

KÉTVÁLTOZÓS EMPIRIKUS ELOSZLÁSFÜGGVÉNYEK ELTÉRÉSÉRŐL

Írta: VINCZE ISTVÁN

Bevezetés

1. Legyen a ξ valószínűségi változó folytonos eloszlásfüggvénye $F(x)$ és legyenek a ξ -re vonatkozó független megfigyelések eredményei $\xi_1, \xi_2, \dots, \xi_n$. E minta empirikus eloszlásfüggvénye

$$F_n(x) = \frac{1}{n} \sum_{\xi_i \leq x} 1,$$

vagyis $F_n(x)$ az x -nél kisebb mintaelemek relatív gyakorisága a mintában. Az elméleti és a tapasztalati eloszlásfüggvények abszolút eltérése maximumának, mint valószínűségi változónak eloszlását KOLMOGOROV 1933-ban publikált nevezetes eredménye világítja meg, mely szerint

$$(1) \quad \lim_{n \rightarrow \infty} \mathbf{P}(\sqrt{n} \sup_{(x)} |F_n(x) - F(x)| < y) = \sum_{-\infty}^{\infty} (-1)^i e^{-2i^2 y^2}, y > 0.$$

A határeloszlás tehát nem függ az $F(x)$ folytonos eloszlásfüggvény speciális alakjától. KOLMOGOROV e tételt rendkívül sokirányú vizsgálatot indított meg, amely vizsgálatok igen nagy jelentőséggel bírnak a sztochasztikus folyamatok elméletében és a matematikai statisztikában mind elméleti kérdések, mind gyakorlati alkalmazások szempontjából. SZMIRNOV meghatározta a maximális — egyoldali — eltérés határeloszlását, melyre a

$$(2) \quad \lim_{n \rightarrow \infty} \mathbf{P}(\sqrt{n} \sup_{(x)} (F_n(x) - F(x)) < y) = 1 - e^{-2y^2}, y > 0$$

adódott.

Újabban véges n -re is kiszámították a maximális eltérés és az abszolút eltérés pontos eloszlását (BLACKMAN [1]), ami szintén független $F(x)$ alakjától, továbbá ismeretes (SCHMID [9]) a KOLMOGOROV-féle eloszlás megfelelője olyan eloszlásfüggvényre, amelynek véges sok szakadási helye van. Ez esetben a határeloszlás már függ az $F(x)$ eloszlásfüggvénytől, vagyis nem eloszlásmentes.

2. Mindezek a vizsgálatok és az eredmények számos további általánosítása egyetlen változó elméleti és empirikus eloszlásának eltérésére vonatkoznak, de a kétváltozós eset megfelelő problémái alig szerepelnek az iro-

dalomban. Így nem ismeretes a

$$A_N^{\pm} = \sup_{(x, y)} (F_N(x, y) - F(x, y))$$

és

$$A_N = \sup_{(x, y)} |F_N(x, y) - F(x, y)|$$

maximális egyoldali, ill. abszolút eltérésének eloszlása vagy határeloszlása sem, ahol itt most $F(x, y) = \mathbf{P}(\xi < x, \eta < y)$ a (ξ, η) valószínűségi változó-pár elméleti eloszlásfüggvénye, míg $F_N(x, y)$ a (ξ, η) -ra vonatkozó N elemű minta, (ξ_i, η_i) $i = 1, 2, \dots, N$ empirikus eloszlásfüggvénye:

$$F_N(x, y) = \frac{1}{N} \sum_{\substack{\xi_i \leq x \\ \eta_i \leq y}} 1,$$

vagyis a $\xi < x, \eta < y$ síknegyedbe eső mintaelemek relatív gyakorisága. Ha ξ és η nem függetlenek, akkor az eloszlás — eltérően az egyváltozós esettől — függeni fog $F(x, y)$ alakjától. Ezen eltérésekre vonatkozólag KIEFER és WOLFOWITZ [5] a következő eredményt érték el:

Léteznek olyan c_0 és c_1 állandók, hogy minden N -re és $F(x, y)$ -ra a a következő relációk állanak:

$$G_N^+(r) = \mathbf{P}(\sqrt{N} A_N^+ < r) > 1 - c_0 e^{-c_1 r^2}$$

és

$$G_N(r) = \mathbf{P}(\sqrt{N} A_N < r) > 1 - c_0 e^{-c_1 r^2}.$$

Ennek az eredménynek érdekességeként, szerzők éppen azt emelik ki, hogy az eloszlásfüggvényeket olyan $r \rightarrow \infty$ -el 1-hez tartó korlát fölé szorítják, amely N -től és $F(x, y)$ -től független, holott a $\lim_{N \rightarrow \infty} G_N^+(r)$ és $\lim_{N \rightarrow \infty} G_N(r)$ függvények, melyek létezését szerzők ugyancsak bebizonyítják, már függeni fognak $F(x, y)$ alakjától. Szerzők tételünket (ξ, η) pár helyett tetszőleges $(\xi^{(1)}, \dots, \xi^{(m)})$ vektorváltozóra bizonyítják.

3. Hasonló a helyzet a kétmintás esetre vonatkozó vizsgálatokkal.

Ha a ξ és ξ' változók folytonos eloszlásfüggvényei $F(x)$, ill. $G(x)$ és a ξ -re vonatkozó $(\xi_1, \dots, \xi_n)$ minta empirikus eloszlásfüggvénye $F_n(x)$, a ξ' -re vonatkozó $(\xi'_1, \xi'_2, \dots, \xi'_m)$ eloszlásfüggvénye $G_m(x)$, akkor az $F(x) \equiv G(x)$ esetben a

$$D_{n, m}^+ = \max_{(x)} (F_n(x) - G_m(x))$$

és a

$$D_{n, m} = \max_{(x)} |F_n(x) - G_m(x)|$$

változókra — $\sqrt{\frac{mn}{m+n}}$ normálással és $\frac{m}{n} \rightarrow c > 0$ feltételezéssel — SZMIRNOV tételei szerint az 1. pontban szereplő (2), ill. (1) határeloszlások adódnak.

A $D_{n,m}^+$ és $D_{n,m}$ valószínűségi változók pontos eloszlása csak $m=n$ (GNEDENKO—KOROLJUK [2]) és $m=kn$ (BLACKMAN [1]) esetben ismeretesek. Más esetben az eloszlás igen bonyolult és legfeljebb módszer ismeretes a meghatározásukra (OZOLSZ [7]), vagy speciális pozíciókban való közelítő meghatározásukra (HODGES [3]). Aszimptotikus formulát KOROLJUK [6] adott.

Ezek az eloszlások annak eldöntése szempontjából jelentősek, hogy két minta azonos eloszlásból származik-e. Arra az esetre, ha n és m keveset különböznek egymástól (pontosabban $m > n$ és $m-n = O(\sqrt{m+n})$) szerző REIMANN JÓZSEFFEL írt cikkében [8] a következő statisztikák vizsgálatát javasolja:

$$B_{n,m}^+ = \max_{(x)} (nF_n(x) - mG_m(x)),$$

$$B_{n,m} = \max_{(x)} \left| nF_n(x) - mG_m(x) + \frac{m-n}{2} \right| - \frac{m-n}{2},$$

amely változók pontos eloszlására eléggé egyszerű formulák adódnak. (A $B_{n,m}^+$ statisztika az alábbi, kétváltozós esetre vonatkozó tételünkben fog szerepet játszani.)

4. Tekintsük most a folytonos $F(x, y)$, ill. $G(x, y)$ eloszlásfüggvényekkel bíró (ξ, η) , ill. (ξ', η') valószínűségi változópárokat. Legyenek a (ξ, η) -ra, ill. (ξ', η') -re vonatkozó független mérési eredmények (ξ_i, η_i) , ill. (ξ'_i, η'_i) , mindkét esetben $i=1, 2, \dots, N$. Jelöljük a megfelelő empirikus eloszlásfüggvényeket $F_N(x, y)$ és $G_N(x, y)$. Az egyváltozós esetnek megfelelően a

$$D_{N,N}^+ = \max_{(x,y)} (F_N(x, y) - G_N(x, y))$$

és

$$D_{N,N} = \max_{(x,y)} |F_N(x, y) - G_N(x, y)|$$

eltérések eloszlásának ismerete bírna érdekességgel, az $F(x, y) \equiv G(x, y)$ esetben, azonban — mint a megfelelő empirikus és elméleti eloszlások maximális eltérésének esetében, azaz az egymintás esetben — itt sem ismeretesek eredmények.

Jelen dolgozatunkban csupán az egyik változóra vonatkozó maximális eltéréssel foglalkozunk, vagyis a

$$D_{N,N}^+(y) = \max_{(x)} (F_N(x, y) - G_N(x, y))$$

és a

$$D_{N,N}(y) = \max_{(x)} |F_N(x, y) - G_N(x, y)|$$

eltérésekkel kapcsolatos eloszlásokkal. Tekintettel a két változó azonos szerepére, eredményeink a megfelelően definiált $D_{N,N}^+(x)$ és $D_{N,N}(x)$ eltérésekre is vonatkoznak.

Most használt jelöléseink felhasználásával a következő tételt fogalmazzuk meg, amit az 1. §-ban bizonyítottunk be.

1. TÉTEL. *Annak valószínűsége, hogy az $F(x, y) \equiv G(x, y)$ feltevés mellett az $F(\infty, y) = H_2(y)$ peremeloszlás-törvény szerint véletlen választott valamely $\eta = y$ -ra nézve a $D_{N,N}^+(y)$ maximális eltérés a $\frac{k}{N}$ értéket ne haladja meg:*

$$(3) \quad P_{N,N}^{(k)} = \mathbf{M}_\eta \left[\mathbf{P} \left(D_{N,N}^+(y) < \frac{k}{N} \right) \right] = \\ = \frac{1}{(2N+1) \binom{2N}{N}} \sum_{n=0}^N \sum_{m=0, n-k}^N \binom{2N-n-m}{N-n} \left[\binom{n+m}{n} - \binom{n+m}{n-k} \right],$$

ahol $\overline{0, n-k} = \max(0, n-k)$ és $k=0, 1, 2, \dots, N$.

A $D_{N,N}(y)$ abszolút eltérésre vonatkozó megfelelő valószínűséget $P_{N,N}^{[K]}$ -val jelölve, ennek értékét a következő kifejezés adja:

$$(4) \quad P_{N,N}^{[K]} = \frac{1}{(2N+1) \binom{2N}{N}} \sum_{n=0}^N \sum_{m=n-k}^{n+k} \binom{2N-n-m}{N-n} \sum_{j=-\infty}^{\infty} (-1)^j \binom{m+n}{n+jk},$$

ahol a második összegezési jelel m értéke az alsó határnál a 0 alá, a felső határnál N fölé nem mehet.

Tételünk egy kombinatorikus értelmezését a bizonyítás 5. pontjában adjuk.

Módszerünkkel az $F(x, y) \equiv G(x, y) = H_1(x)H_2(y)$ független esetben hasonló tétel adható meg a

$$\left\{ D_{N,N}^+(x) < \frac{k}{N}, D_{N,N}^+(y) < \frac{k'}{N} \right\} \quad (k, k' = 0, 1, 2, \dots, N)$$

együttes eseményekre, amire később kívánunk visszatérni.

A 2. §-ban bebizonyítjuk a (3), ill. (4) valószínűségekre vonatkozó következő határeloszlástételt:

2. TÉTEL. *Ha $N \rightarrow \infty$ és $\frac{k}{\sqrt{2N}} \rightarrow r$, akkor*

$$(5) \quad \lim_{N \rightarrow \infty} P_{N,N}^{(k)} = \int_0^1 \Phi \left(\frac{r}{\sqrt{z(1-z)}} \right) dz - e^{-2r^2} \int_0^1 \Phi \left(\frac{2z-1}{\sqrt{z(1-z)}} r \right) dz,$$

$$(6) \quad \lim_{N \rightarrow \infty} P_{N,N}^{[K]} = \\ = \sum_{j=-\infty}^{\infty} (-1)^j e^{-2j^2} \int_0^1 \left[\Phi \left(\frac{1-2j(1-z)}{\sqrt{z(1-z)}} r \right) - \Phi \left(\frac{-1-2j(1-z)}{\sqrt{z(1-z)}} r \right) \right] dz,$$

$$\text{ahol mindkét esetben } r > 0 \text{ és } \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du.$$

1. §. Az eloszlástétel bizonyítása

1. A bizonyításhoz néhány lemmát használunk fel.

Tekintsük az n számú $+1$ -ből és m számú -1 -ből álló véletlen sorozatot: $\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_{n+m}$, amelynél a lehető $\binom{n+m}{n}$ számú lehetséges sorrend mindegyike egyenlően valószínű és legyen $S_i = \mathcal{I}_1 + \dots + \mathcal{I}_i$ ($i = 1, 2, \dots, n+m$). Ekkor a következő reláció érvényes: ha $n < m$, akkor

$$(1.1) \quad P_{n,m}^{(k)} = \mathbf{P}(\max_{(i)} S_i < k) = 1 - \frac{\binom{n+m}{n-k}}{\binom{n+m}{n}}, \quad k = 0, 1, 2, \dots, n;$$

ha $n > m$, akkor $k = 0, 1, 2, \dots, n-m-1$ -re $P_{n,m}^{(k)} = 0$, míg $k = n-m, n-m+1, \dots, n$ -re $P_{n,m}^{(k)}$ értéke az (1.1) formulában adott kifejezéssel egyezik.

A bizonyítás céljából tekintsük a számegyenes egész értékű pontjain történő bolyongást, melynek egy véletlen útja az $S_0 = 0$ -ból indulva az $S_1, S_2, \dots$ pontokon keresztül a $S_{n+m} = n-m$ -ig halad. Minden sorozatnak egy út felel meg, és minden útnak egy sorozat, kölcsönösen egyértelműen. Azon utak számát határozzuk meg, amelyek a k ($k > 0$) magasságot eléri vagy meghaladják. Ha ezeknek az utaknak a k magasságot először elérő pontjuk utáni részét a k pontra tükrözzük, akkor az ugyancsak a 0-ból induló, de az $n+m$ lépésben a $2k-n+m$ pontba jutó utakhoz jutunk oly módon, hogy ez a megfeleltetés egyértelmű. Ilyen út $m+k$ lépést tartalmaz pozitív, s $n-k$ lépést negatív irányban. Ily módon a k -t elérő utak száma $\binom{n+m}{n-k}$ s a k -t el nem érő utak száma $\binom{n+m}{n} - \binom{n+m}{n-k}$, amiből (1.1) formulánk következik. Ha $n > m$, akkor $S_{n+m} = n-m > 0$ lévén, a $k < n-m$ nem lehet S_i -re maximum.

2. Az előző pontban használt jelölésekkel a következő reláció érvényes:

$$(1.2) \quad P_{n,m}^{[k]} = P(\max_{(i)} |S_i| < k) = \frac{1}{\binom{n+m}{n}} \sum_{j=-\infty}^{\infty} (-1)^j \binom{n+m}{n+jk},$$

ha $|n-m| < k < \max(n, m)$; $k < |n-m|$ -re $P_{n,m}^{[k]} = 0$.

E formula speciális esete ELLIS következő formulájának (lásd pl. [4]): Történjék a bolyongás a $(0, s)$ pontok között (s egész) és induljon i -ből ($0 < i < s$), érkezzenek M lépés után j -be. Ekkor azon utak száma, amelyek a 0 és az s pontok egyikét sem érik el,

$$f(s, M, i, j) = \sum_{j=-\infty}^{\infty} \left[\left(\frac{1}{2}(M-i+j) + \gamma s \right) - \left(\frac{1}{2}(M+i+j) + \gamma s \right) \right].$$

Az idézett helyen adott formulából $p = q = \frac{1}{2}$ helyettesítéssel, és 2^M -el — vagyis az ott figyelembe vett összes utak számával — való szorzással kapjuk fenti kifejezésünket, ha még áttérünk használt jelöléseinkre.

Esetünkben az (s, M, i, j) mennyiségeknek a $(2k, n+m, k, k+n-m)$ számnégyes felel meg s ebből már adódik (1.2).

3. Tekintsük a (ξ, η) változó párra vonatkozó (ξ_i, η_i) $i = 1, 2, \dots, N$ mintaelemeket, hasonlóképpen a (ξ'_i, η'_i) $i = 1, 2, \dots, N$ a (ξ', η') párra vonatkozóakat. Jelölje ν_y azon (ξ_i, η_i) -k számát, amelyek η_i koordinátája az y értéket nem éri el és μ_y azon (ξ'_i, η'_i) -k számát, amelyekre $\eta'_i < y$. Ekkor annak valószínűsége, hogy $\nu_y = n$ és $\mu_y = m$ legyen $(n, m = 0, 1, 2, \dots, N)$,

$$(1.3) \quad P(\nu_y = n, \mu_y = m) = \binom{N}{n} \binom{N}{m} [H_2(y)]^{n+m} [1 - H_2(y)]^{2N-n-m},$$

ahol az $F(\infty, y) = H_2(y)$ jelölést használtuk. Annak valószínűsége, hogy ez az esemény *valamely* — a $H_2(y)$ eloszlástörvény szerint véletlenül választott — η -ra bekövetkezzék, a teljes valószínűség tétele szerint

$$(1.4) \quad P_{n,m} = \int_{-\infty}^{\infty} P(\nu_y = n, \mu_y = m) dH_2(y) = \binom{N}{n} \binom{N}{m} \int_0^1 t^{n+m} (1-t)^{2N-n-m} dt = \\ = \frac{1}{2N+1} \frac{\binom{m+n}{n} \binom{2N-n-m}{N-n}}{\binom{2N}{N}}.$$

Ez a valószínűség a következőképpen értelmezhető: Legyen egy urnában N számú piros és N számú fehér golyó; húzzunk M számú golyót oly

módon, hogy a $M = 0, 1, 2, \dots, 2N$ számú golyó húzásának mindegyike egyenlően valószínű. Ekkor (1.4) annak valószínűségét adja, hogy $M = n + m$ legyen a húzás eredménye, ahol n a húzott piros, m a fehér golyók száma. Ennek megfelelően a

$$\sum_{n=0}^N \sum_{m=0}^N P_{n,m} = 1$$

reláció kell álljon, ami — hipergeometrikus eloszlások keverékéről lévén szó — egyszerűen adódik.

4. Most bebizonyítjuk a következő relációt:

$$(1.5) \quad P\left(\max_{(x)} D_{N,N}^+(y) < \frac{k}{N} \mid v_y = n, \mu_y = m\right) = P_{n,m}^{(k)},$$

ahol $P_{n,m}^{(k)}$ -t a jelen paragrafus (1.1) formulája adja. A

$$\{ND_{N,N}^+(y) = N \max_{(x)} (F_N(x, y) - G_N(x, y)) < k\}$$

esemény azt jelenti, hogy az $\eta = y$ magasságú egyenes minden x pontjára a $(\xi < x, \eta < y)$ síknegyedbe az első mintából eső pontok száma legfeljebb $k-1$ -el haladja meg az ugyanott fekvő második mintából származó pontok számát. Ehhez azonban elég ezen érintett pontok ξ_i , ill. ξ'_i koordinátáinak elhelyezkedését vizsgálni, amely $\eta_i < y$, ill. $\eta'_i < y$ feltételeknek eleget tevő változók legyenek:

$$(1.6) \quad \xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_\nu}, \quad \text{ill.} \quad \xi'_{j_1}, \xi'_{j_2}, \dots, \xi'_{j_\mu},$$

ahol $\nu = v_y$ és $\mu = \mu_y$.

Ezek azonban azonos eloszlású független valószínűségi változók, s két mintának tekintve ezeket $F_\nu(x)$, ill. $G_\mu(x)$ eloszlásfüggvényekkel — az $\eta < y$ feltétel mellett — az

$$N(F_N(x, y) - G_N(x, y)) = \nu F_\nu(x) - \mu G_\mu(x)$$

összefüggés áll. Azonban — adott ν és μ mellett — a

$$\{\max_{(x)} (\nu F_\nu(x) - \mu G_\mu(x)) < k\}$$

esemény valószínűsége egyszerűen visszavezethető a jelen paragrafus 1. pontjában adott (1.1) formulára ($n = \nu, m = \mu$). Ugyanis adott x -re a $\nu F_\nu(x) - \mu G_\mu(x)$ megadja, hogy mennyivel haladja meg az x -nél kisebb ξ -k száma a ξ' -k számát az (1.6) sorozatban. Független valószínűségi változókról lévén szó, ezen ξ -k és ξ' -k minden sorrendje egyenlően valószínű, s ha az egyesített sorozatot nagyság szerint rendezve I -vel jelöljük, és bevezetjük a

$$g_i = \begin{cases} +1, & \text{ha } I\text{-ben az } i\text{-edik elem valamely } \xi, \\ -1, & \text{ha } I\text{-ben az } i\text{-edik elem valamely } \xi', \end{cases}$$

akkor $\nu F_\nu(x) - \mu G_\mu(x)$, ha x a $-\infty$ -tól a $+\infty$ -ig növekszik, éppen az $S_0, S_1, \dots, S_{n+m}$ különböző értékeket veszi fel. Tehát

$$\max_{(x)} (\nu F_\nu(x) - \mu G_\mu(x)) = \max_{(i)} S_i,$$

amiből (1.5)-ben adott állításunk következik.

5. Most rátérünk 1. tételünk bizonyítására. Nyilvánvaló, hogy

$$P\left(D_{N,N}^+(y) < \frac{k}{N}\right) = \sum_{n=0}^N \sum_{m=n-k, 0}^N P\left(D_{N,N}^+ < \frac{k}{N} \mid \nu_y = n, \mu_y = m\right) P(\nu_y = n, \mu_y = m),$$

ahol az (1.1) formulára vonatkozó megszorítások indokolják az összegezési határokat; itt az összegben szereplő első valószínűségek (1.5)-ben, a második valószínűségek (1.3)-ban vannak megadva. Ha most annak valószínűségét keressük, hogy valamely $-H_2(y)$ mérték szerint véletlenül választott $- \eta = y$ -ra teljesüljön a $\{D_{N,N}^+(y) < k\}$ esemény, az összetett valószínűségi tétel szerint a

$$P_{N,N}^{(k)} = M_\eta \left[P\left(D_{N,N}^+(\eta) < \frac{k}{N}\right) \right]$$

valószínűségre jutunk, ami (1.4) miatt 1. tételünk 3. formulájához vezet.

A (3) formulából leolvasható annak következő kombinatorikus értelmezése: Tekintsük az N számú $+1$ -ből és N számú -1 -ből álló $(\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_{2N})$ véletlen sorozatot, ahol minden sorrend egyenlően valószínű. Válasszunk ki véletlenszerűen λ számú $\mathcal{G}$ -t: $\mathcal{G}_{i_1}, \mathcal{G}_{i_2}, \dots, \mathcal{G}_{i_\lambda}$, $(i_1 < i_2 < \dots < i_\lambda)$ oly módon, hogy bármely számú kiválasztása egyenlően valószínű:

$$P(\lambda = l) = \frac{1}{2N+1}, \quad \lambda = 0, 1, 2, \dots, 2N.$$

Ekkor a következő reláció érvényes:

$$P(\max_{0 \leq j \leq \lambda} (\mathcal{G}_{i_1} + \mathcal{G}_{i_2} + \dots + \mathcal{G}_{i_j}) < k) = P_{N,N}^{(k)}$$

A $P_{N,N}^{[k]}$ valószínűségekre vonatkozó állítás bizonyítása — a megfelelő változtatásokkal — teljesen azonos módon történik. Ugyancsak érvényes a formula itt megadott kombinatorikus értelmezésének megfelelője. Megjegyezzük, hogy tételünk kombinatorikus értelmezése egyben alternatív bizonyításnak is tekinthető.*

* SARKADI KÁROLY megjegyzése.

2. §. A határeloszlástétel bizonyítása

1. Vezessük be a

$$P_{N,N}^{(k)} = \frac{1}{(2N+1) \binom{2N}{N}} \sum_{n=0}^N \sum_{m=0, n-k}^N \binom{2N-n-m}{N-n} \left[\binom{n+m}{n} - \binom{n+m}{n-k} \right]$$

formulába az indexek következő transzformációját:

$$\begin{aligned} m+n &= t, & m &= \frac{t+s}{2} \\ m-n &= s, & n &= \frac{t-s}{2}. \end{aligned}$$

Ekkor a következő alakhoz jutunk:

$$\begin{aligned} P_{N,N}^{(k)} &= \frac{1}{(2N+1) \binom{2N}{N}} \sum_{s=-k}^{-1} \sum_{t=-s}^{2N+s} \binom{2N-t}{N-\frac{t}{2}-\frac{s}{2}} \left[\binom{t}{\frac{t-s}{2}} - \binom{t}{\frac{t-s}{2}-k} \right] + \\ (2.1) \quad &+ \frac{1}{(2N+1) \binom{2N}{N}} \sum_{s=0}^N \sum_{t=s}^{2N-s} \binom{2N-t}{N-\frac{t}{2}-\frac{s}{2}} \left[\binom{t}{\frac{t-s}{2}} - \binom{t}{\frac{t-s}{2}-k} \right], \end{aligned}$$

ahol a Σ^* csupán a $t = -s, -s+2, \dots$, ill. $t = s, s+2, \dots$ indexekre vonatkozó összegezést jelent.

2. Legyen most $k \sim r\sqrt{2N}$, $t \sim 2Nz$, $s \sim u\sqrt{2N}$. Ekkor

$$k \sim \frac{r}{\sqrt{z}} \sqrt{t}, \quad \frac{s}{2} + k \sim \left(r + \frac{u}{2}\right) \frac{1}{\sqrt{z}} \sqrt{t}, \quad \frac{s}{2} \sim \frac{u}{2\sqrt{1-z}} \sqrt{2N-t}$$

és így az ismert

$$\frac{\binom{2M}{M-l}}{\binom{2M}{M}} \rightarrow e^{-2v^2}, \quad \text{ha } M \rightarrow \infty \text{ és } l \sim v\sqrt{2M},$$

valamint

$$\binom{2M}{M} \sim \frac{2^{2M+\frac{1}{2}}}{\sqrt{2M\pi}}$$

relációk felhasználásával a következőket írhatjuk fel:

$$(2.2) \quad \frac{\binom{2N-t}{N-\frac{t}{2}-\frac{s}{2}}}{\binom{2N-t}{N-\frac{t}{2}}} \rightarrow e^{-\frac{u^2}{2(1-z)}},$$

$$(2.3) \quad \frac{\binom{t}{\frac{t}{2}-\frac{s}{2}}}{\binom{t}{\frac{t}{2}}} \rightarrow e^{-\frac{u^2}{2z}},$$

$$(2.4) \quad \frac{\binom{t}{\frac{t}{2}-\frac{s}{2}-k}}{\binom{t}{\frac{t}{2}}} \rightarrow e^{-\frac{z}{2}\left(r+\frac{u}{2}\right)^2},$$

$$(2.5) \quad \frac{1}{2N+1} \frac{\binom{2N-t}{N-\frac{t}{2}} \binom{t}{\frac{t}{2}}}{\binom{2N}{N}} \sim \frac{1}{\sqrt{2\pi z(1-z)}} \frac{2}{(2N)^{3/2}}.$$

Ekkor (2.2), (2.3) és (2.5)-ből

$$\frac{1}{(2N+1) \binom{2N}{N}} \binom{2N-t}{N-\frac{t}{2}-\frac{s}{2}} \binom{t}{\frac{t-s}{2}} \sim \frac{1}{\sqrt{2\pi z(1-z)}} e^{-\frac{u^2}{2z(1-z)}} \cdot \frac{2}{(2N)^{3/2}},$$

továbbá (2.2), (2.4) és (2.5)-ből

$$\frac{1}{(2N+1) \binom{2N}{N}} \binom{2N-t}{N-\frac{t}{2}} \binom{t}{\frac{t-s}{2}-k} \sim \frac{e^{-2r^2}}{\sqrt{2\pi z(1-z)}} e^{-\frac{[u-2r(1-z)]^2}{2z(1-z)}} \cdot \frac{2}{(2N)^{3/2}}.$$

Figyelembe véve a $\Delta u = \frac{1}{\sqrt{2N}}$, $\Delta z = \frac{1}{N} (\Delta t = 2(!))$ relációkat, továbbá u és z -re vonatkozólag a (2.1) összegezésből adódó határokat — aminek eredményeként a két összeget határértékben összevonhatjuk — a $P_{N,N}^{(k)}$ fenti kifejezése határértékben a következő kétszeres integrálba megy át:

$$(2.6) \quad \frac{1}{\sqrt{2\pi}} \int_0^1 \int_{-r}^{\infty} \frac{1}{\sqrt{z(1-z)}} e^{-\frac{u^2}{2z(1-z)}} dz du - \\ - \frac{1}{\sqrt{2\pi}} e^{-2r^2} \int_0^1 \int_{-r}^{\infty} \frac{1}{\sqrt{z(1-z)}} e^{-\frac{[u+2r(1-z)]^2}{2z(1-z)}} dz du.$$

Figyelembe véve az

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-a}^{\infty} e^{-\frac{v^2}{2\sigma^2}} dv = \Phi(a)$$

összefüggést, (2.6) integrálunk egyszerű transzformációval a 2. tételben adott (5) alakra hozható, amivel határeloszlástételünk egyoldali esetét bebizonyítottuk.

Az abszolút eltérésre vonatkozólag a (2.2), (2.4) és (2.5) relációkat kell alkalmazni, de k helyett jk -t írva, minden rögzített j -re. Így módon, ha az összegezés sorrendjét felcseréljük a j -re vonatkozó összeg általános tagja határértékben (2.6)-nak megfelelően, és a határok figyelembevételével

$$\frac{1}{\sqrt{2\pi}} \int_0^1 \int_{-r}^{+r} \frac{e^{-2j^2r^2}}{\sqrt{z(1-z)}} e^{-\frac{[u-2jr(1-z)]^2}{2z(1-z)}} dz du,$$

ahonnan (2.7) figyelembevételével a 2. tétel (6) formuláját nyerjük.

IRODALOM

- [1] BLACKMAN, J.: Correction to "An extension of the Kolmogorov distribution". *Ann. of Math. Statistics* 29 (1958), 318—324.
- [2] Гнеденко, Б. В.—Королюк, В. С.: О максимальном расхождении двух эмпирических распределений. Доклады Академии Наук СССР 80 (1951), 525—528.
- [3] HODGES, J. L.: The significance probability of the Smirnov two-sample test. *Arkiv för Matematik* 3 (1958), 469—486.
- [4] JORDAN K.: *Fejezetek a klasszikus valószínűségyszámításból*. Akadémiai Kiadó, Budapest, 1956. id. hely: 408. o.

- [5] KIEFER, J.—WOLFOWITZ, J.: On the deviations of the empiric distribution function of vector chance variables. *Transactions of the Amer. Math. Soc.* **87** (1958), 173—186.
- [6] Корольюк, В. С.: Асимптотический анализ распределений максимальных уклонений в схеме Бернулли. Теория вероятностей и ее применения. **4** (1959), 369—397.
- [7] Озолс, В.: О векторандах и непараметрическом критерии согласия для двух конечных выборок. *Latvijas PSR Zinatnu Akademijas Vestis* **8** (1956), 153—158.
- [8] REIMANN, J.—VINCZE, I.: On the comparision of two samples with slightly different sizes. *A MTA Mat. Kut. Int. Közleményei* (sajtó alatt).
- [9] SCHMID, P.: On the Kolmogorov and Smirnov limit theorems for discontinuous functions. *Ann. Math. Statistics* **29** (1958), 1011—1027.

(Beérkezett: 1960. V. 27.)

*A Magyar Tudományos Akadémia
Matematikai Kutató Intézete*