



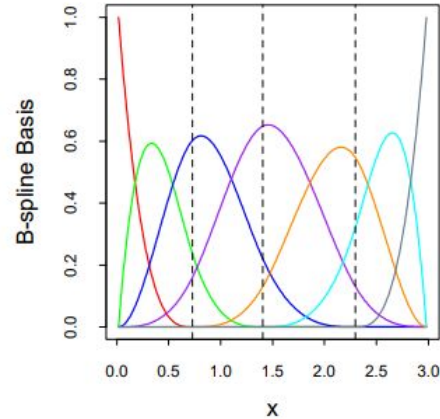
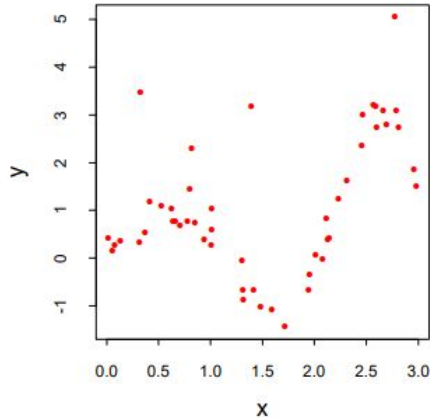
Bootstrap i metody największej wiarygodności

Przykład

dane treningowe $\mathbf{Z} = \{z_1, z_2, \dots, z_N\}$

$$z_i = (x_i, y_i), \quad i = 1, 2, \dots, N$$

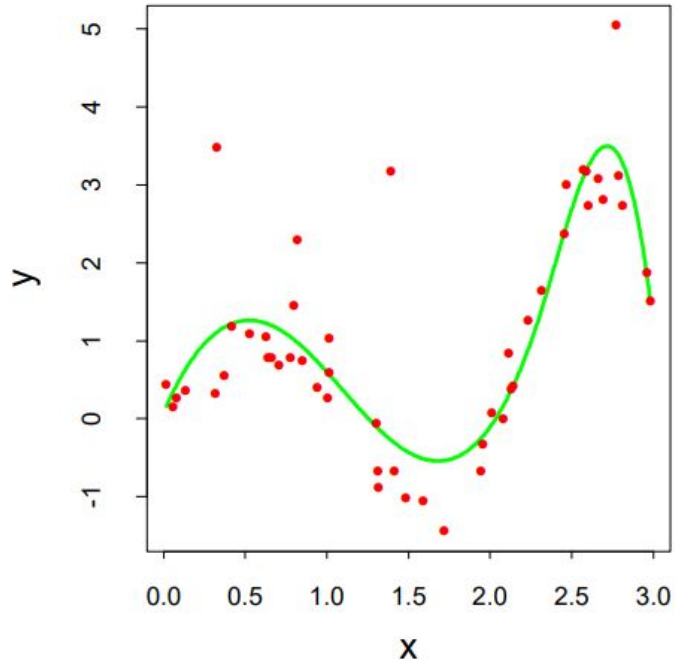
liniowe rozwinięcie funkcji bazowych B-splajn $\mu(x) = \sum_{j=1}^7 \beta_j h_j(x)$



\mathbf{H} - macierz $7 \times N$ z elementami $h_j(x_i)$

$$\hat{\beta} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}$$

$$\hat{\mu}(x) = \sum_{j=1}^7 \hat{\beta}_j h_j(x)$$



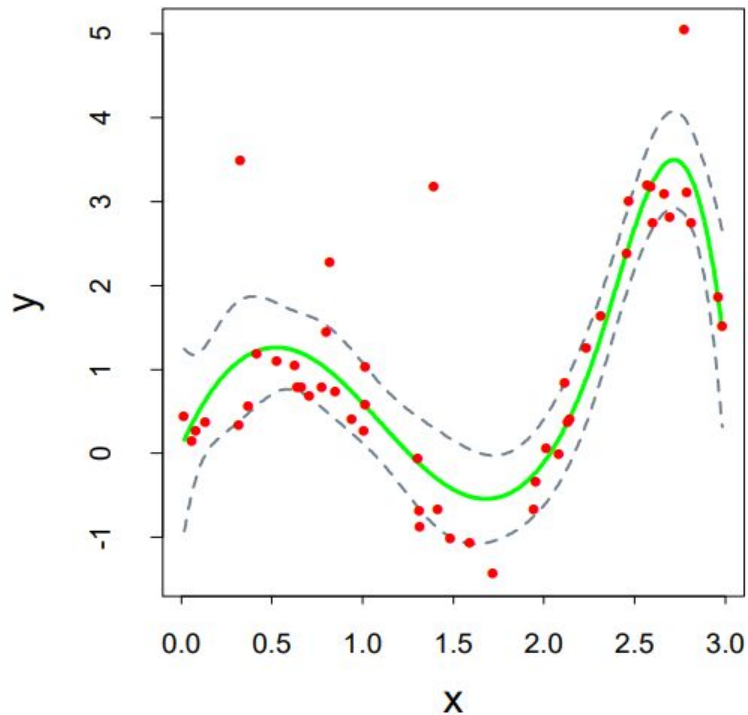
$$\hat{\sigma}^2 = \sum_{i=1}^N (y_i - \hat{\mu}(x_i))^2 / N$$

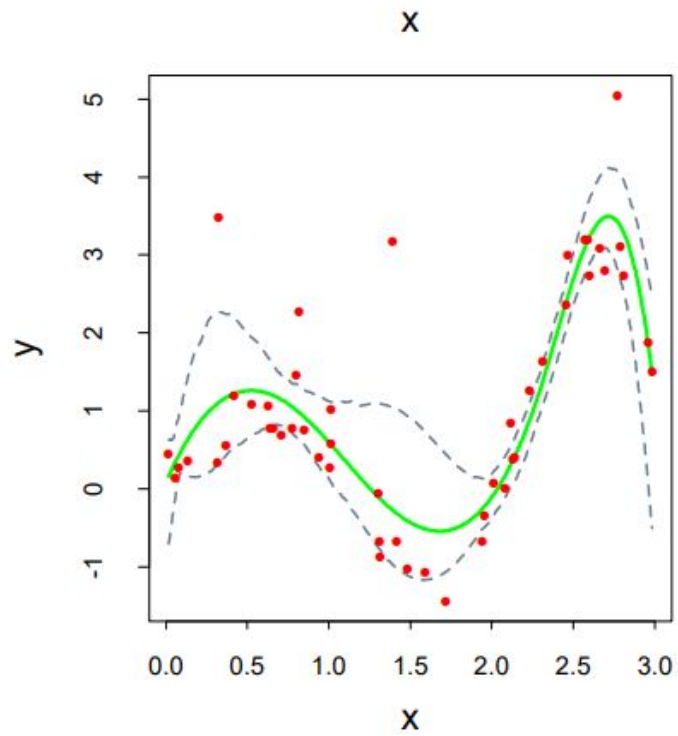
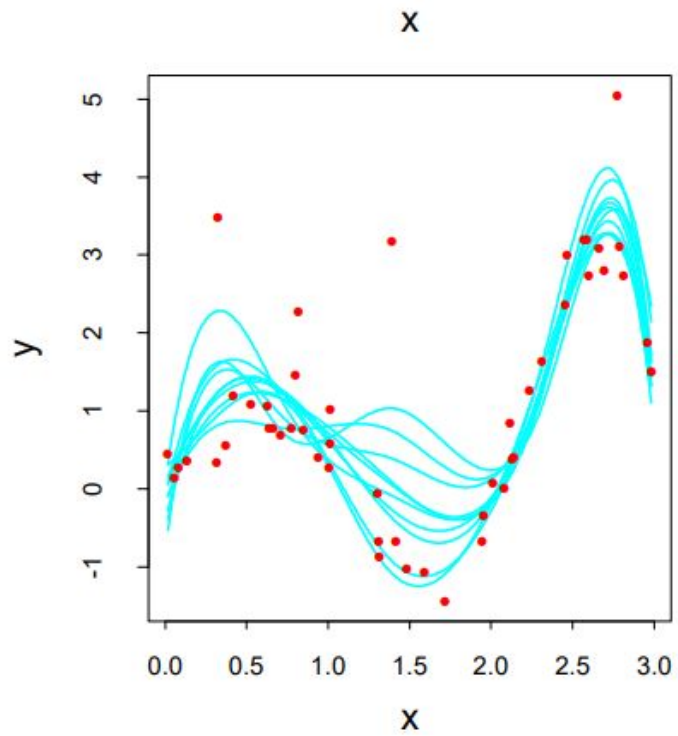
$$\widehat{\text{Var}}(\hat{\beta}) = (\mathbf{H}^T \mathbf{H})^{-1} \hat{\sigma}^2$$

$$h(x)^T = (h_1(x), h_2(x), \dots, h_7(x))$$

$$\widehat{\text{se}}[\hat{\mu}(x)] = [h(x)^T (\mathbf{H}^T \mathbf{H})^{-1} h(x)]^{\frac{1}{2}} \hat{\sigma}$$

$$\hat{\mu}(x) \pm 1.96 \cdot \widehat{\text{se}}[\hat{\mu}(x)]$$





- metoda ładowania początkowego

$$Y = \mu(X) + \varepsilon; \quad \varepsilon \sim N(0, \sigma^2)$$
$$\mu(x) = \sum_{j=1}^7 \beta_j h_j(x).$$

- parametryczna metoda ładowania początkowego

$$y_i^* = \hat{\mu}(x_i) + \varepsilon_i^*; \quad \varepsilon_i^* \sim N(0, \hat{\sigma}^2); \quad i = 1, 2, \dots, N$$

- funkcja oszacowana z próby bootstrapowej

$$\hat{\mu}^*(x) = h(x)^T (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}^*$$
$$\hat{\mu}^*(x) \sim N(\hat{\mu}(x), h(x)^T (\mathbf{H}^T \mathbf{H})^{-1} h(x) \hat{\sigma}^2)$$

Metoda największej wiarygodności

- gęstość prawdopodobieństwa lub funkcja masy prawdopodobieństwa

$$z_i \sim g_\theta(z)$$

- przykładowo, gdy Z jest z rozkładu normalnego

$$\theta = (\mu, \sigma^2)$$

$$g_\theta(z) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}(z-\mu)^2/\sigma^2}$$

- funkcja wiarygodności

$$L(\theta; \mathbf{Z}) = \prod_{i=1}^N g_\theta(z_i)$$

- logarytm funkcji wiarygodności

$$\begin{aligned} \ell(\theta; \mathbf{Z}) &= \sum_{i=1}^N \ell(\theta; z_i) \\ &= \sum_{i=1}^N \log g_\theta(z_i) \end{aligned}$$

- funkcja punktacji (score function)

$$\dot{\ell}(\theta; \mathbf{Z}) = \sum_{i=1}^N \dot{\ell}(\theta; z_i) \quad \dot{\ell}(\theta; z_i) = \partial \ell(\theta; z_i) / \partial \theta$$

- macierz informacji

$$\mathbf{I}(\theta) = - \sum_{i=1}^N \frac{\partial^2 \ell(\theta; z_i)}{\partial \theta \partial \theta^T}$$

- wartość oczekiwana

$$\mathbf{i}(\theta) = \mathbf{E}_{\theta}[\mathbf{I}(\theta)]$$

- prawdziwą wartość θ oznaczamy θ_0

- graniczny rozkład

$$\hat{\theta} \rightarrow N(\theta_0, \mathbf{i}(\theta_0)^{-1})$$

- przybliżony rozkład próbkowy

$$N(\hat{\theta}, \mathbf{i}(\hat{\theta})^{-1}) \quad N(\hat{\theta}, \mathbf{I}(\hat{\theta})^{-1})$$

- oszacowania dla błędów standardowych

$$\sqrt{\mathbf{i}(\hat{\theta})_{jj}^{-1}} \quad \sqrt{\mathbf{I}(\hat{\theta})_{jj}^{-1}}$$

- przedziały ufności

$$\hat{\theta}_j - z^{(1-\alpha)} \cdot \sqrt{\mathbf{i}(\hat{\theta})_{jj}^{-1}} \quad \hat{\theta}_j - z^{(1-\alpha)} \cdot \sqrt{\mathbf{I}(\hat{\theta})_{jj}^{-1}}$$

- dokładniejsze przedziały ufności

$$2[\ell(\hat{\theta}) - \ell(\theta_0)] \sim \chi_p^2$$

- wynikowy przedział ufności

$$2[\ell(\hat{\theta}) - \ell(\theta_0)] \leq \chi_p^{2(1-2\alpha)}$$

- parametr

$$\theta = (\beta, \sigma^2)$$

- funkcja log-wiarogodności

$$\ell(\theta) = -\frac{N}{2} \log \sigma^2 2\pi - \frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - h(x_i)^T \beta)^2$$

- estymator największej warogodności

$$\hat{\beta} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y},$$
$$\hat{\sigma}^2 = \frac{1}{N} \sum (y_i - \hat{\mu}(x_i))^2$$

- blok macierzy informacji

$$\mathbf{I}(\beta) = (\mathbf{H}^T \mathbf{H}) / \sigma^2$$

- oszacowana wariancja

$$(\mathbf{H}^T \mathbf{H})^{-1} \hat{\sigma}^2$$