

A VÁRHATÓ ELTÉRÉSRE VONATKOZÓ SHEPPARD-KORREKCIÓ

SARKADI KÁROLY

Bevezetés

Az ebben a dolgozatban tárgyalt problémával a *Textilipari Minőség-ellenőrző Intézet* felkérésére foglalkoztunk. A kérdés gyakorlati jelentőségét az adja meg, hogy a textiliparban ez idő szerint még mindig a várható eltérés használata az általános a szórás helyett. A közönséges momentumokra vonatkozó Sheppard-korrekciónak elég bő irodalma van, a várható eltérés Sheppard-korrekcijára vonatkozóan azonban semmilyen irodalmi adatot nem találtunk.

A következőkben tehát azzal fogunk foglalkozni, hogy milyen eltérést okoz az, ha a várható eltérés kiszámításához osztálybasorolást alkalmazunk, és megadható-e — a közönséges momentumok Sheppard-korrekcijához hasonlóan — olyan képlet, amely ezt az eltérést jó közelítéssel megadja.

1. §.

A közönséges momentumok esetén a Sheppard-korrekciónak bizonyos eloszlások esetében, jó közelítéssel megadja az eredeti eloszlás és a csoportosított eloszlás momentumai közötti különbséget (lásd például: [1]). Ebben az értelemben a Sheppard-korrekciónak a várható eltérésre nem terjeszthető ki.

Az a bizonyítás, amely az Euler—MacLaurin-formulán alapszik, ebben az esetben ugyanis nem alkalmazható, mert az

$$\left| \xi \right| \int_{\xi - \frac{1}{2}h}^{\xi + \frac{1}{2}h} f(x) dx$$

függvény ξ szerinti deriváltja nem folytonos. (Itt $f(x)$ az eloszlás sűrűségfüggvénye.) A 3. §-ban pedig látni fogjuk, hogy még a normális eloszlás esetében is — amelynél pedig a sűrűségfüggvény valamennyi deriváltja folytonos — az eredeti eloszlás és a csoportosított eloszlás várható eltérése közötti különbség minden esetben felvehet pozitív előjelet is az osztályok elhelyezkedésétől függően.

2. §.

Van a Sheppard-korrekciónak még egy más értelmezése is ([2], p. 74). Jelöljük az osztályintervallum hosszát h -val, az abszolút értékben legkisebb intervallum közepét η -val. η általában $-\frac{h}{2}$ és $\frac{h}{2}$ között minden értéket

felvehet. Ha feltesszük, hogy η a $\left(-\frac{h}{2}, \frac{h}{2}\right)$ intervallumban egyenletes elosz-

lású valószínűségi változó; akkor képezhetjük folytonos alapeloszlás esetén a csoportosított eloszlás és az eredeti eloszlásból számított momentumok különbségének várható értékét. Az erre kapott eredmény számszerűen megegyezik az előző pontban ismertetett Sheppard-korrekciónal, azonban minden folytonos eloszlásra érvényes. Ebben az értelemben a Sheppard-korrekciónál beszélhetünk az abszolút eltérés esetében is.

A csoportosított eloszlás várható eltérése

$$\sum_{j=-\infty}^{\infty} |\eta + jh| \int_{\eta + (j - \frac{1}{2})h}^{\eta + (j + \frac{1}{2})h} f(x) dx.$$

Ennek η szerinti várható értéke:

$$\begin{aligned} & \frac{1}{h} \int_{-\frac{h}{2}}^{\frac{h}{2}} \sum_{j=-\infty}^{\infty} |y + jh| \int_{y + (j - \frac{1}{2})h}^{y + (j + \frac{1}{2})h} f(x) dx dy = \\ (1) \quad & = \frac{1}{h} \sum_{j=-\infty}^{\infty} \int_{(j - \frac{1}{2})h}^{(j + \frac{1}{2})h} \int_{t - \frac{h}{2}}^{t + \frac{h}{2}} |t| f(x) dx dt = \\ & = \frac{1}{h} \int_{-\infty}^{\infty} \int_{t - \frac{h}{2}}^{t + \frac{h}{2}} |t| f(x) dx dt = \frac{1}{h} \int_{-\infty}^{\infty} f(x) \cdot \int_{x - \frac{h}{2}}^{x + \frac{h}{2}} |t| dt dx. \end{aligned}$$

Azonban

$$\int_{x-\frac{h}{2}}^{x+\frac{h}{2}} |t| dt = \begin{cases} \operatorname{sgn} x \cdot \int_{x-\frac{h}{2}}^{x+\frac{h}{2}} t dt = h|x|, & \text{ha } |x| > \frac{h}{2}, \\ \int_0^{x-\frac{h}{2}} t dt + \int_0^{x+\frac{h}{2}} t dt = x^2 + \frac{h^2}{4}, & \text{ha } |x| < \frac{h}{2}; \end{cases}$$

ennélfogva a csoportosított eloszlás várható eltéréseinek várható értéke:

$$\int_{-\infty}^{\infty} x^2 f(x) dx + \int_{-\frac{h}{2}}^{\frac{h}{2}} \left(\frac{x^2}{h} + \frac{h}{4} - |x| \right) f(x) dx$$

lesz. E kifejezés első tagja a korrekt várható eltérés, vagyis a második tag a Sheppard-korrektúra:

$$(2) \quad \int_{-\frac{h}{2}}^{\frac{h}{2}} \left(\frac{x^2}{h} + \frac{h}{4} - |x| \right) f(x) dx.$$

Speciálisan a standardizált normális eloszlás esetében a Sheppard-korrektúra

$$(3) \quad \left(\frac{h}{4} + \frac{1}{h} \right) \left[2 \Phi \left(\frac{h}{2} \right) - 1 \right] + \varphi \left(\frac{h}{2} \right) - \sqrt{\frac{2}{\pi}},$$

ahol $\Phi(x)$ és $\varphi(x)$ a standardizált normális eloszlás eloszlásfüggvénye, illetve sűrűségfüggvénye.

3. §.

Gyakorlati szempontból természetesen az még nem eléggé megnyugtató, ha a korrekt és a csoportosítás útján számított várható eltérés különbségének várható értékét ismerjük. Szükséges azt is tudnunk, hogy η konkrét értéke esetében mekkora lesz ez a hiba. Ha ezt η minden értékére ismerjük,

akkor azt is fogjuk tudni, hogy milyen értékek között mozog, ha η a $-\frac{h}{2}$ és $\frac{h}{2}$

értékek között változik. A közönséges momentumokra vonatkozóan ezt a problémát *R. A. Fisher* oldotta meg ([3], lásd: [2], p. 75). Teljesen

hasonlóan járhatunk el a várható eltérés esetében is. Mivel a csoportosított eloszlás várható eltérése

$$\sum_{-\infty}^{\infty} \eta \int_{\eta - \frac{h}{2}}^{\eta + \frac{h}{2}} f(x) dx$$

η -ban h szerint periodikus, a fenti kifejezést egyenlővé tehetjük az

$$A_0 + A_1 \sin \theta + A_2 \sin 2\theta + \dots \\ + B_1 \cos \theta + B_2 \cos 2\theta + \dots$$

kifejezéssel, ahol

$$(4) \quad \theta = \frac{2\pi\eta}{h},$$

és az $A_0, A_1, B_1, A_2, B_2, \dots$ együtthatók integrálással számíthatók ki:

$$\left. \begin{aligned} A_s &= \frac{1}{\pi} \sum_{j=-\infty}^{\infty} \int_{-\pi}^{\pi} \sin s\theta d\theta \int_{\eta - \frac{h}{2}}^{\eta + \frac{h}{2}} \eta |f(x)| dx \\ B_s &= \frac{1}{\pi} \sum_{j=-\infty}^{\infty} \int_{-\pi}^{\pi} \cos s\theta d\theta \int_{\eta - \frac{h}{2}}^{\eta + \frac{h}{2}} \eta |f(x)| dx \end{aligned} \right\} s = 1, 2, \dots$$

$$A_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} d\theta \int_{\eta - \frac{h}{2}}^{\eta + \frac{h}{2}} \eta |f(x)| dx.$$

Ha A_0 kifejezésében a θ helyébe beírjuk a (4) alatti értéket, az (1)-re jutunk és így az állandó tag a Sheppard-korrekciónhoz vezet.

A periodikus tagok kifejezésében is, ha az integrálás sorrendjét felcseréljük, az egyik integrálás elvégezhető. A kapott kifejezések azonban így is bonyolultak és áttekinthetetlenek, ezért felírásukat mellőzzük.

4. §.

Az alábbiakban egy egyszerűbb közelítő módszert adunk meg az osztálybasorolás által okozott hiba kiszámítására. Írjuk a csoportosított eloszlás és az eredeti eloszlás várható eltéréseinek a különbségét a következő formába:

$$S = \sum_{j=-\infty}^{\infty} \int_{-\frac{h}{2}}^{\frac{h}{2}} (|\eta + jh| - |x + \eta + jh|) f(x + \eta + jh) dx.$$

Ha h elég kicsi, akkor nem követünk el nagy hibát, ha $f(x + \eta + jh)$ helyett az

$$f\left(\eta + \left[j - \frac{1}{2}\right]h\right) + \left(x + \frac{h}{2}\right) \frac{f\left(\eta + \left[j + \frac{1}{2}\right]h\right) - f\left(\eta + \left[j - \frac{1}{2}\right]h\right)}{h}$$

kifejezést írjuk, vagyis az $f(x)$ függvényt az $\eta + \left(j - \frac{1}{2}\right)h$ és $\eta + \left(j + \frac{1}{2}\right)h$ pontok között húrjával helyettesítjük. Az integrálást így elvégezhetjük a $j = 0$ eset kivételével, tehát

$$(5) \quad \left\{ \begin{aligned} S &\approx - \sum_{\substack{j=-\infty \\ j \neq 0}}^{\infty} \operatorname{sgn} j \frac{h^2}{12} \left\{ f\left(\eta + \left[j + \frac{1}{2}\right]h\right) - f\left(\eta + \left[j - \frac{1}{2}\right]h\right) \right\} + \\ &+ \int_{-\frac{h}{2}}^{\frac{h}{2}} (|\eta| - |x + \eta|) f(x + \eta) dx = \\ &= \frac{h^2}{12} \left[f\left(\eta + \frac{h}{2}\right) + f\left(\eta - \frac{h}{2}\right) \right] + \int_{-\frac{h}{2}}^{\frac{h}{2}} (|\eta| - |x + \eta|) f(x + \eta) dx. \end{aligned} \right.$$

Speciálisan a normális eloszlásra közelítőleg

$$S \approx - \frac{h^2}{12 \sqrt{2\pi}} + \frac{|\eta| h}{\sqrt{2\pi}}.$$

Innen rögtön látható, hogy S közelítő értéke a $\left(-\frac{h}{2}, \frac{h}{2}\right)$ intervallumban a minimumát az $\eta = 0$ helyen veszi fel és az

$$(6) \quad S_{\min} \approx -\frac{h^2}{12\sqrt{2\pi}} \approx -0,033 h^2,$$

maximumát pedig az $\eta = \frac{h}{2}$ helyen és ez

$$(7) \quad S_{\max} \approx \frac{h^2}{6\sqrt{2\pi}} \approx 0,066 h^2.$$

A fenti érték egyben nyilván S abszolút értékének maximuma is. Így tehát ennek alapján megállapíthatjuk, hogy milyen kicsire kell h -t választani ahhoz, hogy S az empirikus abszolút középeltérés szórásához képest elhanyagolható kicsiny legyen.

Az (5) képlet hibájának megbecslése céljából tegyük fel, hogy $f''(x)$ folytonos és korlátos, írjuk fel az interpolációs maradéktagot: ez a szummáció j -edik tagjára ($j \neq 0$) vonatkozóan

$$\int_{-\frac{h}{2}}^{\frac{h}{2}} \left(\eta + jh - x + \eta + jh \right) \left(x - \frac{h}{2} \right) \left(x + \frac{h}{2} \right) f''(\xi_j) dx = \operatorname{sgn} j \frac{h^4}{32} y_j'',$$

lesz, ahol y_j'' egy olyan értéket jelöl, amelyet az $f''(x)$ függvény a j -edik intervallum valamely pontjában felvesz. Így az egész összeg maradéktagja

$$\frac{h^4}{32} \sum_{j=1}^{\infty} (y_j'' - y_{-j}'').$$

A fenti összeg megbecslése ismert konkrét eloszlás esetében általában nem jelent nehézséget.

TRODALOM

[1] H. CRAMÉR: *Mathematical methods of statistics*. Princeton University Press, Princeton, 1946. (p. 359—362).

[2] M. S. KENDALL: *The advanced theory of statistics*. Griffin et Comp., London, 1948. (Vol., I., p. 68—85).

[3] R. A. FISHER: „On the mathematical foundations of theoretical statistics” *Philosophical Transactions of the Royal Society of London, Series A*, 22 (1922) 309—368.

ПОПРАВКА ШЕППАРДА ДЛЯ ОЖИДАЕМОГО ОТКЛОНЕНИЯ

К. Шаркади

Резюме

В настоящей работе исследуется вопрос о влиянии группировки данных на ожидаемое отклонение. Если в случае дисперсии, и вообще обыкновенных моментов, поправка Шеппарда для большинства классических распределений дает хорошее приближение с уплотнением интервалов, то для разницы группированного и исходного распределений нельзя в общем задать такую поправку, даже в случае этих классических распределений: изменение поправки при изменении интервалов имеет тот же порядок величины, как и сама поправка. Однако, можно вычислять математическое ожидание этой разницы (2). Можно далее вычислять значение этой разницы как функции расположения интервалов по методу Фишера [3]. Полученный результат для вычислений непригоден, однако интерполяционным приближением можно получить наглядный результат [(4), η — средняя интервала содержащего начало координат]. Для стандартизированного нормального распределения математическое ожидание разницы дается формулой (3), а его минимальное и максимальное значения формулами (6) и (7).

ON SHEPPARD'S CORRECTION CONCERNING THE MEAN DEVIATION

K. SARKADI

Summary

This paper deals with the influence on the mean deviation, exerted by the grouping of data. While in the case of standard deviation and, in general, of the moments calculated from grouped data, Sheppard's correction guarantees a good approximation for the majority of classical distributions if the width of the class intervals is sufficiently small, a correction with similar properties for the mean deviation cannot be found, because the variation, effected by shifting the class intervals of the error committed when replacing the mean deviation of the original distribution by that of the grouped distribution, is of the order of magnitude of the correction itself. One can calculate, however, the expectation (see (2)) of this error. In addition the value of this error as a function of the position of the class intervals can be calculated according to Fisher's method [3]; the obtained result is, however, unsuitable for numerical calculations. Nevertheless an interpolatory approximation gives a simple result [(4); η is the mid-point of the interval which contains the origin]. In the case of a standardized normal distribution the expectation of the error is given by (3); the formulae (6) and (7) give the minimum and maximum values of the error, respectively.