METODE DE PRELUCRARE A DATELOR EXPERIMENTALE

Membru corespondent al AŞM
Evgheni LVOVSCHI

METHODS FOR PROCESSING THE EXPERIMENTAL DATA

The following articles describemethod confluente of processing mathematic-statistic data, alsowith the help of the computer.

În nr. 1 – 2 (9), 2008 al revistei *Akademos* a fost publicat un articol cu titlul de mai sus. Textul dat este o continuare a materialului anterior.

Se ştie că există trei situaţii experimentale:

1. Funcţia *y* şi factorul *x* sunt măsurate exact (fără greşeli).
2. Funcţia *y* este măsurată cu greşeli, iar factorul *x* – fără.
3. Şi funcţia *y*, şi factorul *x* se măsoară cu greşeli.

Primele două situaţii au fost discutate în articolul precedent.

S-a evidenţiat că în primele două situaţii poate fi aplicată metoda pătratelor minime. Spre deosebire de acestea, în cea de-a treia situaţie metoda pătratelor minime nu poate fi aplicată direct. Este utilizat un caz particular al metodei verosimilităţii maxime, şi anume *analiza confluentă*, termen ce provine de la cuvântul francez *confluer* – contopire.

Metoda verosimiliăţii maxime în condiţii speciale poate să fie redusă la metoda pătratelor minime cu rezolvarea prin iteraţii. Aşa dar, problema analizei confluente se reduce la metoda pătratelor minime cu utilizarea unei aproximaţii consecvente.

Aici se discută cel mai simplu caz de dependenţă-pereche, şi anume *y=f(x)*. De fapt, despărţirea acestor două variabile în variabilă dependentă şi variabilă independentă îşi pierde sensul, atunci când ambele variabele sunt supuse oscilaţiei aleatoare.

Probabilitatea faptului că această funcţie în populaţia statistică generală se va găsi lângă x_i şi y_i, când repartizarea este dublu normală:

$$P(x_i y_i;\xi,\eta)=P(x_i\xi)P(y_i\eta)=\frac{1}{2\pi\sigma_{x_i}\sigma_{y_i}}\exp\left[-\frac{1}{2}\left(\frac{x_i-\xi}{\sigma_{x_i}}\right)^2-\frac{1}{2}\left(\frac{y_{i_i}-\eta}{\sigma_{y_i}}\right)^2\right] \quad (1)$$

dacă centrul populaţiei statistice generale este ξ,η.

Probabilitatea de a găsi evenimentul în punctul $x_i y_i$ este proporţională lungimii curbei teoretice şi a unei funcţii $\psi(\xi)$.

Gradul maxim privind probabilitatea găsirii evenimentului în punctul $x_i y_i$ se determină prin integrarea pe arcul curbei:

$$P(x_i y_i)=\int dSP(x_i y_i;\xi,\eta(\xi) \quad \psi(\xi) \quad (2)$$

unde elementul arcului curbei:

$$dS^2=d\xi^2+\frac{\sigma_{x_i}^2}{\sigma_{y_i}^2}d\eta^2 \quad (3)$$

are aceeaşi măsură ca şi $d\xi$.

Funcţia $\psi(\xi)$ reprezintă intensitatea punctelor obţinute prin observaţie şi poate fi determinată prin analiza amănunţită a sensului fizic al gradului de probabilitate:

$$P(x_i y_i)=\int d\xi P(x_i y_i;\xi,\eta(\xi)). \quad (4)$$

Funcţia verosimilităţii este egală cu probabilitatea de a găsi sincronic toate punctele în acele locuri unde ele sunt observate. De aceea, fiindcă observaţiile sunt independente la diferite x_i, acestea au înfăţişarea produsului integralelor (2), luat pe toate evenimentele observate.

În unele condiţii, care de regulă sunt îndeplinite în experiment, problema confluentă poate fi redusă la o succesiune de probleme de regresie obişnuită (discutată anterior). Dacă pe o porţiune de curbă, care se găseşte în intervalul de la $-\sigma_{x_i}$ până la $+\sigma_{x_i}$ şi de la $-\sigma_{y_i}$ până la $+\sigma_{y_i}$ înclinaţia şi curbura variază puţin (curba este destul de netedă), atunci la calculul integralului (2) în caz de repartizare normală se poate limita numai la porţiunea curbei lângă punctul $x_i y_i$. Altă condiţie, care permite de a reduce problema confluentă la o succesiune de probleme de regresie simplă, este cerinţa ca punctul $x_i y_i$ să nu se găsească mai aproape de sfârşitul curbei decât $-\sigma_{x_i}$. În astfel de condiţii, pe o porţiune mică de curbă în apropierea punctului $x_i y_i$ curba teoretică poate fi exprimată prin primii trei membri ai seriei lui Teilor

$$\eta(\xi)=\hat{\eta}(x_i)+(\xi-x_i)\hat{\eta}'(x_i)+ \\ +\frac{(\xi-x)^2}{2}\hat{\eta}''(x_i) \quad (5)$$

unde caracteristicele curbei $\eta(\xi)$ sunt înlocuite cu caracteristicele evaluării ei: $\hat{\eta}$. Dacă (5) se introduce în (1), iar (1) în (4), descompunând (1) la puterile $\hat{\eta}$ în presupunerea că $\frac{\sigma_{x_i}^2}{\sigma_{y_i}^2}\hat{\eta} \prec 1$ şi concomitent realizând operaţiunea de integrare practic în limitele infinite până la valoarea $\hat{\eta}$, atunci prin acesta se obţine valoarea densităţii probabilităţii:

$$P(x_i y_i) = \frac{1}{\sigma_i\sqrt{2\pi}} \times \exp \times \times$$

$$\times\{-\frac{1}{2}\frac{[\hat{\eta}'(x_i) - y_i + \frac{1}{2}\hat{\eta}''\sigma_{xi}^2}{\sigma_i^2}\times \tag{6}$$

$$\times\frac{(\sigma_{yi}^2 - 2\sigma_{xi}^2\hat{\eta}_i'^2)\sigma_i^{-2}]}{\sigma_i^2}\times$$

$$\times\left[1 + \hat{\eta}''(\hat{\eta}(x_i) - y_i)(\hat{\eta}_i'\frac{\sigma_{xi}^2}{\sigma_i^2})^2\right]\}$$

unde

$$\sigma_i^2 = \sigma_{y_i}^2 + \hat{\eta}'^2\sigma_{x_i}^2 \tag{7}$$

dar $\hat{\eta}_i', \hat{\eta}''$ -– sunt valorile primei şi celei de a doua derivate ale evaluării curbei regresiei, când $x = x_i$.

În formula (6) în calitate de greutăţi nenormative sunt introduse

$$w_i = \left[\sigma_{y_i}^2 + \sum_{k=1}^{n}\left(\frac{\partial\eta\left(\vec{x}, \hat{\vec{\beta}}^0\right)}{\partial x_k}\right)^2 \Bigg\uparrow_{x=\vec{x}_i} \sigma_{x_{ik}}^2\right]^{-1} \tag{8}$$

Dincolo de faptul că centrul de repartizare pentru y_i este evaluarea mutată a curbei regresiei: $\eta(x_i) + \alpha_i$, unde

$$\alpha_i = \frac{1/2\,\hat{\eta}''\sigma_{x_i}^2(\sigma_{y_i}^2 - 2\sigma_{x_i}^2\hat{\eta}'^2)}{\sigma_i^2} \tag{9}$$

Fiindcă derivatele $\hat{\eta}_i'$ şi $\hat{\eta}_i''$ precum şi devierile $\Delta y_i = \hat{\eta}(x_i) - y_i$ până la executarea analizei nu sunt cunoscute, la prima vedere se creează impresia unui „cerc fermecat”. Din cercul acesta se poate ieşi cu ajutorul unor succesiuni în baza faptului că funcţia verosimilităţii depinde mai puţin de devierea greutăţilor în procesul aproximaţiei curbei, decât de devierile $y_i - \hat{\eta}(x_i)$.

Aproximaţia „zero” poate fi obţinută trasând „la ochi” curba prin punctele experimentale. Prima derivată se obţine efectuând diferenţierea pe curba aproximaţiei „zero”. În continuare în numitor se depune expresia (8), derivata de gradul doi se înlocuieşte cu zero şi greutăţile obţinute se analizează din nou. Diferenţierea curbei primei aproximaţii permite a găsi valorile $\hat{\eta}_i'$, $\hat{\eta}_i''$ şi Δy cu ajutorul cărora iarăşi se recalculează greutăţile şi se evaluează devierile (9).

Această procedură de precizare a curbei şi greutăţilor continuă până devierea ultimelor nu va fi mai mică decât valoarea cifrei numită prealabil şi care exprimă exactitatea calculelor efectuate. Când valorile $\frac{\sigma_x}{\sigma_y}\eta'$ şi $\frac{\sigma_x^2}{\sigma_y^2}\eta''$ sunt destul de mici, potrivirea procesului de iteraţii trebuie să fie destul de rapidă. Cum s-a mai evidenţiat în privinţa calculului, problema confluentă este echivalentă cu o succesiune de probleme de regresie. Când $\sigma_y \to 0$ şi dacă observaţiile x_i sunt numere fixe, toate formulele analizei confluente automat trec în formulele analizei de regresie (situaţia a doua experimentală).

Să examinăm analiza confluentă în situaţia cu mai mulţi factori x_j.

1) se determină valorile $\hat{\vec{\beta}}^0$,

($\hat{\vec{\beta}}^0$ - evaluarea la zero a vectorului coloanei de coeficienţi de regresie β_j, care aprovizionează minimul funcţiei, greutăţile fiind constante $w_i = \frac{1}{\sigma_{y_i}^2}$;

2) în fiecare punct experimental $\vec{x}_i$ se calculează valorile greutăţilor şi devierilor:

$$\alpha_i^1 = \sum_{k=1}^{n}\frac{\partial^2\eta(\vec{x}, \hat{\vec{\beta}}^0)}{\partial x_k\partial x_k}\Bigg|^{I}_{\vec{x}=x_i}\sigma_{x_{ik}}^2 \tag{10}$$ unde

$\sigma^2_{x_{ik}}$ - dispersia determinării coordonatei k la punctul i pentru $\vec{x}_i$. Derivatele parțiale se obțin prin diferențierea la numerar cu formulele:

$$\frac{\partial \eta(\vec{x}, \overset{0}{\hat{\vec{\beta}}})}{\partial x_k}\Big|_{\vec{x}=\vec{x}_i}^{1} = \frac{1}{2h}[\eta(\overset{0}{\hat{\vec{\beta}}}, x_{ij} k_{ik} \tag{11}$$

$$+h) - \eta(\overset{0}{\hat{\vec{\beta}}}; x_{ij} x_{ik} - h)] - \frac{h^2}{6}\eta^{III}(\overset{0}{\hat{\vec{\beta}}}; x_{ij}\xi_{ik});$$

$$\frac{\partial^2 \eta(\vec{x}, \overset{0}{\hat{\vec{\beta}}})\Big|_{\vec{x}=\vec{x}_i}^{1}}{\partial x_k \partial x_k}$$

$$= \frac{1}{h^2}[\eta(\overset{o}{\hat{\vec{\beta}}}; x_{ij}; x_{ik} - h) + 2\eta(\overset{0}{\hat{\vec{\beta}}}; x_{ij}, x_{ik}) + \tag{12}$$

$$+ \eta(\overset{0}{\hat{\vec{\beta}}}; x_{ij} x_{ik} + h] - \frac{h^2}{12}\eta^{IV}$$

$$\times(\overset{0}{\hat{\vec{\beta}}}; x_{ij}; \xi_{ik})$$

unde $\sigma^2_{x_{ik}}$ – este dispersia determinării coordonatei k punctului i $\vec{x}_i$. Derivatele particulare se determina prin metoda diferențierii la numerar cu formulele derivatelor. Influența pasului de diferențiere h se depistează cu evidența erorii de calcul a pasului la exactitatea calculului, se verifică prin recursul repetat la program cu micşorarea pasului h;

3) se determină valorile $\overset{1}{\vec{\beta}}$, care aprovizionează minimul funcției U_1 cu greutățile w_i constante:

$$U_1 = \sum_{i=1}^{n} W_i \left[y_i - \eta\left(\vec{\beta}; \vec{x}_i\right) - \frac{1}{2}\alpha_i^1 \right]^2 \tag{13}$$

operațiunile 2) şi 3) se repetă relativ la $\overset{1}{\hat{\vec{\beta}}}$ şi $\overset{2}{\hat{\vec{\beta}}}$ şi aşa mai departe, până procesul nu se potriveşte:

$$\max_1 \frac{\overset{m}{\hat{\beta}}_i - \overset{m-1}{\hat{\beta}}_i}{\overset{m}{\hat{\beta}}_i} \subseteq \varepsilon \text{ unde } i = \overrightarrow{1, p} \tag{14}$$

este cantitatea parametrilor necunoscuți, ε este o cifră numită prealabil care e comparabilă cu eroarea posibilă. Valoarea $\overset{m}{\hat{\vec{\beta}}}$, pentru care se îndeplineşte condiția (13), se admite ca evaluare a parametrilor căutați pentru modelul analizei confluente.

Metodele de prelucrare a datelor experimentale, expuse în acest articol, reprezintă o mică parte din amplele cercetări desfăşurate de autor. Ele urmează să fie inserate într-o carte, care promite a fi o noutate absolută în arealul ştiințific românesc.

Literatura

1. Клепикова Н.П., Соколов С.Н. Анализ и планирование экспериментов методом максимума правдоподобия – Москва, Наука, 1964.

2. Николаева Л.С. Программы по регрессионному конфлюэнтному анализу. Препринт. М. Издательство МГУ, 1969, № 9.

3. Федоров В.В. Анализ экспериментов при наличии ошибок в контролируемых переменных. – М., Наука, 1968 №2.

4. Андрукович П.Ф., Николаева Л.С., Федоров В.В. Программы по регрессионному и конфлюэнтному анализу. Издательство МГУ, 1969, №1.

5. Львовский Е.Н. Статистические методы построения эмпирических формул. М. «Высшая школа» 1988.