MÉRÉSI EREDMÉNYEK PONTATLANSÁGÁNAK HATÁSA HISZTOGRAMM FELVÉTELÉNÉL

VINCZE ISTVÁN

Összefoglalás

A cikk azzal a kérdéssel foglalkozik, hogy ha valamely eloszlásból vett minta elemeit — hisztogramm felvétele céljából — osztályokba soroljuk, milyen hatással van a mérési bizonytalanság az egyes osztályokhoz tartozó relatív gyakoriságokra. Feltételezzük, hogy a mérési eredményekhez olyan hiba adódik, amely független magától a mérés eredményétől.

Ha $f(x)$ jelöli a tényleges mérési eredmények eloszlásának sűrűségfüggvényét, míg a hiba eloszlásának sűrűségfüggvénye $g(x)$ (0 várható értékkel és s szórással), akkor az említett relatív hibára a (3) becslést nyerjük; itt ε és δ jelentését az (1) összefüggés adja. Ha a hiba eloszlása szimmetrikus, akkor a (2) egyenlőtlenség érvényes.

A dolgozat további részében két speciális esetet vizsgál, mindkettőben $f(x)$ normális, míg $g(x)$ norm lis, ill. egyenletes eloszlás sűrűségfüggvénye; végül egy gyakorlatban előfordult numerikus példát, ahol a mérési eredmény tapasztalati eloszlása elméletileg ismeretlen jellegű, míg a mérési hiba eloszlása egyenletes.

1. Gyakori eset, hogy valószínűségsűrűség-függvények hisztogramm útján való megközelítésénél a mérési eredmények leolvasása maga is bizonyos pontatlansággal történik. Így pl. golyók törési szilárdságának vizsgálatánál az egyik eljárás csupán egy legfeljebb 50 kg-os bizonytalansággal adja az egyes darabokra nézve az eredményt, vagyis ha a törési szilárdság pontos értéke x, akkor tapasztalat szerint ennél a mérési eljárásnál az $(x - 50, x + 50)$ intervallumban bármely értéket egyenlő valószínűséggel olvashatunk le. Egy ilyen szélsőséges esettől eltekintve is a mérési eljárásnál a leolvasás általában hibával jár, amely gyakori esetben normális, vagy ahhoz közelálló eloszlást követ. Az alábbiakban a mérés eredményét és a leolvasás bizonytalansága által okozott hibát mint valószínűségi változókat egymástól függetleneknek tételezzük fel. Meggondolásaink tehát alkalmazhatók mindazon esetekben, amikor e függetlenségi feltétel teljesül. Vannak olyan esetek is, amikor a mérési hiba növekszik a mért mennyiséggel, azonban legtöbbször a függetlenségi feltevés teljesül, mint azt a tapasztalat felhozott példánk esetében is mutatja.

A hisztogramm felvételénél mármost a változó értékének számbavehető közét intervallumokra osztjuk és megállapítjuk az egyes intervallumokba eső mérési eredmények számát. Kérdés, az említett bizonytalanság milyen mértékben torzítja az egyes intervallumokban a relatív gyakoriságot, vagyis az illető intervallumba eső mérési eredmények várható számának és az összes mérési eredmények számának hányadosát. — E kérdés jelentőséggel bírhat pl. akkor, ha a χ^2 módszert alkalmazzuk, amikor is az egy-egy intervallumba eső mérések számára megbízható gyakoriság-érték szükséges.

Az alábbiakban bizonyos feltételek mellett megállapítjuk azt az eltérést, amely valamely intervallumba eső tényleges gyakoriság és a mérési bizonytalanság következtében mutatkozó gyakoriság várható értékei között fennáll, majd megbecsüljük ennek és a tényleges gyakoriságnak a hányadosát, másszóval az értékek leolvasásának bizonytalansága következtében fellépő relatív hibát.

A szereplő mennyiségek elméleti értékek és a konkrét számolás ezeknek a mennyiségeknek az adatokból történő megbecslése útján vihető keresztül, ami elegendő nagyszámú mérésnél jó közelítésnek tekinthető.*

2. Jelentse a ξ valószínűségi változó valamely mérés eredményét, jelentse továbbá az η valószínűségi változó a mérési eredmény leolvasásának hibáját. Ekkor a mérés általunk észlelt eredménye $\xi + \eta$. Legyenek ξ és η függetlenek, és jelöljük ξ sűrűségfüggvényét $f(x)$-szel, amely legyen kétszer differenciálható, az η sűrűségfüggvénye $g(x)$, várható értéke 0, szórása s. Tekintsük továbbá azt a $(-\delta, \delta)$ $(\delta > 0)$ intervallumot a 0 körül, amelyben az η változó értékeit $1 - \varepsilon$ valószínűséggel veszi fel, vagyis legyen

$$\int_{-\delta}^{\delta} g(x)\,dx = 1 - \varepsilon. \tag{1}$$

Fenti jelöléseinkkel a $\xi + \eta$ sűrűségfüggvénye

$$h(x) = \int_{-\infty}^{\infty} f(x-y)\,g(y)\,dy.$$

Tegyük most fel, hogy N számú értéket mérünk, amelyek valamely (a, b) intervallumban helyezkednek el és osszuk fel az (a, b) intervallumot (pl. ekvidisztans) részintervallumokra. Egy ilyen részintervallum legyen (α, β), amelybe essék a N közül *ténylegesen* — vagyis ξ-re nézve — n számú mérési eredmény, míg a *leolvasott* értékek — vagyis $\xi + \eta$ értékeinek — száma legyen n'. Ekkor a relatív gyakoriságok várható értékei a következők:

$$M\left(\frac{n}{N}\right) = p = \int_{\alpha}^{\beta} f(x)\,dx$$

$$M\left(\frac{n'}{N}\right) = P = \int_{\alpha}^{\beta} h(x)\,dx.$$

* A kérdést gyakorlati feladat kapcsán *Kollár Károly* vetette fel; az itt tárgyalt általánosabb problémát *Rényi Alfréd* fogalmazta meg.

A relatív gyakoriságok eltérésének várható értéke

$$P - p = \int_{\alpha}^{\beta} \left(\int_{-\infty}^{\infty} [f(x-y) - f(x)]\, g(y)\, dy \right) dx.$$

Bontsuk ezt a kifejezést három részre a következő módon:

$$P - p = \int_{\alpha}^{\beta} \left(\int_{-\infty}^{-\delta} [f(x-y) - f(x)]\, g(y)\, dy \right) dx +$$

$$+ \int_{\alpha}^{\beta} \left(\int_{\delta}^{\infty} [f(x-y) - f(x)]\, g(y)\, dy \right) dx +$$

$$+ \int_{\alpha}^{\beta} \left(\int_{-\delta}^{\delta} [f(x-y) - f(x)]\, g(y)\, dy \right) dx.$$

Ha most az $f(x)$ sűrűségfüggvény maximumát m-el jelöljük, akkor (1)-re való tekintettel az első két integrál együttes értéke abszolút értékben legfeljebb $m(\beta - \alpha)\,\varepsilon$.

A harmadik tag megbecslésére fejtsük az $f(x - y)$ függvényt hatványsorba y szerint:

$$f(x-y) = f(x) - yf'(x) + \frac{y^2}{2} f''(x^*),$$

ahol x^* valamely alkalmasan választott érték az $(x - |y|,\ x + |y|)$ intervallumban. Ekkor

$$\int_{-\delta}^{\delta} [f(x-y) - f(x)]\, g(y)\, dy =$$

$$= f'(x) \int_{-\delta}^{\delta} - y\, g(y)\, dy + \frac{1}{2} f''(x^*) \int_{-\delta}^{\delta} y^2\, g(y)\, dy.$$

Itt mindenesetre

$$\int_{-\delta}^{\delta} y^2\, g(y)\, dy < s^2$$

és

$$\left| \int_{-\delta}^{\delta} y\, g(y)\, dy \right| \leq \left| \int_{\delta}^{\infty} y\, g(y)\, dy \right| + \left| \int_{-\infty}^{-\delta} y\, g(y)\, dy \right| \leq s \sqrt{\varepsilon}$$

és felírhatjuk P relatív hibájára a következő becslést :

$$(2) \qquad \frac{P-p}{p} < \frac{\beta-\alpha}{p}\,[\varepsilon\cdot\max f(x) +$$

$$+\, s\sqrt{\varepsilon}\max_{(\alpha-\delta,\beta+\delta)}|f'(x)| + \frac{1}{2}s^2\max_{(\alpha-\delta,\beta+\delta)}|f''(x)|\,].$$

Ha az η eloszlás szimmetrikus, akkor

$$\int_{-\delta}^{\delta} y\,g(y)\,dy = 0$$

és egyenlőtlenségünk a következő egyszerűbb alakot ölti :

$$(2') \qquad \frac{P-p}{p} < \frac{\beta-\alpha}{p}\,[\varepsilon\max f(x) + \frac{1}{2}s^2\max_{(\alpha-\delta,\alpha+\delta)}|f''(x)|\,].$$

Megjegyezzük, hogy figyelemmel kell lenni a P relatív gyakoriság szórására is, ami $\sqrt{\dfrac{P(1-P)}{N}}$ és ami N növelésével tetszőlegesen kicsinnyé tehető.

3. *Rényi Alfréd* jegyezte meg, hogy eredményünk alkalmazható abban az esetben is, ha az elkövetett hiba arányos a mérési eredmény nagyságával : $\eta = \xi\eta'$, ahol az η' változó független a mérés hibamentes ξ eredményétől. Ekkor a mérés végső eredménye $\xi + \xi\eta' = \xi\,(1+\eta')$. Meggondolásainkat ugyanis alkalmazhatjuk a $\log\xi + \log(1+\eta')$ független valószínűségi változók összegére.

4. Lássuk, hogyan becsülhető meg konkrét esetben számszerűen a relatív hiba, illetve hogyan tehető a gyakorlati követelményeknek megfelelően kicsinnyé.

Ki kell indulnunk mindenekelőtt az alapintervallum valamely felosztásából. Ez lehet egyenlő hosszúságú intervallumok választása, de megjegyezzük, hogy a fenti egyenlőtlenség is utal arra — amit más meggondolásból is következtetni szoktak, — hogy ahol kis valószínűségű intervallumok lépnének fel (vagyis a gyakorlati esetekben a szélső intervallumok), ezek hosszabbakra választandók, ami által p értéke nem válik túlságosan kicsinnyé. Az intervallumok megválasztása után p értéke P-vel becsülhető, amivel tulajdonképen az elkövetett hibát a tapasztalati relatív gyakorisághoz viszonyítjuk, s ennek kellő kicsiny volta ugyancsak megnyugtató a felvetett kérdés szempontjából.

Az $|f'(x)|$ és $|f''(x)|$ maximumának megbecslése általában kétféle módon történhet. Ha ismerjük az eloszlás jellegét, akkor elméleti meggondolásokból a tapasztalati paraméterekkel nyerhetünk becslést ; így pl. ha tudjuk, hogy normális, vagy ahhoz közeleső eloszlásról van szó, akkor az átlag és a tapasztalati szórás segítségével ez a becslés könnyen elvégezhető. Ha előttünk ismeretlen eloszlással állunk szemben, akkor a sűrűségfüggvény első közelítésének tekinthetjük azt a függvényt, amely az egyes intervallumok középpontjában a $\dfrac{P_k}{x_{k+1}-x_k}$ értéket veszi fel. Itt x_k $(k = 0, 1, 2, \ldots r)$ jelöli az intervallumok osztópontjait, P_k a relatív gyakoriság megfelelő tapasztalati értékét az (x_k, x_{k+1}) intervallumban, és pl. $f''(x)$ értékei közelesnek a következő értékhez:

$$\frac{P_{k+1} - 2P_k + P_{k-1}}{(x_{k+1} - x_k)^3} \tag{3}$$

feltéve, hogy itt a szomszédos közöket egyenlő hosszúságúaknak választottuk. A gyakorlatban előforduló eseteknél a sűrűségfüggvény görbületének túlságosan gyors megváltozásával nem kell számolni, s így az ily módon kapott érték tájékoztat $f''(x)$ nagyságrendje felől.

Az alábbiakban megvizsgáljuk azt az esetet, amikor η normális, majd amikor egyenletes eloszlást követ. Az előző eset a gyakorlatban leginkább előforduló, míg az utóbbira vonatkozó alábbi eredményünk megnyugtató olyan szélsőséges esetek szempontjából, amilyet bevezetésünkben említettünk. Mindkét esetben ξ-t normális eloszlásúnak választjuk. Végül egy, a gyakorlatban előfordult példát ismertetünk, ahol az eloszlás normálistól különböző, ismeretlen jellegű.

5. Legyen ξ normális eloszlású 0 várható értékkel és σ szórással, η ugyancsak normális eloszlású 0 várható értékkel és s szórással. Tekintsünk ezret meg nem haladó mérési eredményt, amelyek — ha s kicsiny σ-hoz képest, ami a tárgyalt probléma gyakorlati vonatkozása szempontjából elsősorban bír érdekességgel — nagy valószínűséggel a $(-3\sigma, +3\sigma)$ intervallumba esnek; osszuk ezt fel 12 egyenként $\frac{\sigma}{2}$ hosszúságú intervallumra. Ekkor

$$\beta - \alpha = x_{k+1} - x_k = \frac{\sigma}{2}, \quad k = 0, 1, 2, \dots 11.$$

Válasszuk δ-t $3s$-nek, ekkor $\varepsilon = 0{,}0027$; tekintsük továbbá azt az esetet, amikor a leolvasási hiba szórásának hatszorosa egy egész intervallumot tesz ki: $6s = \frac{\sigma}{2}$.

Ha (2') formulánk segítségével meghatározzuk a P relatív gyakoriságok viszonylagos hibáinak becslését a 12 intervallumra a következő értéket kapjuk (itt k jelenti balról kezdve az intervallum sorszámát):

k	A relatív hiba kisebb, mint
1 és 12	0,146
2 és 11	0,094
3 és 10	0,019
4 és 9	0,009
5 és 8	0,007
6 és 7	0,009

Az első és utolsó intervallumra kapott 14,6%-os relatív hiba nem jelentős, tekintettel arra, hogy az ide eső értékek száma a gyakorlatban nem haladja meg a 10-et.

Megjegyezzük, hogy ebben az esetben közvetlen (és valamivel pontosabb) becslést kapunk, ha figyelembe vesszük a $\xi + \eta$ sűrűségfüggvényét, ami normális eloszlásról lévén szó:

$$h(x) = \frac{1}{\sqrt{2\pi(\sigma^2+s^2)}} e^{-\frac{x^2}{2(\sigma^2+s^2)}}.$$

6. Tekintsük most azt az esetet, amikor ξ eloszlása normális, míg η eloszlása egyenletes. Legyen η sűrűségfüggvénye

$$g(x) = \begin{cases} 0, & \text{ha } x \leq x_0 - \delta \\ \frac{1}{2\delta}, & \text{ha } x_0 - \delta < x \leq x_0 + \delta \\ 0, & \text{ha } x_0 - \delta < x. \end{cases}$$

Ekkor $s^2 = \frac{\delta^2}{3}$. Válasszuk most $2\,\delta$-t egy teljes intervallum hosszúságúnak: $2\delta = \frac{s}{2}$. Ebben az esetben $\varepsilon = 0$.

Kiszámítva az egyes intervallumokban a relatív hibákra a (2′) egyenlőtlenségből kapható becsléseket, a következő táblázatot írhatjuk fel:

k	A relatív hiba kisebb, mint
1 és 12	0,103
2 és 11	0,051
3 és 10	0,022
4 és 9	0,010
5 és 8	0,010
6 és 7	0,015

7. Csapágygolyók törőszilárdságára végzett összesen 271 mérés az alábbi táblázatban feltüntetett eredményeket szolgáltatta. Tapasztalat szerint a mérési eljárásnál a valódi értéktől számítva $\pm$ 50 kg-ig minden érték leolvasása egyenlő valószínűséggel előfordulhat. Az értékek csoportosításánál a közök hosszát 500 kg-nak választottuk.

Az osztályközök sorszáma	Törőszilárdság értéke 100 kg-ban	Az értékek gyakorisága	Az értékek relatív gyakorisága
1	20 —25	3	0,0111
2	25,1—30	2	0,0074
3	30,1—35	14	0,0516
4	35,1—40	20	0,0738
5	40,1—45	18	0,0664
6	45,1—50	41	0,1513
7	50,1—55	71	0,2620
8	55,1—60	45	0,1661
9	60,1—65	36	0,1328
10	65,1—70	20	0,0738
11	70,1—75	1	0,0037
		271	1,0000

A (2′) formula alkalmazásánál $x_{k+1} - x_k = 5$, $\delta = 0{,}5$, $s^2 = 0{,}083$. Ha az $f''(x)$-re nézve a (3) alatt megjelölt közelítést használjuk, akkor azt találjuk, hogy az 1—10-ig terjedő intervallumokban a relatív gyakoriságok viszonylagos hibájának nagyságrendje 1%, míg a 11. intervallumban 4%. Ilyen módon a leolvasási hiba ellenére az értékek itt megválasztott osztályba sorolása és az egyes osztályokban kapott gyakoriságok további felhasználása teljesen indokolt.

ВЛИЯНИЕ НЕТОЧНОСТИ РЕЗУЛЬТАТОВ ИЗМЕРЕНИЙ ПРИ ПОСТРОЕНИИ ГИСТОГРАММ

И. Винце

Резюме

Статья рассматривает вопрос, какое влияние оказывает неточность измерения на частоты групп, при группировке элементов выборки с целью построения гистограмм, взятых из некоторого распределения? Предполагаем, что погрешность получающаяся при измерениями не зависит от самого результата измерений.

Если f(x) обозначает плотность распределения фактических результатов измерений, когда плотность распределения погрешности — g(x), со средним значением 0 и дисперсией s то относительно упомянутой погрешности получаем оценку (3); здесь значение ε и δ дается зависимостью (1.).

Если распределение погрешности является симметричным, то имеет место неравенство (2).

В дальнейшем работа занимается двумя специальными случаями. В обоих случаях распределение f(x) является нормальным, а распределение g(x) нормальным или равномерным. В заключение рассматривается один, — взятый из практики — пример, где характер распределения f(x) неизвестен, а распределение g(x) является равномерным.

DIE WIRKUNG DER FEHLER VON MESSERGEBNISSEN BEI DER AUFNAHME EINES HISTOGRAMMS

ST. VINCZE

Zusammenfassung

Die Elemente einer Stichprobe seien zwecks Aufnahme eines Histogrammes in Klassen eingeteilt. Dann wird untersucht, welche Wirkung die Messungenauigkeit auf die zu den einzelnen Klassen gehörigen relativen Häufigkeiten hat. Es wird angenommen, dass zu den Messergebnissen ein von ihnen unabhängiger Fehler hinzukommt.

Es bezeichne $f(x)$ die Dichtefunktion der richtigen Messergebnisse; die Dichtefunktion des Fehlers mit dem Mittelwert 0 und Streuung s sei mit $g(x)$ bezeichnet. Wir erhalten für den obenerwähnten relativen Fehler die Abschätzung (3), wobei die Bedeutung von ε und δ aus (1) abgelesen werden kann. Ist die Verteilungsfunktion des Fehlers symmetrisch, so gilt die Ungleichung (2).

Im Weiteren werden zwei Spezialfälle behandelt. $f(x)$ ist in beiden Fällen die Dichtefunktion der normalen Verteilung, $g(x)$ ist in einem Falle die Dichtefunktion der normalen Verteilung, im anderen Falle die Dichtefunktion der Gleichverteilung.

Zum Schluss wird ein numerisches Beispiel aus der Praxis berechnet: $f(x)$ ist eine empirische Verteilung, während $g(x)$ die Dichtefunktion der Gleichverteilung.