A KOLMOGOROV–SZMIRNOV ÉS MÁS NEMPARAMÉTERES PRÓBÁK ERŐFÜGGVÉNYÉRŐL[1]

Írta: VINCZE ISTVÁN

Bevezetés

Próbák hatásosságának, különböző ellenhipotézisekkel szemben mutatott viselkedésének vizsgálata, ugyanezen tulajdonságoknak két próbára vonatkozó összehasonlítása az erőfüggvény segítségével történhet. Míg számos paraméteres próba erőfüggvénye egyszerűen előállítható, addig nemparaméteres próbák esetén ez az előállítás sok esetben nem egyszerű. Az alábbiakban explicite megadjuk a kétmintás *Szmirnov*-próba erőfüggvényét egyenlő mintadarabszámok esetén, vagyis a *Gnyegyenko—Koroljuk* esetben éspedig, ha nullhipotézisként mindkét változóra a [0, 1] intervallumban az egyenletes eloszlást választjuk, míg ellenhipotézisként ugyanezen intervallumban az egyik eloszlásfüggvény az egyenletes, a másik pedig szakaszonként lineáris, folytonos[2] eloszlásfüggvény.

A próbák tulajdonságainak vizsgálata szempontjából a [0, 1] intervallumra való szorítkozás nem lényeges. Ugyanakkor meggondolásainkból kitűnik, hogy módszerünk alkalmazható más nemparaméteres próbák esetén is és az erőfüggvény explicit, számolásra alkalmas felírhatóságának feltétele lényegében az, hogy a próbánál alkalmazott statisztika eloszlása nullhipotézisre explicite felírható legyen. Esetünkben a *Gnyegyenko—Koroljuk* eloszlások, illetve azok bizonyos módosított alakjai kerülnek alkalmazásra, amely eloszlásokat a szerző REIMANN JÓZSEFfel közös cikkében [6] használt.

A kétmintás *Szmirnov*-próba erőfüggvényének explicit előállítását HOEFFDING [3] adta meg azáltal, hogy a két minta elemei minden lehető sorrendjének valószínűségét felírta adott folytonos ellenhipotézisre. E valószínűségeket kell összegezni a kritikus tartományra, amely előállítás azonban igen bonyolult, aszimptotikus vizsgálatra nem alkalmas. LEHMANN [4] alkalmazta e formulát egyszerű ellenhipotézisekre és kisszámú mintadarabokra ($m=n=4,6$) kiszámította az erőfüggvényt.

Mind az egymintás *Kolmogorov*-próba esetén, mind a kétmintás *Kolmogorov—Szmirnov* esetben MASSEY [5] a következőképpen járt el: Tekintette a változónak olyan értékét, amelyre a nullhipotézist és az alternatívát adó két eloszlásfüggvény (illetve kétmintás esetben az alternatívát adó két eloszlásfüggvény) eltérése maximális. Annak valószínűsége már most, hogy az empirikus és az elméleti eloszlásfüggvény, illetve a két empirikus eloszlásfüggvény eltérése *e kiválasztott helyen* nagyobb, mint egy adott érték, kisebb, minthogy a megfelelő szuprémum, tehát a különbség *valahol* nagyobb, mint az illető érték. Ez utóbbi esemény azonban — az állandó kellő megválasztásával — a kritikus tartományt adja és ily módon

[1] A szerző e cikkben foglalt eredményeit először az *Oberwolfach*ban 1964 aug. 2—7-ig tartott *valószínűségszámítás és matematikai statisztika* tárgyú kollokviumon adta elő.

[2] Megjegyezzük, hogy ami tételünk bizonyításánál is kitűnik, a folytonosság nem lényeges kikötés, véges sok szakadáspont megengedhető volna.

az erőfüggvényt alulról becsültük egyetlen pontbeli esemény valószínűségével, amely valószínűség binomiális kifejezések összege. E szellemes módszert azóta is sokszor alkalmazták. Legutóbb J. ROSENBLATT [7] mutatott rá, hogy MASSEY, aki a binomiális eloszlást normálissal közelítette, jelentősen eltért a pontos értéktől és — igen bonyolult iterációval — javította is a módszert.

Lényegében e két módszerben fellelhető gondolatokat alkalmazták az erőfüggvény vizsgálatánál más szerzők is, mint Z. W. BIRNBAUM [1], I. R. SAVAGE [8], D. G. CHAPMAN [2]. E szerzők vizsgáltak szakaszonként lineáris alternativákat is elsősorban bizonyos, az alternatívák terében értelmezett távolságokhoz tartozó extrém helyzeteknek megfelelő maximum és minimum alternatívákat. Érdekessége van azonban más típusú, szakaszonként lineáris függvényekből előállítható alternatívák vizsgálatának is és erre irányulnak szerző jelen és további vizsgálatai.

A következőkben az általános esetre érvényes formula felírása után, „igen közeli" alternatívákat vizsgálunk, majd bizonyos alternatívákra — kis minták esetén — első- és másodfajú hibákat adunk meg. Tekintettel arra, hogy a 20—50-es mintadarabszám a gyakorlatban eléggé sokszor kerül alkalmazásra, e számított értékek tájékoztatást nyújtanak, hogy milyen ellenhipotézisekre van értelme ilyen nagyságú minták esetén a próbát egyáltalán alkalmazni. Szerző e kérdések további vizsgálatára vissza kíván térni.

1. §. Az erőfüggvény a Gnyegyenko—Koroljuk esetben lineáris szakaszokból álló ellenhipotézis esetén

Legyenek ξ és η valószínűségi változók $F(x)$, ill. $G(x)$ folytonos eloszlásfüggvényekkel és vizsgáljuk a

$$H_0: G(x) \equiv F(x) = \begin{cases} 0, & \text{ha } x \leqq 0 \\ x, & \text{ha } 0 \leqq x \leqq 1 \\ 1, & \text{ha } 1 \leqq x \end{cases}$$

nullhipotézist a

$$H_1^{(r)}: \begin{cases} F(x) \text{ a } H_0 \text{ alatti egyenletes eloszlás} \\ G'_{(x)} = g(x) = \begin{cases} 0, & \text{ha } x \leqq 0 \\ g_i, & \text{ha } z_{i-1} \leqq x < z_i. \qquad i = 1, 2, \dots r \\ 0, & \text{ha } 1 \leqq x \end{cases} \end{cases}$$

alternatívával szemben, ahol $0 = z_0 < z_1 < z_2 < \dots < z_r = 1$. A g_i $(i = 1, 2, \dots, r)$ értékekre természetesen az

$$\int_0^1 g(x)\,dx = \sum_{i=1}^{r} g_i(z_i - z_{i-1}) = 1$$

összefüggés teljesül.

Tekintsük ξ-re és η-ra a $\xi_1, \xi_2, \dots, \xi_n$, ill. $\eta_1, \eta_2, \dots, \eta_n$ független elemekből álló két mintát, jelöljék az ezekhez tartozó empirikus eloszlásfüggvényeket $F_n(x)$, ill. $G_n(x)$. A statisztikai probléma eldöntésére a *Szmirnov*-féle

$$D_{n,n} = \frac{1}{n} \max |F_n(x) - G_n(x)|$$

próbafüggvényt választjuk.

Ha az alkalmazott próba terjedelme α $(0<\alpha<1)$ és $\mathbf{P}(D_{n,n}<d_\alpha|H_0) = 1-\alpha$, akkor a kritikus tartományt ama minták összessége adja, amelyekre a

$$\{D_{n,n}\geqq d_\alpha\}$$

esemény teljesül.

A próba erőfüggvénye,

$$K_n(H_1, \alpha) = \mathbf{P}(D_{n,n}\geqq d_\alpha|H_1).$$

Tételünk megfogalmazásához még egy további esemény valószínűségére vezetünk be jelölést. Tekintsük a két minta rendezett alakját:

$$\xi_1^*<\xi_2^*< \dots <\xi_n^*,$$

$$\eta_1^*<\eta_2^*< \dots <\eta_n^*$$

és ezeknek egyetlen rendezett sorozatba való egyesítését:

$$\zeta_1^*<\zeta_2^*<\dots<\zeta_{2n}^*.$$

Vezessük be a ϑ_i $(i=1, 2, \dots, 2n)$ valószínűségi változókat a szokásos módon:

$$\vartheta_i = \begin{cases} +1, & \text{ha} \quad \zeta_i^* = \xi_j^* \\ -1, & \text{ha} \quad \zeta_i^* = \eta_l^* \end{cases}$$

teljesül valamely j, illetve l indexre.

Ha most s_j jelöli a ϑ_i sorozat j-edik részletösszegét, vagyis $s_0=0$,

$$s_j = \vartheta_1+\vartheta_2+\dots+\vartheta_j, \qquad j=1, 2, \dots, 2n,$$

$s_{2n}=0$, akkor mint ismeretes és egyszerűen belátható:

$$D_{n,n} = \frac{\max\limits_{(i)} |s_i|}{n}.$$

Ugyancsak jól ismert és egyszerű tény, hogy H_0 fennállása esetén a $(\vartheta_1, \vartheta_2, \dots, \vartheta_{2n})$ valószínűségi változókból álló sorozatban az n számú $+1$ és n számú -1 minden lehetséges sorrendje egyenlően valószínű, vagyis minden sorrend valószínűsége $\binom{2n}{n}^{-1}$. Tekintsük most a sík (i, s_i) koordinátájú pontjait $i=0, 1, \dots, 2n$-re, vagyis a két mintához rendelt bolyongás egy útját. Rögzítsük ezen út $0 = i_0\leqq i_1\leqq\dots$ $\dots\leqq i_r=2n$ abszcisszájú pontjaihoz tartozó ordinátákat az $s_{i_0}=0$, $s_{i_1}=a_{i_1}, \dots, s_{i_j}=$ $=a_{i_j}, \dots, s_{i_r}=0$ értékekben, ahol feltesszük, hogy a megadott értékek egy lehetséges úthoz tartoznak, vagyis a

$$(1.1) \qquad -(i_m-i_{m-1}) \leqq a_{i_m}-a_{i_{m-1}} \leqq i_m-i_{m-1} \qquad m=1, 2, \dots, r$$

feltételek teljesülnek. Legyen végül

$$\mathbf{p}_n(k; a_{i_{m-1}}, a_{i_m}) = \mathbf{P}\big(\max_{i_{m-1}<j<i_m} |s_j|<k\,|s_{i_{m-1}} = a_{i_{m-1}}, s_{i_m} = a_{i_m}\big),$$

ahol e valószínűségek a H_0 hipotézis mellett értendők. A következő tételünkben az (i_m-i_{m-1}) szám azt adja meg, hogy összesen hány mintaelem esik a (z_{m-1}, z_m)

intervallumba, amit $\nu_m+\mu_m$ fog jelölni, ti. ν_m a ξ_i-k, μ_m az η_j-k számát fogja adni a megjelölt intervallumban; vagyis $i_m=\sum_{j=1}^{m}(\mu_j+\nu_j)$. A $p_n(k; a_{i_{m-1}}, a_{i_m})$ valószínűség értéke természetesen 0, ha $|a_{i_{m-1}}|$ vagy $|a_{i_m}|$ valamelyike $\geqq k$.

Bevezetett jelöléseinkkel érvényes a következő

TÉTEL: *Az α terjedelmű kétmintás Kolmogorov—Szmirnov-próba ereje a fent definiált $H_1^{(r)}$ alternatíva esetén,*

$$K_n(H_1^{(r)}, \alpha) = 1-(n!)^2 \sum_{\{\nu\}}{}^{*} \sum_{\{\mu\}}{}^{*} \prod_{m=1}^{r} \frac{g_m^{\mu_m}(z_m-z_{m-1})^{\nu_m+\mu_m}}{\nu_m!\,\mu_m!}\,\cdot$$

$$\cdot p_n\left(d_\alpha;\ s_{i_{m-1}}=\sum_{j=1}^{m-1}(\nu_j-\mu_j),\ s_{i_m}=\sum_{j=1}^{m}(\nu_j-\mu_j)\right),$$

ahol Σ^* r-szeres összegezést jelent minden olyan $(\nu_1, \nu_2, \ldots, \nu_r)$, ill. $(\mu_1, \mu_2, \ldots, \mu_r)$ értékrendszerre, amelyre a $0\leqq\nu_m, \mu_m\leqq n$, $m=1, 2, \ldots, r$, továbbá a $\sum_{m=1}^{r}\mu_m=\sum_{m=1}^{r}\nu_m=n$ feltétel teljesül.

Megjegyzés: E formula egyszerű szerkezetű, nem túl nagy r esetén számolásra alkalmas és kezelhetőbb, mint a többszörös integrálokon alapuló kifejezések; természetesen meghatározásának feltétele a H_0 érvényességét feltételező p_n valószínűségek ismerete. A bizonyításban, mint látni fogjuk, nem használjuk ki lényegesen, hogy éppen a *Kolmogorov—Szmirnov*-próbáról van szó, vagyis más rendezett mintás próbára is hasonló jellegű az erőfüggvény. A p_n valószínűségeket, melyek egyszerűen meghatározhatók, itt nem adjuk meg, csak a 2. §-ban tárgyalt speciális esetekben.

A bizonyításhoz néhány jól ismert egyszerű lemmát használunk fel. Nem szorul bizonyításra a következő

1. 1. SEGÉDTÉTEL: *Legyen a ξ folytonos valószínűségi változó sűrűségfüggvénye állandó az (a, b) $(0\leqq a<b\leqq 1)$ intervallumban. Ekkor ama feltétel mellett, hogy ξ az (a, b) intervallumba esik, ott egyenletes eloszlású.*

E segédtétel egyszerű következménye az

1. 2. SEGÉDTÉTEL: *Legyenek a ξ és η folytonos valószínűségi változók sűrűségfüggvényei a $(0, 1)$ intervallum (a, b) részintervallumában (nem szükségképpen azonos értékű) állandók. Essenek a ξ-re vonatkozó $\xi^{(1)}, \xi^{(2)}, \ldots, \xi^{(\nu)}$, valamint az η-ra vonatkozó $\eta^{(1)}, \eta^{(2)}, \ldots, \eta^{(\mu)}$ független mintaelemek az (a, b) intervallumba. E feltétel mellett az (a, b) intervallumba eső ezen $\nu+\mu$ számú mintaelem minden sorrendje egyenlő valószínűségű.*

Tekintsük most a tételben szereplő ξ-re és η-ra vonatkozó egyenként n elemű mintákat és jelöljük azt az eseményt, hogy az első minta elemei közül ν_m számú esik a (z_{m-1}, z_m), $(m=1, 2, \ldots, r)$ intervallumba $\{\nu\}$-vel; hasonló jelentést tulajdonítunk a második mintára vonatkozóan a $\{\mu\}$ jelölésnek. Ekkor érvényes a következő reláció:

$$\mathbf{P}(D_{n,n}<d_\alpha\,|H_1) = \sum_{(\nu)}{}^{*} \sum_{(\mu)}{}^{*} \mathbf{P}(D_{n,n}<d_\alpha|\,\{\nu\}, \{\mu\}, H_l)\,\mathbf{P}(\{\nu\}, \{\mu\}, H_1).$$

A ν_m-ek együttes eloszlása polinomiális, ugyanez áll a μ_m-ekre is, tehát

$$\mathbf{P}(\nu_1 = n_1, \nu_2 = n_2, \dots \nu_r = n_r | H_1) = n! \prod_{m=1}^{r} \frac{(z_m - z_{m-1})^{n_m}}{n_m!},$$

tekintettel arra, hogy a ξ_j-k $(z_m - z_{m-1})$ valószínűséggel esnek a (z_{m-1}, z_m) intervallumba; ugyanez a valószínűség az η_l-ekre $(z_m - z_{m-1})g_m$, és így

$$\mathbf{P}(\mu_1 = n_1, \mu_2 = n_2, \dots \mu_r = n_r | H_1) = n! \prod_{m=1}^{r} \frac{[(z_m - z_{m-1})g_m]^{n_m}}{n_m!}.$$

Továbbá a $\{\nu\}$ és $\{\mu\}$ események függetlensége miatt

$$\mathbf{P}(\{\nu\}, \{\mu\} | H_1) = \mathbf{P}(\{\nu\} | H_1), \mathbf{P}(\{\mu\} | H_1).$$

Határozzuk most meg a $\mathbf{P}(D_{n,n} < d_\alpha | \{\nu\}, \{\mu\}, H_1)$ valószínűségeket. Tekintsük evégből a

$$D_{n,n}(x) = F_n(x) - G_n(x), \qquad 0 \leqq x \leqq 1$$

folyamatot. Nyilvánvaló a

$$\{D_{n,n} < d_\alpha, \{\nu\}, \{\mu\}\} = \bigcap_{m=1}^{r} \{|D_{n,n}(x)| < d_\alpha, z_{m-1} \leqq x < z_m, \{\nu\}, \{\mu\}\}$$

összefüggés. *A jobb oldalon álló események rögzített $\{\nu\}$, $\{\mu\}$ értékrendszerek mellett függetlenek.* Ezt beláthatjuk a bolyongási modell segítségével. Ugyanis a (0, 0)-ból induló és $2n$ lépés után a $(2n, 0)$-ba visszatérő bolyongás r részre tagozódik, minden rész elején és végén a lépésszám, valamint a bolyongó pont helyzete adott; éspedig az m-edik $(m = 1, 2, \dots, r)$ intervallum utolsó mintaelemének a $\sum_{j=1}^{m} (\nu_j + \mu_j)$ lépésszám és a $\sum_{j=1}^{m} (\nu_j - \mu_j)$ helyzet (origótól való eltávolodás) felel meg. Márpedig 1. 2 segédtételünk szerint két ilyen lépés között minden lehetséges út (sorrend) egyenlő — $\binom{\nu_m + \mu_m}{\nu_m}^{-1}$ — valószínűségű, függetlenül az előző és következő intervallumban fellépő eseményektől. Eszerint felírhatjuk, hogy

$$\mathbf{P}(D_{n,n} < d_\alpha | \{\nu\}, \{\mu\}) = \prod_{m=1}^{r} \mathbf{P}\{|D_{n,n}(x)| < d_\alpha, z_{m-1} \leqq x < z_m, \{\nu\}, \{\mu\}\}.$$

Megjegyzésünk szerint azonban ez a valószínűség megegyezik az

$$\prod_{m=1}^{r} p_n\Big(d_\alpha, s_{i_{m-1}} = \sum_{j=1}^{m-1} (\nu_j - \mu_j), \ s_{i_m} = \sum_{j=1}^{m} (\nu_j - \mu_j)\Big)$$

kifejezéssel és így tételünket bebizonyítottuk.

2. §. Két lineáris szakaszból álló alternatívák vizsgálata

1. Legyen $r=2$ és $z_0=0, z_1=z, z_2=1$ $(0<z<1)$, továbbá $g_1=g$ $(0\leqq g<1)$, amiből $g_2=\frac{1-gz}{1-z}$ adódik; ebben az esetben a mintaelemek megoszlására a két intervallumban a $\nu_1=\nu$, $\nu_2=n-\nu$, $\mu_1=\mu$, $\mu_2=n-\mu$ jelöléseket alkalmazhatjuk. Minthogy most $G(x)<F(x)$, egyoldali ellenhipotézisről van szó, ennélfogva egyoldali próbát alkalmazhatunk, amikor is a $d_\alpha=\frac{k}{n}$ jelöléssel a

$$p_n(k;\, s_{i_0}=0,\, s_{i_1}=\nu-\mu)=1-\frac{\binom{\nu+\mu}{\nu-k}}{\binom{\nu+\mu}{\nu}},$$

$$p_n(k;\, s_{i_1}=\nu-\mu,\, s_{i_2}=0)=1-\frac{\binom{2n-\nu-\mu}{n-\nu+k}}{\binom{2n-\nu-\mu}{n-\nu}}$$

kifejezések adják a megfelelő — k magasságot el nem érő — utak valószínűségét az első, ill. második intervallumban a $\{\nu, n-\nu\}$, $\{\mu, n-\mu\}$ feltételek mellett. Ha a

$$C_{\nu,\mu}^{n,k}=\left(1-\frac{\binom{\nu+\mu}{\nu-k}}{\binom{\nu+\mu}{\nu}}\right)\left(1-\frac{\binom{2n-\nu-\mu}{n-\nu+k}}{\binom{2n-\nu-\mu}{n-\nu}}\right)$$

jelölést bevezetjük, akkor az erőfüggvény a következő:

$$1-\sum_{\nu}{}^{*}\sum_{\mu}{}^{*}\binom{n}{\nu}z^{\nu}(1-z)^{n-\nu}\binom{n}{\mu}(gz)^{\mu}(1-gz)^{n-\mu}C_{\nu,\mu}^{n,k}, \tag{2.1}$$

ahol az összegezés a $0\leqq\nu\leqq n$, $0\leqq\mu\leqq n$, $\nu-\mu<k$ feltételeknek eleget tevő ν, μ indexekre veendő.

E formulából máris érdekes következtetést vonhatunk le: ha $n\to\infty$, akkor egyrészt $k\sim y\sqrt{2n}$, másrészt $\nu\sim nz+\lambda\sqrt{nz(1-z)}$, $\mu\sim nzg+\lambda'\sqrt{ngz(1-gz)}$, vagyis a

$$\nu-\mu\sim nz(1-g)+\lambda\sqrt{nz(1-z)}-\lambda'\sqrt{nzg(1-zg)}<y\sqrt{2n} \tag{2.2}$$

követelmény csak végtelen felé tartó λ-kra teljesíthető, amikor is a megfelelő valószínűségek (valószínűség-sűrűségek) már összességükben 0-hoz tartanak. Ily módon ν az n nagyságrendet el sem érheti és így a binomiális valószínűségek nem közelíthetők — ν esetében — normális eloszlással. Ezt a megjegyzést, mint említettük — MASSEY eljárásával szemben — már J. ROSENBLATT kifejezésre juttatta.

Ha a $\max\big(F(x)-G(x)\big) = \Delta$ jelölést vezetjük be, akkor $\Delta = z-zg = z(1-g)$ és az erőfüggvény a következő alakban írható:

$$(2.3)\qquad 1-\sum_{\nu}{}^{*}\sum_{\mu}{}^{*}\binom{n}{\nu}\binom{n}{\mu}C_{\nu,\mu}^{n,k}z^{\nu+\mu}(1-z)^{2n-\nu-\mu}\left(1-\frac{\Delta}{z}\right)^{\mu}\left(1+\frac{\Delta}{1-z}\right)^{n-\mu}.$$

Minthogy $\Delta=0$-ra $F(x)\equiv G(x)$, ennélfogva z-ben a következő azonosságot kapjuk:

$$1-\sum_{\substack{\nu\\ \nu-\mu<k}}{}^{*}\sum_{\mu}{}^{*}\binom{n}{\nu}\binom{n}{\mu}C_{\nu,\mu}^{(n,k)}z^{\nu+\mu}(1-z)^{2n-\nu-\mu}\equiv 1-\frac{\binom{2n}{n-k}}{\binom{2n}{n}}.$$

2. Tekintsünk most igen közeli alternatívát, pontosabban legyen $\Delta\sim c/n$. Minthogy ebben az esetben (2. 2)-ből

$$\nu-\mu\sim c+\lambda\sqrt{nz(1-z)}-\lambda'\sqrt{nz(1-z)\left(1-\frac{c}{nz}\right)\left(1+\frac{c'}{n(1-z)}\right)},$$

a binomiális valószínűségek kivételesen normális eloszlással közelíthetők. Ahelyett, hogy e kissé bonyolult, de egyszerűen keresztülvihető számításba bocsátkoznánk, megjegyezzük, hogy a ν-re és μ-re (2. 2)-ben alkalmazott közelítéssel az

$$\left(1-\frac{\Delta}{z}\right)^{\mu}\left(1+\frac{\Delta}{1-z}\right)^{n-\mu}\sim 1$$

reláció adódik és így az erőfüggvény értéke $n\to\infty$ esetén $-c$-től függetlenül, α-hoz tart. A *Kolmogorov—Szmirnov*-próba tehát ebben az esetben is torzítatlan, ami egyoldali próbánál általánosabb tételből is adódik (l. CHAPMAN [2]).

3. Legyen most $g=0$, amikor is $\max\big(F(x)-G(x)\big) = F(z)-G(z) = z$. Ebben az esetben $\mathbf{P}(\mu=0|H_1) = 1$ és így az erőfüggvény a következő egyszerű alakú

$$(2.4)\qquad 1-\sum_{\nu=0}^{k-1}\binom{n}{\nu}z^{\nu}(1-z)^{n-\nu}\left(1-\frac{\binom{2n-\nu}{n-\nu+k}}{\binom{2n-\nu}{n-\nu}}\right).$$

A $z=\Delta\sim c/n$ közeli alternativa ismét az előbbi eredményre vezet. Ekkor ugyanis a binomiális eloszlás *Poisson*-közelítését alkalmazhatjuk:

$$\binom{n}{\nu}z^{\nu}(1-z)^{n-\nu}\sim\frac{c^{\nu}}{\nu!}e^{-c},$$

továbbá minden véges ν-re és $k\sim y\sqrt{2n}$ esetén

$$\frac{\binom{2n-\nu}{n-\nu+k}}{\binom{2n-\nu}{n-\nu}}\sim e^{-2y^2}.$$

Ezekből az erőfüggvény határértéke $n \to \infty$ esetén

$$1-\left(1-e^{-2y^2}\right)\sum_{\nu=0}^{\infty}\frac{c^\nu}{\nu!}e^{-c}=e^{-2y^2}=\alpha.$$

4. Térjünk a $g=0$, tehát $\Delta=z$ esetben numerikus értékekre. Ez az alternatíva a Δ-hoz tartozó maximum és minimum alternatívák között helyezkedik el és így tájékoztatást kapunk egy „közbeeső" alternatíva esetében a *Szmirnov*-próba erejéről. Az erőfüggvény (2. 4) kifejezése az

$$\binom{n}{\nu}\frac{\binom{2n-\nu}{n-\nu+k}}{\binom{2n-\nu}{n-\nu}}=\frac{\binom{2n}{n-k}}{\binom{2n}{n}}\binom{n+k}{\nu}=\alpha\binom{n+k}{\nu}$$

összefüggés alapján a következőképpen írható:

$$1-\sum_{\nu=0}^{k-1}\binom{n}{\nu}z^\nu(1-z)^{n-\nu}+\frac{\alpha}{(1-z)^k}\sum_{\nu=0}^{k-1}\binom{n+k}{\nu}z^\nu\cdot(1-z)^{n+k-\nu}= \tag{4. 1}$$

$$=1-B(1-z;\, n-k+1, k)+\frac{\alpha}{(1-z)^k}B(1-z;\, n+1,\, k),$$

ahol

$$B(z;\, p, q)=\frac{\int\limits_0^z t^{p-1}(1-t)^{q-1}\,dt}{\int\limits_0^1 t^{p-1}(1-t)^{q-1}\,dt}$$

a nem teljes Beta-függvényt jelenti. A (4. 1) formulával a számolás könnyűszerrel keresztülvihető a binomiális együtthatók és a nem teljes Beta-függvény táblázatának használatával. Tekintettel az $n \sim 20-50$ nem nagy értékeire az elsőfajú hiba nem volt eleve tetszőlegesen megadható és oly módon választottuk, hogy értéke a 10% közelében legyen; ehhez számítottuk azután a $\Delta=z=0{,}05$; $0{,}1$; $0{,}2$; $0{,}3$ távolságoknak megfelelő másodfajú hibákat. Az eredményeket az alábbi táblázat tünteti fel.

A másodfajú hiba nagysága, ha $\Delta=z$ (n = 50-re extrapoláció útján nyert értékek)

n	k	α	0,05	0,1	0,2	0,3
20	6	0,1684		0,6842	0,4332	0,1837
20	7	0,0873		0,8177	0,6163	0,3363
25	7	0,1428		0,7026	0,4154	0,1446
30	8	0,1197		0,7230	0,4039	0,1164
35	9	0,0996		0,7438	0,3961	0,0950
40	9	0,1331		0,6587	0,2640	0,0377
45	10	0,1091	0,8178	0,6890	0,2670	0,0319
50	11	0,0893	0,8430	0,7190	0,2700	0,0265

IRODALOM

[1] Birnbaum, Z. W.: On the power of a one-sided test of fit for continuous distribution functions, *Ann. Math. Stat.*, **24** (1953) 284—289.
[2] Chapman, D. G.: A comparative study of several one- sided goodness-of-fit tests, *Ann. Math. Stat.* **29** (1958) 655—674.
[3] Hoeffding, W.: Optimum nonparametric tests, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Berkeley* (1951) 83—92.
[4] Lehmann, E. L.: The power of rank tests, *Ann. Math. Stat.* **24** (1953) 23—42.
[5] Massey, F. J.: A note on the power of a nonparametric test, *Ann. Math. Stat.* **21** (1950) 440—443.
[6] Reimann J.—Vincze I.: On the comparison of two samples with slightly different sizes, *MTA Mat. Kut. Int. Közleményei* (sajtó alatt).
[7] Rosenblatt, J.: Some modified Kolmogorov—Smirnov tests of approximate hypotheses and their properties, *Ann. Math. Stat.* **33** (1962) 513—524.
[8] Savage, I. R.: Contributions to the theory of rank order statistics — the two sample case, *Ann. Math. Stat.* **27** (1956) 590—615.

(Beérkezett: 1965. II. 25.)