PRIMJENA LINEARNOG PROGRAMIRANJA U REGRESIONOJ ANALIZI

1. UVOD

U statistici postoji više problema koji se mogu formulirati kao problemi matematičkog programiranja. Takav pristup statističkim problemima već se koristi u međusektorskoj analizi i operativnom istraživanju. Radi se o nekim problemima iz teorije uzoraka i regresione analize. Vjerojatno je da će taj krug problema uskoro biti mnogo širi.

Uzmimo, na primjer, problem optimalne alokacije uzorka na stratume i etape izbora. Kad se iz uzorka želi dobiti informacija samo o jednom obilježju, optimalna alokacija se određuje klasičnim metodama, tako da se minimizira iznos resursa za fiksnu preciznost ocjene, izraženu varijancom, ili se minimizira varijanca ocjene uz date resurse. Kad dva ili više obilježja treba ocijeniti simultano, problem alokacije postaje jako kompliciran. Optimalni raspored uzorka za jedno obilježje nije nužno optimalan za neko drugo obilježje. Neki autori kao R. Jagannathan [7, 8] rješavaju taj problem tako da minimiziraju funkciju troškova uz uvjet da varijanca ocjene svakog obilježja ne prelazi datu gornju granicu i uz još neka ograničenja (npr. na veličinu uzorka u svakom stratumu). Radi se, dakle, o problemu vezanog ekstrema s ograničenjima u formi nejednadžbi, koji po svojoj prirodi spada u matematičko programiranje.

Slični problemi javljaju se u regresionoj analizi kad se po kriteriju najmanjih kvadrata traži najbolje slaganje sa datim podacima a zna se a priori da neki parametri zadovoljavaju linearne nejednadžbe. Na takve probleme primjenjuju se metode kvadratnog programiranja (vidjeti P. Wolfe [18]).

Linearno programiranje u regresionoj analizi je posebna tema koju ćemo ovdje nastojati što potpunije obraditi.

U 2. odjeljku pokazat ćemo kako se problem linearne višestruke regresije svodi na ekvivalentni problem linearnog programiranja. Ukazat ćemo i na mogućnosti proširenja takvog postupka na neke slučajeve nelinearne regresije. U 3. odjeljku bavit ćemo se Čebiševim kriterijem optimalnosti ocjene. Pokazat ćemo kako se problem ocjene parametara linearne funkcije regresije po tom kriteriju svodi na jedan specijalni problem intervalnog linearnog programiranja pa ćemo u vezi s tim upozoriti na neke nove metode u linearnom programiranju. U nastavku rada zabilježit ćemo dva primjera iz prakse, jedan iz operativnog istraživanja a drugi iz međusektorske analize.

2. KRITERIJ NAJMANJIH APSOLUTNIH DEVIJACIJA

Dobro je poznato da metoda najmanjih kvadrata nije jedina i u svakoj prilici najbolja metoda. Da bi ta metoda dala optimalne ocjene, treba da budu ispunjene neke pretpostavke (naročito slučajni izbor i normalna distribucija). U protivnom slučaju metoda najmanjih kvadrata može dati pretjeranu težinu ekstremno velikim devijacijama.

Jedna alternativa minimiziranju sume kvadrata je minimiziranje sume apsolutnih devijacija. Ta alternativa ima prednost u tome što se može izraziti kao problem linearnog programiranja.

Uzmimo problem ocjene parametara u linearnoj regresiji

$$Y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k \tag{1}$$

iz ovih podataka

$$\begin{bmatrix} y_1 & x_{11} & x_{12} \ldots x_{1k} \\ y_2 & x_{21} & x_{22} \ldots x_{2k} \\ \ldots & \ldots & \ldots \quad \ldots \\ y_n & x_{n1} & x_{n2} \ldots x_{nk} \end{bmatrix}$$

gdje x_{ij}, $(i = 1, 2, \ldots, n; j = 1, 2, \ldots, k)$ predstavlja i-to opažanje na j-toj »nezavisnoj« varijabli dok je y_i odgovarajuća vrijednost »zavisne« varijable.

Žele se ocijeniti parametri b_j, $(j = 1, 2, \ldots, k)$ tako da minimiziraju ovu sumu

$$S\,(b_1, b_2, \ldots, b_k) = \sum_{i=1}^{n} |\, y_i - Y_i \,| \tag{2}$$

Budući da je funkcija S konveksna, problem nije jednostavan. Tim problemom bavilo se više matematičara (Edgeworth, Rhodes, Singleton, Karst, — da spomenemo samo autore novijih radova) ali bez osobitog uspjeha, jer su se njihove metode pokazale jednostavne i praktične samo u dvodimenzionalnom slučaju. Tek 1955. je pošlo za rukom Charnesu, Cooperu i Fergusonu [4] da svedu taj problem na problem linearnog programiranja.

»Vertikalna« devijacija može se ovako prikazati:

$$y_i - Y_i = u_i - v_i, \ (i = 1, 2, \ldots, n) \tag{3}$$

pri čemu je $u_i \geq 0$, $v_i \geq 0$. Kad se uvrsti (1) u (3), dobije se ovaj sistem

$$b_0 + \sum_{j=1}^{k} x_{ij}\, b_j + u_i - v_i = y_i, \quad (i = 1, 2, \ldots, n) \tag{4}$$

od n linearnih jednadžbi sa $2n + k + 1$ nepoznanica u_i, v_i i b_j. Sistem (4) je takve prirode da njegovo bazično rješenje ne sadrži $u_i > 0$ i $v_i > 0$ ni za jedno i. Naime, vektori koeficijenata tih varijabli su jedinični vektori suprotnog predznaka, dakle linearno zavisni. Prema tome, ako je u bazi u_i ne može u njoj biti v_i i obratno. Ograničimo li se, dakle, na bazična rješenja, možemo pisati

$$|u_i - v_i| = u_i + v_i$$

pa svaku vrijednost koju uzima funkcija S iz (2) može uzeti ova funkcija:

$$\sum_{i=1}^{n} (u_i + v_i) \tag{5}$$

Na taj način ta funkcija postaje kriterij optimalnosti rješenja sistema (4). Minimiziranje funkcije (5) uz uvjete (4) i uvjet nenegativnosti $u_i, v_i \geq 0$ je problem linearnog programiranja ekvivalentan problemu minimuma funkcije (2).[1])

Ako parametre b_j izrazimo ovako:

$$b_j = b_j^+ - b_j^-, \quad (b_j^+, \ b_j^- \geq 0)$$

i to uvrstimo u (4), dobijemo sistem od n linearnih jednadžbi u $2(n + k + 1)$ nenegativnih varijabla. Optimalno rješenje može se tada dobiti po simpleks metodi. Inicijalna baza je odmah pri ruci a sastoji se od n vektora koeficijenata varijabla u_i i v_i. Naime, kad je $y_i \geq 0$, u bazi je u_i a kad je $y_i < 0$, u bazi je v_i.

Ako je funkcija regresije nelinearna i aditivna u smislu da se dade napisati u ovome obliku:

$$Y = \sum_{j=1}^{k} f_j(x_j),$$

može se transformirati u jednu linearnu funkciju nekom transformacijom nezavisnih varijabli kao što se često radi u regresiji po metodi najmanjih kvadrata. To prosto znači da je $x_j = \log z_j$ ili $x_j = z^j$ i slično. Na primjer $z = ax^b y^c$ je takva funkcija. Ona je linearna u logaritmima od z, x i y, tj. $\log z = \log a + b \log x + c \log y$. M. H. Wagner [17] je išao korak dalje pa je pokazao da se nelinearni model regresije može tretirati metodama linearnog programiranja kad su ispunjeni stanoviti blagi uvjeti na prirodu nelinearnosti, kao što su monotonost i konkavnost.

Treba upozoriti na jednu teškoću u provođenju opisane procedure, koja nastaje kad je n velik broj. Može se, naime, desiti da n broji hiljadu i više opservacija a broj parametara k je relativno malen. U tom slučaju sistem (4) postaje pregolem i računski teško savladiv. Izlaz je u tome da se ide na dualni problem.

$$\max \sum_i y_i d_i$$

$$\sum_i d_i = 0$$

---

[1]) Taj način linearizacije konveksnog funkcionala primijenio je nedavno B. Contini [5] u jednom problemu optimalne proizvodne sekvencije u strojogradnji.

(6) $$\sum_i x_{ij} d_i = 0, \ (j = 1, 2, \ldots, k)$$

$$d_i \leq 1, \ (i = 1, 2, \ldots, n)$$

$$-d_i \leq 1$$

ili, ako pišemo $f_i = d_i + 1$, $(i = 1, 2, \ldots, n)$, i uvrstimo gore, dobijemo

(7) $$\max \sum_i (y_i f_i - y_i)$$

$$\sum_i f_i = n$$

$$\sum_i x_{ij} f_i = \sum_i x_{ij}, \quad (j = 1, 2, \ldots, k)$$

$$0 \leq f_i \leq 2, \quad (i = 1, 2, \ldots, n)$$

Taj problem sa n odozgo omeđenih varijabli d_i ima samo $k + 1$ ograničenja. Može se riješiti npr. Wagnerovom dualnom metodom i kao nusprodukt daje optimalne vrijednosti parametara b_j.

3. KRITERIJ ČEBIŠEVA

Čebiševljev princip aproksimacije je ekvivalentan minimiziranju maksimalne apsolutne vrijednosti devijacija, to jest

(8) $$\min_{b_j} \left\{ \max_i \mid Y_i - y_i \mid \right\}$$

ili

(8′) $$\min_{b_j} \left\{ \max_i \left| \sum_{j=0}^{k} x_{ij} b_j - y_i \right| \right\},$$

gdje je $x_{io} \equiv 0$ za svako i.

Ako uvedemo ovu oznaku:

(9) $$\max_i \left| \sum_j x_{ij} b_j - y_i \right| = \varepsilon ,$$

možemo pisati

(10) $$\left| \sum_j x_{ij} b_j - y_i \right| \leq \varepsilon \qquad (i = 1, 2, \ldots, n)$$

Odatle slijedi

(11) $$-\varepsilon \leq \left(\sum_j x_{ij} \ b_j - y_i \right) \leq \varepsilon \qquad (i = 1, 2, \ldots, n)$$

pri čemu se traži

(12) $$\min \ \varepsilon$$

uz

(13) $$\varepsilon \geq 0$$

Problem (11) (12) i (13) je specijalni problem intervalnog linearnog programiranja (Vidjeti npr. Lj. Martić [10]). Ako uvedemo novu varijablu

$$v = \frac{1}{\varepsilon},$$

možemo taj problem ovako pisati:

(14) $$\text{Max } v$$

(15) $$-1 \leq \left(\sum_j x_{ij}\, b'_j - y_i v\right) \leq 1, \quad (i = 1, 2, \ldots, n)$$

(16) $$v \geq 0,$$

gdje je

$$b'_j = \frac{b_j}{\varepsilon}.$$

Sada vidimo da doista imamo posla sa intervalnim linearnim programom. Problem (14), (15) i (16) može se riješiti standardnim metodama (vidjeti H. M. Wagner [16]). Od 1968. god. raspolažemo sa dvije posebne iterativne metode za takve slučajeve. Autor je P. D. Robers [12]. A. Charnes i A. Ben-Israel [2] dali su eksplicitno rješenje problema intervalnog linearnog programiranja, ali za sada to rješenje ima samo teoretske prednosti pred dobro poznatim iterativnim metodama linearnog programiranja.

Interesantna je veza između kriterija najmanjih kvadrata i Čebiševljeva kriterija.

Ako devijacije poslije primjene metode najmanjih kvadrata označimo sa

$$\rho_1, \rho_2, \ldots, \rho_n,$$

to jest

(17) $$Min \sum_i (Y_i - y_i)^2 = \sum_i \rho_i^2,$$

tada je očigledno

(18) $$\sum_i \rho_i^2 \leq \sum_i r_i^2,$$

gdje je r_i oznaka za devijaciju empiričke od teoretske vrijednosti, izračunatu na neki drugi način.

Iz (18) slijedi

(19) $$\sum_i \rho_i^2 \leq \sum_i r_i^2 \leq n r_{max}^2$$

a odatle

(20) $$\sqrt{\frac{\sum \rho_i^2}{n}} \leq r_{max}$$

Ako je r_{max} proizašlo iz primjene Čebiševljeva kriterija, to jest,

(21) $$r_{max} = \min_{b_j} \left\{ \max_{i} \left| Y_i - y_i \right| \right\},$$

tada je

(22) $$r_{max} \leq \rho_{max}$$

No, iz (20) i (22) slijedi

(23) $$\sqrt{\frac{\sum_i \rho_i^2}{n}} \leq r_{max} \leq \rho_{max}$$

ili

(23') $$\sigma_\rho \leq r_{max} \leq \rho_{max}$$

Na taj način u stanju smo ograditi Čebiševljevu »distancu« sa veličinama što proizlaze iz primjene Gausove metode najmanjih kvadrata.

4. PRIMJENE U OPERATIVNOM ISTRAŽIVANJU I MEĐUSEKTORSKOJ ANALIZI

1. *Nagrađivanje osoba na rukovodećem položaju u poduzeću.* A. Charnes u suradnji sa W. W. Cooperom i R. O. Fergusonom razvio je jednu eksplicitnu formulu koja je trebala da služi kao uputstvo za nagrađivanje rukovodioca s obzirom na njihov nivo odgovornosti i na njihovu potencijalnu vrijednost i izglede u organizaciji. U tu svrhu klasificirani su svi poslovi odnosno položaji u organizaciji, određen iznos svakog faktora (inicijativa, iskustvo, obrazovanje itd.), koji se zahtijeva na svakom nivou posla, utvrđene donje i gornje granice za primanja itd. Konačno postavljena je formula u ovome obliku:

(24) $$S = \sum_{i=1}^{n} a_i \, y_i,$$

gdje je

S = plaća;

y_i = iznos faktora i koga posjeduje osoba čiju plaću treba odrediti;

a_i = težina (ponder) koji se dodjeljuje faktoru i.

Nadalje

x_{ik} je poznati iznos faktora i izražen u bodovima, koji se traži na nivou posla $k = 1, 2, \ldots, L$. Poslovi su rangirani u opadajućem redoslijedu od 1 do L, tako da k označava poziciju posla u organizacijskoj hijerarhiji. Plaće pridružene tim poslovima su isto tako u opadajućem redoslijedu.

S_M i S_m su gornja odnosno donja granica plaće. Ako je potrebno, mogu se upotrebiti i neke međuvrijednosti.

Uvjeti problema mogu se ovako formulirati:

(25)
$$\sum_{i=1}^{n} a_i x_{i1} \leq S_M$$
$$\sum_{i=1}^{n} a_i x_{i2} \leq \sum_{i=1}^{n} a_i x_{i1}$$
$$— — — — — —$$
$$\sum_{i=1}^{n} a_i x_{iL} \leq \sum_{i=1}^{n} a_i x_{i,L-1}$$
$$S_m \leq \sum_{i=1}^{n} a_i x_{iL}$$
$$a_i \geq 0 \text{ za svako } i$$

Problem se sastoji u tome da se nađu težine a_i, tako da suma apsolutnih devijacija od tih nivoa bude minimalna, to jest

(26)
$$\min \sum_{k \varepsilon K} \left| \sum_{i=1}^{n} a_i x_{ik} - S_k \right|,$$

gdje je K skup indeksa od plafona, osnovice i intermedijarnih nivoa plaća. Kad se tako ocijene težine a_i, uvrste se u formulu (24) i plaća je određena (usporediti A. Charnes and W. W. Cooper [3], Chapter *X*).

2. *Ocjenjivanje strukturalnih koeficijenata.* U međusektorskoj analizi ocjenjuju se »input-output« koeficijenti a_{ij} za jednu godinu iz ove relacije

(27)
$$x_{ij} = a_{ij} x_j + u_j,$$

gdje x_{ij} predstavlja tok iz sektora i u sektor j, x_j je »output«-j-tog sektora a u_j je slučajna varijabla sa srednjom vrijednosti 0.

Druga metoda ocjene koeficijenata a_{ij} se svodi na prilagođavanje relacije

(28)
$$x_i = \sum_{j=1}^{n} a_{ij} x_j + f_i$$

podacima o »outputima« x_j i finalnoj potražnji f_i u nekom periodu vremena uz pretpostavku da se a_{ij} ne mijenjaju u tom periodu. Primjena ove metode zahtijeva dovoljno duge vremenske serije »outputa« i finalne potražnje.

Kenneth J. Arrow i Marvin Hoffenberg [1] sa grupom istraživača iz poznate RAND korporacije pošli su od pretpostavke da za sektore industrije potrošnih dobara koeficijenti a_{ij} variraju u vremenu u skladu sa ovom relacijom:

(29)
$$a_{ij}(t) = b_{ij} + c_{ij}(t) + d_{ij}w(t) + e_{ij}y(t) + f_{ij}[v_j(t)/x_j(t)],$$

gdje je t oznaka za vrijeme a w, y i v_j su mjere participacije ekonomije u ratu, realni raspoloživi dohodak po glavi odnosno stupanj obrazovanja radne snage. Arrow i Hoffenberg ocijenili su strukturalne koeficijente b_{ij}, c_{ij}, d_{ij}, e_{ij} i f_{ij} metodama linearnog programiranja, tako da minimiziraju sumu apsolutnih devijacija u bilancnim jednadžbama (28), to jest

$$min \sum_t \left| x_i(t) - \sum_{j=1}^{n} a_{ij}(t) \; x_j(t) - f_i(t) \right| \tag{30}$$

uz još neke uvjete na varijable.

Fakultet ekonomskih nauka, Zagreb

Ljubomir MARTIĆ

LITERATURA

1. K. J. Arrow and M. Hoffenberg, *A Time Series Analysis of Interindustry Demands*, Amsterdam, 1959.
2. A. Ben-Israel and A. Charnes, »An Explicit Solution of a Special Class of Linear Programming Problems«, *Operations Research*, Vol. 16, 1968, No. 6, 1166—1175.
3. A. Charnes and W. W. Cooper, *Management Models and Industrial Applications of Linear Programming*, Vol. 1, New York, 1961.
4. A. Charnes, W. W. Cooper and R. Ferguson, »Optimal Estimation of Executive Compensation by Linear Programming«, *Management Science*, Vol. 1, 1955, No. 2, 138—151.
5. B. Contini, »Un modello di sequenze di produzione risolvibile con la programmazione lineare«, *Metroeconomica*, Vol. 20, 1968, No. 1, 36—41.
6. W. D. Fisher, »A Note on Curve Fitting with Minimum Deviations by Linear Programming«, *Journal of the American Statistical Association*, Vol. 56, 1961, No. 294, 359—362.
7. R. Jagannathan, »The Programming Approach in Multiple Character Studies«, *Econometrica*, Vol. 33, 1965, No. 1, 236—7.
8. R. Jagannathan, »A Method for Solving a Nonlinear Programming Problem in Sample Surveys«, *Econometrica*, Vol. 33, 1965, No. 4, 841—846.
9. O. Y. Karst, »Linear Curve Fitting Using Least Deviations«, *Journal of the American Statistical Association*, Vol. 53, 1958, No. 281, 118—132.
10. Lj. Martić, »Linearni intervalni programi«, *Ekonomska analiza*, god. 3, 1969., br. 1—2, 80—83.
11. P. Rabinowitz, »Applications of Linear Programming to Numerical Analysis«, *SIAM Review*, Vol. 10, 1968, No. 2, 121—159.
12. P. D. Robers, *Interval Linear Programming*, Evanston, 1968.
13. E. Stiefel, »Note on Jordan elimination, linear programming and Tchebycheff approximation«, *Numerische Mathematik*, 2. Band, 1960, 1. Heft, 1—17.
14. E. L. Stiefel, *An Introduction to Numerical Mathematics*, New York, 1963.
15. W. Vogel, *Lineares Optimieren*, Leipzig, 1967.
16. H. M. Wagner, »Linear Programming Techniques for Regression Analysis«, *Journal of the American Statistical Association*, Vol. 54, 1959, No. 285, 206—212.
17. H. M. Wagner, »Non-Linear Regression with Minimal Assumptions«, *Journal of the American Statistical Association*, Vol. 57, 1962, No. 299, 572—578.
18. P. Wolfe, »The Simplex Method for Quadratic Programming«, *Econometrica*, Vol. 27, 1959, No. 3, 382—398.