UNIVERZITA PARDUBICE

Licenční Studium Archimedes

Statistické zpracování dat a informatika

4.4 Aproximace křivek a vyhlazování křivek

Mgr. Jana Kubátová

Endokrinologický ústav

V Praze, leden 2012

Obsah

Otázka 1. Napište matici pro případ lineárního regresního spline vyjádřeného přes useknuté polynomy pro případ dvou uzlových bodů ksí 1 = 1 a ksí 2 = 4. Experimentální body jsou x = [0.2 0.4 0.6 1.5 2.0 3.0 3.5 6.0 8.0 10.0], y = [1 1 2 3 4 6 6 4 2 2].

Použitý program: Adstat

Zdrojová data

x	y
0.2	1
0.4	1
0.6	2
1.5	3
2.0	4
3.0	6
3.5	6
6.0	4
8.0	2
10.0	2

V programu Adstat byl na data aplikován lineární spline z modulu Kalibrace. Byly zadány 2 uzlové body ξ_1 a ξ_2.

Výsledky

Pomocí programu Adstat jsem proložila zadanými body 3 přímky následovně: osa x se pomocí 2 zadaných uzlových bodů rozdělila na 3 části. Na těchto úsecích bylo poté potřeba pomocí metody nejmenších čtverců nalézt přímky, které nejlépe prokládají dané body (tyto přímky na sebe musí navazovat). Výsledek lze reprezentovat dvěma způsoby: buď pomocí matematického zápisu tří nezávislých přímek, který je redundantní, nebo lépe pomocí vzorce pro useknuté polynomy:

$$S_m(x) = \sum_{j=0}^{m} a_j x^j + \sum_{j=1}^{n} b_j * (x - \xi_j)_+^m \quad \text{kde } (x)_+ = x \text{ pro } x \geq 0 \text{ resp 0 jinde}$$

$$\mathrm{y_i} = S_m(x_i) + \varepsilon_i$$

V našem případě, kdy průběh aproximujeme pomocí přímek (polynomy 1. stupně), je parametr m=1, kde $(x)_+$ zajistí, že pro každý z úseků vyjde postupně popis přímky s jinými parametry. Pro popis pomocí tohoto vzorce je potřeba obecně (n+m+1) parametrů, nicméně při využití omezujících podmínek nám postačí pouze n parametrů.

Ve výstupu programu Adstat byly uvedeny parametry pro popis aproximace prvním způsobem. Tyto parametry tedy v matici popisují koeficienty 3 nezávislých přímek ve tvaru y=g[i]*x+h[i] pro každý úsek zvlášť, kde *g* popisuje počáteční bod a *h* sklon přímky. Ve výsledné matici je koeficient *k* obdoba funkce $(x)_+$, tedy určuje definiční obor funkce. Výsledná matice je uvedena v následující tabulce. Analýza reziduí a další parametry jsou uvedeny v další tabulce.

(1) PARAMETRY KALIBRACE:

Koeficienty rovnice : g[i]*x+h[i] pro k[i-1] < x <= k[i]

k[i]	g[i]	h[i]
1.0000E+00	2.4745E+00	3.4240E-01
4.0000E+00	1.1706E+00	1.6463E+00
1.0000E+01	-8.5614E-01	9.7532E+00

(2) ANALÝZA REZIDUÍ:

Bod	Měřená hodnota	Predikovaná hodnota	Absolutní reziduum	Relativní reziduum
i	yexp[i]	yvyp[i]	e[i]	er[i]
1	1.0000E+00	8.3730E-01	-1.6270E-01	-1.9431E+01
2	1.0000E+00	1.3322E+00	3.3220E-01	2.4936E+01
3	2.0000E+00	1.8271E+00	-1.7290E-01	-9.4632E+00
4	3.0000E+00	3.4022E+00	4.0219E-01	1.1822E+01
5	4.0000E+00	3.9875E+00	-1.2515E-02	-3.1386E-01
6	6.0000E+00	5.1581E+00	-8.4193E-01	-1.6322E+01
7	6.0000E+00	5.7434E+00	-2.5663E-01	-4.4683E+00
8	4.0000E+00	4.6164E+00	6.1638E-01	1.3352E+01
9	2.0000E+00	2.9041E+00	9.0409E-01	3.1132E+01
10	2.0000E+00	1.1918E+00	-8.0819E-01	-6.7812E+01

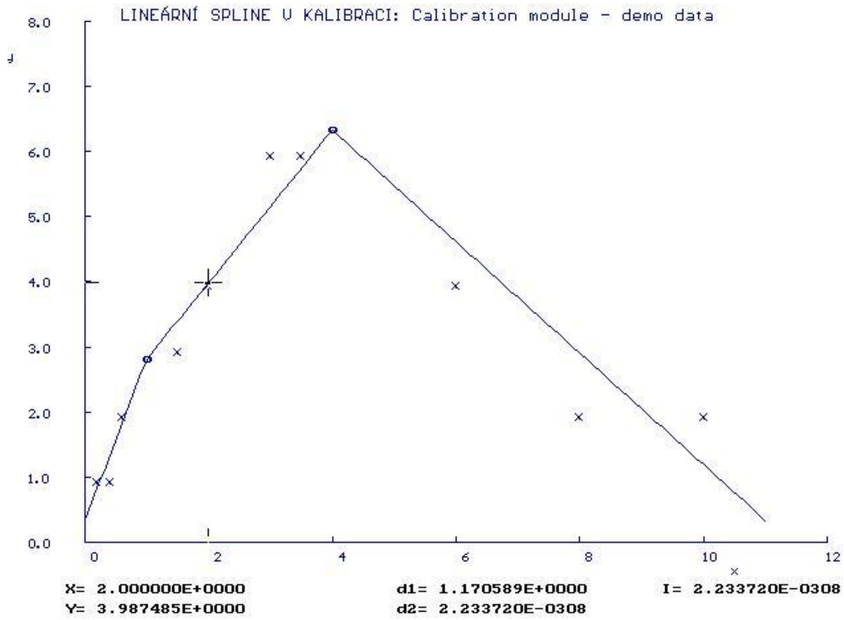
Reziduální součet čtverců, RSC : 2.9538E+00

Průměr absolutních hodnot reziduí, Me : 4.5097E-01

Průměr relativních reziduí, Mer[%] : 19.905

Odhad reziduálního rozptylu, s^2(e) : 4.9230E-01

Odhad směrodatné odchylky reziduí, s(e) : 7.0164E-01



Graf č.1

Závěr

Pomocí programu Adstat jsem vypočítala matici koeficientů 3 přímek prokládající zadané vstupní body. Reprezentaci řešení pomocí druhého způsobu (useknuté polynomy) bychom dostali adekvátním odečtením parametrů. Kvalitu výsledné aproximace si můžeme ověřit na grafu č. 1.

Otázka 2. Spočítejte integrál funkce $1/(1+x^2)$ pro interval [0,0.8] programem Spaeth

V intervalu od 0 do 0.8 jsem zvolila dalších 7 bodů, dosadila do funkce a vypočítala pro ně hodnoty.

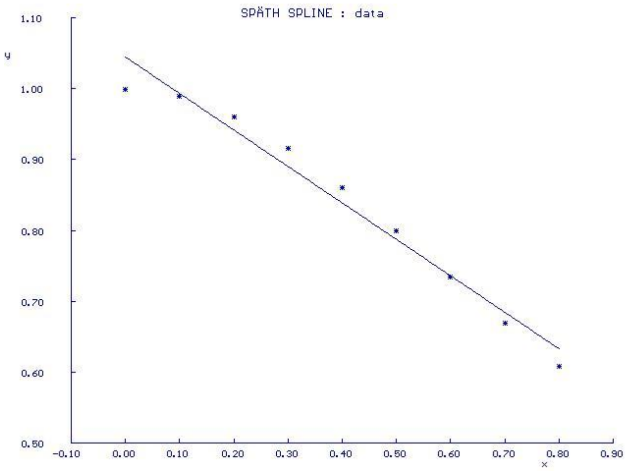
Použitý program: Adstat

Zdrojová data

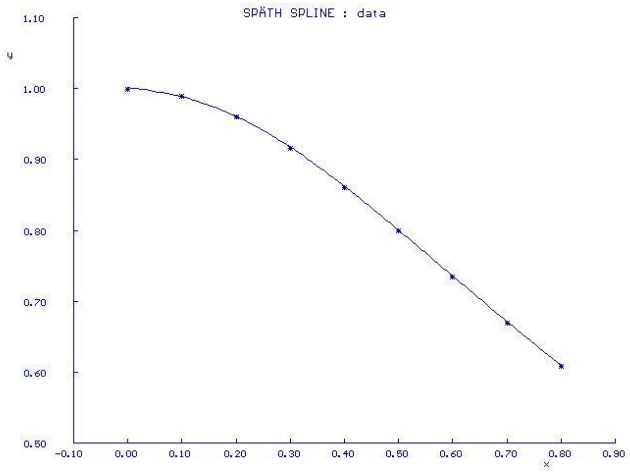
x	y
0	1
0.1	0.9901
0.2	0.9615
0.3	0.9174
0.4	0.8621
0.5	0.8
0.6	0.7353
0.7	0.6711
0.8	0.6098

Výstup

V programu Adstat jsem zadala vstupní data a v modulu Vyhlazování zvolila algoritmus Spaeth. Zde se uvádí lokální parametr vyhlazení P. Platí, že pro P blížící se nule je potlačena podmínka hladkosti a výsledkem je regresní přímka a pro P blížící se nekonečnu prochází spline všemi body. Pro ukázku na grafech č. 1 a 2 jsou ukázány grafy splinů s parametrem P=0,02, resp. P=10000.



Graf č.1 Spaeth spline, P=0.02



Graf č.2 Spaeth spline, P=10000

Proložení na grafu č. 2 je lepší, pro výpočet integrálu použiji výsledky se zadaným parametrem P=10000.

Výsledky

TABULKA DERIVACÍ A INTEGRÁLŮ:

Bod	Nezávisle proměnná	Závisle proměnná	První derivace	Druhá derivace	Integrál
i	x[i]	yexp[i]	der1[i]	der2[i]	int[i]
1	0.0000E+00	1.0000E+00	-1.0464E-01	0.0000E+00	0.0000E+00
2	1.0000E-01	9.9010E-01	-1.8949E-01	-1.6970E+00	9.9576E-02
3	2.0000E-01	9.6150E-01	-3.5927E-01	-1.6985E+00	1.9719E-01
4	3.0000E-01	9.1740E-01	-5.0257E-01	-1.1676E+00	2.9123E-01
5	4.0000E-01	8.6210E-01	-5.9443E-01	-6.6962E-01	3.8028E-01
6	5.0000E-01	8.0000E-01	-6.3989E-01	-2.3952E-01	4.6343E-01
7	6.0000E-01	7.3530E-01	-6.4701E-01	9.7170E-02	5.4022E-01
8	7.0000E-01	6.7110E-01	-6.2886E-01	2.6576E-01	6.1055E-01
9	8.0000E-01	6.0980E-01	-6.1557E-01	0.0000E+00	6.7459E-01

Výpočet integrálu

$$\int_{0}^{0.8} \frac{1}{1 + \mathrm{x}2} = 0.6746 - 0 = 0.6746$$

Závěr

Integrál funkce $1/(1+x^2)$ pro interval [0,0.8] se rovná číslu 0.6746.

Otázka 3. Uveďte základní rozdíly mezi spline vyhlazováním a spline regresí.

Spline vyhlazování je součást numerického vyhlazování, které se používá pro odstranění náhodných šumů ε_i v modelu $y_i = g(x_i) + \varepsilon_i$. Pro spline vyhlazování je charakteristické, že uzlové body ξ_i jsou totožné s *x*-ovými souřadnicemi zadaných experimentálních dat $\{x_i, y_i\}$ i = 1, ..., *n*. Spliny tedy nemusí přímo procházet uzlovými body. Požadavkem je, aby vyhlazující funkce byla dostatečně hladká a spojitá ve zvoleném počtu derivací. Nejčastějším případem je, když požadujeme, aby funkce *g*(*x*) byla spojitá v 1 a 2. derivaci. Ve spline vyhlazování nám nejde až tak nalezení tvaru aproximující funkce g(x), ale o rekonstrukci bezšumové závislosti. Ve spline regresi snažíme se proložit body po částech definované funkce. Uzlové body tvoří hranice intervalů, kde jsou definovány jednotlivé funkce, uzlovými body tedy křivka prochází.

Otázka 4. Co můžeme říct o filtru 53H.

Filtr 53H je robustní nelineární filtr. Je necitlivý k hrubým chybám, zařazujeme ho tedy tam, kde se očekává výskyt hrubých nenáhodných chyb. Je dán výrazem:

$$Z_i = \frac{S(5, i-2)}{4} + \frac{S(5, i-1)}{2} + \frac{S\ (5, i)}{4}$$

Z rovnice je vidět, že k vypočtení mediánu *i*-tého bodu počítá s dvěma body předchozími.
Filtr 53H užívá mediánu 5. Stupně (v=5), což lze vypočítat z jednoduchého vztahu:
$u = (v-1)/2$, kde u=2, protože filtr zohledňuje 2 poslední hodnoty.
Patří do skupiny rekurzivních filtrů, tedy fyzikálně realizovatelných - nepoužívá k vypočtení *i*-tého bodu následující hodnoty.