MINTÁK OPTIMÁLIS ELHELYEZÉSI STRATÉGIÁJA, HA A MINTAÉRTÉK EGY SÍKTARTOMÁNYON FOLYTONOSAN VÁLTOZÓ FÜGGVÉNY

JANOSITZ JÁNOS*

[Beérkezett 1969. december 10-én]

Az ásványi előfordulások egy vizsgált területen belül igen változékonyak lehetnek. A kutató fúrásokkal ezen változékonyságot igyekszünk maximális módon felderíteni. Ez olyan telepítéssel lehetséges, amelynél a fúrásokat úgy helyezzük el, hogy onnan kapjunk mindig újabb információt, ahol legnagyobb a bizonytalanság. A dolgozat olyan módszert ismertet, melynél a rendelkezésre álló adatok alapján mindig meghatározható az a fúrási pont, ahol egy újabb fúrás a leggyorsabban csökkenti az előfordulás egy vizsgált adatának bizonytalanságát.

A kettős integrálok Monte-Carlo módszerrel való közelítésénél és a föld mélyében levő ásványkincsek mélyfúrásokkal való kutatásánál és minden olyan feladatnál ahol egy síktartományban kell meghatározni a tartományon belül folytonosan változó függvényértékek átlagértékét, a tartomány különböző pontjaiban vett minták alapján felvetődik az a következő probléma, hogy a síktartományban vizsgált érték átlagáról melyik mintavételi stratégia szolgáltatja a legtöbb információt kötött mintaszám esetén.

Ha a síktartományról semmiféle előzetes információ nem áll rendelkezésre, úgy előzetes mintavétellel megtudhatjuk azt, hogy a tartomány mely részében nagyobb és mely részében kisebb a vizsgált érték változása, illetőleg szórása.

Az előzetes mintavételnél az egyes mintákat szabályos hálópontokba célszerű elhelyezni [1, 5].

Ha a síktartomány területe A és az egyes minták egymástól mért távolsága a, valamint az előzetesen tervezett mintaszám n, úgy szabályos háromszögháló esetén az

$$A = n \cdot \frac{a^2 \cdot \sqrt{3}}{4}, \tag{1}$$

négyzetháló esetén pedig az

$$A = n \cdot a^2 \tag{2}$$

összefüggés alapján határozható meg az a értéke.

* Dr. Janositz János, Nehézipari Műszaki Egyetem, Miskolc-Egyetemváros.

Az [5] alapján az előzetes mintavétel után az újabb mintákat olyan romboid-hálópontokba célszerű elhelyezni, amelyeknél a romboid oldalak iránya és nagysága a maximális, illetőleg minimális szórásirányától és azok nagyságaitól függ. Az egyes irányokba eső szórások pedig úgy vannak értelmezve, hogy az adott irányokba mutató párhuzamos egyenesekkel sávokra osztjuk a vizsgált tartományt és a sávokon belüli szórások átlaga adja meg az adott irányú szórás értékét.

A módszer ellen viszont felhozható az, hogy a szórás értéke általában az egyes irányok mentén a területen belül is változik. Ilyen esetet vizsgálva célunk egy kötött számú mintával azok optimális elhelyezésével a vizsgált síktartományról a legtöbb információt összegyűjteni.

A vizsgálatainknál az egyes mintaelemek költségeitől eltekintünk. Ha a területről semmiféle információnk nincs, úgy a mintákat a következő módon helyezzük el a területen.

Előzetes információszerzés érdekében n mintát tervezünk.

Az n értéke általában a tervezett N végleges mintaszám $10 \div 20\%$-a legyen, de nem kevesebb mint $4 \div 5$. Az n rögzítése után ezeket a mintákat olyan szabályos háromszögháló pontjaiba helyezzük el, hogy az egyes háromszögek területe közelítőleg a terület n-ed részével egyezzék (1).

A szabályos háromszögháló a vizsgált területre való elhelyezésekor lehetőség szerint minél több minta kerüljön a vizsgált síktartomány határvonalainak közelébe. Vagyis az előzetes mintavétel után a tartományban egy tetszőleges helyen a vizsgált jellemző becslésekor a lehetőség szerint maximális esetben legyen módunk interpolációra és minimális esetben legyen szükséges extrapoláció.

A megmaradó N-n számú mintaelem helyének megállapításakor pedig a következőképpen járunk el.

Természetes, hogy egy újabb információ mindig onnan a legsürgetőbb, ahol a legnagyobb a bizonytalanság. Vagyis az újabb mintaelemet olyan területrészen belül helyezzük el, ahol a területrészen belül a mintaértékek átlagának szórása a legnagyobb. A legkisebb területrész, ami itt még figyelembe vehető, az n előzetes minták alkotta háló egy-egy háromszöge.

Legyen a háromszög csúcsaiban levő minták értéke

$$x_1 \leq x_2 \leq x_3. \tag{3}$$

Ha

$$\frac{x_2 - x_1}{x_3 - x_1} = \lambda, \tag{4}$$

akkor a területen belül a szórásnégyzet közelítő értéke:

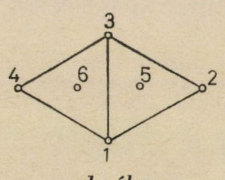
$$\sigma^2 = \frac{x_1^2 + x_2^2 + x_3^2 - \frac{1}{3}(x_1 + x_2 + x_3)^2}{2} = \frac{(x_3 - x_1)^2}{3}\left[\left(\lambda - \frac{1}{2}\right)^2 + \frac{3}{4}\right] \tag{5}$$

Az (5) összefüggésből látható, hogy a szórásnégyzet értékét három mintaérték esetében elsődlegesen a legnagyobb és legkisebb mintaérték különbsége határozza meg. A közbenső harmadik mintaérték a szórás értékét pedig csak $0 \div 25\%$-ig befolyásolja. Ebből következik, hogy a maximális szórású helyeket ott kell keresni, ahol két szomszédos mintaérték között legnagyobb az eltérés.

A fentiek alapján a maximális szórású helyek megkeresése meggyorsítható. Ha ui. a maximális szórású háromszög már adott, úgy az újabb mintaelemet ennek a súlypontjában helyezzük el. Legyen az így nyert mintaérték x_4, és a (3)-ban szereplő értékek nélkül a mintaértékek átlaga $\bar{x}_1$, úgy az x_4 figyelembevételével számítható „pontosabb átlagérték" $\bar{x}'$:

$$\bar{x}' = \frac{(n-3)\bar{x}_1 + \frac{3}{4}\sum_{i=1}^{4} x_i}{n}. \tag{6}$$

Természetes, hogy további mintákat mindig azon háromszög súlypontjába helyezünk, melyekben a szórás értéke éppen maximális. Ha így két egymás mellett levő háromszög súlypontjában is már új minta került (1. ábra), úgy az új nagyobb valószínűségű átlagérték $\bar{x}''$:



1. ábra

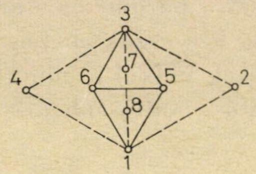
$$\bar{x}'' = \frac{(n-4)\bar{x}_2 + 4\,\frac{4}{6}\sum_{i=1}^{6} x_i}{n}, \tag{7}$$

ahol $\bar{x}_2$ a két háromszög területén kívül eső minták alapján számítható átlag.

Ha több egymás mellett levő háromszögben is adott újabb mintaérték, úgy az átlagszámításnál ezt az (5, 6)-hoz hasonló módon kell figyelembe venni.

A megmaradó minták fokozatos elhelyezésével előfordulhat, hogy a képződött kisebb — a kiinduláskori háromszögek területének harmadrészével egyező területű egyenlő oldalú — háromszögek valamelyikében a számítható szórás egyharmada nagyobb lesz, a megmaradt nagy háromszögekben számítható legnagyobb szórásnál. Ilyenkor a legközelebbi mintát az adott kis háromszög súlypontjában helyezzük el. Az átlag korrigálása ebben az esetben megbízhatóbb átlagértékhez vezet.

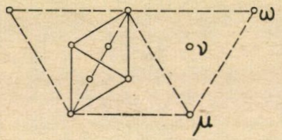
Az átlag korrekciója a következőképpen történik. Az egymás mellett levő kis háromszögekben, melyek súlypontjába már új minta is került, a háromszög csúcspontjaiban és súlypontjaiban levő minták alapján nyert értékekből számítunk egy átlagot, és ezt olyan súllyal vesszük figyelembe, ahány nagy háromszögbeli csúcspontot tartalmaz. A 2. ábra alapján, ha az 1. ábra és (7) összefüggés jelöléseit megtartjuk:



2. ábra

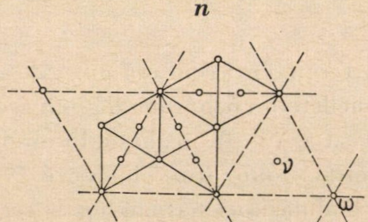
$$x=\frac{(n-4)\,\bar{x}_2+x_2+x_4+2\cdot\dfrac{2}{6}\,(x_1+x_8+x_5+x_6+x_7+x_3)}{n}. \qquad (8)$$

Ha esetleg így valamely nagy háromszög súlypontból nyert mintaérték mellől a csúcsponti mintaérték közül egy, kettő vagy mindhárom már más számításokban szerepelne, úgy azt a megmaradó kettőhöz, ill. egyhez vagy a legutóbbi esetben ahhoz rendeljük, melynek környezetében legkevesebb adat áll rendelkezésre (3., 4., 5. ábra).



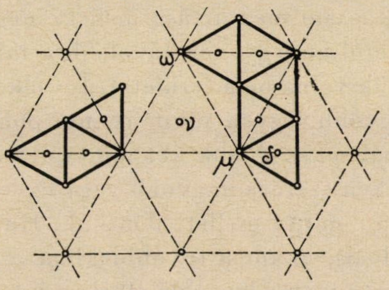
3. ábra

$$x^{(1)}=\frac{(n-2)\,\bar{x}_3+2\cdot\dfrac{2}{3}\,(x_\nu+x_\omega+x_\mu)}{n},$$



4. ábra

$$\bar{x}^{(k)}=\frac{(n-1)\,\bar{x}_4+\dfrac{1}{2}\,(x_\nu+x_\omega)}{n}.$$



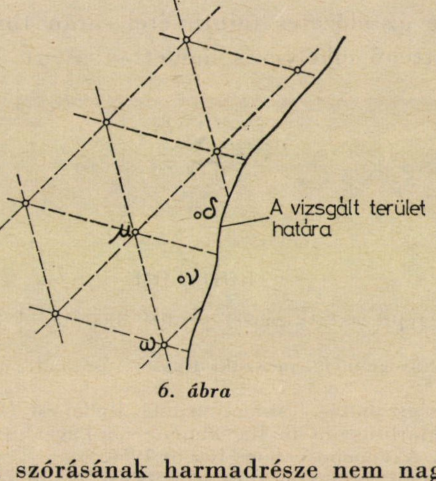
5. *ábra*

A bevonalkázott háromszögekben az átlag számításánál az x'_ω-t vesszük figyelembe:

$$x'_\omega = \frac{x_\omega + x_\nu}{2}.$$

Ha x_δ adat nem állna rendelkezésre, úgy x_μ helyett is x'_μ-t kellene figyelembe venni a számításoknál:

$$x'_\mu = \frac{x_\nu + x_\mu}{2}.$$



6. *ábra*

Ha a kis háromszögek szórásának harmadrésze nem nagyobb lényegesen a nagy háromszögbeli szórásnál, akkor a kisebb szórás ellenére is a nagy háromszög súlypontjában célszerű egy újabb mintát elhelyezni. Itt ugyanis a minták távolabb vannak egymástól és így a számított szórás bizonytalansága is nagyobb.

Hasonlóképpen kell kezelni a terület peremrészeit. Amint már említettük az extrapoláció általában nagyobb bizonytalansággal jár mint az interpoláció. Ennek a bizonytalanságnak a csökkentése érdekében az előzőekben a háló elhelyezéskor említetteken felül célszerű az (1)-ben számított háromszög oldalakat az előzetes mintaszám megtartása mellett megnövelni, hogy az előzetes minták a terület szélétől ne legyenek távolabbra mintha N számú mintákból készített hálóval fedtük volna be a területet. Így elkerülhető az, hogy az extrapolációs távolság nagyobb legyen mint az interpolációs. A széleken elhelyezendő újabb mintáknál sokszor csak két minta alapján dönthetünk (6. ábra).

Ha x_ω és x_μ értékek szórása nagyobb vagy egyező egy nagy háromszögön belüli szórással, úgy a ν minta mellett döntünk. Ha ugyanez áll x_μ és x_ν értékekre (egységnyi súllyal), akkor a továbbiakban a δ minta mellett döntünk.

Az N számú minta ily módon való elhelyezése után az ismertetett módon számítható a legvalószínűbb átlagérték. Ennek szórása a következőképpen számítható.

Legyen az összefüggően egyező „sűrűségű" minták száma $n_2, n_2, \ldots, n_\nu$ ezen i $(i = 1, 2, \ldots, \nu)$ részterületen belül a szórásnégyzet σ_i^2:

$$\sigma_i^2 = \frac{\sum_{j=1}^{n_i} [x_j - \bar{x}_i]^2}{n_i - 1} \qquad [i - 1, 2, \ldots, \nu]. \tag{9}$$

ahol $\bar{x}_i$ a részterületen belül az egyező sűrűségű mintákból számítható átlag. Legyen az egyes részterületen belül k_i $(i = 1, 2, \ldots, \nu)$ számú olyan minta, melynek elhelyezése az előzetes mintavétel során történt. Így az N számú minta ily módon történő elhelyezés mellett a nyert átlag érték szórásnégyzete $\sigma_{\bar{x}}^2$:

$$\sigma_{\bar{x}}^2 = \sum_{i=1}^{\nu} \frac{k_i^2}{n^2} \sigma_i^2. \tag{10}$$

IRODALOM

1. JANOSITZ, J.: Ocena reprezentatywnosci próbek Przegla, d Geologiczny Nr. 8. serpien. 1969. Varsó.
2. PRÉKOPA A.: Valószínűségelmélet műszaki alkalmazásokkal. Műszaki Könyvkiadó, Budapest 1962.
3. RÉNYI A.: Valószínűségszámítás. Tankönyvkiadó, Budapest 1954, 1966.
4. STAMMBERGER F.: Einfürhrung in die Berechnung von Lagerstättenvorräten fester mineralischer Rochstoffe. Akademie-Verlag, Berlin 1956.
5. TREMBECKI. A. S.: Zasady projektowania selektiwnej explolatacji zlóz surowców mineralnych, Wydawnictwo „Slask" Katowice 1966.

Strategy of Optimum Placing of Samples, if the Sampled Value is a Function Continuously Varying in a Plane Area. Within an investigated territory, occurrence of minerals may vary widely. By carrying out trial drillings, as thorough an exploration as possible is aimed at. This is possible by placing the drillings in such a way that fresh information is obtained from where the uncertainty is the greatest. The paper presents a method for determining, from the available data, always that drilling point where a new drilling will maximally reduce the uncertainty about the occurence of an examined characteristic data.

Optimale Strategie für die Verteilung der Proben, wenn der untersuchte Wert eine in einem ebenen Bereich kontinuierlich veränderliche Funktion ist. Innerhalb eines untersuchten Bereichs können die Mineralvorkommen stark veränderlich sein. Durch Probebohrungen wird angestrebt, diese Veränderlichkeit maximal aufzudecken. Dies ist mit einer solchen Plazierung möglich, bei der die Bohrungen so angeordnet werden daß immer dort neue Informationen erhalten werden wo die Unsicherheit am größten ist. Die Arbeit beschreibt eine derartige Methode, mit welcher auf Grund der verfügbaren Daten immer derjenige Bohrpunkt bestimmt werden kann, wo eine neue Bohrung die Unsicherheit einer untersuchten Charakteristik des Vorkommens am schnellsten verringert.