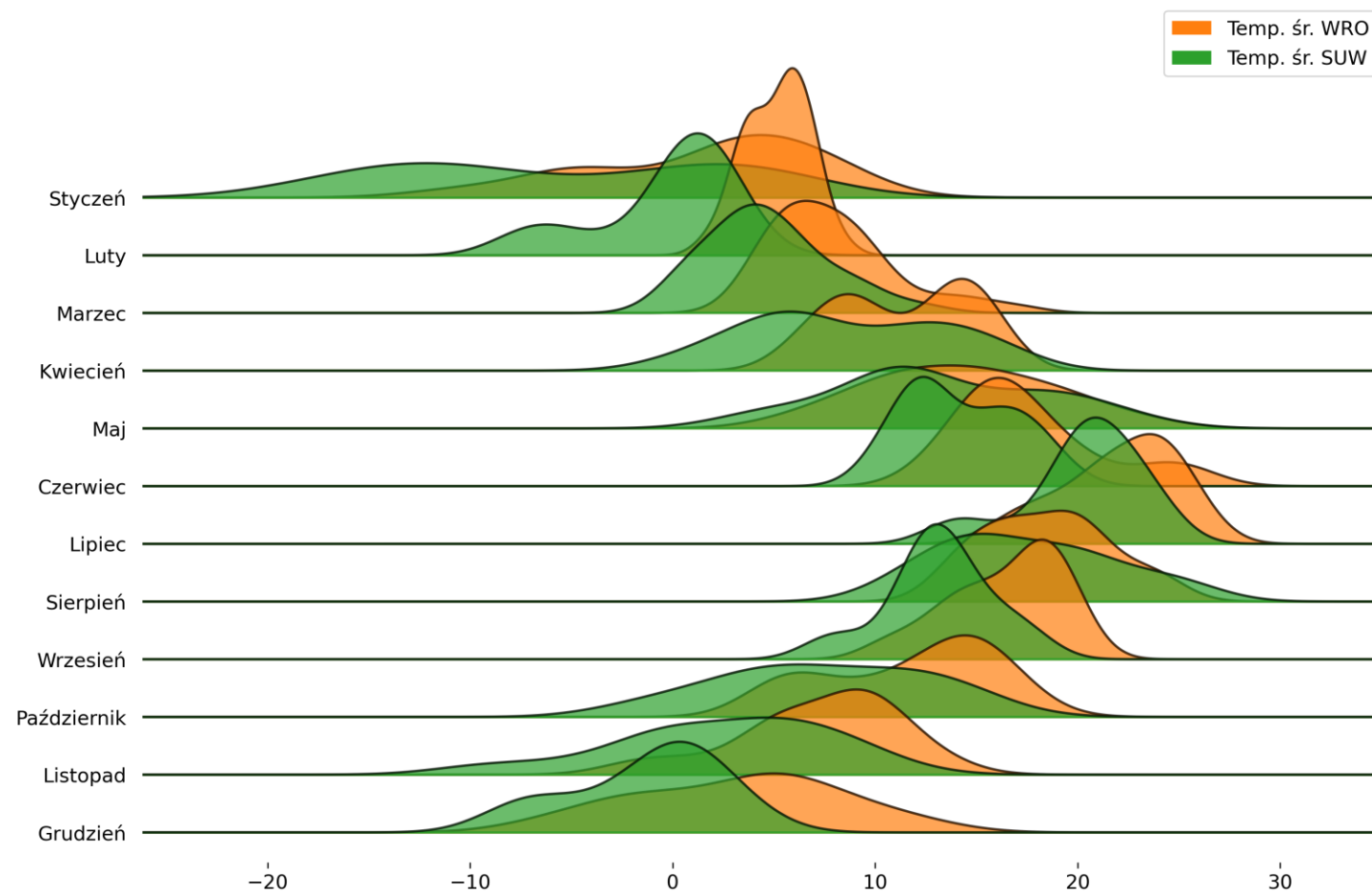
Przypominajka statystyczna



$$P(\text{Temp śr 2003}|\text{miesiac})$$

$$P(\text{Temp śr 2020}|\text{miesiac})$$

Przypominajka statystyczna

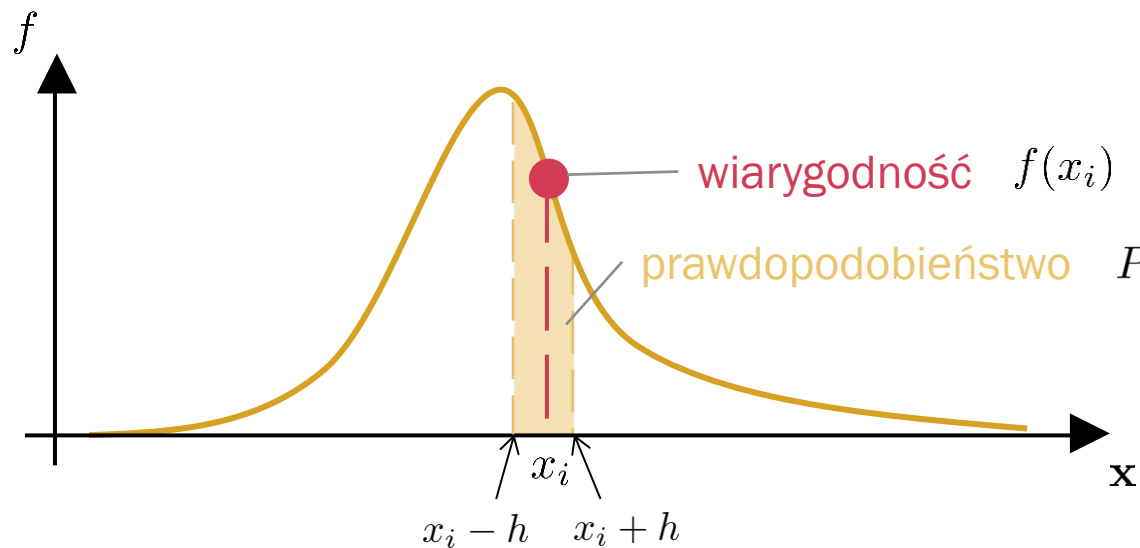


$P(\text{Temp } \acute{s}r \text{ Wroc}l\acute{a}w | \text{miesiac})$

$P(\text{Temp } \acute{s}r \text{ Suwa}lki | \text{miesiac})$

Prawdopodobieństwo a wiarygodność

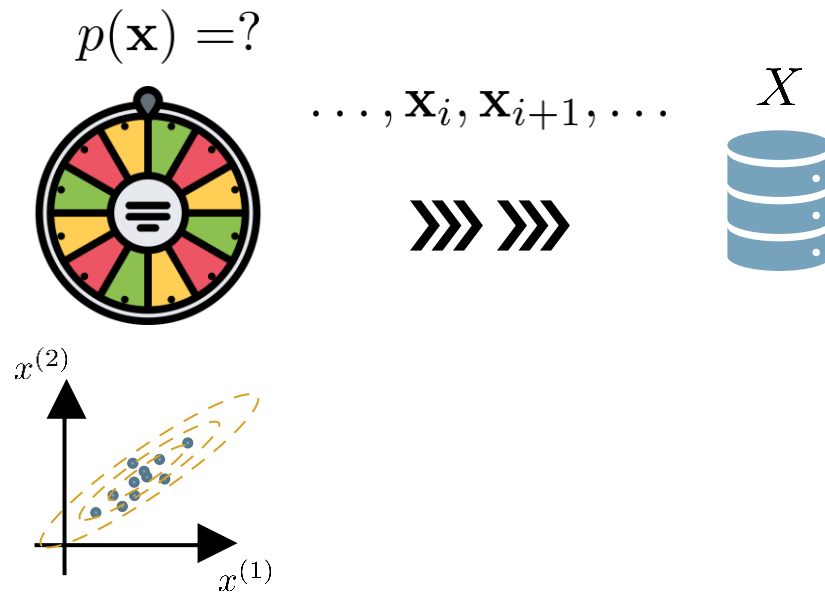
ang. *probability vs likelihood*



prawdopodobieństwo $P(x_i - h \leq x \leq x_i + h) = \int_{x_i - h}^{x_i + h} f(x) dx$

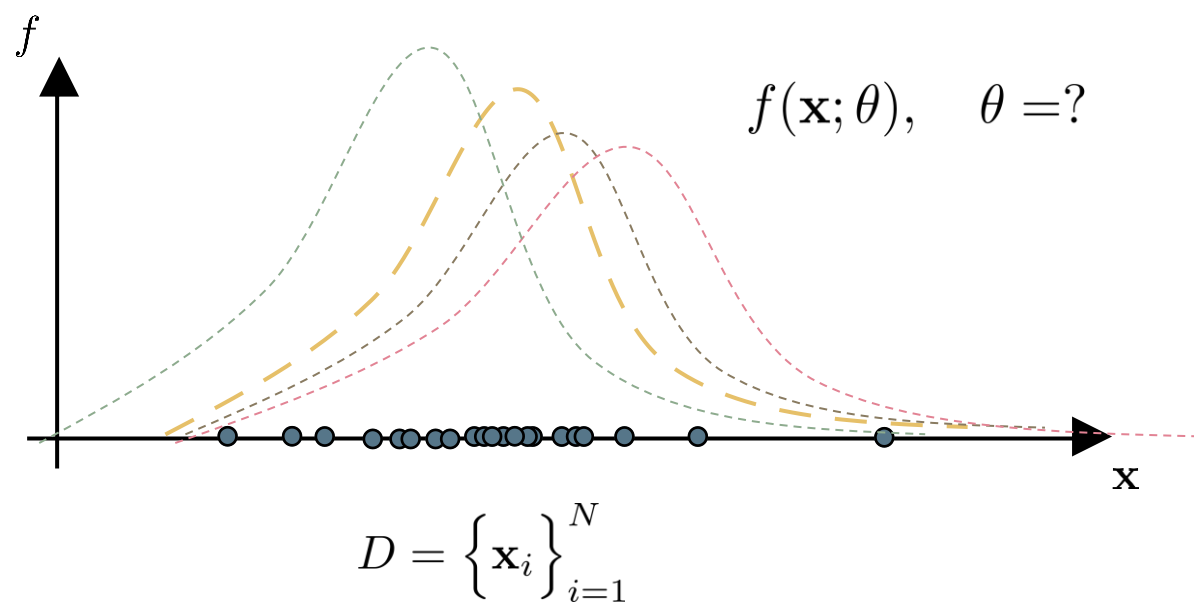
$$\int_{-\infty}^{\infty} f(x) dx = 1$$

Rozkład prawdopodobieństwa jako model generatywny



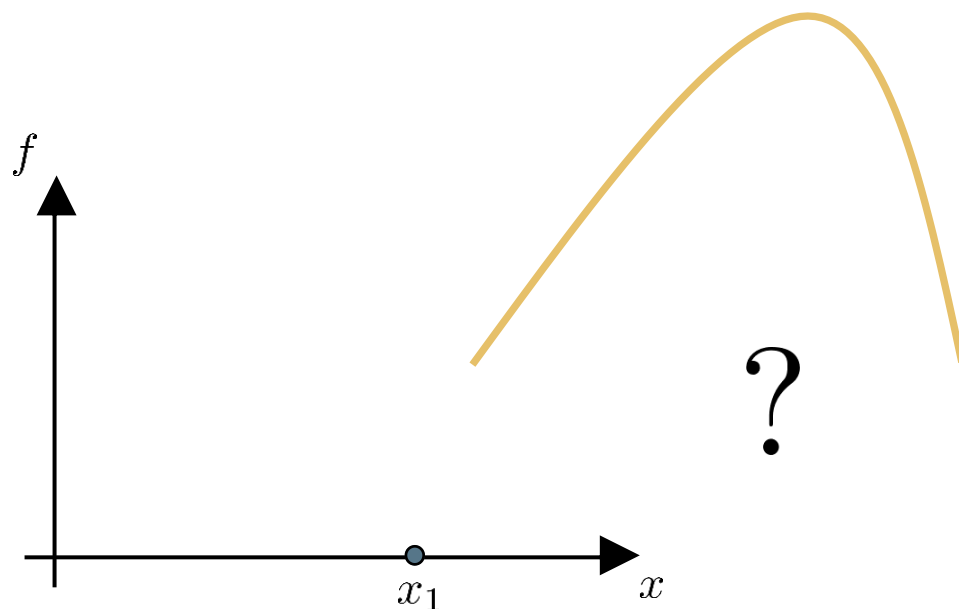
Dane pomiarowe traktujemy, jakby pochodziły z odpowiednio przygotowanego generatora liczb losowych

Dopasowanie rozkładu do danych



Dopasowanie rozkładu do danych

Zrób to sam



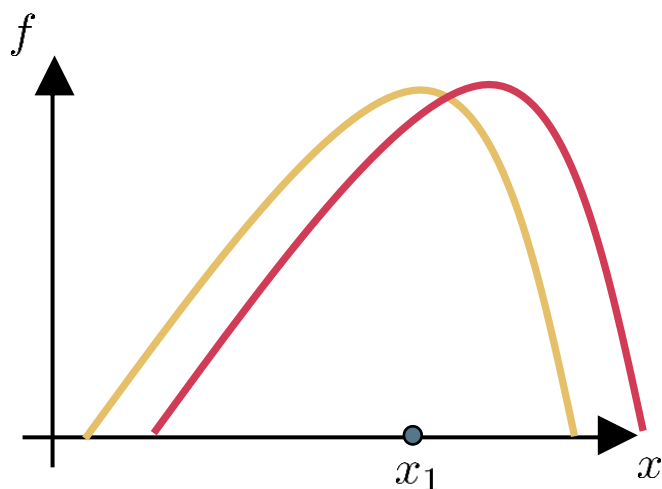
to jest pysk gorylcy

Gdzie go umieścisz na tle pojedynczego pomiaru?

Czym się będziesz kierować?

Dopasowanie rozkładu do danych

Zrób to sam



?



to jest pysk gorylcy

Gdzie go umieścisz na tle pojedynczego pomiaru?

Czym się będziesz kierować?

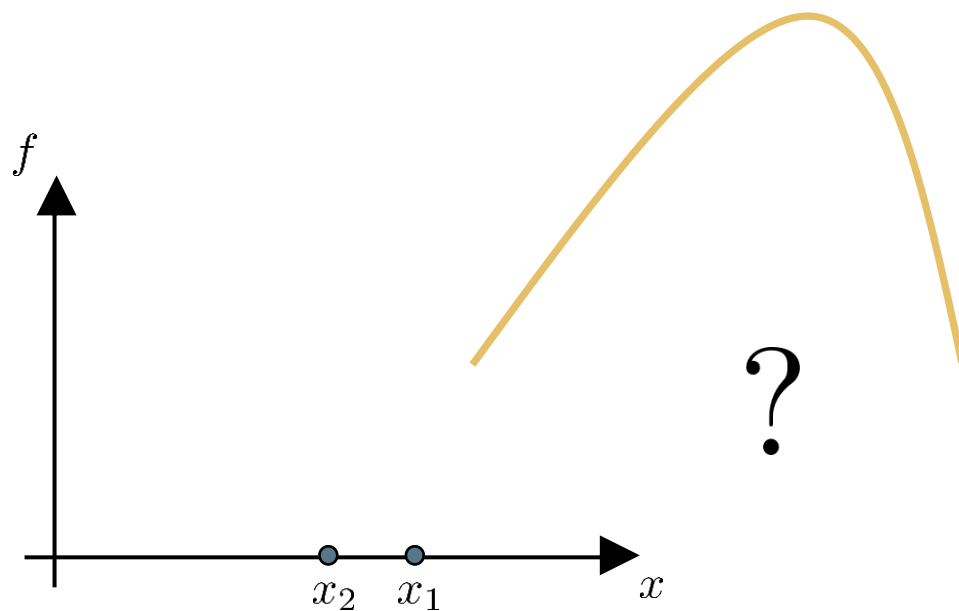
$$\hat{\theta}_{\text{ML}} = \arg \max_{\theta} f(x_1; \theta)$$

Metoda maksymalnej wiarygodności to **heurystyka**, czyli **zalgorytmizowana intuicja** / **droga na skróty**

Dopasowanie rozkładu do danych

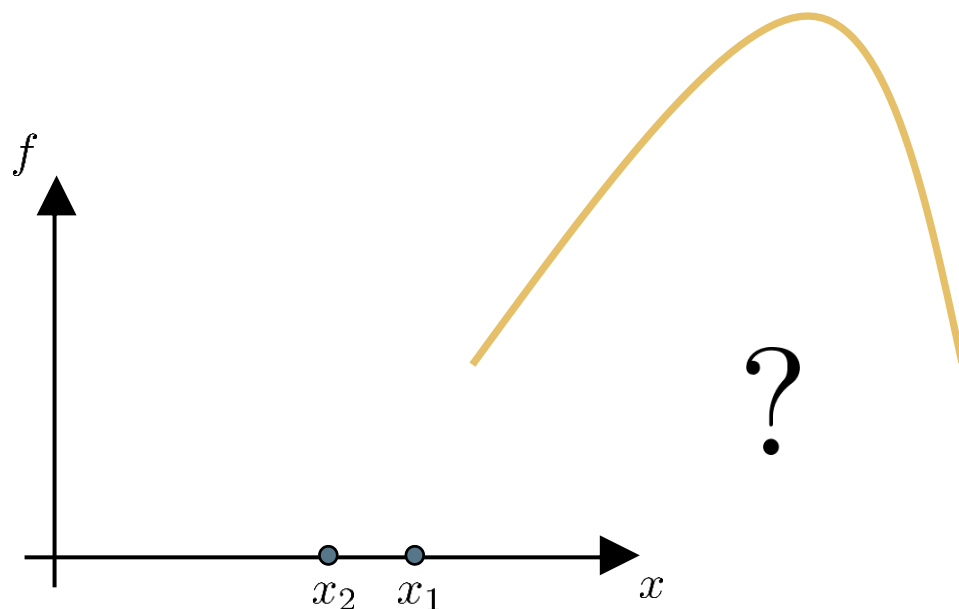
Zrób to sam

A co zrobisz teraz?



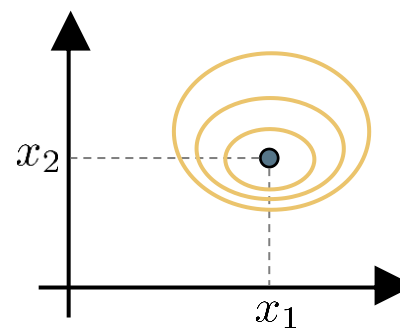
Dopasowanie rozkładu do danych

Zrób to sam



A co zrobisz teraz?

Może raczej tak:

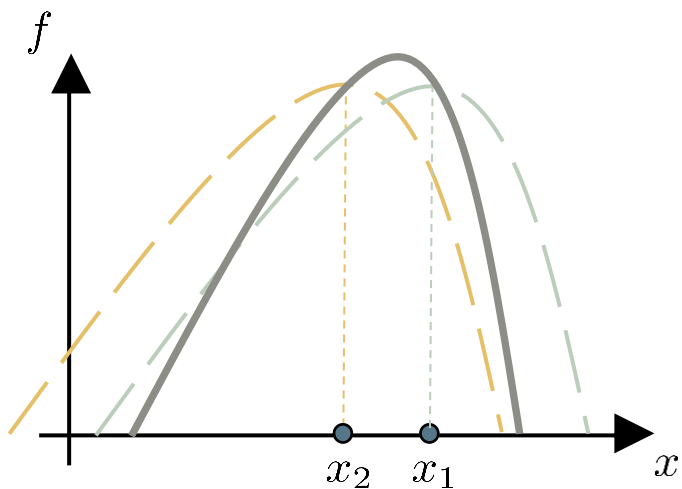


$$f(x_1, x_2; \theta), \quad \theta = ?$$

Dopasowanie rozkładu do danych

Zrób to sam

- To może lepiej tak



Zamiast $f(x_1, x_2; \theta)$

upraszczamy $f(x_1|\theta) \cdot f(x_2|\theta) \equiv p(\mathbf{X}|\theta)$

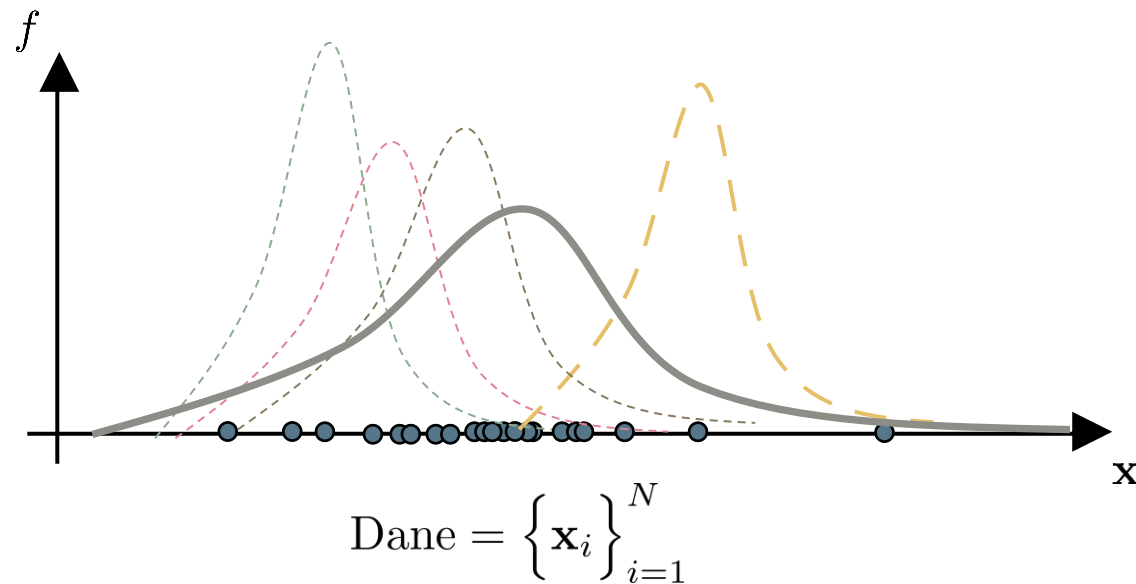
zachowujemy się więc, tak, jakby pomiary były niezależne:

$$\text{Cov}(x_1, x_2) = 0$$

... może są?

Funkcja wiarygodności

- Dla dowolnego zbioru danych



$$L(\theta) \equiv p(\mathbf{X}|\theta) = \prod_{i=1}^N p(\mathbf{x}_i|\theta)$$

$$\hat{\theta}_{\text{ML}} = \arg \max_{\theta} L(\theta)$$

Dwa problemy:

- niedomiar (ang. *underflow error*)
- pochodna iloczynu funkcji

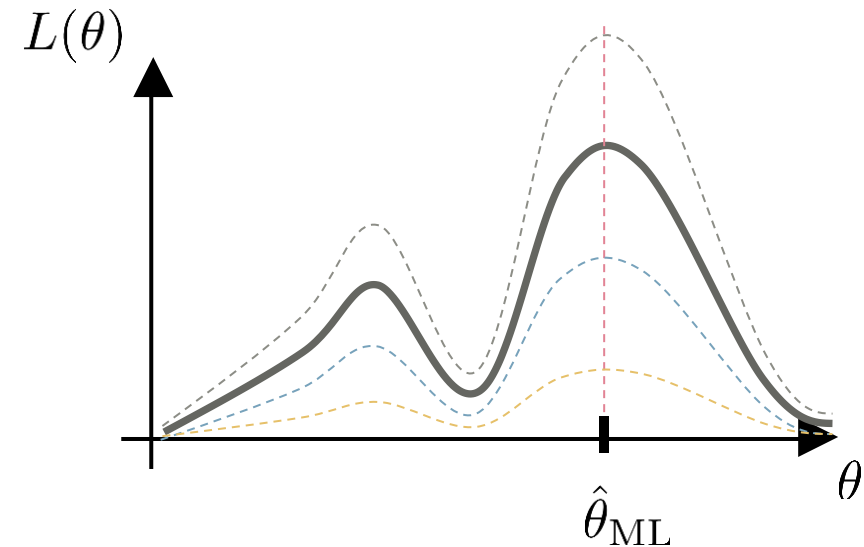
Funkcja wiarygodności

- Zmiana skali na logarytmiczną

$$\log \prod_{i=1}^N \heartsuit_i = \sum_{i=1}^N \log \heartsuit_i$$

$$\log L(\theta) = \sum_{i=1}^N \log p(\mathbf{x}_i|\theta)$$

$$\hat{\theta}_{\text{ML}} = \arg \max_{\theta} L(\theta) = \arg \max_{\theta} \log L(\theta)$$



Przykład zadania

- Dane pochodzą z rozkładu normalnego, ale nie znamy jego parametrów

$$\text{Dane} = \left\{ x_i \right\}_{i=1}^N$$

$$p(x; \theta) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right], \quad \theta = \begin{bmatrix} \mu \\ \sigma \end{bmatrix}$$

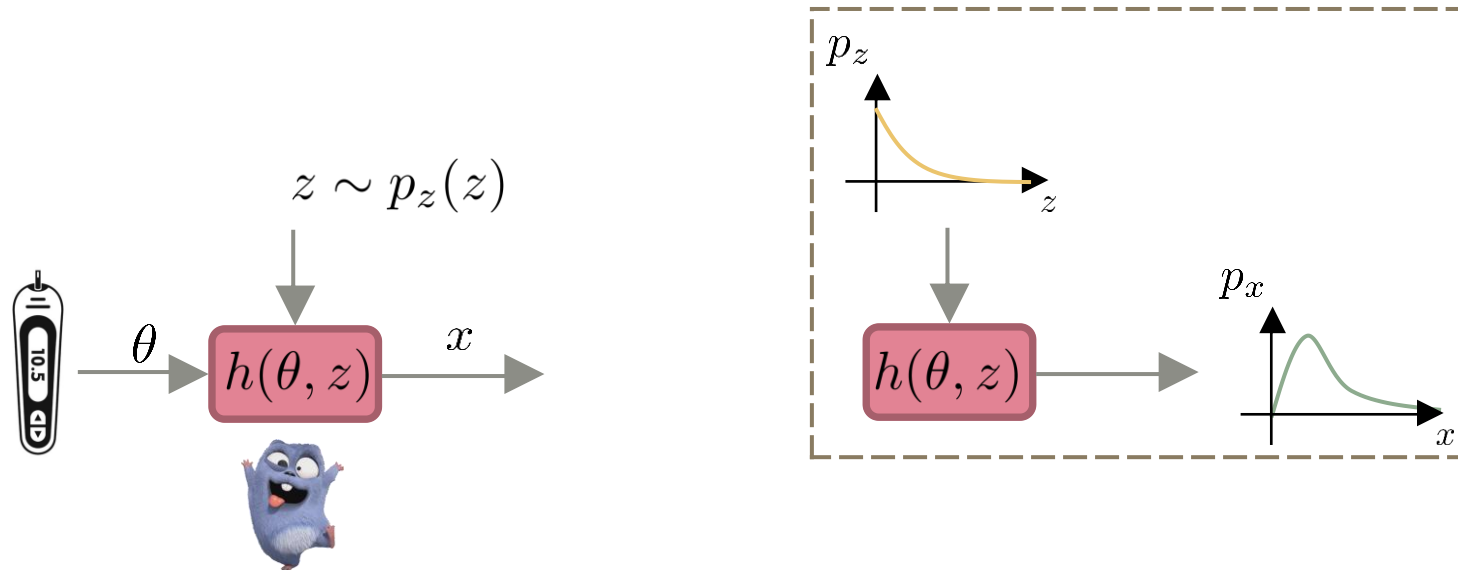
- Należy wyznaczyć wzór analityczny na parametry rozkładu tak, by maksymalizować funkcję wiarygodności

ESTYMACJA MIERZONEJ WIELKOŚCI

Metoda maksymalnej wiarygodności



Estymacja mierzonej wielkości



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

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

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

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

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

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

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

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

Zaczynamy rozwiązywać i okazuje się, że ...

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

Założmy, że $\mu_z = 0$

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

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

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

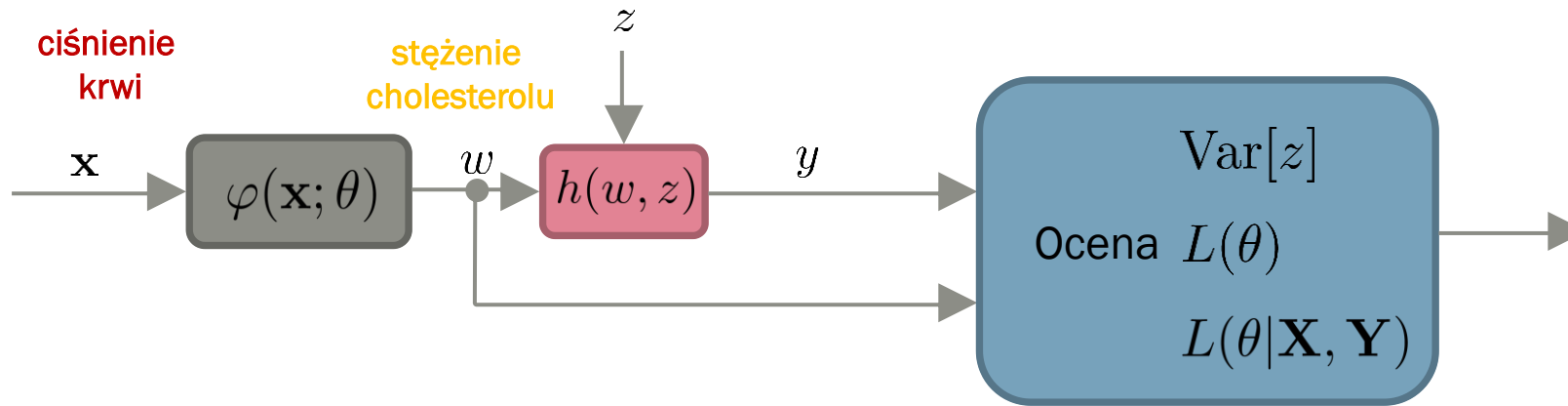
$$\hat{\theta}_{\text{ML}} = \hat{\theta}_{\text{LS}} = \arg \min_{\theta} \sum_{i=1}^N (x_i - \theta)^2$$

ESTYMACJA PARAMETRU MODELU

Metoda maksymalnej wiarygodności



Estymacja parametrów „ubranych” w model



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

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

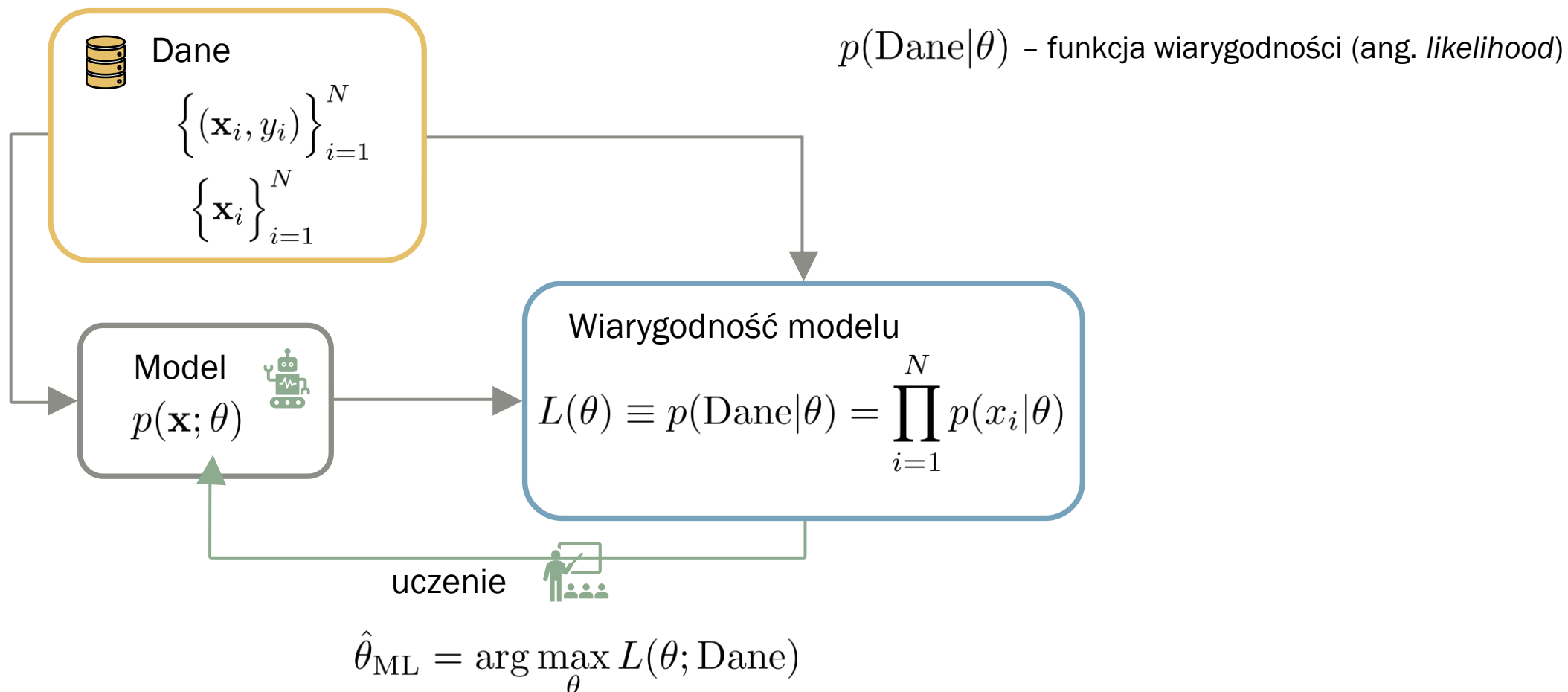
$$z = h^{-1}(w, y)$$

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

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

Dopasowanie modelu do danych

Metoda maksymalnej wiarygodności (ang. *Maximum Likelihood 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 naiwnie użyta 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 naiwnie użyta MNK a jak MMW?

UWAGI DODATKOWE



Sposoby wyrażenia modelu

W uczeniu maszynowym

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

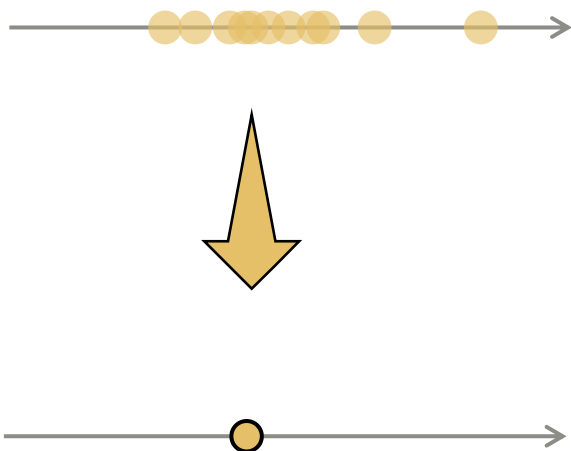
W modelowaniu statystycznym

$$y = \varphi(\mathbf{x}; \theta) + z$$

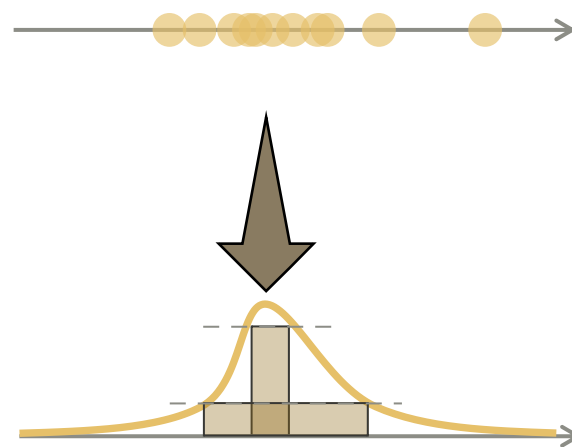
reszty (ang. *residuals*)

$$e_i = y_i - \hat{y}_i$$

Estymacja punktowa



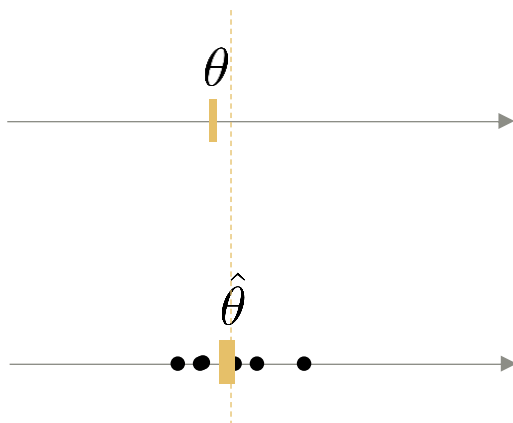
Estymacja przedziałowa



Jak myśleć o estymacji ?

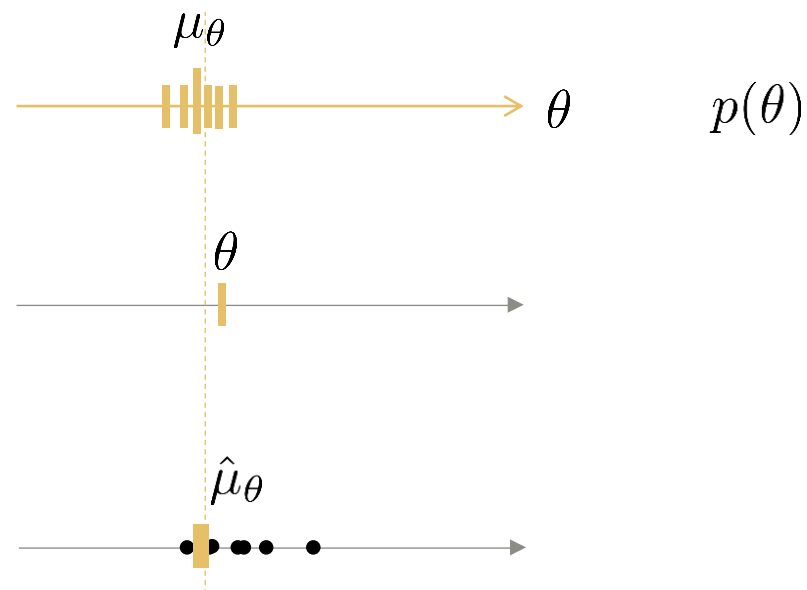
- Podejście częstościowe
(ang. *frequentist approach*)

estymuję nieznaną, ale ustaloną wartość



- Podejście bayesowskie
(ang. *bayesian approach*)

estymuję wartość wylosowaną z pewnego worka





THE END