

OSZTÁLYKÖZÖK MEGVÁLASZTÁSA MINTÁK ELEMEINEK OSZTÁLYBASOROLÁSÁNÁL

SARKADI KÁROLY

Összefoglalás

A szerző adott $f(x)$ sűrűségfüggvénnyel bíró eloszlásra nézve kimutatja, hogy a (3) egyenletrendszer teljesülése szükséges ahhoz, hogy az osztálybasorolással és a pontos módon kiszámított középérték közötti eltérés négyzetének várható értéke minimális legyen. Ha az egyenletrendszernek csak egy megoldása van, a feltétel elegendő is. Az ellenkező esetre ellenpéldát ad. Kimutatja továbbá, hogy ha $\log f(x)$ konkáv, akkor a megoldás egyértelmű. A normális eloszlás esetére a táblázatot is megadja különböző számú intervallumokra.

A problémának csillagászati vonatkozása is van, mert a Smidt-elmélet szerint a bolygóknak a Naptól való távolságát meghatározó egyenletrendszer pontosan meg-
egyezik a (3) egyenletrendszerrel.

1. Ha nagy minták középértékét vagy szórását ki akarjuk számítani, a számítások pontos elvégzése az adatok nagy száma miatt igen nagy munkát jelent. Ezért ilyenkor az egymáshoz közeleső értékeket össze szoktuk foglalni úgy, hogy a számegyenest, vagy azt a tartományát, amely a valószínűségi változó lehetséges értékeit magában foglalja, több intervallumnak a közép-pontjával (vagy más meghatározott belső pontjával) helyettesítjük, amelybe az adat esik.

Természetesen, önként felmerül a kérdés, hogyan legcélszerűbb az intervallumok határait, valamint az egyes meghatározott pontokat kiválasztani?

Ebben a dolgozatban ennek a kérdésnek egy speciális változatára fogunk választ adni. A célszerűség alatt azt fogjuk érteni, hogy a mintabeli adat eltéréseinek négyzete az öt helyettesítő értéktől minimális legyen. Feltesszük, hogy az elméleti eloszlásfüggvényt ismerjük; feltesszük azt is, hogy az intervallumok határai, és a »helyettesítő« pontok megválasztásánál egyéb szempontok nem játszanak közre.

2. Legyen ξ folytonos valószínűségi változó, sűrűségfüggvénye $f(x)$, legyen továbbá

$$(1) \quad -\infty = c_0 < a_1 < c_1 < a_2 < c_2 < \dots < a_{k-1} < c_{k-1} < a_k < c_k = \infty$$

és foglalkozzunk azzal a kérdéssel, hogy adott k esetén a minta középértékének kiszámítása szempontjából az (1) sorozat tagjainak milyen megválasztásánál követünk el legkisebb hibát avval, hogy ξ helyett mindig azzal az

a_i -vel számolunk, amelyekkel közös (c_{i-1}, c_i) intervallumban van. Keressük az így számított $(\xi - a_i)^2$ várható értékének minimumát; ez ugyanakkor lesz minimum, amikor a közelítéssel és a pontosan kiszámított közép eltérése négyzetének várható értéke. Az utóbbi az előbbi várható értéknek n -edrésze, ha n a megfigyelések száma, tekintve, hogy $\xi - a_i$ megfigyelésből származó értékei függetlenek egymástól.

Tehát vizsgáljuk, hogy

$$(2) \quad \sum_{i=1}^k \int_{c_{i-1}}^{c_i} (x - a_i)^2 f(x) dx$$

az (1) sorozat tagjainak milyen értékei mellett lesz minimális.

Analitikus módszerrel megállapíthatjuk, hogy a (2) alatti kifejezésnek az értelmezési tartomány olyan belső pontjában lehet csak szélső értéke, ahol

$$(3) \quad a_i = \frac{\int_{c_{i-1}}^{c_i} xf(x) dx}{\int_{c_{i-1}}^{c_i} f(x) dx} \quad \text{és} \quad c_i = \frac{a_i + a_{i+1}}{2} \\ (i = 1, 2, \dots, k-1). \\ (i = 1, 2, \dots, k)$$

Hogy lokális minimum általában csak ilyen pontban lehet, az elemi úton is belátható.

Ha $f(x)$ -nek nincs azonosan eltűnő szakasza, akkor, ha a (3) egyenletrendszer nem teljesül, a (2) valamelyik paraméter megváltoztatásával mindig csökkenthető. Éspedig ha a baloldali egyenletek közül pl. az i -edik nem teljesül, változtassuk meg a_i értékét úgy, hogy teljesüljön (ez nyilván mindig lehetséges és csak egyféleképpen), ezáltal a (2) i -edik tagja csökken, a többi változatlan marad. Ha viszont a (2) jobboldali egyenletei közül valamelyik, pl. az i -edik nem teljesül, akkor c_i értékét változtassuk meg úgy, hogy teljesüljön s ekkor a (2)-ben az integrálnak c_i régi és új értéke közé eső részén az integrandus csökken.

Tehát ha $f(x)$ nem tűnik el valamilyen intervallumban, akkor a (3) szükséges feltétele annak, hogy a (2) minimális értéket vegyen fel.

Ha $f(x)$ eltűnik valamilyen szakaszban, akkor a fenti következtetés nyilván úgy módosul, hogy ha van minimum, akkor azt a (2) egy olyan pontban is felveszi, ahol a (3) teljesül; de bármelyik c_i szabadon változtatható olyan intervallumon belül, ahol $f(x)$ azonosan 0, a nélkül, hogy ezáltal a (2) alatti érték megváltoznék.

Mivel a (2) összes paramétereinek folytonos függvénye, van minimuma vagy az értelmezési tartomány belsejében, vagy annak határán. Az értelmezési tartományt az (1) szabja meg; határpontjaira (1) azzal a módosítással áll fenn, hogy egy vagy több $<$ jel helyébe egyenlőségi jel lép. Az értelmezési tartomány határpontjaiban (3) nem állhat fenn; könnyen utána számolhatnánk, hogy ez a feltevés arra vezetne, hogy az a_i és c_i számok mind egyenlők,

ami lehetetlen. Tehát az előbbieket szerint minimum sem lehet határpontban; ezek szerint (2)-nek tehát az (1) által meghatározott értelmezési tartományban van minimuma és azt olyan helyen veszi fel, ahol (3) teljesül.

Kimutatható még, hogy az ilyen helyen nemcsak a (2) alatti kifejezés értéke lesz minimális, hanem a pontosan számított és az osztálybasorolással számított minta középérték eltérése négyzetének várható értéke. Jelöljük a mintaadatokat $\xi_1, \xi_2, \dots, \xi_n$ -nel és legyen i_j annak az intervallumnak a sorszáma, amelybe a ξ_j esik ($j = 1, 2, \dots, n$), vezessük be továbbá az $\eta_j = \xi_j - a_{i_j}$ jelölést. Az η_j valószínűségi változók nyilván egyforma eloszlásúak és egymástól függetlenek, tehát

$$\mathbf{M} \left(\frac{\sum_{j=1}^n \xi_j}{n} - \frac{\sum_{j=1}^n a_{i_j}}{n} \right)^2 = \frac{1}{n^2} \mathbf{M} \left(\sum_{j=1}^n \eta_j \right)^2 = \frac{1}{n} \mathbf{M} (\eta^2) + \frac{n-1}{n} \mathbf{M}^2 (\eta).$$

Az első tag a (2) alatti kifejezés $\frac{1}{n}$ -szerese. Ez minimális értékét olyan helyen veszi fel, ahol a (3) teljesül. Ugyanakkor azonban a második tag is felveszi minimális értékét. Ugyanis a (3) baloldali egyenleteinek fennállása esetén $\mathbf{M} (\eta) = 0$, tekintve, hogy η -nak bármelyik intervallumhoz tartozó feltételes várható értéke 0. Tehát a fenti összeg második tagja ilyenkor 0, negatív pedig sohasem lehet. Így tehát az összeg is ugyanott veszi fel minimumát, ahol a (2).

3. A (3) alatti feltételek az (1) alatti értékeket nem mindig határozzák meg egyértelműen.

Erre lássuk az alábbi ellenpéldát:

Legyen $k = 2$ és

$$f(x) = \begin{cases} \frac{1}{6}, & \text{ha } |x| \leq 1 \\ \frac{1}{18}, & \text{ha } 1 < |x| \leq 7 \\ 0, & \text{ha } |x| > 7 \end{cases}$$

A $c_1 = -1, 0, 1$ értékek mindegyike meghatározza (3) egy-egy megoldását. A három megoldás közül a $c_1 = \pm 1$ értékek valóban minimummá teszik (2)-t $\left(\frac{61}{18}\right)$, a $c_1 = 0$ -hoz tartozó megoldás nem $\left(\frac{100}{27}\right)$.

4. A következőkben egy elégséges feltételt adunk arra, hogy (3)-ból az a_i -k és c_i -k egyértelműen meghatározhatók legyenek, és pedig bebizonyítjuk, hogy az előállítás egyértelmű, ha $f(x)$ a következő tulajdonsággal rendelkezik:

(4) $f(x)$ vagy mindenütt pozitív, vagy egy véges vagy végtelen intervallumban pozitív, másutt 0; és ahol $f(x)$ pozitív, ott $\log f(x)$ konkáv.

Megjegyzendő, hogy a (4) feltételnek a normális eloszlás és több fontos eloszlás eleget tesz.

A bizonyításhoz a következő *segédtelepre* lesz szükségünk :
Ha az $f(x)$ a (4) feltételnek eleget tevő függvény és $\gamma_1 \leq \gamma_2$ továbbá

$$(5) \quad \alpha_i = \frac{\int_{\gamma_i}^{\delta_i} x f(x) dx}{\int_{\gamma_i}^{\delta_i} f(x) dx} \quad (i = 1, 2)$$

$$\alpha_2 - \alpha_1 \geq \gamma_2 - \gamma_1$$

(itt $\gamma_2 - \gamma_1 = 0$ értendő, ha $\gamma_2 = \gamma_1 = -\infty$) — akkor, ha δ_1 véges,

$$(6) \quad \delta_2 - \delta_1 \geq \alpha_2 - \alpha_1.$$

Az egyenlőségi jel (6)-ban csak abban az esetben állhat fenn, ha az (5)-ben is fennáll, vagy ha $\gamma_1 = \gamma_2 = -\infty$, és ezekben az esetekben is csak akkor, ha $\log f(x)$ lineáris a (γ_1, δ_2) intervallumban. Ugyanez a feltétele annak is, hogy δ_1 végtelen lehessen.

Bizonyítás: (4)-re való tekintettel $h > 0$ esetén

$$\log f(\alpha_1 - h) + \log f(\alpha_2) \leq \log f(\alpha_1) + \log f(\alpha_2 - h)$$

(7) és

$$\log f(\alpha_1 + h) + \log f(\alpha_2) \geq \log f(\alpha_1) + \log f(\alpha_2 + h).$$

Mindkét egyenlőtlenség abból következik, hogy a konkáv ív minden pontja a húr felett van.

Ha a (7) alatti egyenlőtlenségek közül az elsőben a jobboldalon a $\log f(x)$ függvény helyébe $(\alpha_1 - h, \alpha_2)$ intervallum feletti húrjának megfelelő ordinátákat írjuk, a jobboldal csökkentésével egyenlőségre jutunk s ez az egyenlőséget igazolja. Ugyanígy módon igazolható az alsó egyenlőtlenség is. A mondottakból az is következik, hogy az egyenlőség a (7) egyenlőtlenségeiben csak akkor állhat fenn, ha $\log f(x)$ lineáris $\alpha_1 - h$ és α_2 , illetve α_1 és $\alpha_2 + h$ között.

Ha a (7)-ben h helyébe $(\alpha_1 - x)$ -et, majd $(x - \alpha_1)$ -et írunk és a

$$g(x) = \frac{f(\alpha_1)}{f(\alpha_2)} f(x + \alpha_2 - \alpha_1)$$

jelölést vezetjük be, akkor a (7)-ből következik, hogy

$$g(x) \geq f(x), \text{ ha } x < \alpha_1$$

és

$$g(x) \leq f(x), \text{ ha } x > \alpha_1.$$

Itt egyenlőség csak abban az esetben állhat fenn, ha $f(x) = e^{mx+b}$ alakú x és a_2 , illetve a_1 és $x + a_2 - a_1$ között. Nyilván

$$\frac{\int_{\gamma'_2}^{\delta'_2} xg(x) dx}{\int_{\gamma'_2}^{\delta'_2} g(x) dx} = \alpha_1,$$

ahol

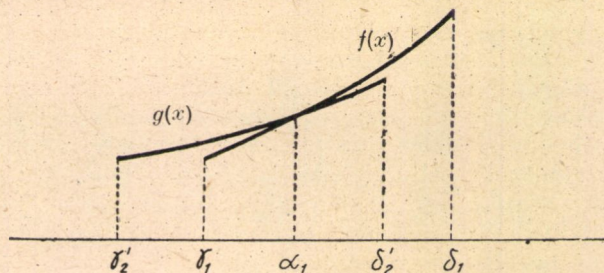
$$\gamma'_2 = \gamma_2 + \alpha_1 - a_2$$

és

$$\delta'_2 = \delta_2 + \alpha_1 - a_2$$

és így az (5) szerint $\gamma'_2 \leq \gamma_1$.

Az $j(x)$ görbe alatti (γ_1, δ_1) intervallumba eső terület súlypontjának abszcisszája megegyezik a $g(x)$ görbe alatti γ'_2, δ'_2 terület súlypontjával. Ha az előbb említett területet a súlyponttól balra eső pozitív területrésszel kiegészítjük,



1. ábra

szítjuk, jobboldalt pedig pozitív területrészt elveszünk belőle, a súlypont balra tolódik el, tehát nem maradhat ugyanaz. Éppen ezért a fentiek figyelembevételével lehetetlen, hogy $\delta'_2 < \delta_1$ legyen (l. az ábrát), sőt $\delta_1 = \delta'_2$ is csak abban az esetben lehetséges, ha a két terület teljesen megegyezik egymással, vagyis $\gamma'_2 = \gamma_1$ és $f(x) \equiv g(x)$ ($\gamma_1 < x < \delta_1$). Mindebből pedig következik bebizonyítandó segédteételünk.

Most áttérünk eredeti állításunk bebizonyítására. Tegyük fel, hogy a (3)-nak két különböző gyökrendszere van, a_{i1}, c_{i1} , illetve a_{i2}, c_{i2} . Ha $a_{11} = a_{12}$ volna, akkor a többi ismeretlenek a (3)-ból rendre egyértelműen meghatározhatók volnának, így a két gyökrendszer nem volna különböző. Tehát feltehetjük, hogy $a_{11} < a_{12}$. Ekkor segédteételünk értelmében $c_{12} - c_{11} \geq a_{12} - a_{11}$, vagyis a (3)-at alkalmazva $a_{22} - a_{21} \geq c_{12} - c_{11}$. Erre ismét alkalmazhatjuk a segédteételt s. í. t. Az ismételt alkalmazások során a k -adik lépésben segédteételünk értelmében az eredeti feltételezés szerinti $c_{k1} = c_{k2} = \infty$ értékekre csak akkor juthatunk, ha $\log f(x)$ lineáris $-\infty$ és ∞ között. Ekkor

azonban az $f(x)$ alatti terület véges nem lehet, mert $\int_{-\infty}^{\infty} e^{mx+b} dx$ semilyen

m, b értékpár esetén sem korlátos. Tehát ilyen sűrűségfüggvény nincsen. Ezzel ellentmondásra jutottunk, s így állításunk igaz.

5. Normális eloszlás esetére kiszámítottuk az intervallumok határait és súlypontjait, a (3) egyenletrendszer megoldása útján, $k = 2$ -től $k = 15$ -ig. A számítás közelítő módszerrel történt. A normális eloszlás esetén egyszerűbb a probléma, mint az általános esetben, egyrészt mert az eloszlás szimmetrikus,

másrészt mert normális eloszlás esetén $\int_{-\infty}^x xf(x) dx = -f(x)$ és ennek a

függvénynek, valamint az eloszlásfüggvénynek az értékei táblázatokba vannak foglalva. További egyszerűsítést a módszerben úgy értünk el, hogy milliméterpapiroson ábrázoltuk az $f[F^{-1}(x)]$ függvényt, ahol $F(x)$ az eloszlásfüggvény, ennek segítségével a (3) baloldali egyenleteinek a megoldása sokkal gyorsabban történhetik, mert az ábrázolt függvénynek a c_i, c_{i+1} ívhez tartozó differenciahányadosa a_i -vel egyenlő. A differenciahányados mérése állítható vonalzóval történt.

A számításokat intézetünkben Tekse Kálmán alkalmazott matematika szakos hallgató termelési gyakorlati munkájaként végezte, a számítás egyszerűsítéséhez néhány ötlettel hozzájárult. A kiszámított értékeket a függelékként közölt táblázat tartalmazza. Ezeket alkalmazhatjuk abban az esetben, ha normális eloszlásból veszünk ki nagy mintát. Ha ennek várható értékére és szórására a számítások elvégzése előtt még tájékoztató adatunk sincsen, akkor a két szélső értékből nyerhetünk ezekre előzetes becslést.

Az esetek nagy részében azonban a gyakorlati alkalmazást bizonyos körülmények egyelőre még megnehezítik. Így pl. nehézséget jelent az, hogy feltételeztük az elméleti eloszlás ismeretét, továbbá, hogy az így kapott értékek nem kerek számok. A további feladat lesz tehát a módszernek a gyakorlati alkalmazás megkönnyítése irányában való továbbfejlesztése.

6. Mint érdekes ténytet megemlíjtük, hogy az itt tárgyalt probléma egy csillagászati problémával is kapcsolatban áll.

O. J. Smidt szovjet akadémikus ismert kozmogóniai elméletében [2] többek között foglalkozik a bolygótávolságok törvényével is. Ismeretes, hogy elmélete szerint a bolygók a Nap által befogott meteoritrajból keletkeztek. Felteszi, hogy a raj befogás utáni q fajlagos impulzusnyomatékának sűrűségfüggvénye valamilyen $f(q)$ függvény volt. Megállapítja, hogy ebben az esetben az i -edik bolygó fajlagos impulzusnyomatékát q_i -vel jelölve:

$$(8a) \quad q_i = \frac{\int_{c_{i-1}}^{c_i} qf(q) dq}{\int_{c_{i-1}}^{c_i} f(q) dq},$$

ahol

$$(8b) \quad c_i = \frac{q_i + q_{i+1}}{2}.$$

q_i -ből az i -edik bolygónak a Naptól való távolsága kiszámítható. ($R_i = kq_i^2$, ahol k adott állandó.) Smidt azt írja, hogy minden egyes eloszlási törvénynek, azaz minden $f(x)$ függvénynek megfelel a bolygótávolságoknak egy törvénye.

Az intervallumok határai és súlypontjai 0 várható értékű és egységnyi szórású normális eloszlás esetére.

k	Az intervallumok határa (<i>dőlt szedéssel</i>) és súlypontjai																			
1	0																			
2	—0,71 0 0,71																			
3	—1,22 —0,61 0 0,61 1,22																			
4	—1,52 —0,99 —0,45 0 0,45 0,99 1,52																			
5	—1,72 —1,24 —0,76 —0,38 0 0,38 0,76 1,24 1,72																			
6	—1,90 —1,46 —1,02 —0,67 —0,32 0 0,32 0,67 1,02 1,46 1,90																			
7	—2,04 —1,62 —1,20 —0,88 —0,56 —0,28 0 0,28 0,56 0,88 1,20 1,62 2,04																			
8	—2,13 —1,73 —1,33 —1,04 —0,75 —0,50 —0,25 0 0,25 0,50 0,75 1,04 1,33 1,73 2,13																			
9	—2,23 —1,84 —1,45 —1,17 —0,89 —0,66 —0,43 —0,22 0 0,22 0,43 0,66 0,89 1,17 1,45 1,84 2,23																			
10	—2,33 —1,95 —1,57 —1,31 —1,05 —0,83 —0,61 —0,40 —0,20 0 0,20 0,40 0,61 0,83 1,05 1,31 1,57 1,95 2,33																			
11	—2,42 —2,05 —1,68 —1,43 —1,18 —0,97 —0,76 —0,57 —0,38 —0,19 0 0,19 0,38 0,57 0,76 0,97 1,18 1,43 1,68 2,05 2,42																			
12	—2,50 —2,15 —1,80 —1,55 —1,30 —1,10 —0,90 —0,72 —0,53 —0,35 —0,18 0 0,18 0,35 0,53 0,72 0,90 1,10 1,30 1,55 1,80 2,15 2,50																			
13	—2,57 —2,23 —1,89 —1,65 —1,41 —1,21 —1,02 —0,84 —0,66 —0,50 —0,34 —0,17 0 0,17 0,34 0,50 0,66 0,84 1,02 1,21 1,41 1,65 1,89 2,23 2,57																			
14	—2,65 —2,31 —1,97 —1,74 —1,51 —1,31 —1,12 —0,95 —0,78 —0,62 —0,46 —0,31 —0,15 0 0,15 0,31 0,46 0,62 0,78 0,95 1,12 1,31 1,51 1,74 1,97 2,31 2,65																			
15	—2,71 —2,37 —2,04 —1,81 —1,58 —1,40 —1,21 —1,05 —0,89 —0,73 —0,58 —0,43 —0,29 —0,14 0 0,14 0,29 0,43 0,58 0,73 0,89 1,05 1,21 1,40 1,58 1,81 2,04 2,37 2,71																			

Megjegyzés: A táblázat k -adik sora rendre a megfelelő $a_1, c_1, a_2, c_2, \dots, c_{k-1}, a_k$ értékeket adja meg, a c_i értékeket *dőlt szedéssel* ($c_0 = -\infty, c_k = \infty$). Például $k = 4$ esetén $c_0 = -\infty, a_1 = -1,52, c_1 = -0,99, a_2 = -0,45, c_2 = 0, a_3 = 0,45, c_3 = 0,99, a_4 = 1,52, c_4 = \infty$.

A Magyar Tudományos Akadémia által a Smidt-elméletről rendezett ankéton Rényi Alfréd hozzászólásában [1] bebizonyította, hogy bizonyos $f(x)$ függvények esetén a bolygótávolságoknak nem egy, hanem több különböző törvénye lehet. Bebizonyította továbbá, hogy bizonyos esetekben [ha $f(x) = Cx^\alpha$, ahol $\alpha < 0$, továbbá, ha $f(x)$ monoton növekvő és konkáv függvény] a bolygótávolságok törvénye egyértelműen meghatározott. Felvetette a kérdést, hogy általában mi a szükséges és elégséges feltétele az $f(x)$ függvényre nézve, hogy a (8) megoldása egyértelmű legyen.

Látható, hogy (8) megegyezik (3)-mal, bár a határfeltételek különbözőek az (1)-ben megadott határfeltételektől. Smidt felteszi ugyanis, hogy a két szélső helyzetű bolygó fajlagos impulzusmomentuma adott. Ez a feltevés nyilván önkényes a számítás egyszerűbbé tétele céljából; hiszen a szélső bolygók fajlagos impulzusnyomatéka sem lehet az eredeti eloszlástól független. Mindenesetre több joggal tehetjük fel az (1)-ben foglalt határfeltételt, természetesen egyidejűleg feltételezve, hogy $f(x)$ csak egy véges intervallumban lehet pozitív, egyébként 0.

Tehát a jelen dolgozat 4. pontjában közölt eredmény a Smidt-elméletre is alkalmazható.* Megjegyezzük azonban, hogy segédteételünk akkor is alkalmazható, és ugyanarra az eredményre vezet, mint eredetileg, ha a határfeltételeket Smidt tárgyalásának megfelelően módosítjuk.

Eredményünkéből tehát következik, hogy ha $\log f(x)$ konkáv, akkor a bolygótávolság egyértelműen meghatározott.

Ez az eset magában foglalja azt az esetet is, mikor $f(x) = cx^\alpha$, ahol $\alpha > 0$, továbbá azt az esetet is, ha $f(x)$ konkáv, függetlenül attól, hogy monoton-e vagy nem.

Továbbra is eldöntetlen marad azonban az a probléma, hogy mi a szükséges és elegendő feltétele a (3) egyértelmű megoldhatóságának az $f(x)$ függvényre nézve.

IRODALOM

[1] Ankét O. J. Smidt: »Négy előadás a Föld keletkezésének elméletéről« c. könyvről. Rényi Alfréd hozzászólása. MTA. III. oszt. Közleményei. III/4. (1953) 595. o.

[2] O. J. Smidt: Négy előadás a Föld keletkezésének elméletéről. Akadémiai Kiadó, 1952.

ВЫБОР ИНТЕРВАЛОВ ПРИ ГРУППИРОВКЕ ЭЛЕМЕНТОВ ВЫБОРКИ

К. Шаркади

Резюме

В работе доказано, что для того, чтобы среднее значение квадрата отклонения между средними значениями, вычисленными путем группировки и путем точного расчета, было минимальным, необходимо выполнение системы уравнений (3). Если эта система уравнений имеет только одно решение, то это условие является также достаточным. На примере показано, что в противном случае это не имеет места. В дальнейшем

* Nyilvánvaló, hogy teljesen mellékes, hogy az $f(x)$ függvény az $\int_{-\infty}^{+\infty} f(x) dx = 1$ feltétellel normálva van-e vagy nem.

доказывается, что если $\log f(x)$ — вогнутая, то решение однозначно определено. Для случая нормального распределения даны также таблицы для различного числа интервалов.

Проблема связана также с астрономией, так как — по теории О. Ю. Шмидта — система уравнений, определяющая расстояние между планетами и солнцем в точности совпадает с системой уравнений (3).

CHOICE OF INTERVALS FOR GROUPING OF DATA

K. SARKADI

Summary

The following statement is proved: the expected value of the square of the difference between the averages calculated by grouping of data and by the exact method cannot be minimal unless the system of equations (3) is fulfilled. If (3) has only one solution, the condition itself is sufficient. An example for the opposite case is given. Furthermore, it is stated that if $\log f(x)$ is concave the solution is unique. To cover the case of normal distribution a table for several numbers of intervals is constructed.

The problem also has some astronomical bearings, since the system of equations, which according to the theory of O. J. Schmidt determines the distance of the planets from the sun, agrees exactly with the system of equations (3).