Multidiszciplináris tudományok, 9. kötet. (2019) 4 sz. pp. 455-459 **https://doi.org/10.35925/j.multi.2019.4.47**

A LEGGYAKORIBB ÉRTÉK ÉS DIHÉZIÓ MÓDOSÍTÁSA

Fegyverneki Sándor
egyetemi docens, Miskolci Egyetem, Alkalmazott Matematikai Intézeti Tanszék
3515 Miskolc, Miskolc-Egyetemváros, e-mail: matfs@uni-miskolc.hu

Absztrakt

Ebben a cikkben definiáljuk a klasszikus hely- és skálaparaméter problémát. Leírjuk a leggyakoribb értéket és dihéziót. Javaslunk egy megoldást a problémára az alapegyenletek módosításával.

Kulcsszavak: *robusztus becslések, leggyakoribb érték, dihézió, Cauchy eloszlás*

Abstract

In this paper we define the classical location and scale parameter problem. The most frequent value and dihesion were described. We suggest a solution for the problem with modification of the basic equations of the solution.

Keywords: *robust estimators, most frequent value, dihesion, Cauchy distribution*

1. Bevezetés

Adott az $x_1, x_2, \dots, x_n$ mérési sorozat. Általános statisztikai feladatnak tekinthető egy középérték és egy szóródási jellemző keresése. Határozzunk meg T_n és s_n értékeket, azaz középpontot és skálázást. Milyen szempontból jellemzik ezek az értékek a halmazt? A feladat egyszerű és nehéz, mert a probléma általános. Hogyan választunk szempontokat és milyen feltételezett ismereteket tudunk? Ezekre próbálunk válaszolni.

1.1. A paraméterbecslési feladat

Az F és G eloszlásfüggvények típusa megegyezik, ha valamely $a > 0$ és b konstansok mellett

$$G(x) = F(ax + b).$$

Ez egy ekvivalenciarelációt, osztályozást határoz meg. Ezután megfogalmazhatjuk a matematikai statisztikai feladatot.
Adott a

$$\xi_1, \xi_2, \dots, \xi_n, \quad \xi_i \sim F$$

független minta, a mintaelemek eloszlása az F_0 eloszlásfüggvénnyel reprezentált eloszlástípusba tartozik.
A paraméterbecslési feladat: A μ, σ paraméterek becslése úgy, hogy az

$$F_0\left(\frac{x-\mu}{\sigma}\right) = F(x)$$

egyenlőség teljesüljön.

1.2. Megoldási módszerek

1. Ha $E(\xi^2)$ létezik, akkor

$$E(\eta) = \sigma E(\xi) + \mu,$$
$$D^2(\eta) = \sigma^2 D^2(\xi).$$

Ez az ún. momentumok módszere, amellyel számos probléma adódik: létezési problémák, robusztusság, meghatározási nehézségek stb.
2. A másik általánosan használt módszer a maximum likelihood. Legyen a minta sűrűségfüggvénye

$$\xi_i \sim f_0\left(\frac{x-\mu}{\sigma}\right) \quad (i = 1,2,\dots,n).$$

$$\min_{\mu,\sigma}\left\{-\sum_{i=1}^{n} \ln\, f_0\left(\frac{\xi_i-\mu}{\sigma}\right)\right\}$$

Ha létezik f_0', akkor a loglikelihood függvény deriválsása után kapjuk, hogy

$$\sum_{i=1}^{n} \frac{f_0'\left(\frac{\xi_i-\mu}{\sigma}\right)}{f_0\left(\frac{\xi_i-\mu}{\sigma}\right)} = 0$$

$$\sum_{i=1}^{n} \frac{\xi_i-\mu}{\sigma} \frac{f_0'\left(\frac{\xi_i-\mu}{\sigma}\right)}{f_0\left(\frac{\xi_i-\mu}{\sigma}\right)} = 0$$

Itt is számos probléma adódik: létezés, robusztusság, meghatározási nehézségek, érzékenység stb.
3. Huber-típusú jelölés (általánosította az előző, likelihood alapú formulákat):

$$\sum_{i=1}^{n} \psi\left(\frac{\xi_i - T}{s}\right) = 0,$$
$$\sum_{i=1}^{n} \chi\left(\frac{\xi_i - T}{s}\right) = 0,$$

ahol ψ és χ a hely- illetve skálaparaméterhez tartozó ún. ψ függvény [5], [6].
Néhány becslés leírása a Huber-féle jelölésekkel:

1. Maximum likelihood:

$$\psi(x) = \frac{f'(x)}{f(x)}, \chi(x) = -x\frac{f'(x)}{f(x)} - 1.$$

2. Medián és MAD:

$$\psi(x) = sgn(x), \chi(x) = sgn(|x| - 1).$$

3. Átlag és szórás:

$$\psi(x) = x, \chi(x) = x^2 - 1.$$

4. Huber-féle becslés (1964):

$$\psi_b(x) = x \cdot \min \left\{1, \frac{b}{|x|}\right\},$$

$$\chi(x) = {\psi_b}^2(x) - \int {\psi_b}^2(y) d\Phi(y),$$

ahol Φ a standard Gauss-eloszlásfüggvénye.

2. Leggyakoribb érték, dihézió

A javasolt egyenletrendszert a

$$\psi(x) = \frac{x}{1+x^2}, \quad \chi(x) = \frac{3x^2-1}{(1+x^2)^2}$$

függvények definiálják. A kapott paraméterértékeket elnevezték leggyakoribb értéknek és dihéziónak. Honnan kapjuk ezeket a függvényeket? Gyakorlatilag azt a két értéket határozza meg az eljárás, amelyekhez tartozó Cauchy-eloszlás "legközelebb" van az adott mintához. A Cauchy-eloszlás

$$f(x) = \frac{s}{\pi(s^2 + (x-c)^2)},$$

ahol a c helyparaméter nem ismert. Ekkor nem létezik a várható érték. A Csernyák, Steiner megoldás [2], [10], [11].
Az I-divergencia értelmezése a következő:

$$I(f||g) = \int_{-\infty}^{+\infty} f(x) \ln \frac{f(x)}{g(x)} dx,$$

ahol f és g sűrűségfüggvények. Ha az I-divergenciát mini−malizáljuk, amikor a helyparamétert változtatjuk és

$$g(x;\vartheta) = \frac{1}{\pi(1 + (x-\vartheta)^2)},$$

azaz a Cauchy-eloszlás sűrűségfüggvénye ismeretlen helyparaméterrel, akkor a következőket kapjuk:

$$\int_{-\infty}^{+\infty} \frac{\partial g(x;\vartheta)}{\partial \vartheta} \frac{f(x)}{g(x;\vartheta)} dx = 0.$$

$$\int_{-\infty}^{+\infty} \left[\frac{\partial g(x;\vartheta)}{\partial \vartheta} \frac{1}{g(x;\vartheta)}\right]^2 f(x)dx -$$

$$- \int_{-\infty}^{+\infty} \frac{\partial^2 g(x;\vartheta)}{\partial \vartheta^2} \frac{f(x)}{g(x;\vartheta)} dx > 0,$$

ekkor kapjuk a minimumot. Ha feltételezzük, hogy az utóbbi kifejezés második fele 0, akkor az egyenlőtlenség biztosan teljesül. Tehát a kapott megoldás minimalizálja az I-divergenciát és az utóbbi felté-

telezés megad egy a skálaparaméter meghatározására alkalmas egyenletet, s ezek a Cauchy-eloszlás esetén pontosan a ψ, χ párral megadott esethez vezetnek.
A numerikus meghatározások során hamar kiderült, hogy az összes eset néhány százalékában a javasolt sorozatok nem konvergálnak. Születtek módosítások és javítások, de végleges megoldás nem [10], [11]. A valódi probléma abban keresendő, hogy ha csak a helyparamétert tekintjük ismeretlennek, akkor az I-divergencia minimalizálásnál vannak olyan esetek, amikor nem csak egy megoldás van [1], [9].

3. A leggyakoribb érték és dihézió módosítása

Minimalizáljuk az I-divergenciát ha f a minta sűrűségfüggvénye és

$$g(x) = \frac{s}{\pi(s^2 + (x-c)^2)},$$

azaz a Cauchy-eloszlás sűrűségfüggvénye ismeretlen hely- és skálaparaméterrel.
A deriválások alapján kapjuk, hogy a javasolt egyenletrendszert a

$$\psi(x) = \frac{x}{1+x^2}, \quad \chi(x) = \frac{x^2-1}{1+x^2}$$

függvények definiálják. Nagy különbség, hogy a dihéziót meghatározó függvény változik.
Fontos, hogy az új problémának pontosan egy megoldása van. Ezenkívül az elméleti megoldás mellett létezik numerikus algoritmus is. Jelölje c_n és s_n a kapott becsléseket n elemű mintára [3]. Maximum likelihood becslés [8].
A kapott becslések néhány fontosabb tulajdonsága:
B-, V- és kvalitatív robusztus, amelyre a katasztrófapontok

$$\varepsilon^*(c_n) = 0.5.$$

$$\varepsilon^*(s_n) = \frac{-\chi(0)}{\chi(-\infty) - \chi(0)} = \frac{1}{3}.$$

Ezenkívül (c_n, s_n) együttes eloszlása aszimptotikusan normális, azaz

$$\sqrt{n}((c_n, s_n) - (\mu, \sigma)) \to^d N(0, \Sigma),$$

ahol μ, σ az eredeti paraméterek és a kovariancia mátrix $\Sigma = C^{-1}S[C^{-1}]^T$.

$$C = \begin{pmatrix} E\left(\frac{\partial}{\partial\mu}\psi\left(\frac{\xi-\mu}{\sigma}\right)\right) & E\left(\frac{\partial}{\partial\sigma}\psi\left(\frac{\xi-\mu}{\sigma}\right)\right) \\ E\left(\frac{\partial}{\partial\mu}\chi\left(\frac{\xi-\mu}{\sigma}\right)\right) & E\left(\frac{\partial}{\partial\sigma}\chi\left(\frac{\xi-\mu}{\sigma}\right)\right) \end{pmatrix},$$

$$S = \begin{pmatrix} E(\psi^2(\eta)) & E(\psi(\eta)\chi(\eta)) \\ E(\psi(\eta)\chi(\eta)) & E(\chi^2(\eta)) \end{pmatrix} = \begin{pmatrix} \frac{1}{8} & 0 \\ 0 & \frac{1}{2} \end{pmatrix},$$

ahol az η valószínűségi változó F_0 eloszlásfüggvényű [3], [4].

4. Összefoglalás

A cikkben a matematikai statisztika egy alapvető feladatával foglalkoztunk a hely- és skálaparaméter probléma megoldásával robusztus oldalról. Leírjuk a leggyakoribb értéket és dihéziót. A módszer (a becslés) a Miskolci Egyetemen született. Csernyák László és Steiner Ferenc professzorok munkája. Bekerült szócikk formájában (készítette: Adrienne W. Kemp) a nagy statisztikai enciklopédiába [7]. Miután Nagy Ferenc a Miskolci Egyetem Matematikai Intézetének oktatója megoldotta a két paraméteres Cauchy paraméterbecslési problémát, így a cikkben javasolt módszer végleges megoldás a leggyakoribb érték és dihézió módosítására.

5. Köszönetnyilvánítás

A cikkben ismertetett kutatómunka a GINOP-2.2.1-15-2017-00090 - "E-mobility Miskolcról: Hűtővíz keringető szivattyú és motorhűtő ventilátor továbbfejlesztése az elektromos járművekben elvárt magasabb minőségi követelmények figyelembevételével" projekt keretében valósul meg.

Irodalom

[1] Clarke, B.R.: *Uniqueness and Frechet Differentiability of Functional Solutions to Maximum Likelihood Type Equations*, Ann. Statist., 11 (1983), pp. 1196-1205. **https://doi.org/10.1214/aos/1176346332**

[2] Csernyák, L.: *On the most frequent value and cohesion of probability functions*, Acta Geodaet., Geophys. et Mont. Acad. Sci. Hung., 8 (1973) pp. 397-401.

[3] Fegyverneki, S.: *A special joint estimation of location and scale with applications*, Publ. Univ. of Miskolc, Series D. Natural Sciences, Mathematics, 39 (1999) pp. 21-27.

[4] Fegyverneki, S.: *Robust estimators and probability integral transformation*, Math. Comput. Modelling, 38 (2004) pp. 803-814. **https://doi.org/10.1016/S0895-7177(03)90065-3**

[5] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A.: *Robust statistics: the approach based on influence functions*, Wiley, New York, 1986.

[6] Huber, P.J.: *Robust statistics*, Wiley, New York, 1981. **https://doi.org/10.1002/0471725250**

[7] Kotz, S, Read, C.B., Balakrishnan, N., Vidakovic, B., Johnson, N.L.(Editor-in-Chiefs): *Encyclopedia of Statistical Sciences* Wiley, New York, 16 Volume Set, 2nd Edition, (2005), pp. 1-9686., ISBN: 978-0-471-15044-2, Volume 12, STEINER'S MOST FREQUENT VALUE, pp. 8161-8162.

[8] Nagy, F.: *Parameter Estimation of the Cauchy Distribution in Information Theory Approach*, Journal of Universal Computer Science, 12 (2006) pp. 1332-1344.

[9] Reeds, J.A.: *Asymptotic Number of Roots of Cauchy Location Likelihood Equations*, Ann. Statist., 13 (1985), pp. 775-784. **https://doi.org/10.1214/aos/1176349554**

[10] Steiner, F.: *Most frequent value procedures (A short monograph)*, Geophysical Transactions, 34 (1988) pp. 139-260.

[11] Steiner, F.: *A geostatisztika alapjai*, Tankönyvkiadó, 1990., Budapest.