RŮZNÉ PŘÍSTUPY K JÁDROVÉMU ODHADU REGRESNÍ FUNKCE

MARIE BUDÍKOVÁ, IVANA HOROVÁ

Cílem příspěvku je poukázat na souvislosti mezi aparátem teorie aproximací funkcí a neparametrickým odhadem regresní funkce.

I. VLASTNOSTI LINEÁRNÍCH POZITIVNÍCH OPERÁTORŮ

V roce 1968 definoval D. D. Stancu následující třídu lineárních pozitivních operátorů ([5]):

$$P_n^{\alpha}(f;x) = \sum_{k=0}^{n} w_{n,k}^{\alpha}(x) f\left(\frac{k}{n}\right); \quad x \in [0,1],\ f \in C[0,1], \tag{1}$$

kde α je reálný parametr, který může záviset pouze na přirozeném čísle n,

$$w_{n,k}^{\alpha}(x) = \frac{\binom{n}{k} x^{(k,-\alpha)} (1-x)^{(n-k,-\alpha)}}{1^{(n,-\alpha)}} \tag{2}$$

a

$$x^{(k,-\alpha)} = \begin{cases} 1, & k = 0 \\ x(x+\alpha)\dots(x+(k-1)\alpha), & k \geq 1. \end{cases}$$

Pro vhodné hodnoty parametru α definují operátory $P_n^{\alpha}(f;x)$ známé typy lineárních pozitivních operátorů, jako např. Bernštejnovy polynomy, Szász-Mirakjanovy operátory, Baskakovovy operátory a pro $\alpha = -1/n$ dostáváme Lagrangeův interpolační polynom v uzlech $x_i = \dfrac{i}{n}$, $i = 0, 1, \dots, n$.

Funkce $w_{n,k}^{\alpha}(x)$, $k = 0, 1, 2, \dots, n$, budeme nazývat vahovými funkcemi. Je zřejmé, že platí:

a) $w_{n,k}^{\alpha}(x) \geq 0 \ \forall x \in [0,1]$, k= 0, 1,2, ... , n

b) $\sum_{k=0}^{n} w_{n,k}^{\alpha}(x) = 1 \ \forall x \in [0,1]$

Připomeňme větu o konvergenci posloupnosti operátorů $P_n^{\alpha}(f;x)$.

Věta 1. *Nechť $f \in C[0,1]$ a nechť $\alpha = \alpha(n)$, $\alpha(n) \to 0$ pro $n \to \infty$. Pak posloupnost $\{P_n^{\alpha}(f;x)\}$, $n = 0, 1, 2, \dots$, konverguje stejnoměrně k funkci $f(x)$ na intervalu $[0,1]$.*

Důkaz lze najít v [5].

Poznámka. Položíme-li v (1) $\alpha = 0$, dostáváme známé *Bernštejnovy polynomy*, tj.

$$P_n^0(f;x) = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} f\left(\frac{k}{n}\right)$$

a věta 1 je v tomto případě známou Weierstrassovou větou o stejnoměrné aproximaci spojité funkce polynomem.

Nyní uvedeme některé vlastnosti funkcí $w_{n,k}^{\alpha}(x)$ ([3]):

1. $x \in (0,1)$, $x < \alpha$, $\alpha > 1/2 \Rightarrow$

$$w_{n,k-1}^{\alpha}(x) \geq w_{n,k}^{\alpha}(x), \text{ je-li } 1 \leq k \leq \left[\frac{(n+1)(\alpha - x)}{2\alpha - 1}\right]$$

$$w_{n,k-1}^{\alpha}(x) \leq w_{n,k}^{\alpha}(x), \text{ je-li } \left[\frac{(n+1)(\alpha - x)}{2\alpha - 1}\right] < k \leq n,$$

kde $[a]$ značí celou část čísla a.

Jelikož se dále budeme převážně zabývat hodnotami vahových funkcí v bodě $x = 1/2$, uvedeme vlastnosti posloupnosti těchto funkcí v bodě $x = 1/2$.

2. $x = 1/2$, $\alpha > 1/2 \Rightarrow$

$$w_{n,k-1}^{\alpha}\left(\frac{1}{2}\right) \geq w_{n,k}^{\alpha}\left(\frac{1}{2}\right), \quad \text{je-li } 1 \leq k \leq \left[\frac{n+1}{2}\right]$$

$$w_{n,k-1}^{\alpha}\left(\frac{1}{2}\right) \leq w_{n,k}^{\alpha}\left(\frac{1}{2}\right), \quad \text{je-li } \left[\frac{n+1}{2}\right] < k \leq n$$

3. $x = 1/2$, $\alpha = 1/2 \Rightarrow$

$$w_{n,k-1}^{1/2}\left(\frac{1}{2}\right) = w_{n,k}^{1/2}\left(\frac{1}{2}\right), \quad \text{je-li } k = 1, 2, \ldots, n.$$

4. $x = 1/2$, $0 < \alpha < 1/2 \Rightarrow$

$$w_{n,k-1}^{\alpha}\left(\frac{1}{2}\right) \leq w_{n,k}^{\alpha}\left(\frac{1}{2}\right), \quad \text{je-li } 1 \leq k \leq \left[\frac{n+1}{2}\right]$$

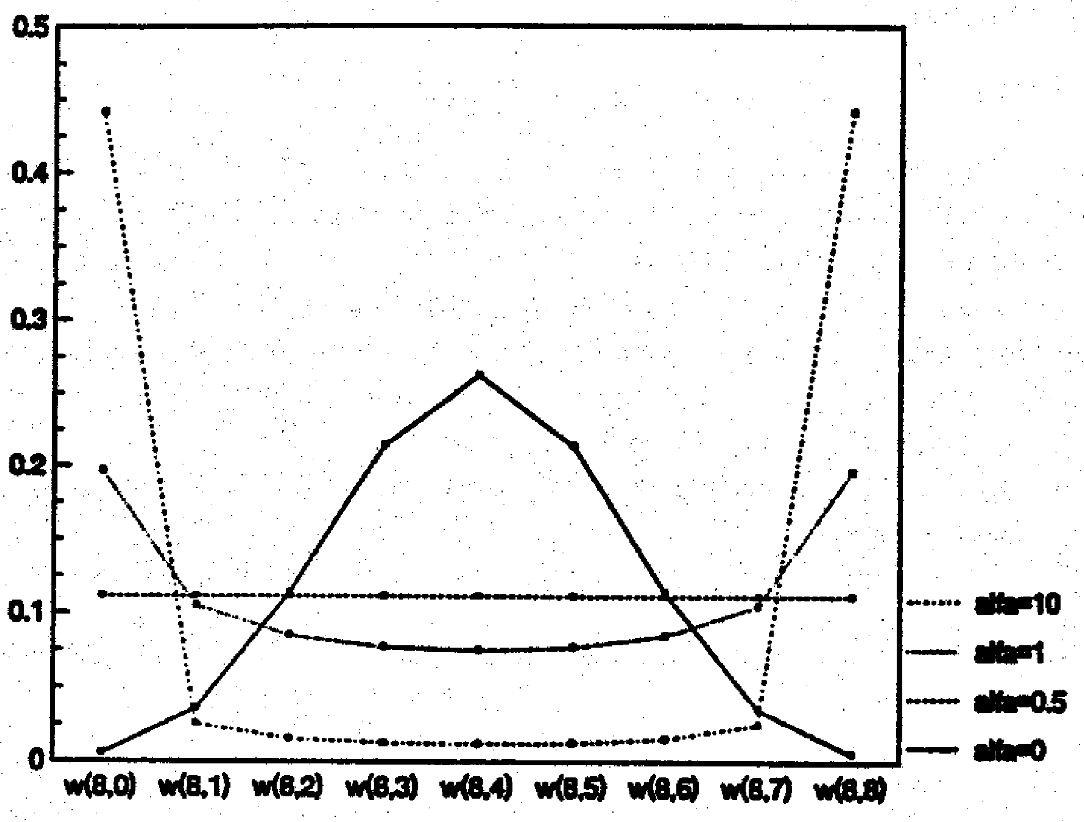
$$w_{n,k-1}^{\alpha}\left(\frac{1}{2}\right) \geq w_{n,k}^{\alpha}\left(\frac{1}{2}\right), \quad \text{je-li } \left[\frac{n+1}{2}\right] < k \leq n$$

5. $x \in (0,1)$, $\alpha = 0$, $\Rightarrow$

$$w_{n,k-1}^{0}(x) \leq w_{n,k}^{0}, \quad \text{je-li } 1 \leq k \leq [(n+1)x]$$

$$w_{n,k-1}^{0} \geq w_{n,k}^{0}(x), \quad \text{je-li } [(n+1)x] < k \leq n.$$

Z výše uvedeného plyne, že vlastnosti funkcí $w_{n,k}^{\alpha}$ jsou pro $\alpha > 1/2$ zcela odlišné od vlastností těchto funkcí v případě, že $\alpha \leq 1/2$. Tuto skutečnost také ilustruje obr. 1.



Obr. 1. Průběhy vahových funkcí pro $n = 8$ a různé hodnoty α.

II. Jádro s minimálním rozptylem

Nechť

$$\mathcal{M}_{\nu,k} = \left\{ f \in \mathrm{Lip}([-1,1]), \int_{-1}^{1} f(x)x^j \, dx = \left\{ \begin{array}{l} 0, \; 0 \le j < k, \; j \neq \nu \\ (-1)^\nu \nu! \, , j = \nu \end{array} \right. \right\}$$

Říkáme, že funkce $K_\nu(x)$ je jádrem řádu k, jestliže $K_\nu \in \mathcal{M}_{\nu,k}$ a $\beta_k = \int_{-1}^{1} K_\nu(x)x^k \, dx \neq 0$.

Předpokládáme dále, že $0 \le \nu \le k-2$ a čísla ν, k jsou obě buď lichá nebo sudá. *Jádra s minimálním rozptylem* jsou řešením variační úlohy

$$\int_{-1}^{1} K_\nu^2(x) \, dx \to \min\,, \quad K_\nu \in \mathcal{M}_{\nu,k}$$

Řešení této úlohy je založeno na vlastnostech Legendreových polynomů a je následující:

Věta 2. *Jádra K_ν řádu k s minimálním rozptylem na intervalu $[-1,1]$ jsou polynomy stupně $(k-2)$ s $(k-2)$ různými reálnými kořeny v $(-1,1)$.*

Poznámka. Jádro $K_0(x)$ řádu 2 s minimálním rozptylem je obdélníkové jádro $K_0(x) \equiv \frac{1}{2}$, $\forall x \in [-1,1]$.

III. Vyhlazování

Uvažujme regresní model

$$y_i = m(x_i) + \varepsilon_i\,, \quad i = 1,2,\dots,N\,,$$

kde $y_1, y_2, \dots, y_N$ jsou realizace závisle proměnné veličiny Y, $x_1, x_2, \dots, x_N$ jsou realizace vysvětlující proměnné X, m je teoretická *regresní funkce* a veličiny $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N$ representují náhodnou složku.

Budeme se zabývat výhradně *modelem s pevným plánem*. V tomto případě X představuje deterministickou proměnnou, která je volena ekvidistantně. Předpokládejme dále, že $x_i = i$, $i = 1,2,\dots,N$ a náhodné veličiny $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N$ jsou stochasticky nezávislé, $E(\varepsilon_i) = 0$, $D(\varepsilon_i) = \sigma^2$, $i = 1,2,\dots,N$.

Neznámou regresní funkci m budeme odhadovat neparametricky *metodou vážených klouzavých průměrů*.

Nechť $(2l+1)$ je počet vah (tj. šířka vyhlazovacího okna). Váhy $w_0, w_1, \dots, w_{2l}$ jsou reálná čísla taková, že

$$\sum_{k=0}^{2l} w_k = 1\,, \quad w_k = w_{2l-k},\ k = 0,1,\dots,l\,. \tag{3}$$

Odhad regresní funkce m v bodě $x_i = i$ označme $\hat{y}_i$. Prvních a posledních l hodnot neodhadujeme. Lze snadno ukázat, že pro odhad regresní funkce ve středu vyhlazovacího okna (o šířce $2l+1$) platí

$$\hat{y}_{l+i} = \sum_{k=0}^{2l} w_k y_{i+k}\,, \quad i = 0,1,2,\dots,N-2l\,. \tag{4}$$

Podívejme se nyní na předchozí úvahy z jiného hlediska, a to z hlediska vyhlazování pomocí jader.

Nejdříve sestrojíme pomocí vah $w_0, w_1, \dots, w_{2l}$ jádro K_0 řádu 2.

Definujme funkci K_0 takto:

a) K_0 je spojitá a po částech lineární v intervalu $[-1,1]$.

b) $K_0(x) \equiv 0$ vně intervalu $[-1,1]$.

c) v bodech $x_k = -1 + \frac{k}{l}$, $k = 0,1,2,\dots,2l$ nabývá hodnot $\frac{lw_k}{1-w_0}$, $k = 0,1,\dots 2l$ (za předpokladu $w_0 \neq 1$), tj.

$$K_0\left(-1 + \frac{k}{l}\right) = \frac{lw_k}{1-w_0}\,, \quad k = 0,1,\dots,2l\,. \tag{5}$$

Snadno ověříme, že takto definovaná funkce $K_0 \in \mathcal{M}_{0,2}$. Je totiž

$$\int_{-1}^{1} K_0(x)\,dx = \frac{l}{1-w_0} \sum_{k=1}^{2l} \int_{x_{k-1}}^{x_k} \bigl((w_k - w_{k-1})l(x - x_k) + w_k\bigr)\,dx$$

$$= \frac{l}{1-w_0}\,\frac{1}{2l} \sum_{k=1}^{2l} (w_k + w_{k-1})\,.$$

Užijeme-li nyní (3), dostaneme

(6) $$\int_{-1}^{1} K_0(x)\,dx = 1\,,$$

což znamená, že pro $j = \nu = 0$ je

$$\int_{-1}^{1} K_0(x)\,x^j\,dx = 1\,.$$

Nechť nyní $j = 1$, tj. $0 \le j < k$, $k = 2$. Pak zřejmě

(7) $$\int_{-1}^{1} K_0(x)\,x\,dx = 0\,,$$

neboť $K_0(x)$ je sudá funkce a dále

(8) $$\beta_2 = \int_{-1}^{1} K_0(x)\,x^2\,dx \neq 0\,.$$

Je rovněž jasné, že takto sestrojená funkce $K_0 \in \mathrm{Lip}([-1,1])$. Tedy $K_0 \in \mathcal{M}_{0,2}$ a je jádrem řádu 2.

Použijme nyní tohoto postupu k získání jádra pomocí operátorů $P_n^{\alpha}(f;x)$.

Položme

$$w_k = w_{n,k}^{\alpha}(\frac{1}{2}),\, n = 2l\,, \quad k = 0,1,\ldots,2l\,.$$

V tomto případě jsou všechny váhy w_k kladné a splňují předepsané podmínky (3).

Položíme-li $\alpha = 0$, tj. pro případ Bernštejnových polynomů, nabývá funkce K_0 v bodech $x_k = -1 + k/l$, $k = 0,1,\ldots,2l$, hodnot $\dfrac{l}{2^{2l}-1}\dbinom{2l}{k}$, $k = 0,1,\ldots,2l$.

Je-li $\alpha = 1$, je

$$K_0(x_k) = \frac{(2k-1)!!\,(2(2l-k)-1)!!}{(2l)!2^{2l} - (4l-1)!!}\,l\binom{2l}{k}\,.$$

Z vlastnosti (3) pro váhové funkce $w_{n,k}^{\alpha}(x)$ plyne, že pro $\alpha = 1/2$, $x = 1/2$ je posloupnost $w_{n,k}^{1/2}$ konstantní. Snadno se ukáže, že

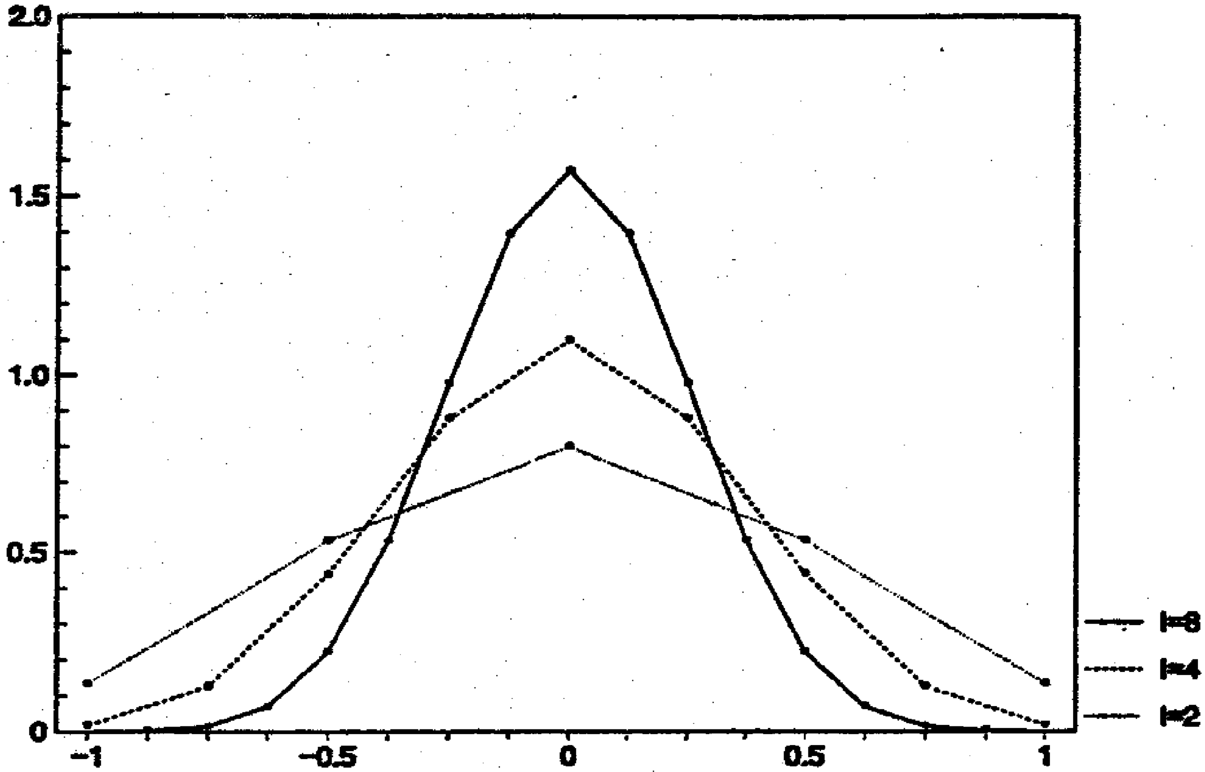
$$w_{n,k}^{1/2}(\frac{1}{2}) = \frac{1}{n+1}\,, \quad k = 0,1,\ldots n\,.$$

Užijeme-li nyní těchto vah ke konstrukci jádra, tj.

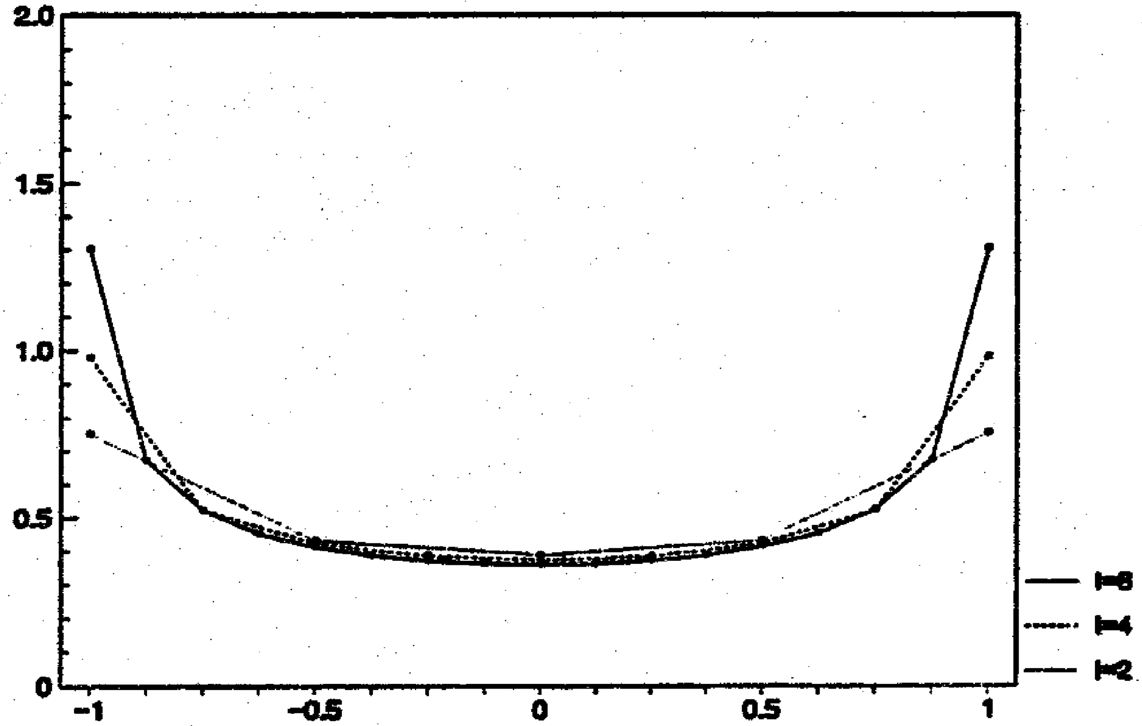
$$w_k = \frac{1}{2l+1}\,, \quad k = 0,1,\ldots,2l\,,$$

dostaneme obdélníkové jádro $K_0(x) \equiv 1/2$, $\forall x \in [-1,1]$. To znamená, že výše uvedená konstrukce zahrnuje jádro s minimálním rozptylem na intervalu $[-1,1]$.

Následující obrázky ukazují tvary jader pro různé hodnoty l a α.



Obr. 2. Tvar jádra $K_0(x)$ pro $x = \frac{1}{2}$, $\alpha = 0$, $l = 2, 4, 8$.



Obr. 3. Text viz obr. 2, $\alpha = 1$

Připomeňme, že jádrový odhad regresní funkce v bodě $x_i = i$ je dán vzorcem

(9) $$\hat{y}_i = \frac{\frac{1}{N}\sum_{i=1}^{N} K(i-k)y_k}{\frac{1}{N}\sum_{i=1}^{N} K(i-k)}, \quad i = 1, 2, \ldots, N,$$

kde $K(x)$ je jádro s nosičem $[-1,1]$ a $\int_{-1}^{1} K(x)\,dx = 1$.

Ve jmenovateli je odhad hustoty pravděpodobnosti nezávisle proměnné veličiny X. Vzhledem k tomu, že uvažujeme model s pevným plánem s ekvidistantními vzdálenostmi, je tato hustota konstantní. Její odhad označme c. Protože prvních a posledních l hodnot regresní funkce neodhadujeme, vzorec (9) nabývá tvaru

$$\hat{y}_{l+i} = \frac{1}{c}\sum_{i=1}^{N} K\,(l+i-k)y_k\,, \quad i = 1,2,\ldots,N-2l \tag{10}$$

Porovnáním (4) a (10) dostaneme

$$\frac{1}{c}K(l+i-k) = w_k\,, \quad k = 1,2,\ldots,N-2l$$

a současně podle (5) $K_0\Big(-1+\dfrac{k}{l}\Big) = \dfrac{l}{1-w_0}\,w_k$, $\quad k = 0,1,\ldots,2l$, tudíž $c = \dfrac{l}{1-w_0}$,

$$K_0\Big(-1+\frac{k}{l}\Big) = K(l+i-k)\,, \quad k = 0,1,\ldots,2l,\ i = 1,2,\ldots,N-2l\,.$$

Položíme-li $k = j-i$, můžeme vzorec (10) přepsat do tvaru

$$\hat{y}_{l+i} = \frac{1-w_0}{l}\sum_{j=1}^{N} K_0\Big(\frac{-l+j-i}{l}\Big)\,y_j\,, \quad i = 1,2,\ldots,N-2l\,.$$

V tabulce je uvedena závislost velikosti $\int_{-1}^{1} K_0^2(x)\,dx$ na šířce vyhlazovacího okna $2l+1$ pro různé hodnoty parametru α.

$2l+1$	$\alpha = 0$	$\alpha = \frac{1}{2}$	$\alpha = 1$
3	0,5185	0,5	0,5067
5	0,5748	0,5	0,5255
7	0,6647	0,5	0,5406
9	0,7623	0,5	0,5530
11	0,8560	0,5	0,5634

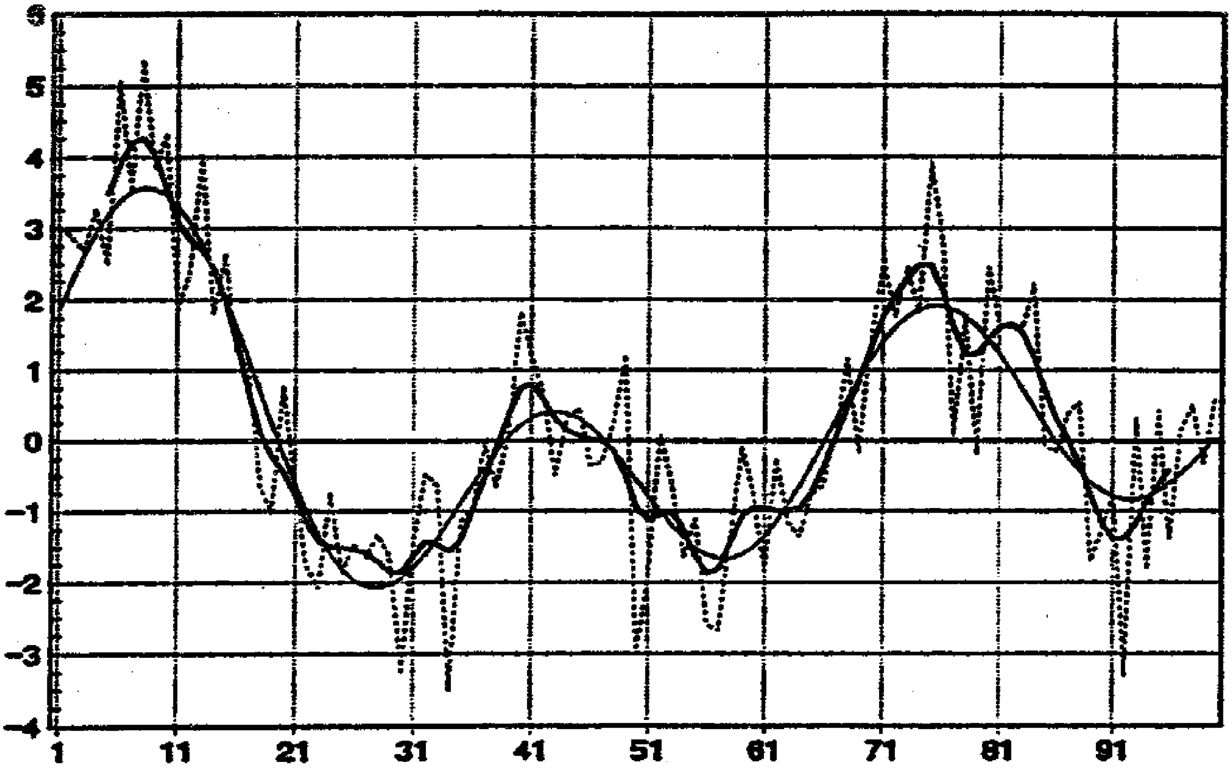
IV. Aplikace

Vliv jádra $K_0(x)$ pro $\alpha = 0,\ \frac{1}{2},\ 1$ na odhad regresní funkce jsme zkoumaly v simulačním experimentu s regresní funkcí uvažovaného modelu tvaru

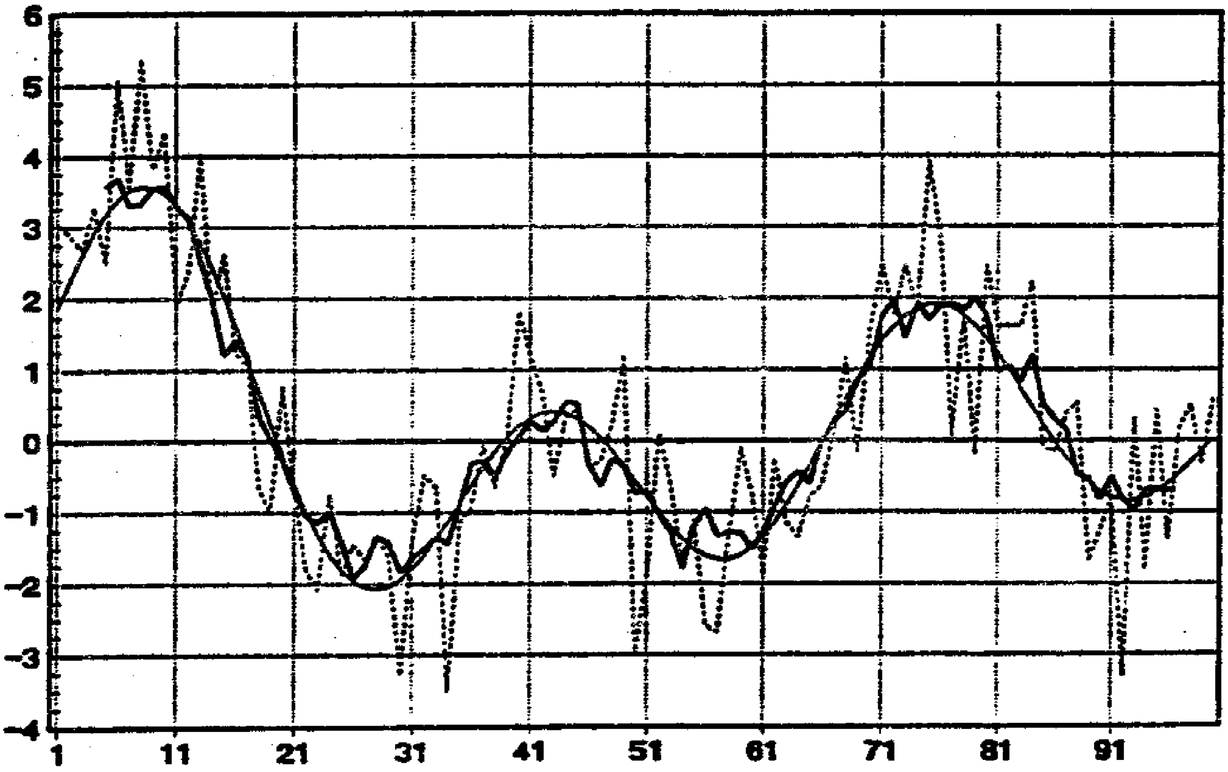
$$m(x) = \sin(0,2x - \frac{\pi}{6}) + \sin(0,12x + \frac{\pi}{8}) + \sin(0,16x + \frac{\pi}{3}) + \sin(0,08x + \frac{\pi}{4})\,,$$

kde proměnná x nabývala hodnot $1,2,\ldots,100$. Náhodné chyby měly standardizované normální rozložení. Šířka vyhlazovacího okna byla zvolena $2l+1 = 9$.

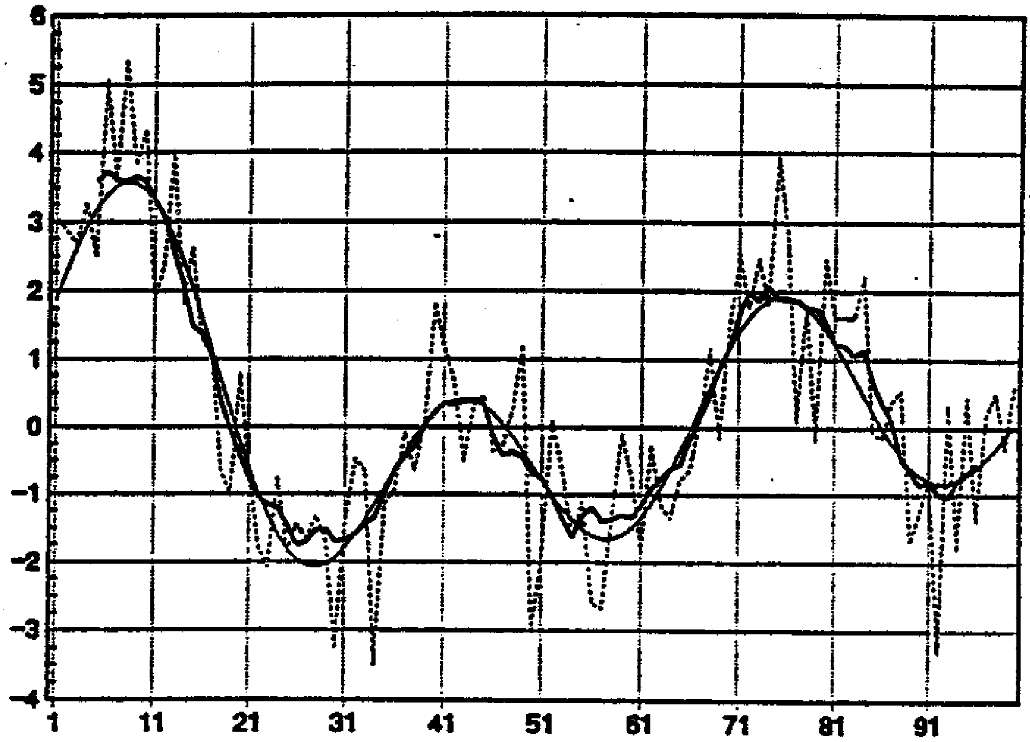
Výsledky simulačního experimentu ukazují obrázky.



Obr. 4. Vyhlazování simulovaných dat pomocí jádra $K_0(x)$, $x = \frac{1}{2}$ s parametry $\alpha = 0$, $l = 4$.

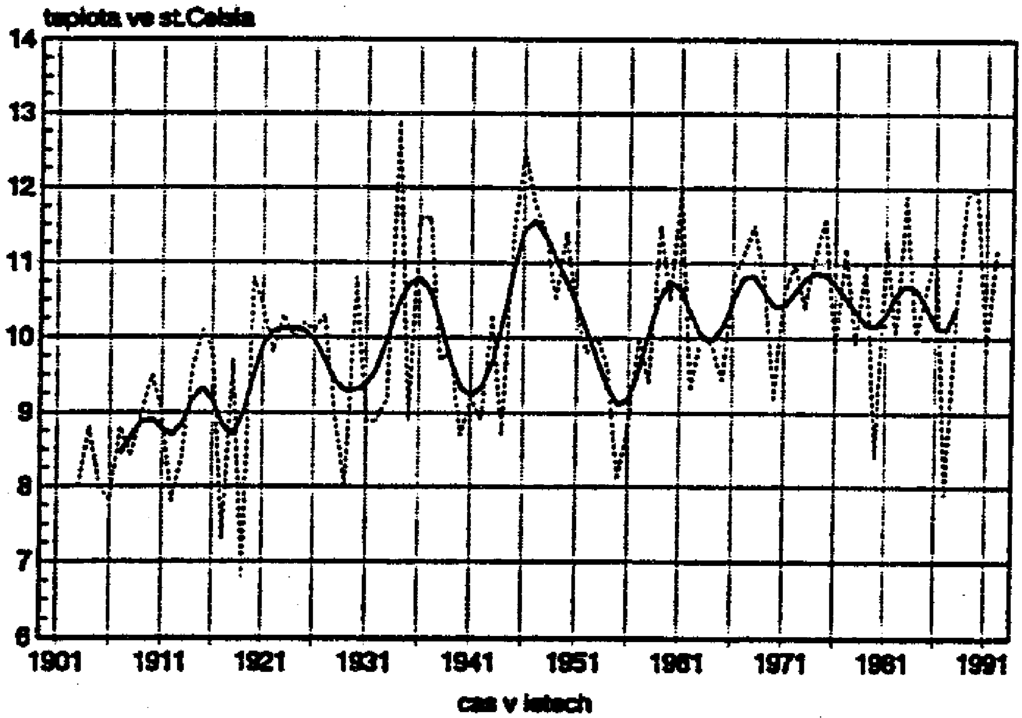


Obr. 5. Text viz obr. 4, $\alpha = 1$.

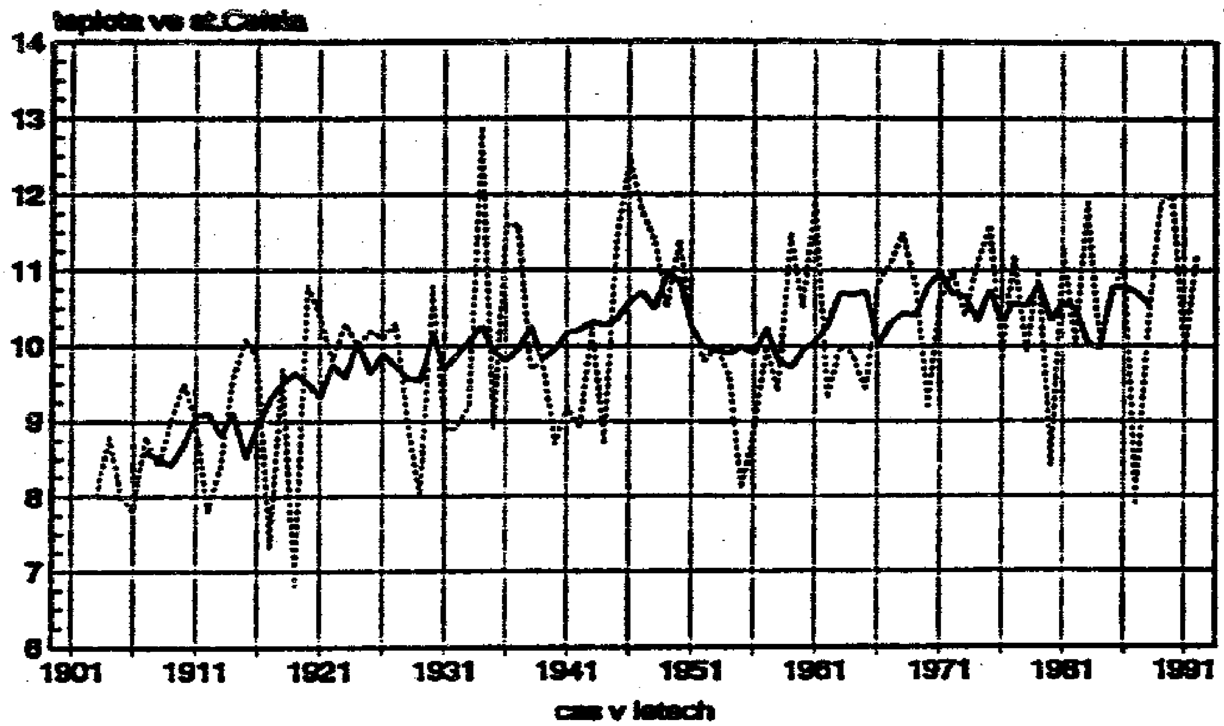


Obr. 6. Text viz obr. 4, $\alpha = \frac{1}{2}$.

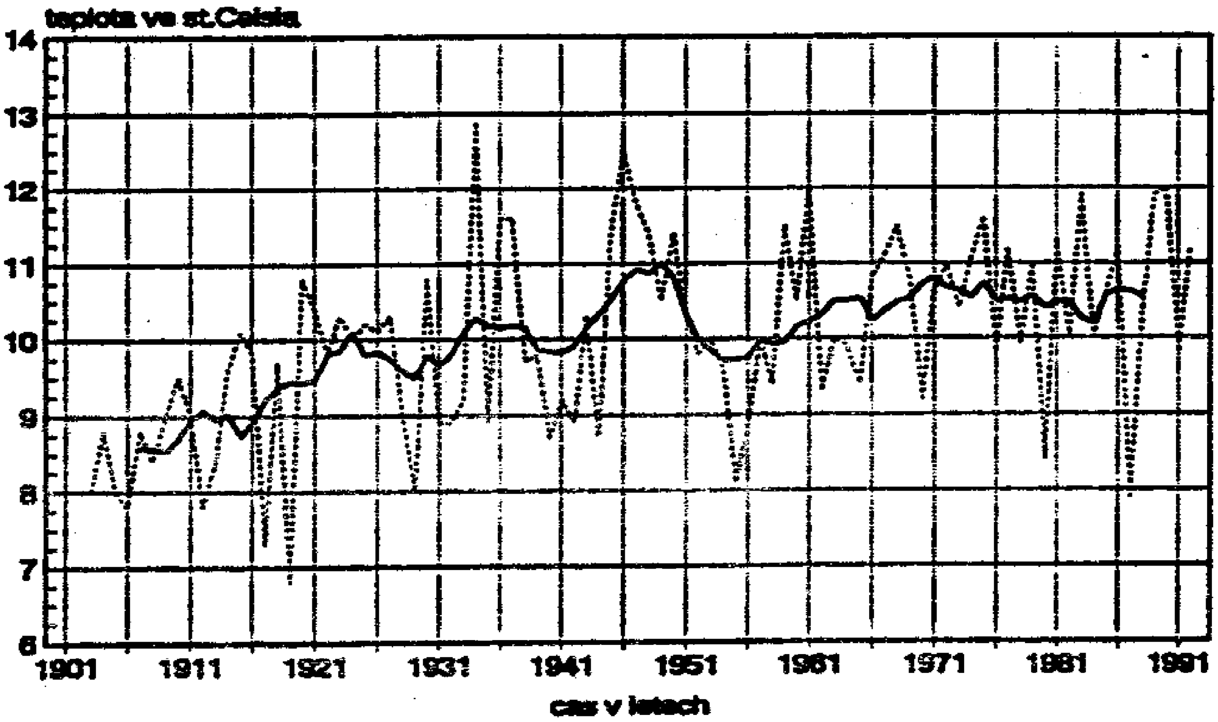
Jádro $K_0(x)$ s parametry $\alpha = 0, \frac{1}{2}, 1$ jsme rovněž použily k vyhlazení průměrných jarních teplot v Hurbanově v letech 1903–1992.



Obr. 7. Chod průměrných jarních teplot v Hurbanově v letech 1903-1992, vyhlazení jádrem $K_0(x)$, $x = \frac{1}{2}$ s parametry $\alpha = 0$, $l = 4$.



Obr. 8. Text viz obr. 7, $\alpha = 1$.



Obr. 9. Text viz obr. 7, $\alpha = \frac{1}{2}$.

Práce byla uskutečněna za finanční podpory GA ČR reg. č. 201/93/2408.

Literatura

[1] Cipra, T., *Analýza časových řad s aplikacemi v ekonomii*, SNTL/Alfa, Praha 1986.

[2] Härdle, W., *Applied Nonparametric Regression*, Econometric Society Monographs, Cambridge University Press, 1990.

[3] Horová, I., Budíková, M., *A Note on D. D. Stancu Operators*, zasláno do tisku.

[4] Müller, H. G., *Nonparametric Regresion Analysis of Longitudinal Data*, Lecture notes in statistics 46, Springer-Verlag, Berlin 1988.

[5] Stancu, D. D., *Approximation of functions by a new class of linear positive operators*, Rev. Roumaine Math. Pures Appl. 13 (1968), 1173-1194.

Adresy autorů

Marie Budíková & Ivana Horová

Katedra aplikované matematiky PřF MU

Janáčkovo nám. 2a

662 95 Brno