XXVIII. MAGYAR OPERÁCIÓKUTATÁSI KONFERENCIA,BALATONŐSZÖD, 2009. JÚN. 8-10.

Kalibrálás és konvex programozás – egy határterület áttekintése

Mihályffy László, KSH

Konvex programozási feladat

$$F(\mathbf{x}) = \min!$$

$$A\mathbf{x} = \mathbf{b}$$

$$\mathbf{x}^1 \le \mathbf{x} \le \mathbf{x}^2,$$

ahol $\mathbf{x}, \mathbf{x}^1, \mathbf{x}^2 \in \mathrm{R}^n$, $\mathbf{b} \in \mathrm{R}^m$, A $m \times n$-es,

$F: \mathrm{R}^n \to \mathrm{R}^1$, konvex, folytonosan differenciálható

Kalibrálási feladat

$$F(w_1,\dots,w_n,w_1^{o},\dots,w_n^{o})$$
$$=\sum_{j=1}^{n} G(w_j / w_j^{o}) = \min!$$

$$\sum_{j=1}^{n} x_{ij} w_j = X_i\,, \quad i = 1, 2, \dots, m \ (< n)$$

$$(\ L \le w_j / w_j^{o} \le U\,, \quad j = 1, 2, \dots, n\ )$$

$w_1^{o},\dots,w_n^{o}$ „design” súlyok

$w_1, w_2, \dots, w_n$ kalibrált súlyok

$0 \le L < 1$, $1 < U$, F: távolságfüggvény,

konvex, folyt. diffh., F=0 $\Leftrightarrow w_j \equiv w_j^{o}$

Kalibrálási feladat

Mátrix-vektor írásmód:

X*w* **=** *X*, ahol

X $m \times n$

$\boldsymbol{w} = (w_1, w_2, ..., w_n)^T$

és $X = (X_1, X_2, ..., X_m)^T$;

Négy feladat-típus

Távolságfüggvények:

Kvadratikus: $G(w, w^{o}) = ½(w - w^{o})^{2} / w^{o}$

Információ-divergencia::

$$G(w, w^{o}) = w \log \frac{w}{w^{o}} - w + w^{o}$$

I. típus: kvadratikus, individuális korlát nincs

II. típus: inf. div., individuális korlát nincs

III. típus: kvadratikus + individuális korlát

IV. típus: inf. div. +individuális korlát

I. típus

Lagrange-függvény:

$$\boldsymbol{\Psi} = ½\textstyle\sum_{j=1}^{n}(w_j - w_j^{o})^2 / w_j^{o} \text{ - } \lambda^{\mathrm{T}}(\mathbf{Xw} \text{ - } X)$$

$$\lambda = (\lambda_1, \ldots, \lambda_m)^T$$

Kalibrált súlyok::

$$\mathbf{w} = \mathbf{w}^{o} + \Omega\mathbf{X}^T(\mathbf{X}\Omega\mathbf{X}^T)^{-1}(X - \hat{X})$$

Célváltozó becsült értékösszege:

$$\hat{Y}^{\mathrm{kal}} = \hat{Y} + \mathbf{y}^T\Omega\mathbf{X}^T(\mathbf{X}\Omega\mathbf{X}^T)^{-1}(X - \hat{X})$$

$$= \hat{Y} + \textstyle\sum_{i=1}^{m} b_i(X_i - \hat{X}_i)$$

Ω=diag($w_1^{o}, \ldots, w_n^{o}$),

$\hat{X} = (\hat{X}, \ldots, \hat{X}_m)^T \leftarrow$ segédvált. becsült ért. össz.

I. típus tulajdonságai

- explicit előállítás,
- kalibrált becslés = regressziós becslés
- kalibrált súlyok lehetnek negatívok v. túl nagyok

II. típus:Iteratív arányos közelítések, „raking”

(Darroch-Ratcliff, 1972).

Kezdő érték: w_j^{o}

Iteráció:

(1) $r_i \sum_{j=1}^{n} x_{ij} w_j = X_i \quad i = 1, 2, \ldots, m$

(2) $w'_j = w_j \prod_{i=1}^{m} r_i^{x_{ij}/x_{.j}}$, ahol $x_{.j} = \sum_{i=1}^{m} x_{ij}$

Tulajdonságok konvergens, minimalizálja a távolságfüggvényt, $w_j \geq 0$, de lehet túl nagy

(2)-ben számtani közép is használható

III-IV. típus: Newton módszer

Laagrange-függvény:

$$\Psi = \sum_{j=1}^{n} G(w_j, w_j^{\text{o}}) - \lambda^T (\mathbf{Xw} - X)$$

$\Rightarrow \ \partial G / \partial w_j = \mathbf{X}_{.j}^T \lambda, \quad \mathbf{X}_{.j}$ az $\mathbf{X}$ j-edik oszlopa

$\Rightarrow \ w_j = h(\mathbf{X}_{.j}^T \lambda), \quad h$ a $\partial G / \partial w_j$ inverze

Módosítás az individuális korlátok miatt: ahol w_j kilépne a korlátok közül, a h függvényt a megfelelő alsó- v. felső korláttal tesszük egyenlővé

III-IV. típus: Newton módszer

Megoldandó egyenlet:

$X - \hat{X} - \phi(\lambda) = 0$, $\quad \phi(\lambda) = \sum_{j=1}^{n} h(\mathbf{X}_{.j}^{T}\lambda)\mathbf{X}_{.j} - \hat{X}$

Módosított célfüggvény a IV. típusnál:

$$G(v) = \frac{1}{A}\left[(v-L)\log\frac{v-L}{1-L} + (U-v)\log\frac{U-v}{U-1}\right],$$

$$A = (U-L)/((1-L)(U-1)), \qquad v = w_j / w_j^{o}$$

III-IV. típus: „Raking” módosítása

A (2) lépésben

$$w'_j = w_j \prod_{i=1}^{m} r_i^{\,x_{ij}/x_{.j}}$$

lehatároljuk a változó értékét, amennyiben túllép a korláton

Akkor is lehetséges, ha számtani középpel dolgozunk

Eljárások összehasonlítása

Négy eljárás:

(1) iteratív arányos közelítések, mértani átlaggal,

(2) iteratív arányos közelítések számtani átlaggal (KSH)

(3) Newton módszer, információ div. távolság-függvénnyel

(4) Newton módszer kvadratikus távolságfüggvénnyel

plusz mindegyiket a konvex programozással (Wolfe redukált gradiens módszer)

Összehasonlítási szempontok

- a konvex programozás minimumának megközelítése
- az eljárás hatása néhány célváltozóra
- gépidő-igény

Felhasznált szoftver:
(1)-(2) módszer: saját fejlesztés (SAS /Interactive Matrix Language)
(3)-(4) módszer: CALMAR SAS macro, O. Sautory, INSEE
Wolfe redukált gradiens módszere: saját fejlesztés (SAS IML)

Adatállományok

KSH munkaerő-felmérés, 2006. febr. és 2009. jan (kb. 10000-10000 háztartás)

Megyénkénti kalibrálás $\Rightarrow$ összesen 40 feladat, egyenként 22 segédváltozóval (2x10 korcsoportos létszám, háztartás-szám, megyei jogú városok lakossága)

Számítási eredmények (részlet) 1. Táblázat

Távolság-függvény	Távolság-fgv. értéke a minimum %-ában			
	1.	2.	3.	4.
55910	106.9	106.8	100.0	100.0
17143	108.1	108.3	100.3	102.6
9871	111.6	111.9	100.0	101.1
5160	106.2	106.2	100.0	100.5
9537	109.9	110.0	100.0	100.9
7613	115.3	115.3	100.0	102.3
6859	106.3	105.9	100.0	100.0
4914	110.3	110.9	100.0	100.0
3803	108.9	108.9	100.0	100.0
2735	110.2	109.9	100.0	100.1
4756	109.4	109.5	100.0	100.0
8034	108.5	108.5	100.0	100.1
7921	112.8	113.0	100.2	100.5

Számítási eredmények (részlet) 2a Táblázat

M	Foglalkoztatott					Munkanélküli				
	Létszám	1.	2.	3.	4.	Létszám	1.	2.	3.	4.
01	755618	0.00	-0.21	-0.22	-0.33	32183	0.02	0.57	1.00	0.73
02	145528	-0.07	-0.81	-0.41	0.05	14335	-0.03	-1.90	-1.60	-1.37
03	189137	0.02	-0.33	-0.33	-0.44	16732	-0.15	0.04	0.01	0.99
04	128268	0.03	-0.61	-0.66	-0.52	14398	0.01	3.41	3.13	2.50
17	91181	-0.00	-0.17	-0.15	-0.11	8641	0.03	0.50	0.90	0.04
18	114347	-0.01	0.03	0.02	0.04	9903	0.14	-0.24	-0.54	-1.20
20	131226	0.01	0.42	0.45	0.36	10420	-0.06	-2.12	-1.21	1.09
T	3865062	0.00	0.01	-0.06	-0.05	331249	-0.04	-0.44	0.52	0.27

Számítási eredmények (részlet)   2b Táblázat

M	Ráták, %						Rel. eltérések, %			
		1.	2.	3.	4.		1.	2.	3.	4.
01	4.09	4.09	4.12	4.13	4.13	.	0.02	0.75	1.17	1.02
09	9.16	9.15	9.26	9.29	9.23	.	-0.09	1.13	1.36	0.78
10	9.35	9.33	9.46	9.49	9.59	.	-0.25	1.14	1.47	2.48
11	10.81	10.81	10.40	10.41	10.34	.	0.06	-3.75	-3.64	-4.27
13	5.50	5.50	5.54	6.10	5.90	.	-0.01	0.86	10.88	7.27
14	9.43	9.40	9.42	9.41	9.52	.	-0.25	-0.13	-0.20	0.93
15	14.30	14.31	14.35	14.35	14.36	.	0.01	0.35	0.29	0.42
20	7.36	7.35	7.18	7.24	7.41	.	-0.07	-2.35	-1.54	0.67
T	7.89	7.89	7.86	7.94	7.92	.	-0.04	-0.42	0.53	0.30

Konklúziók

Az (1)-(2) módszerek megoldásánál a táv. függvény átlagosan 8-9 %-kal marad el a konvex programozás optimumától, a két módszere eredménye közel van egymáshoz

A (3)-(4) módszerek gyakorlatilag az optimálissal megegyező megoldást adnak (max.eltérés 1-2 %).

A négy módszer gyakorlatilag azonos eredményt ad a foglalkoztatott és a munkanélküli létszámra, valamint a munkanélküliségi rátára. Az átlagos relatív eltérés 0,5%, ill. 2-3 % a létszámadatoknál, és 0,5 százalékpont alatt van a rátánál.

<u>Az eltérések a mintavételi hiba határain belül vannak</u>

Konklúziók

Deville-Särndal (1992): a kalibrálási eljárások aszimptotikusa ekvivalensek

Műveleti sebesség: a CALMAR programmal a Newton módszer kb kétszer gyorsabb az iteratív arányos közelítéseknél. A konvex programozás lényegesen lassabb.

Végeredmény: mind a Newton módszernek, mind pedig az iteratív arányos közelítések módszerének indokolt a használata kalibrálási feladatok megoldására