

Návrhy experimentů v neparametrické regresi

Zdeněk Hlávka

Univerzita Karlova v Praze
Matematicko-fyzikální fakulta
Katedra pravděpodobnosti a matematické statistiky

2.5.2012

Návrhy experimentů v neparametrické regresi

- Parametrická (lineární) regrese:
 - problémy při chybné specifikaci regresní funkce.
- Neparametrická regrese:
 - odhad regresní funkce,
 - odhad nulových bodů (polohy extrémů),
ZH: zobecnění pro slabě závislá pozorování.
 - optimální návrhy experimentů:
 - přehled literatury (odhady regresní funkce),
 - ZH: návrh experimentu pro odhad nulového bodu.

Historická poznámka

- Lineární regrese (metoda nejmenších čtverců):

Legendre (1805) *Nouvelles méthodes pour la détermination des orbites des comètes*.

Gauss (1809) *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*.

Galton (1886) Regression towards mediocrity in hereditary stature. *Journ. Anthropol. Inst.* 15, 246–263.

- Neparametrická regrese:

Nadaraya (1964), Watson (1964), Gasser & Müller (1984), Fan & Gijbels (1996).

- Nulové body: Müller (1985), Ziegler (2003), Wieczorek & Ziegler (2010).
- Optimální návrh experimentu: H.-G. Müller (1984), Faraway (1990), W.G. Müller (1996), Cheng et al (1998), Fedorov et al (1999), Titterington (2001), Biedermann & Dette (2001a,b), Hlávka (2011, 2012).

Regresní model

Pozorujeme dvojice (X_i, Y_i) , $i = 1, \dots, n$.

Předpokládáme, že

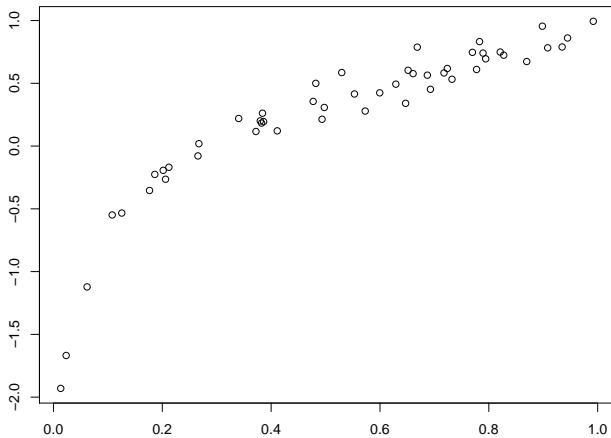
$$Y_i = m(X_i) + \varepsilon_i,$$

kde ε_i jsou náhodné chyby ($E\varepsilon_i = 0$, $\text{Var}\varepsilon_i = \sigma^2$, X_i a ε_i nezávislé), a chceme odhadnout funkci $m(\cdot)$.

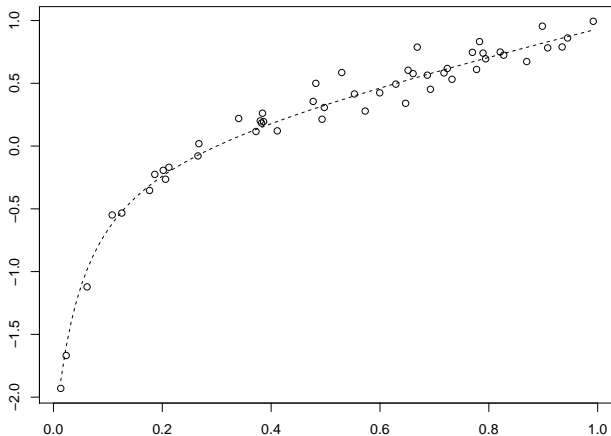
Parametrická regrese:

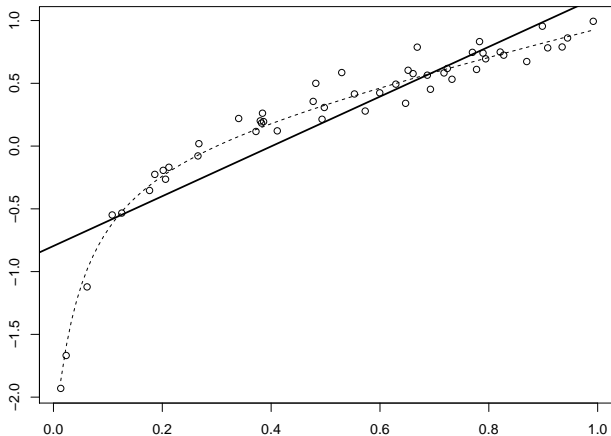
Nejčastěji se předpokládá, že $m(\cdot)$ je lineární a závisí pouze na několika parametrech (lineární model).

Příklad: simulovaná data



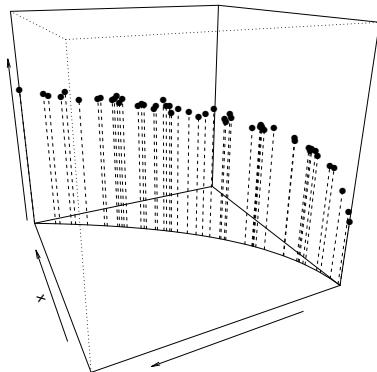
Příklad: simulovaná data



Příklad: simulovaná dataLineární regrese: $\hat{Y}_i = \alpha + \beta X_i, i = 1, \dots, n.$ 

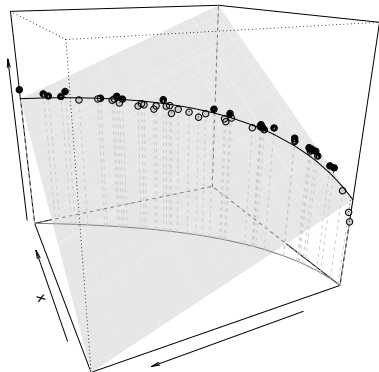
Příklad: simulovaná data

Kvadratická regrese: $\hat{Y}_i = \alpha + \beta_1 X_i + \beta_2 X_i^2, i = 1, \dots, n.$



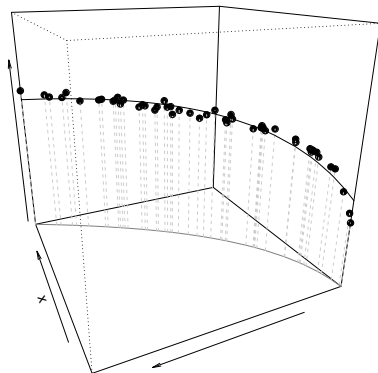
Příklad: simulovaná data

Kvadratická regrese: $\hat{Y}_i = \alpha + \beta_1 X_i + \beta_2 X_i^2$, $i = 1, \dots, n$.



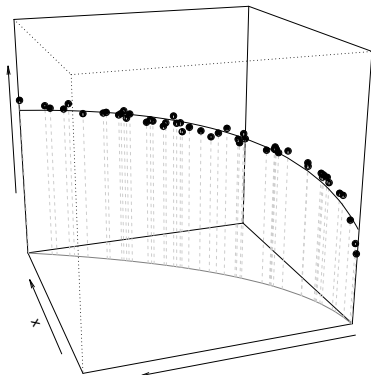
Příklad: simulovaná data

Kvadratická regrese: $\hat{Y}_i = \alpha + \beta_1 X_i + \beta_2 X_i^2, i = 1, \dots, n.$



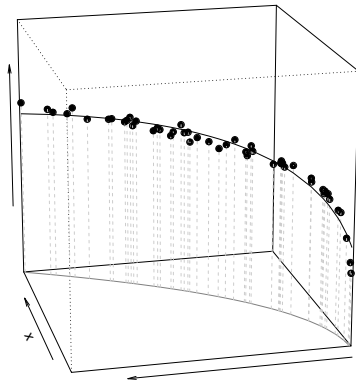
Příklad: simulovaná data

Kvadratická regrese: $\hat{Y}_i = \alpha + \beta_1 X_i + \beta_2 X_i^2$, $i = 1, \dots, n$.



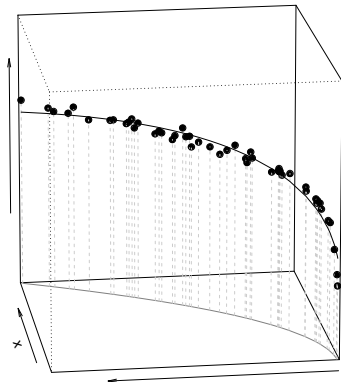
Příklad: simulovaná data

Kvadratická regrese: $\hat{Y}_i = \alpha + \beta_1 X_i + \beta_2 X_i^2$, $i = 1, \dots, n$.



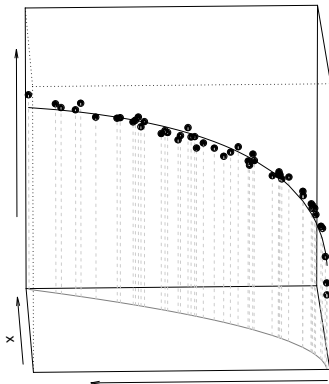
Příklad: simulovaná data

Kvadratická regrese: $\hat{Y}_i = \alpha + \beta_1 X_i + \beta_2 X_i^2$, $i = 1, \dots, n$.



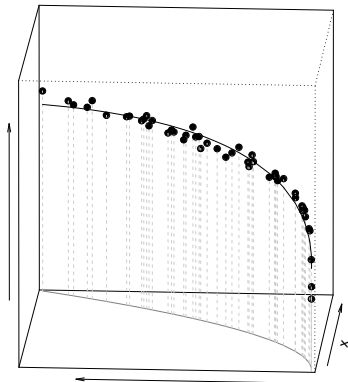
Příklad: simulovaná data

Kvadratická regrese: $\hat{Y}_i = \alpha + \beta_1 X_i + \beta_2 X_i^2$, $i = 1, \dots, n$.



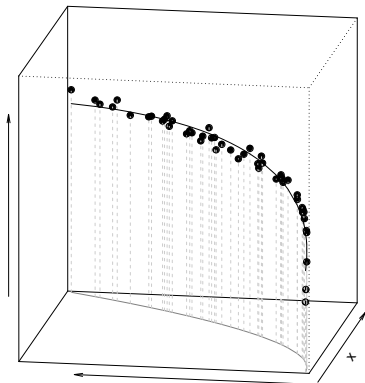
Příklad: simulovaná data

Kvadratická regrese: $\hat{Y}_i = \alpha + \beta_1 X_i + \beta_2 X_i^2, i = 1, \dots, n.$



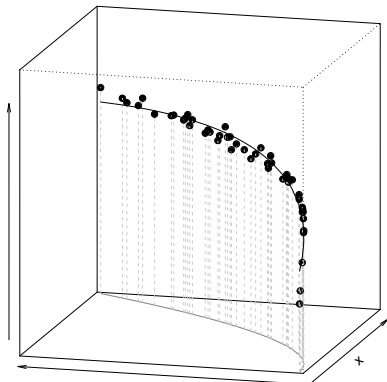
Příklad: simulovaná data

Kvadratická regrese: $\hat{Y}_i = \alpha + \beta_1 X_i + \beta_2 X_i^2, i = 1, \dots, n.$



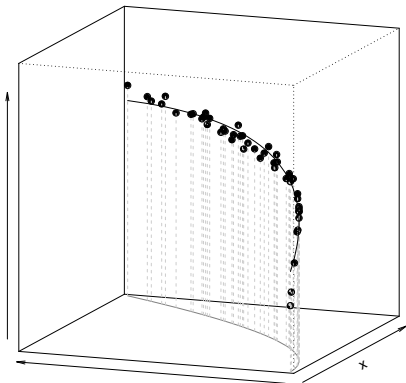
Příklad: simulovaná data

Kvadratická regrese: $\hat{Y}_i = \alpha + \beta_1 X_i + \beta_2 X_i^2, i = 1, \dots, n.$



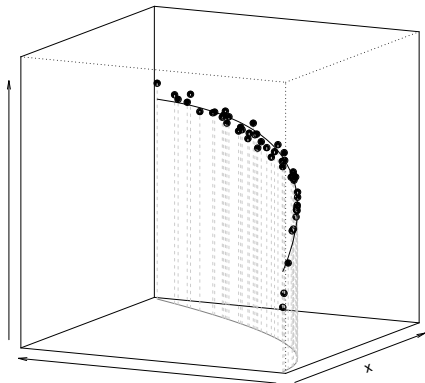
Příklad: simulovaná data

Kvadratická regrese: $\hat{Y}_i = \alpha + \beta_1 X_i + \beta_2 X_i^2$, $i = 1, \dots, n$.



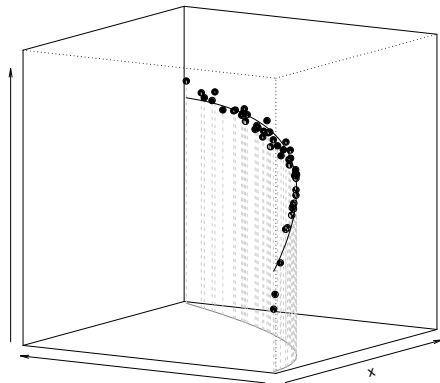
Příklad: simulovaná data

Kvadratická regrese: $\hat{Y}_i = \alpha + \beta_1 X_i + \beta_2 X_i^2, i = 1, \dots, n.$



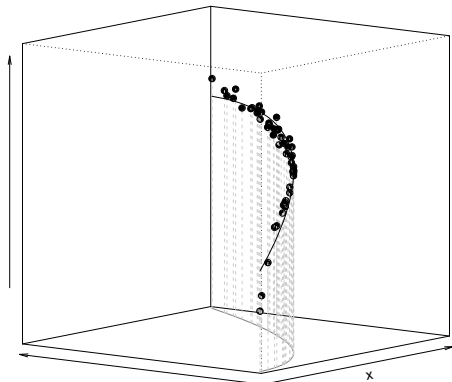
Příklad: simulovaná data

Kvadratická regrese: $\hat{Y}_i = \alpha + \beta_1 X_i + \beta_2 X_i^2, i = 1, \dots, n.$



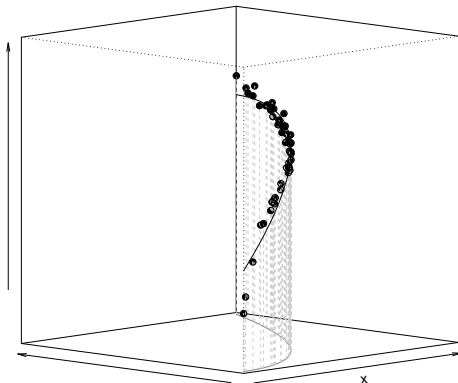
Příklad: simulovaná data

Kvadratická regrese: $\hat{Y}_i = \alpha + \beta_1 X_i + \beta_2 X_i^2, i = 1, \dots, n.$



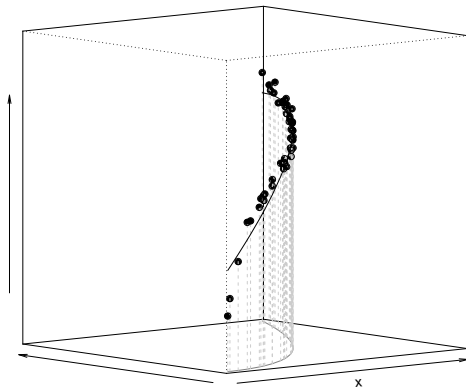
Příklad: simulovaná data

Kvadratická regrese: $\hat{Y}_i = \alpha + \beta_1 X_i + \beta_2 X_i^2$, $i = 1, \dots, n$.



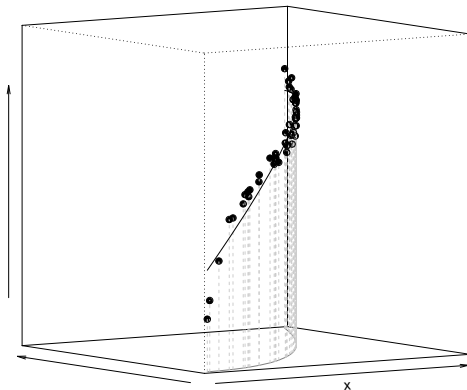
Příklad: simulovaná data

Kvadratická regrese: $\hat{Y}_i = \alpha + \beta_1 X_i + \beta_2 X_i^2, i = 1, \dots, n.$



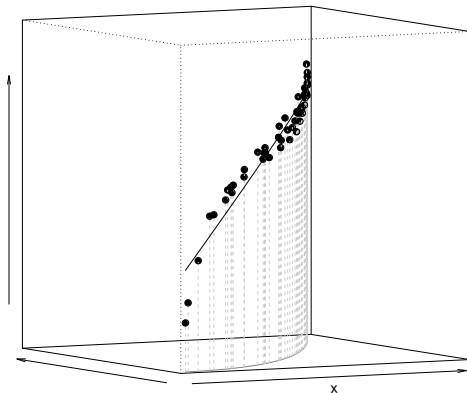
Příklad: simulovaná data

Kvadratická regrese: $\hat{Y}_i = \alpha + \beta_1 X_i + \beta_2 X_i^2, i = 1, \dots, n.$



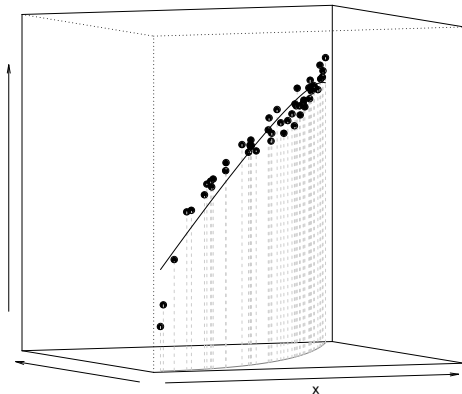
Příklad: simulovaná data

Kvadratická regrese: $\hat{Y}_i = \alpha + \beta_1 X_i + \beta_2 X_i^2, i = 1, \dots, n.$



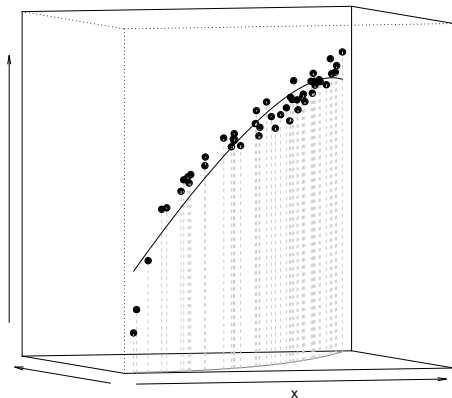
Příklad: simulovaná data

Kvadratická regrese: $\hat{Y}_i = \alpha + \beta_1 X_i + \beta_2 X_i^2, i = 1, \dots, n.$



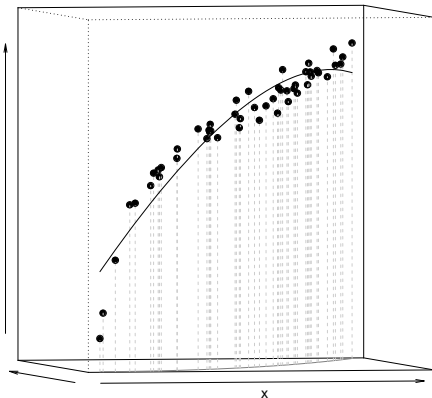
Příklad: simulovaná data

Kvadratická regrese: $\hat{Y}_i = \alpha + \beta_1 X_i + \beta_2 X_i^2, i = 1, \dots, n.$



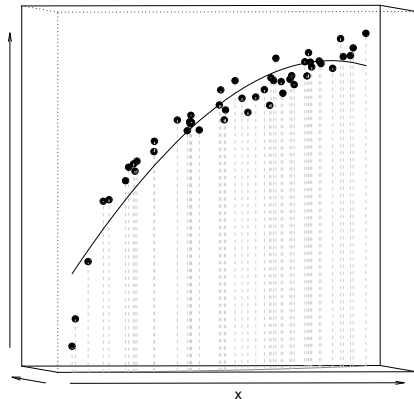
Příklad: simulovaná data

Kvadratická regrese: $\hat{Y}_i = \alpha + \beta_1 X_i + \beta_2 X_i^2, i = 1, \dots, n.$



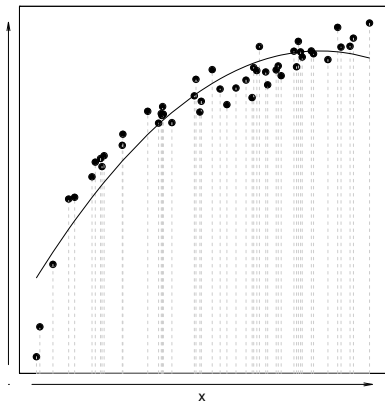
Příklad: simulovaná data

Kvadratická regrese: $\hat{Y}_i = \alpha + \beta_1 X_i + \beta_2 X_i^2, i = 1, \dots, n.$



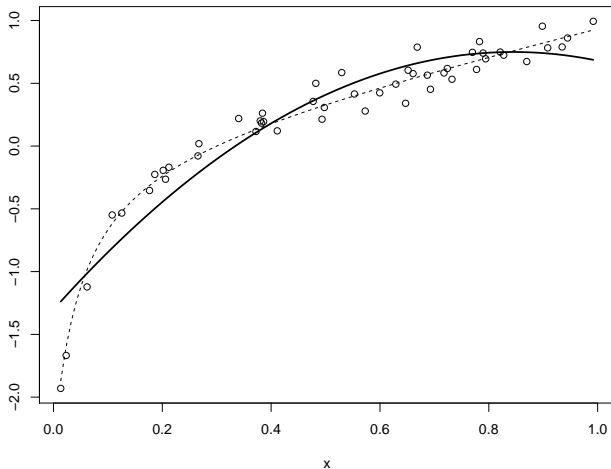
Příklad: simulovaná data

Kvadratická regrese: $\hat{Y}_i = \alpha + \beta_1 X_i + \beta_2 X_i^2, i = 1, \dots, n.$



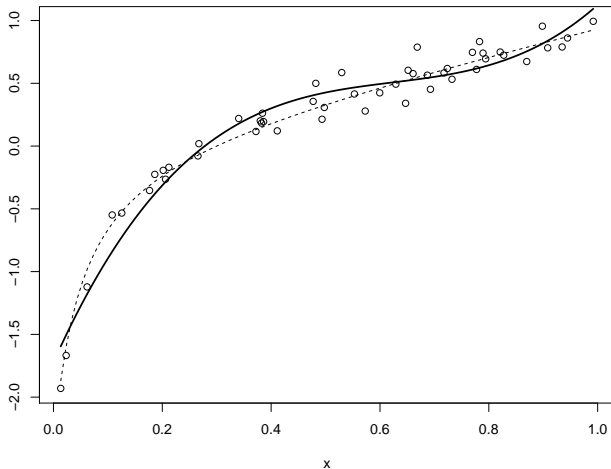
Příklad: simulovaná data

Kvadratická regrese: $\hat{Y}_i = \alpha + \beta_1 X_i + \beta_2 X_i^2$, $i = 1, \dots, n$.



Příklad: simulovaná data

Kubická regrese: $\hat{Y}_i = \alpha + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3, i = 1, \dots, n.$



Neparametrické regresní modely

Flexibilní modely, které nevyžadují předpoklady o tvaru regresní funkce:

- spliny,
- **jádrové vyhlazování** (lokálně konstantní a lokálně polynomické),
- regresní stromy, neuronové sítě atd.

Jádrové odhady navrhli Nadaraya (1964) a Watson (1964):

$$m_h^{NW}(x) = \frac{\sum_{i=1}^n K_h(x - X_i) Y_i}{\sum_{i=1}^n K_h(x - X_i)} = \sum_{i=1}^n \frac{K_h(x - X_i)}{\sum_{j=1}^n K_h(x - X_j)} Y_i,$$

kde $K_h(x) = K(x/h)/h$, $K(\cdot)$ je *jádrová funkce* a h je *bandwidth*.

Nadaraya (1964) On estimating regression, *Theory of Probability and its Applications* 9, 141—142.

Watson (1964) Smooth regression analysis, *Sankyā A* 26, 359—372.

My se nyní zaměříme na Gasser-Müllerův odhad vhodný pro plánované experimenty (pevně zvolené x_i), ale obdobné výsledky lze odvodit i pro jiné odhady.

Gasser-Müllerův odhad

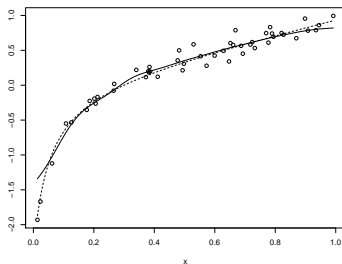
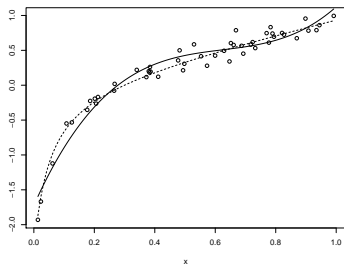
$$m_h^{GM}(x) = \sum_{i=1}^n \int_{s_{i-1}}^{s_i} K_h(x-u) du Y_i,$$

kde $s_i = (x_{(i)} + x_{(i+1)})/2$ a $K(\cdot)$ je jádrová funkce řádu (ν, k) .

Jedná se opět o vážené průměry Y_i , protože

$$\sum_{i=1}^n \int_{s_{i-1}}^{s_i} K_h(x-u) du = \int K(x) dx = 1.$$

Gasser & Müller (1984) Estimating regression functions and their derivatives by the kernel method, *Scand J Statist* 11, 171–185.

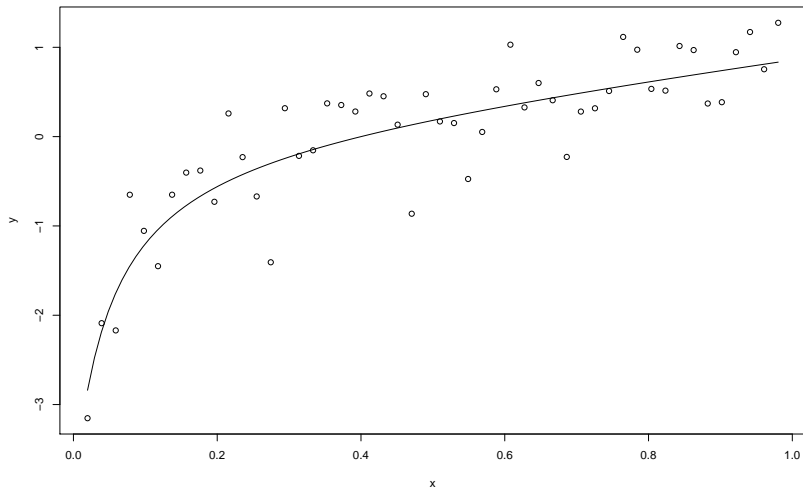


O neparametrické regresi byly napsány desítky monografií.

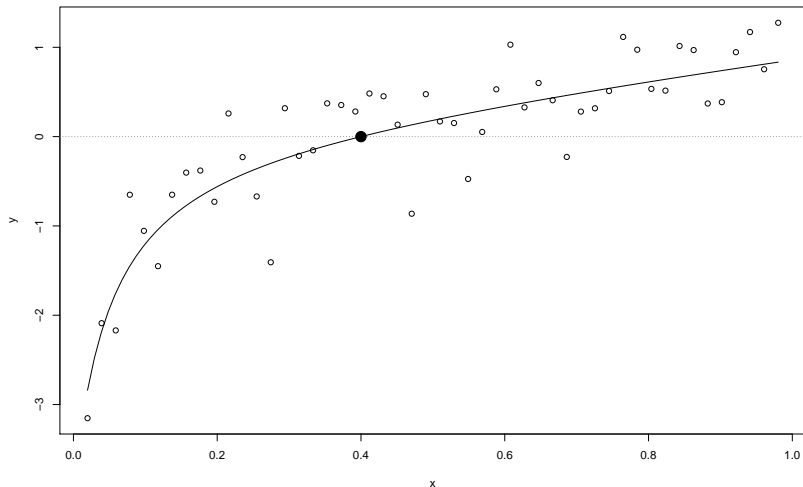
Zajímavý problém: odhad polohy bodů, ve kterých regresní funkce (nebo její derivace) nabývá dané hodnoty.

Odhadujeme **nulový bod**, t.j. polohu bodu, ve kterém funkce $m(\cdot)$ protíná osu x .

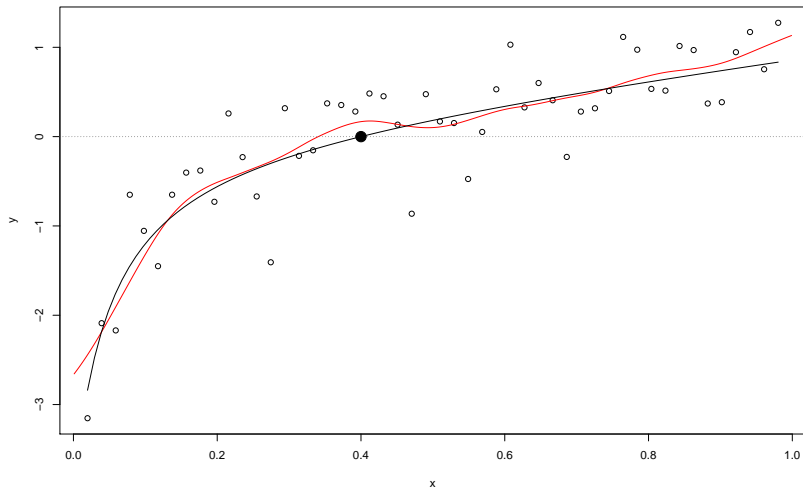
Odhadujeme **nulový bod**, t.j. polohu bodu, ve kterém funkce $m(\cdot)$ protíná osu x .



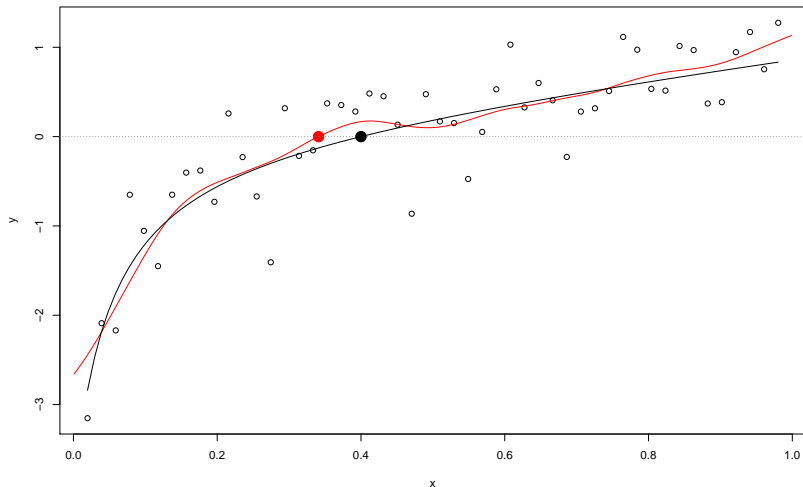
Odhadujeme **nulový bod**, t.j. polohu bodu, ve kterém funkce $m(\cdot)$ protíná osu x .



Odhadujeme **nulový bod**, t.j. polohu bodu, ve kterém funkce $m(\cdot)$ protíná osu x .



Odhadujeme **nulový bod**, t.j. polohu bodu, ve kterém funkce $m(\cdot)$ protíná osu x .



Neparametrický odhad nulového bodu

Neznámý parametr (nulový bod):

$$\xi : m(\xi) = 0$$

Odhad (empirický nulový bod):

$$\xi_n : m_h^{GM}(\xi_n) = 0$$

Vlastnosti ξ_n jako odhadu ξ jsou v článku Müller (1985) odvozeny pro iid náhodné chyby.

Müller, H.-G. (1985). Kernel estimators of zeros and of location and size of extrema of regression functions, *Scand. J. Statist.* 12: 221–232.

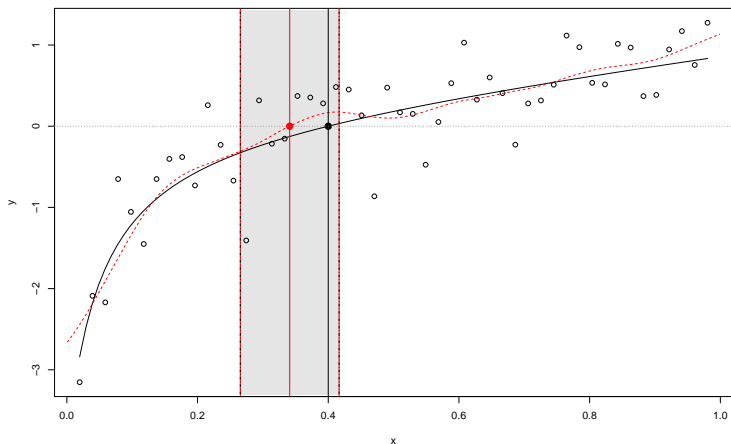
- 1 Regresní funkce $m(\cdot)$ je $3 \times$ spojitě diferencovatelná, $m'(\xi) \neq 0$.
Jádrová funkce $K(\cdot)$ je diferencovatelná a $K'(\cdot)$ je Lipschitzovsky spojitá.
- 2 $\liminf_{n \rightarrow \infty} nh_n^3 > 0$, $nh_n^3 / \log n \rightarrow \infty$, ε_i jsou iid, $E\varepsilon_i = 0$, $E|\varepsilon_i|^r < \infty$ pro nějaké $r > 2$ a $\liminf_{n \rightarrow \infty} h_n n^{1-2/r} / \log n > 0$.
- 3 Existují $x_l < \xi < x_u$, $c > 0$ a $\tau \geq 1$ takové, že $m(\cdot)$ je ryze monotónní na intervalu $\langle x_l, x_u \rangle$ a $|m(t)| > c|t - \xi|^\tau$ pro $t \in \langle x_l, x_u \rangle$.

Müller (1985), Theorem 3.1: Předpokládejme 1–3 a že body měření x_i jsou rovnoměrně rozložené na intervalu $(0, 1)$. Pokud $nh_n^5 \rightarrow d^2 \geq 0$, pak

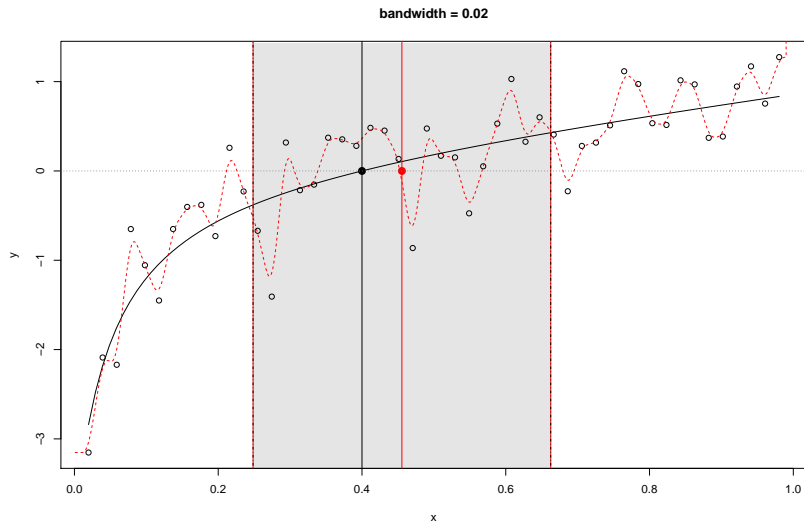
$$(nh_n)^{1/2}(\xi_n - \xi) \xrightarrow{\mathcal{D}} N \left(-\frac{dm^{(2)}(\xi)B_2}{m'(\xi)}, \frac{\sigma^2 V}{\{m'(\xi)\}^2} \right),$$

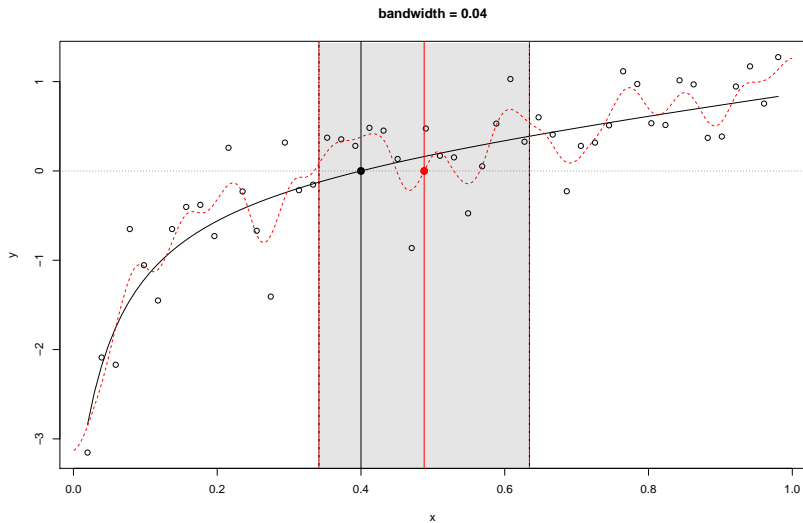
kde $B_2 = (1/2) \int K(s)s^2 ds$ a $V = \int K^2(s) ds$.

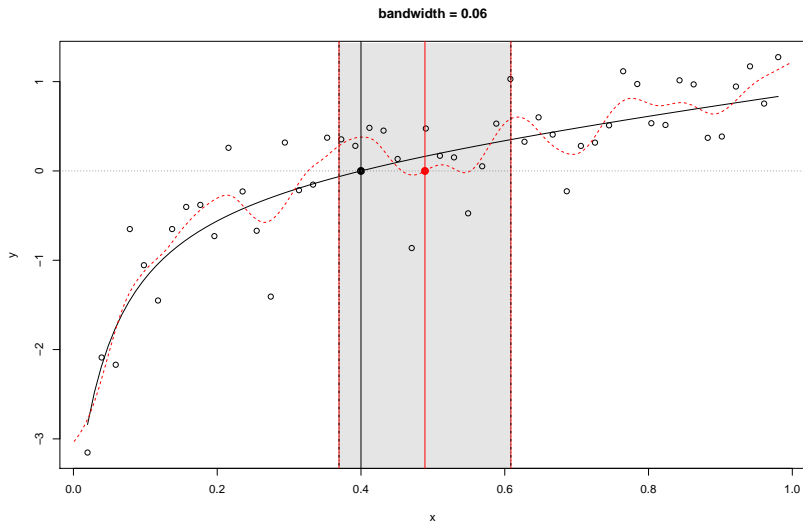
$$P \left\{ \left(\xi_n - \frac{u_{1-\alpha/2} \sigma \sqrt{V}}{\sqrt{nh_n} m'(\xi)}, \xi_n + \frac{u_{1-\alpha/2} \sigma \sqrt{V}}{\sqrt{nh_n} m'(\xi)} \right) \ni \xi \right\} \approx 1 - \alpha$$

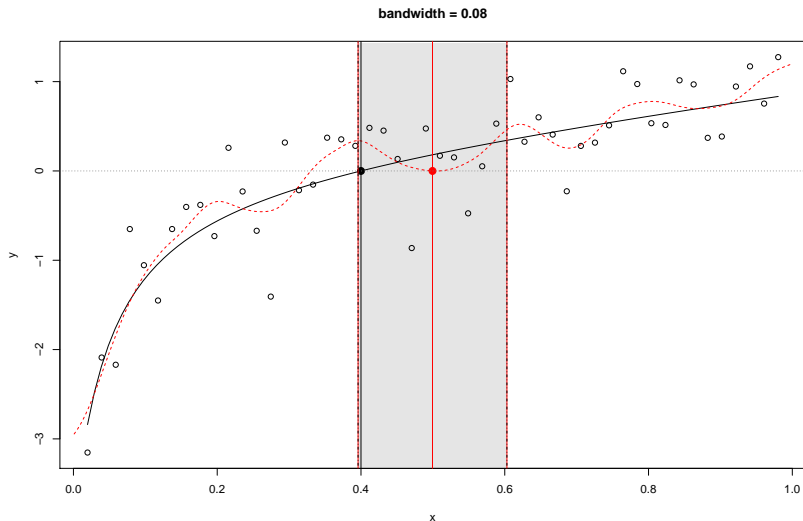


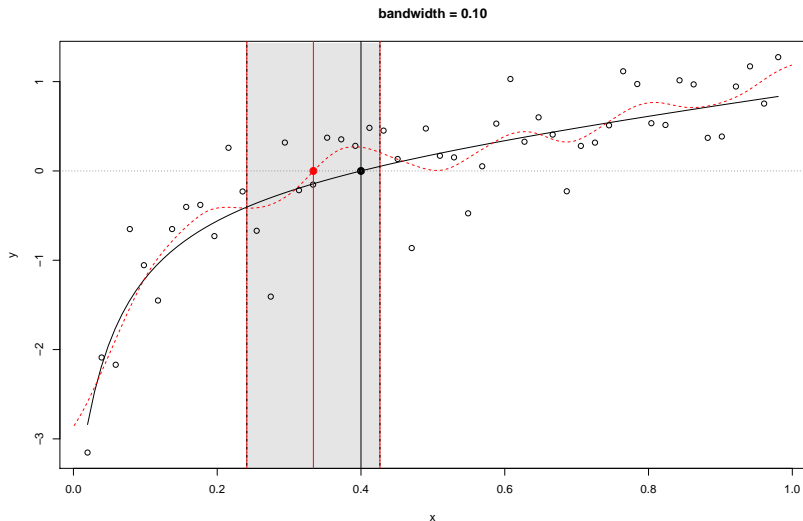
Závislost na bandwidth h_n

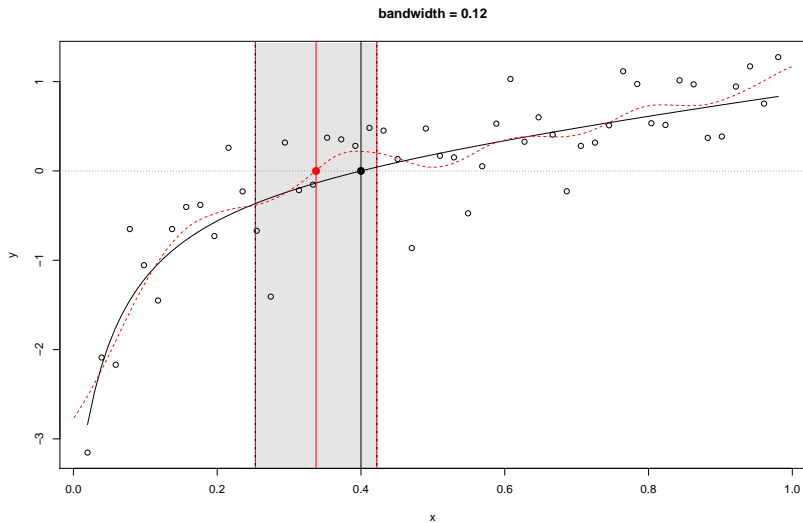
Závislost na bandwidth h_n 

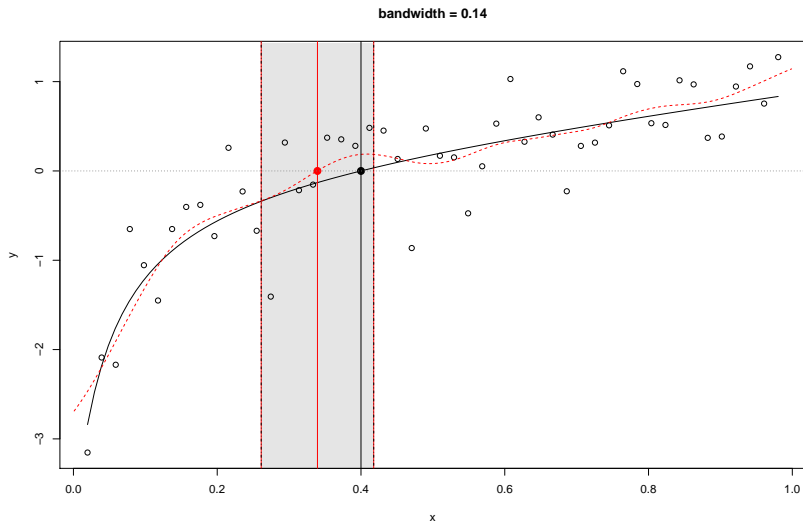
Závislost na bandwidth h_n 

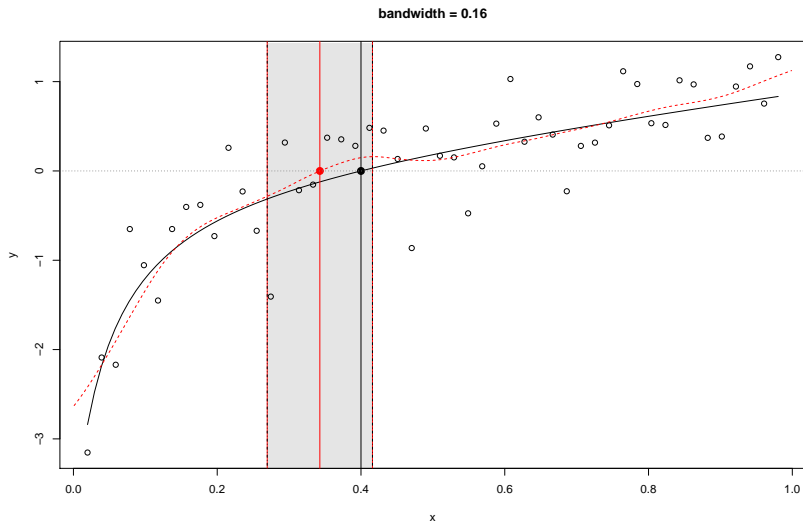
Závislost na bandwidth h_n 

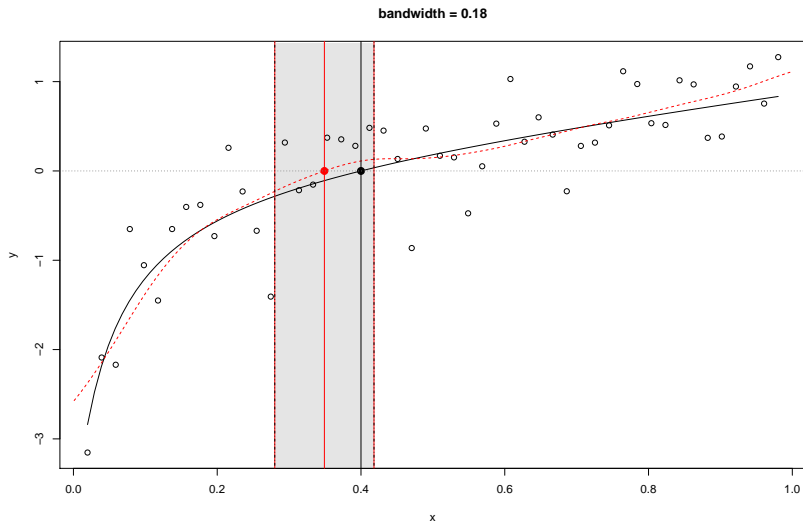
Závislost na bandwidth h_n 

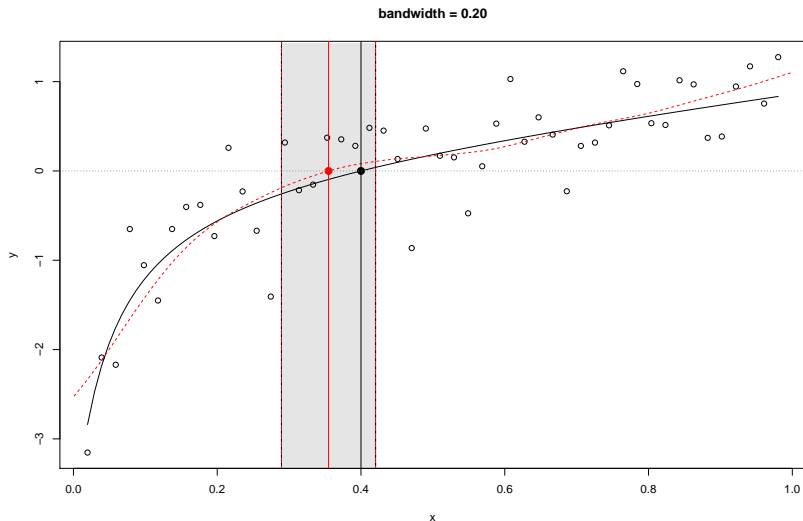
Závislost na bandwidth h_n 

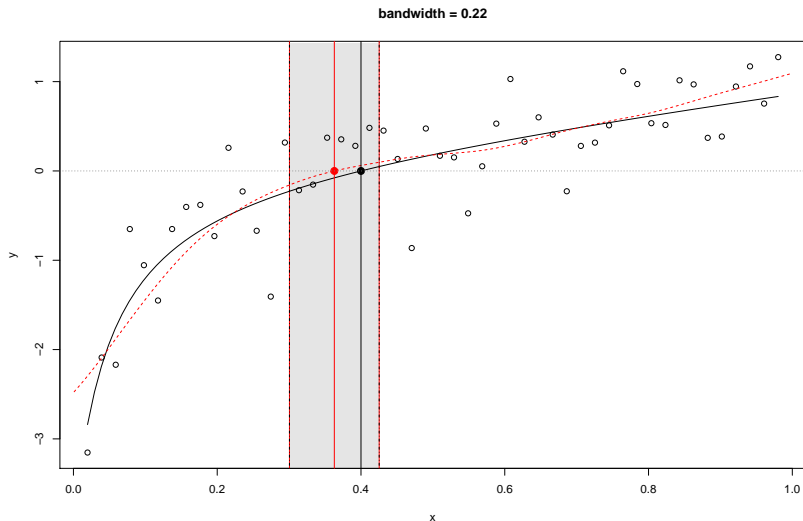
Závislost na bandwidth h_n 

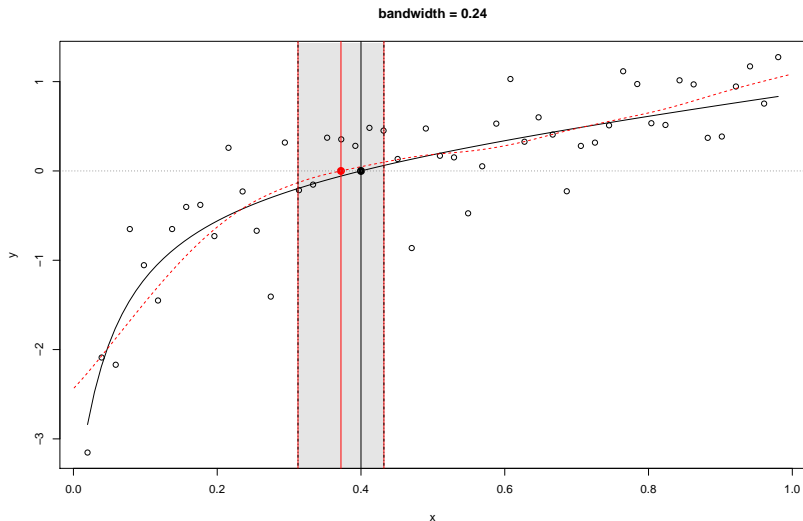
Závislost na bandwidth h_n 

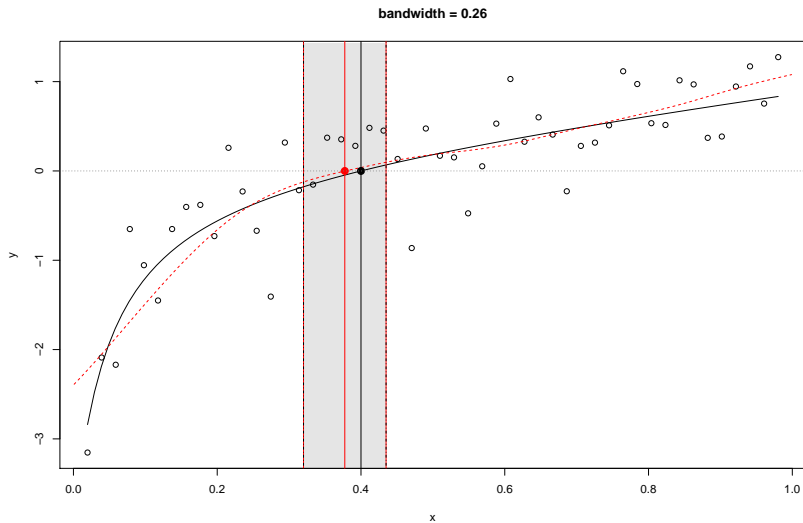
Závislost na bandwidth h_n 

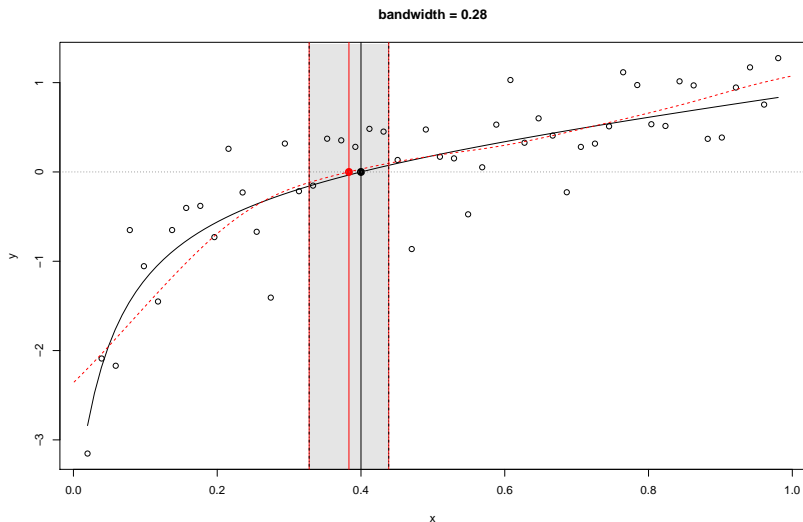
Závislost na bandwidth h_n 

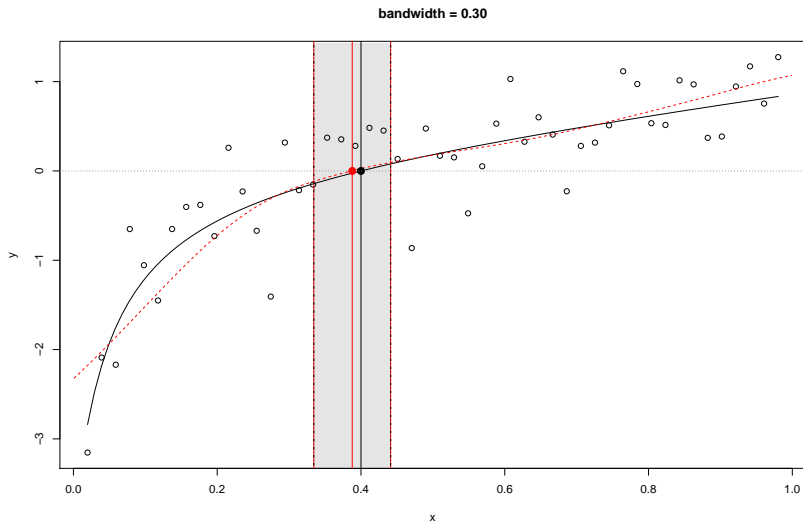
Závislost na bandwidth h_n 

Závislost na bandwidth h_n 

Závislost na bandwidth h_n 

Závislost na bandwidth h_n 

Závislost na bandwidth h_n 

Závislost na bandwidth h_n 

Optimální bandwidth

Teoreticky: asymptoticky optimální bandwidth plyne z porovnání vychýlení (rostoucí funkce h_n) a rozptylu (klesající funkce h_n):

- konstantní bandwidth $h_n = O(n^{-1/(2k+1)})$,
- lokální bandwidth $h_n(x) \propto [1/\{nf_X(x)\}]^{1/(2k+1)}$, kde $f_X(x)$ je hustota měření,
- lokální bandwidth v heteroskedastickém modelu $h_n(x) \propto [\sigma^2(x)/\{nf_X(x)\}]^{1/(2k+1)}$.

V praxi se bandwidth často volí až po ukončení experimentu (křížové ověřování).

Jaká je ale optimální hustota měření $f_X(\cdot)$?

Optimální návrhy experimentů pro odhad $m(\cdot)$

Obvyklé kritérium kvality odhadu: MSE (střední čtvercová chyba).

Přehledový článek:

Titterington (2001)

V literatuře jsou popsány dva různé přístupy:

Minimalizace IMSE (integrovaná MSE):

Müller (1984), Faraway (1990), Cheng, Hall & Titterington (1998),
Biedermann & Dette (2001a,b)

Klasický přístup (kritéria optimality):

W. G. Müller (1996), Fedorov, Montepiedra & Nachtsheim (1999)

Optimální návrhy experimentů (minimalizace AIMSE)

H.-G. Müller (1984). Optimal designs for nonparametric kernel regression, *Statistics & Probability Letters* 2, 285–290

$$\text{Var}\{m^{GM}(x)\} = \frac{\sigma^2}{nh_n} \left[\int K^2(s) ds \{f_X(x)\}^{-1} + o(1) \right]$$

Je-li $h(\cdot)$ kladná a spojitá pravděpodobnostní hustota na $\langle 0, 1 \rangle$, pak

$$f_X^*(x) = \frac{h(x)^{1/2}}{\int h(u)^{1/2} du} \propto h(x)^{1/2}$$

minimalizuje:

$$\text{AIMSE} = E \int \{m^{GM}(x) - m(x)\}^2 dH(x).$$

Návrhy experimentů pro odhad nulového bodu

V praxi často můžeme získat alespoň přibližnou apriorní informaci o poloze nulového bodu (nebo maxima) díky:

- pilotní studii,
- expertním znalostem a literatuře.

Naše očekávání (apriorní informaci) lze vyjádřit pomocí apriorní hustoty $a(\cdot)$ neznámého parametru ξ .

Další postup: závislost rozptylu ξ_n na hustotě měření (i pro slabě závislá měření) a odvození optimální hustoty měření $f_X(\cdot)$.

Vlastnosti empirického nulového bodu

Zobecnění článku Müller (1985) pro nerovnoměrně rozložené body měření a pro slabě závislé náhodné chyby (motivace: situace, ve kterých po sobě následující měření mohou být závislá).

Nástroje: CLV pro nestacionární slabě závislé náhodné veličiny (Peligrad, 1996) a Hoeffdingova nerovnost pro “strongly mixing” posloupnosti náhodných veličin (Roussas, 1996).

V heteroskedastickém modelu se slabě závislými náhodnými chybami je ξ_n asymptoticky normální a

$$\text{Var } \xi_n \doteq \frac{\sigma^2(\xi) V}{nh_n^{2\nu+1} \{m^{(\nu+1)}(\xi)\}^2 f_X(\xi)}.$$

Délka konfidenčního intervalu je přímo úměrná $\sigma(\xi)/\{f_X(\xi)\}^{1/2}$.

Optimální návrh experimentu

V homoskedastickém modelu ($\sigma^2(\cdot) \equiv \sigma^2$) lze optimální návrh experimentu vyjádřit ve tvaru: $f_X^*(\cdot) \propto a^{r_{\text{opt}}}(\cdot)$.

Konstantní bandwidth:

- $r_{\text{opt}}(MSE) = 1/2$ minimalizuje očekávaný rozptyl $\int \text{Var}(\xi_n|\xi)a(\xi)d\xi$,
- $r_{\text{opt}}(MAD) = 2/3$ minimalizuje očekávanou délku konfidenčního intervalu.

Lokální bandwidth:

- $r_{\text{opt}}(MSE) = (2k + 1)/(4k - 2\nu + 1)$ minimalizuje očekávaný rozptyl,
- $r_{\text{opt}}(MAD) = (2k + 1)/(3k - \nu + 1)$ minimalizuje očekávanou délku konfidenčního intervalu.

Optimální návrh experimentu

V heteroskedastickém modelu s konstantní bandwidth je optimální návrh experimentu:

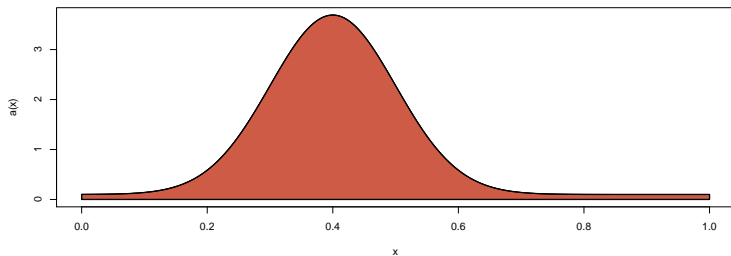
- $f_X^*(.) \propto \{\sigma^2(.)a(.)\}^{1/2}$ pro očekávaný rozptyl,
- $f_X^*(.) \propto \{\sigma(.)a(.)\}^{2/3}$ pro očekávanou délku konfidenčního intervalu.

Pro lokální bandwidth:

- $f_X^*(.) \propto \{\sigma^2(.)\}^{(2k-2\nu)/(4k-2\nu+1)} \{a(.)\}^{(2k+1)/(4k-2\nu+1)}$ minimalizuje očekávaný rozptyl,
- $f_X^*(.) \propto \{\sigma(.)\}^{(2k-2\nu)/(3k-\nu+1)} \{a(.)\}^{(2k+1)/(3k-2\nu+1)}$ minimalizuje očekávanou délku konfidenčního intervalu.

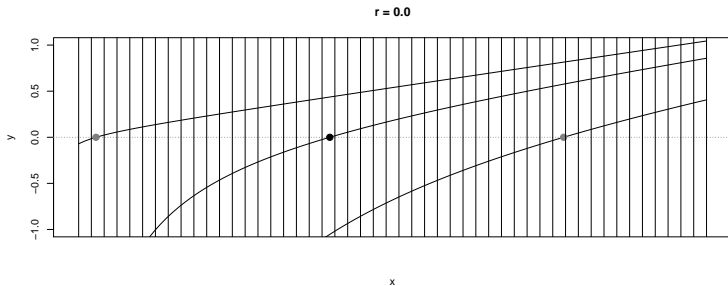
Simulace

- 1 Apriorní rozdělení ξ : $a(\xi) = 0.1 + 0.9N(0.4, 0.1^2)$.
- 2 Regresní funkce: $m(x) = \{(x + 0.05)^2 - (\xi + 0.05)^2\} / (x + 0.05)$.
- 3 Hustota měření $f_X(x) \propto a^r(x)$, $r \in \langle 0, 1 \rangle$.
- 4 Lokální bandwidth $h_f(x) = bf^{-1/5}(x)$, $b \in \langle 0.005, 0.5 \rangle$.



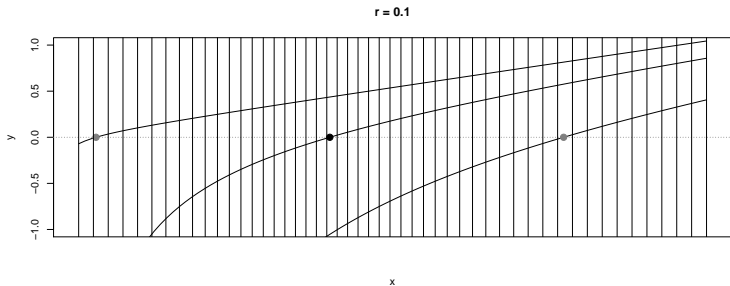
Simulace

- 1 Apriorní rozdělení ξ : $a(\xi) = 0.1 + 0.9N(0.4, 0.1^2)$.
- 2 Regresní funkce: $m(x) = \{(x + 0.05)^2 - (\xi + 0.05)^2\} / (x + 0.05)$.
- 3 Hustota měření $f_X(x) \propto a^r(x)$, $r \in \langle 0, 1 \rangle$.
- 4 Lokální bandwidth $h_f(x) = bf^{-1/5}(x)$, $b \in \langle 0.005, 0.5 \rangle$.



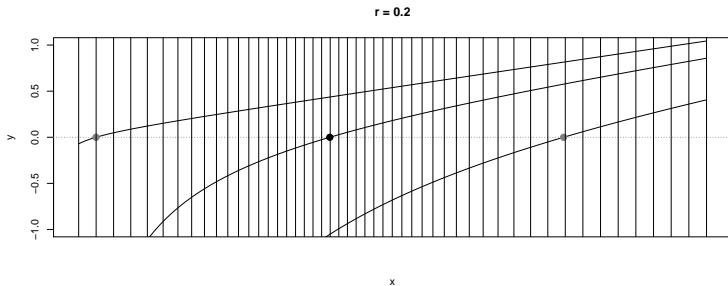
Simulace

- 1 Apriorní rozdělení ξ : $a(\xi) = 0.1 + 0.9N(0.4, 0.1^2)$.
- 2 Regresní funkce: $m(x) = \{(x + 0.05)^2 - (\xi + 0.05)^2\} / (x + 0.05)$.
- 3 Hustota měření $f_X(x) \propto a^r(x)$, $r \in \langle 0, 1 \rangle$.
- 4 Lokální bandwidth $h_f(x) = bf^{-1/5}(x)$, $b \in \langle 0.005, 0.5 \rangle$.



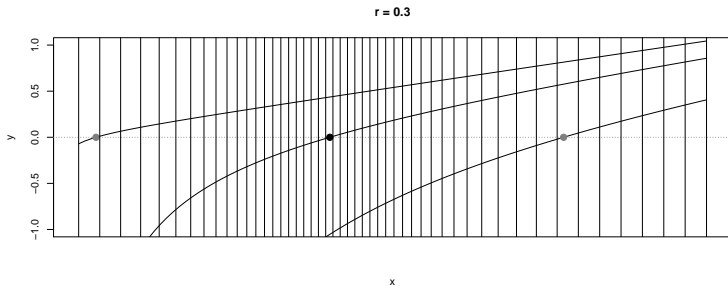
Simulace

- 1 Apriorní rozdělení ξ : $a(\xi) = 0.1 + 0.9N(0.4, 0.1^2)$.
- 2 Regresní funkce: $m(x) = \{(x + 0.05)^2 - (\xi + 0.05)^2\} / (x + 0.05)$.
- 3 Hustota měření $f_X(x) \propto a^r(x)$, $r \in \langle 0, 1 \rangle$.
- 4 Lokální bandwidth $h_f(x) = bf^{-1/5}(x)$, $b \in \langle 0.005, 0.5 \rangle$.



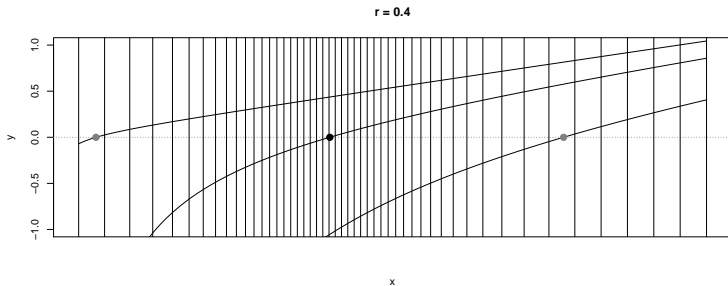
Simulace

- 1 Apriorní rozdělení ξ : $a(\xi) = 0.1 + 0.9N(0.4, 0.1^2)$.
- 2 Regresní funkce: $m(x) = \{(x + 0.05)^2 - (\xi + 0.05)^2\} / (x + 0.05)$.
- 3 Hustota měření $f_X(x) \propto a^r(x)$, $r \in \langle 0, 1 \rangle$.
- 4 Lokální bandwidth $h_f(x) = bf^{-1/5}(x)$, $b \in \langle 0.005, 0.5 \rangle$.



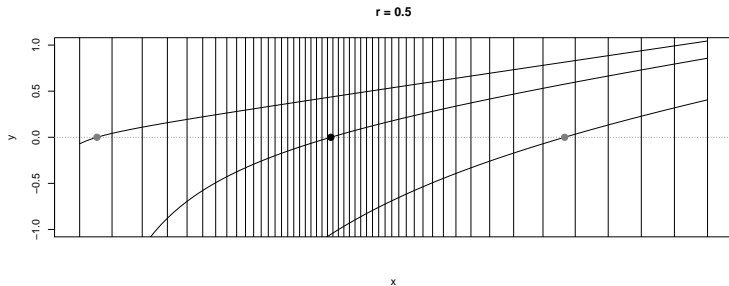
Simulace

- 1 Apriorní rozdělení ξ : $a(\xi) = 0.1 + 0.9N(0.4, 0.1^2)$.
- 2 Regresní funkce: $m(x) = \{(x + 0.05)^2 - (\xi + 0.05)^2\} / (x + 0.05)$.
- 3 Hustota měření $f_X(x) \propto a^r(x)$, $r \in \langle 0, 1 \rangle$.
- 4 Lokální bandwidth $h_f(x) = bf^{-1/5}(x)$, $b \in \langle 0.005, 0.5 \rangle$.



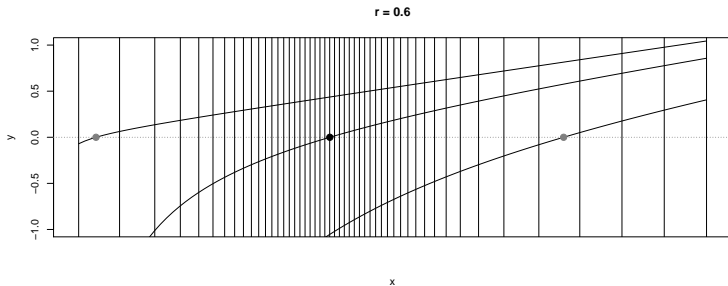
Simulace

- 1 Apriorní rozdělení ξ : $a(\xi) = 0.1 + 0.9N(0.4, 0.1^2)$.
- 2 Regresní funkce: $m(x) = \{(x + 0.05)^2 - (\xi + 0.05)^2\} / (x + 0.05)$.
- 3 Hustota měření $f_X(x) \propto a^r(x)$, $r \in \langle 0, 1 \rangle$.
- 4 Lokální bandwidth $h_f(x) = bf^{-1/5}(x)$, $b \in \langle 0.005, 0.5 \rangle$.



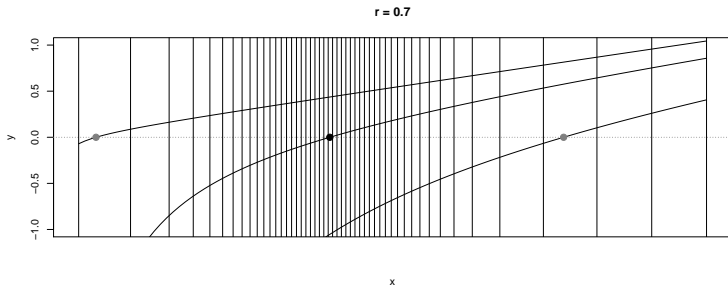
Simulace

- 1 Apriorní rozdělení ξ : $a(\xi) = 0.1 + 0.9N(0.4, 0.1^2)$.
- 2 Regresní funkce: $m(x) = \{(x + 0.05)^2 - (\xi + 0.05)^2\} / (x + 0.05)$.
- 3 Hustota měření $f_X(x) \propto a^r(x)$, $r \in \langle 0, 1 \rangle$.
- 4 Lokální bandwidth $h_f(x) = bf^{-1/5}(x)$, $b \in \langle 0.005, 0.5 \rangle$.



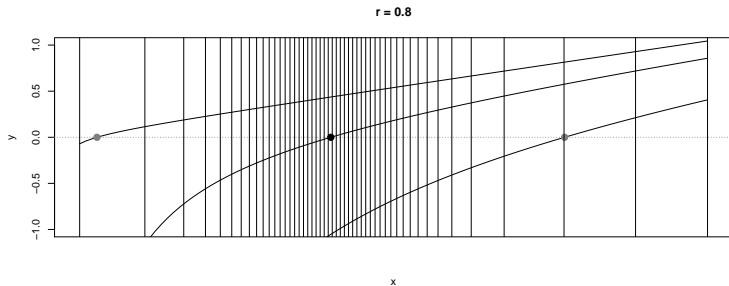
Simulace

- 1 Apriorní rozdělení ξ : $a(\xi) = 0.1 + 0.9N(0.4, 0.1^2)$.
- 2 Regresní funkce: $m(x) = \{(x + 0.05)^2 - (\xi + 0.05)^2\} / (x + 0.05)$.
- 3 Hustota měření $f_X(x) \propto a^r(x)$, $r \in \langle 0, 1 \rangle$.
- 4 Lokální bandwidth $h_f(x) = bf^{-1/5}(x)$, $b \in \langle 0.005, 0.5 \rangle$.



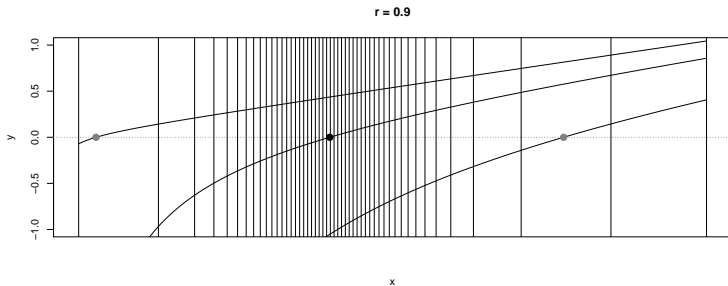
Simulace

- 1 Apriorní rozdělení ξ : $a(\xi) = 0.1 + 0.9N(0.4, 0.1^2)$.
- 2 Regresní funkce: $m(x) = \{(x + 0.05)^2 - (\xi + 0.05)^2\} / (x + 0.05)$.
- 3 Hustota měření $f_X(x) \propto a^r(x)$, $r \in \langle 0, 1 \rangle$.
- 4 Lokální bandwidth $h_f(x) = bf^{-1/5}(x)$, $b \in \langle 0.005, 0.5 \rangle$.



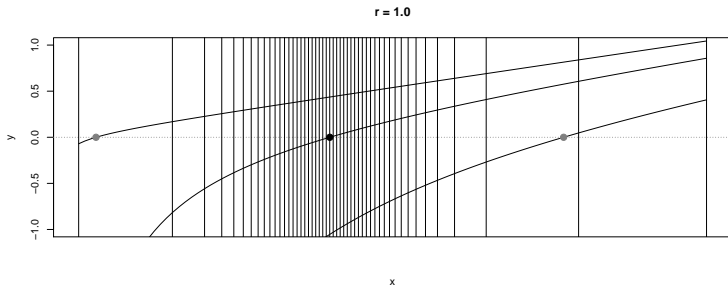
Simulace

- 1 Apriorní rozdělení ξ : $a(\xi) = 0.1 + 0.9N(0.4, 0.1^2)$.
- 2 Regresní funkce: $m(x) = \{(x + 0.05)^2 - (\xi + 0.05)^2\} / (x + 0.05)$.
- 3 Hustota měření $f_X(x) \propto a^r(x)$, $r \in \langle 0, 1 \rangle$.
- 4 Lokální bandwidth $h_f(x) = bf^{-1/5}(x)$, $b \in \langle 0.005, 0.5 \rangle$.



Simulace

- 1 Apriorní rozdělení ξ : $a(\xi) = 0.1 + 0.9N(0.4, 0.1^2)$.
- 2 Regresní funkce: $m(x) = \{(x + 0.05)^2 - (\xi + 0.05)^2\} / (x + 0.05)$.
- 3 Hustota měření $f_X(x) \propto a^r(x)$, $r \in \langle 0, 1 \rangle$.
- 4 Lokální bandwidth $h_f(x) = bf^{-1/5}(x)$, $b \in \langle 0.005, 0.5 \rangle$.



Simulační studie

Vyhodnocení:

- MSE (Mean Squared Error), optimální hodnota $r_{\text{opt}}(MSE)$,
- MAD (Mean Absolute Deviation), optimální hodnota $r_{\text{opt}}(MAD)$

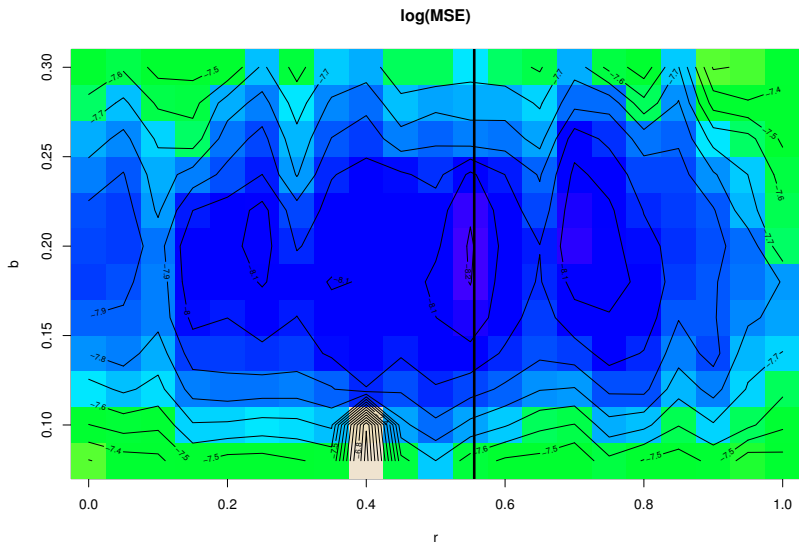
jako funkce parametru bandwidth b a hustoty pozorování r .

Teoreticky optimální hodnota r pro homoskedastická data a lokální bandwidth:

	$m(\cdot)$	$m'(\cdot)$
$r_{\text{opt}}(MSE)$	$5/9 \doteq 0.56$	$7/11 \doteq 0.63$
$r_{\text{opt}}(MAD)$	$5/7 \doteq 0.71$	$7/9 \doteq 0.78$

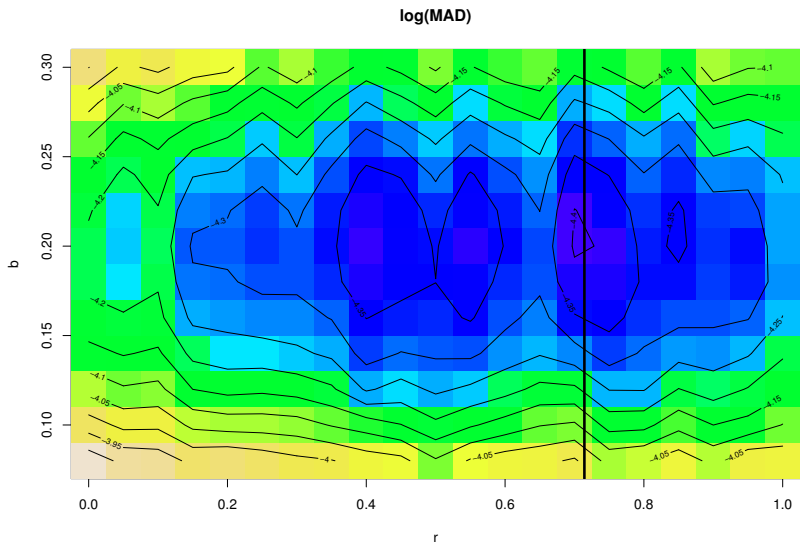
Simulace: lokální bandwidth, $n = 800$, $\sigma = 0.5$

MSE (optimal value $\doteq 0.55$)



Simulace: lokální bandwidth, $n = 800$, $\sigma = 0.5$

MAD (optimal value $\doteq 0.71$)



Shrnutí

- Odvození asymptotické normality ξ_n i pro slabě závislá pozorování.

Shrnutí

- Odvození asymptotické normality ξ_n i pro slabě závislá pozorování.
- Nový přístup k navrhování experimentů v neparametrické regresi (AIMSE optimální návrhy experimentů jsou MSE optimální i pro odhad nulového bodu).

Shrnutí

- Odvození asymptotické normality ξ_n i pro slabě závislá pozorování.
- Nový přístup k navrhování experimentů v neparametrické regresi (AIMSE optimální návrhy experimentů jsou MSE optimální i pro odhad nulového bodu).
- Další výzkum: lokálně polynomické modely a modely s více vysvětlujícími pozorování (motivace: závislost GHBC na čase od diagnózy, věku a pohlaví, spolupráce Z. Šumník, 2. LF).

Shrnutí

- Odvození asymptotické normality ξ_n i pro slabě závislá pozorování.
- Nový přístup k navrhování experimentů v neparametrické regresi (AIMSE optimální návrhy experimentů jsou MSE optimální i pro odhad nulového bodu).
- Další výzkum: lokálně polynomické modely a modely s více vysvětlujícími pozorování (motivace: závislost GHBC na čase od diagnózy, věku a pohlaví, spolupráce Z. Šumník, 2. LF).

Dále: neparametrické regresní odhady pro korelovaná data s omezeními (aplikace na ceny opcí, spolupráce W. Härdle: JoE, CSDA) a ověřování předpokladů (“ověřování stability modelu v čase”, spolupráce S. Meintanis, C. Kirch, M. Hušková: JMVA, TEST, CSDA).

Shrnutí

- Odvození asymptotické normality ξ_n i pro slabě závislá pozorování.
- Nový přístup k navrhování experimentů v neparametrické regresi (AIMSE optimální návrhy experimentů jsou MSE optimální i pro odhad nulového bodu).
- Další výzkum: lokálně polynomické modely a modely s více vysvětlujícími pozorování (motivace: závislost GHBC na čase od diagnózy, věku a pohlaví, spolupráce Z. Šumník, 2. LF).

Dále: neparametrické regresní odhady pro korelovaná data s omezeními (aplikace na ceny opcí, spolupráce W. Härdle: JoE, CSDA) a ověřování předpokladů (“ověřování stability modelu v čase”, spolupráce S. Meintanis, C. Kirch, M. Hušková: JMVA, TEST, CSDA).

DĚKUJI ZA POZORNOST