# A SELEJTARÁNY BAYES-FÉLE VALÓSZÍNŰSÉGI HATÁRAIRA VONATKOZÓ DUALITÁSI ELVRŐL

SARKADI KÁROLY

*Összefoglalás*

A cikk *Steinhaus* [6] és *Oderfeld* ([3], [5]) cikkeihez kapcsolódik. Ezek a cikkek kritika tárgyává teszik azt az álláspontot, hogy a Bayes-szabály alkalmazásán alapuló módszert ki kell rekeszteni a matematikai statisztikai módszerek közül. Rámutatnak arra, hogy a Bayes-szabály alkalmazása sok esetben ugyanarra az eredményre vezet, mint a megbízhatóság elvén alapuló módszer. Az *Oderfeld* említett cikkeiben tárgyalt dualitási elv is egy ilyen jellegű megállapítást mond ki, a selejtvizsgálatra vonatkozóan. A szerző itt ezt a dualitási elvet általánosítja a hipergeometrikus, polihipergeometrikus, polinomiális és Poisson-eloszlás esetére.

*Bevezetés*

*J. Oderfeld* több munkájában (lásd pl. [3] és [4]) foglalkozik azzal a kérdéssel, hogy adott mintavételi eljárás eredményéből hogyan állapíthatók meg adott valószínűséggel a tétel selejtarányának határai. Ezeket a határokat retrospektív paramétereknek nevezzük szemben az általánosan használt prospektív paraméterekkel, amelyek ismertnek feltételezett selejtarány esetére adnak tájékoztatást a mintavételi eljárás eredményére nézve. *Oderfeld* [3] munkájában kimutatja, hogy a retrospektív paraméterek meghatározása a priori egyenletes eloszlás feltevése mellett visszavezethető a prospektívekére. *Oderfeld* ezt az eredményt dualitás elvének nevezi.

A dualitás elvének gyakorlati jelentőségét jól megvilágítják *Steinhaus* [6] és *Oderfeld* ([3], [4]). Ismeretes, hogy a matematikai statisztika alkalmazásában az adott megbízhatóságú konfidencia-intervallum *J. Neyman*tól származó módszere (a prospektív módszer) szinte teljesen kiszorította a Bayes-tétel alkalmazásán alapuló (retrospektív) módszert. Ennek az az oka, hogy a retrospektív problémára szabatos választ csak akkor lehet adni, ha az a priori eloszlás ismert. Ezen felül olyan esetekben, amikor a meghatározandó mennyiség nem valószínűségi változó, hanem állandó szám, akkor a priori eloszlásról egyáltalában, elvileg nem lehet beszélni, ilyen esetekben tehát csak *Neyman* módszere alkalmazható. A selejtarány esetében azonban nem ez a helyzet, azt joggal tekinthetjük valószínűségi változónak, viszont az a priori eloszlás általában nem ismeretes. A régebben helyesnek tekintett Bayes-szabály szerint abban az esetben, ha az a priori eloszlás nem ismeretes, a priori egyenletes eloszlással kell dolgoznunk. Ez nyilvánvaló önkényesség és joggal kifogásolható. Ezzel szemben a prospektív kérdésre mindig szabatos választ tudunk adni.

*Steinhaus* és *Oderfeld* említett cikkeikben rámutatnak arra, hogy ennek ellenére a prospektív módszer sem mentes az önkényességtől. A módszer exakt feleletet ad egy bizonyos kérdésre, de az önkényességet azzal követi el, hogy nem az eredetileg feltett kérdésre válaszol. Miután a mintát mindig azért veszik, hogy abból a tételre következtetést vonjanak le, így a mintavétel alkalmából a retrospektív kérdés természetes módon merül fel. Minthogy tehát a prospektív módszer sem mentes az önkényességtől, nem lehet azt egyedül helyes módszernek tekinteni.

A prospektív módszer nem körültekintő alkalmazása éppen úgy okozhat hibát, mint a retrospektív módszeré. Az alternatív mintavételnél a gyakorlati szakemberek, különösen az átvevő, rendszerint a retrospektív kérdést szokták feltenni: ha az $n$ elemű mintában $k$ selejtes darabot találtam, milyen határok között lehet a tétel selejtaránya? A modern módszer szerint erre a kérdésre nem szabad közvetlenül választ adni, hanem meg kell adni azokat a selejtarányhatárokat, amelyeknél a mintavételi eljárás már majdnem biztosan elfogadással, illetve elutasítással végződik. A gyakorlati szakember így kénytelen ezzel a válasszal beérni és nem teljesen megnyugodva ugyan, de többnyire úgy használja fel a megadott selejtarányokat, mintha azokat az eredeti kérdésre kapta volna válaszul.

Az említett dualitási elvből tudjuk, hogy a priori egyenletes eloszlás feltevése mellett — hacsak $n$ nem nagyon kicsi — a retrospektív paraméterek igen közel megegyeznek a prospektívekkel, az egyszeres tervek esetében. A matematikus tehát a retrospektív problémát megkerülő válaszával a gyakorlati szakembert tulajdonképpen az a priori egyenletes eloszlás alkalmazására készteti. Ez bizonyosan helytelen, mert az a priori egyenletes eloszlás általában távolról sem egyezik meg a valósággal.

A dualitási elv gyakorlati jelentősége kettős: egyrészt rámutat erre a hibára, amely a mintavételi gyakorlatban sűrűn előfordul, másrészt lehetővé teszi módszer kidolgozását ennek a hibának a kiküszöbölésére. Ez a módszer — *Oderfeld* a priori béta eloszláson alapuló módszere — nem önkényes a priori eloszlással dolgozik, hanem az elmult átvételek tapasztalatai alapján állapítja meg az a priori eloszlást.

A jelen értekezés 1—5. pontja a dualitási elv kiterjesztésével foglalkozik. Látni fogjuk, hogy az igaz akkor is, ha nem tételezzük fel a minta selejtarányának binomiális eloszlását, hanem a véges tétel esetén valóságban fennálló hipergeometrikus eloszlást vesszük figyelembe. Megmutatjuk továbbá, hogy dualitási elv állítható fel akkor is, ha egyszerű alternatíva helyett több osztályozási lehetőség van, végül akkor is, ha a folytonos eloszlású tételben a hibák száma Poisson-eloszlást követ.*

A dualitási elvnek ez a kiterjesztése elvi jelentőségű, rámutat a tétel matematikai alapjára.

A 6. pont a dualitási elvnek a nem egyenletes a priori eloszlás esetére való alkalmazhatóságának kérdésével foglalkozik. *Oderfeld* [3] kimutatta, hogy a priori béta eloszlás esetére alkalmazható a dualitási elv s megállapítja, hogy az a priori béta eloszlás munkahipotézisül a gyakorlatban általában megfelelő.

* Az ebben a cikkben tárgyalt problémákat *Rényi Alfréd* vetette fel.

Itt látni fogjuk, hogy ha az a priori eloszlás béta-eloszlások keveréke, akkor is fennáll a dualitási elvnek egy — bár kissé bonyolultabb — formája. Ha az összetevő béta-eloszlások száma nem nagy, úgy az összefüggés gyakorlati szempontból kezelhető. Ez azért jelentős, mert előfordulhat, hogy a selejtarány tapasztalati eloszlása béta-eloszlással nem közelíthető meg, de elég jól meg tudjuk közelíteni két, illetve három béta-eloszlás keverékével. (Mint ismeretes, bármely folytonos sűrűségfüggvénnyel rendelkező (0, 1) intervallumbeli valószínűségeloszlás tetszés szerinti mértékben megközelíthető véges sok béta-eloszlás keverékével.) Felhasználhatjuk továbbá az eredményt úgy is, hogy figyelembe vesszük a gyártási folyamat megváltozásának a lehetőségét úgy, hogy felteszünk egy eloszlást arra az esetre, ha nem változik a gyártás, egy eloszlást arra az esetre, ha a gyártás megváltozik, megbecsüljük a két lehetőség valószínűségét, képezzük az ezekkel, mint súlyokkal vett keverék-eloszlást és ezt tekintjük a priori eloszlásnak.

## *1. A dualitás elve*

A következőkben azt a mintavételi *eredményt*, hogy $n$ elemű mintában $k$ darab selejtest találtunk, röviden $(k, n)$ eredménynek fogjuk nevezni. Azt a mintavételi *tervet*, amely úgy intézkedik, hogy $n$ elemű mintát kell venni, és a tételt el kell fogadni, ha a mintában legfeljebb $k$ selejtes darab van, ellenkező esetben pedig el kell utasítani, a $k \| n$ jellel jelöljük.

Legyen $0 < \beta_2 < \beta_1 < 1$ ; $\beta_2$ 0-hoz, $\beta_1$ 1-hez közeleső szám legyen.

Jelöljük $R_i$ $(k, n)$-nel $(i = 1, 2)$ azt a selejtarányt, hogy $\beta_i$ legyen annak a posteriori valószínűsége a $(k, n)$ eredmény alapján, hogy a tétel selejtaránya kisebb $R_i$ $(k, n)$-nél. Ezt a két számot nevezzük a $(k, n)$ eredmény retrospektív paraméterének.

Jelöljük $P_i$ $(k\|n)$-nel $(i = 1, 2)$ azt a selejtarányt, amely mellett a $k$ $n$ mintavételi eljárás a tételt $\beta_i$ valószínűséggel visszautasítja. Ezt a két $P_i$ $(k\|n)$ számot nevezzük a $k\|n$ terv prospektív paramétereinek.

*Oderfeld* kimutatta, hogy a priori egyenletes eloszlás feltevése mellett

$$R_i\,(k, n) = P_i\,(k \,\|\, n + 1) \qquad (i = 1, 2)$$

és így a retrospektív paraméterek meghatározása visszavezethető a prospektívekére. A fenti összefüggés bizonyítása lényegében a

$$\mathbf{P}\,(\zeta < p | \varkappa_n = k) = \mathbf{P}\,(\varkappa_{n+1} \geq k + 1 | \zeta = p) \tag{1}$$

összefüggés igazolásán alapszik. Itt $\zeta$ a selejtarány, $\varkappa_n$ az $n$ elemű mintában talált hibás darabok száma ; $k$ pozitív egész szám, $p$ 1-nél kisebb pozitív szám. Ez az összefüggés már *Castelnuovo* valószínűségszámításról szóló könyvében is szerepel [1].

Az (1) összefüggést *Oderfeld* a Bayes-tétel és az

$$\int_0^p (n + 1) \binom{n}{k} x^k (1 - x)^{n-k}\, dx = \sum_{i=k+1}^{n+1} \binom{n+1}{i} p^i (1 - p)^{n+1-i} \tag{2}$$

azonosság felhasználásával mutatja ki. Az utóbbi azonosságot analitikus úton, integrál maradéktagos Taylor-sorral igazolja, *Kendall* nyomán. Megjegyezzük, hogy mint ismeretes, a (2) valószínűségszámítási úton is igazolható. Így tulajdonképpen a (2) azonosságra nincs is szükség, csak a baloldalon lévő kifejezés valószínűségi értelmezésére.

Mint látható, *Oderfeld* feltételezi, hogy a minta darabszáma elhanyagolható kicsiny a tétel darabszámához képest, és így a selejtdarabok számát a mintában binomiális eloszlásúnak tekinthetjük.

A következőkben tehát az (1) összefüggés általánosításával foglalkozunk.

## 2. *Hipergeometrikus eloszlás*

Ha a tétel véges, elemeinek száma $N$, akkor, mint azt az alább bebizonyítjuk, az (1) helyett az alábbi pontosabb összefüggés van érvényben:

$$\mathbf{P}\left(\zeta \leq \frac{M}{N} \,\middle|\, \varkappa_{nN} = k\right) = \mathbf{P}\left(\varkappa_{n+1,N+1} \geq k+1 \,\middle|\, \zeta = \frac{M+1}{N+1}\right). \tag{3}$$

Itt $\varkappa$ második indexe a tétel darabszámát jelzi, amelyből a mintát vettük, $M$ egész számot jelent, egyébként az előző pontban használt jelöléseket alkalmazzuk. Az a priori feltevés természetesen ebben az esetben az, hogy a selejtarány 0-tól 1-ig bármelyik $N$ nevezőjű tört értékét egyforma valószínűséggel veheti fel.

Ez az összefüggés ebben a formában szerepel már *Simon* minőségellenőrzési kézikönyvében is [5], azzal, hogy *Molina* ezt bebizonyította. Az idézett helyen [2] azonban sem a tétel, sem bizonyítása nem található.

A (3) bebizonyítása céljából először gondoljuk el a következő modellt: egy $N+1$ darabból álló tétel darabjai legyenek 1-től $N+1$-ig sorszámmal ellátva. Vegyünk ki ebből egy $n+1$ elemű mintát, s jelentsék $\xi_1, \xi_2, \ldots, \xi_{n+1}$ a minta elemeinek tételbeli sorszámait. Megjegyezzük, hogy minta alatt ebben a cikkben mindig visszatevés nélküli mintát értünk. Tehát a jelen esetben a $\xi$-k között egyenlők nem fordulhatnak elő. Rendezzük a $\xi$-ket nagyság szerint s a rendezett minta elemeit jelöljük $\eta_1, \eta_2, \ldots, \eta_{n+1}$-gyel.

Könnyen belátható, hogy fennáll a következő összefüggés:

$$\mathbf{P}\left(\eta_{k+1} \leq M+1 \mid \eta_{k+1} = \xi_1\right) = \mathbf{P}\left(\eta_{k+1} \leq M+1\right). \tag{4}$$

A $\xi_1$ ugyanis sem az eloszlása szempontjából, sem az $\eta_{k+1}$ definíciója szempontjából nincs kitüntetve a többi $\xi$ közül, helyére bármelyiket írhatnánk, a (4) baloldalán álló valószínűség nem változik, tehát ugyanakkora, mint a feltétel nélküli valószínűség.

Bebizonyítható továbbá, hogy

$$\mathbf{P}\left(\eta_{k+1} \leq M+1\right) = \mathbf{P}\left(\varkappa_{n+1,N+1} \geq k+1 \,\middle|\, \zeta = \frac{M+1}{N+1}\right). \tag{5}$$

Előbbi $(N+1)$ elemű sorszámozott tételünkben nevezzük az első $(M+1)$-et selejtesnek, a többit jónak. Hogy ilyen feltétel mellett a tételből kihúzott $(n+1)$ elemű mintában legalább $k+1$ a selejtes, az ugyanazt jelenti, minthogy a nagyságszerinti $(k+1)$-ik, legfeljebb $(M+1)$-ik sorszámú az eredeti tételben, tehát még selejtes. Az (5) két oldala azonban ennek a két egymással megegyező eseménynek a valószínűségét tünteti fel. Tehát az (5) igaz, mert természetes, hogy az (5) jobboldalának értéke nem függ attól, hogy azt az így elgondolt tételre, vagy egy tetszőleges, a feltételnek megfelelő $N+1$ elemű tételre vonatkoztatjuk.

Végül bebizonyítjuk, hogy

$$\mathbf{P}\,(\xi_1 \leq M+1 \mid \eta_{k+1} = \xi_1) = \mathbf{P}\left(\zeta \leq \frac{M}{N} \,\Big|\, \varkappa_{nN} = k\right). \tag{6}$$

Hogy a (6)-ot belássuk, tekintsünk ismét egy $N+1$ elemű sorszámozott tételt és vegyük ki ebből találomra a $\xi_1$ mintaelemet, s a tételnek maradt még $N$ eleme. Ebben az $N$ elemű tételben nevezzük a $\xi_1$-nél kisebb sorszámúakat selejteseknek, a többit jónak. Ezáltal nyertünk egy olyan tételt, amelynek a selejtaránya $\frac{\xi_1 - 1}{N}$ 0 és 1 között bármely $N$ nevezőjű tört értékét egyforma valószínűséggel veheti fel, tehát a priori feltevésünk szerinti. Ezután kihúzzuk még az $N$ elemű tételből a $\xi_2, \ldots, \xi_{n+1}$ $n$ elemű mintát. Akkor egyrészt az az esemény, hogy a $\xi_1, \ldots \xi_{n+1}$ $(n+1)$ elemű mintában $\xi_1$ nagyság szerint a $(k+1)$-ik lesz, megegyezik azzal az eseménnyel, hogy a $\xi_2, \ldots \xi_{n+1}$ $n$ elemű mintában pontosan $k$ darab lesz selejtes, tehát az $\eta_{k+1} = \xi_1$ esemény megegyezik a $\varkappa_{nN} = k$ eseménnyel. Másrészt, mivel a tétel selejtaránya $\frac{\xi_1 - 1}{N}$, tehát az, hogy $\xi_1 \leq M+1$ azt jelenti, hogy az $N$ elemű tétel selejtaránya $\zeta \leq \frac{M}{N}$. A (6) tehát igaz és természetesen igaz marad akkor is, ha azt nem modell segítségével képzett tételre, hanem tetszőleges olyan $N$ elemű tételre vonatkoztatjuk, amely az a priori feltevésnek és a feltételnek megfelel.

A (4), (5) és (6) egyenletekből következik a (3). Ezzel tehát a dualizmus elvét bebizonyítottuk véges tétel esetére is. Az (1) a (3)-ból is nyerhető határátmenet segítségével.

A (3) összefüggést még más úton is igazoljuk. Ha a Bayes-tételt alkalmazzuk, arra jutunk, hogy (3) igazolásához a következő azonosság bizonyítása szükséges:

$$\frac{\sum\limits_{i=k}^{M} \binom{i}{k} \binom{N-i}{n-k}}{\sum\limits_{i=k}^{N-n+k} \binom{i}{k} \binom{N-i}{n-k}} = \frac{\sum\limits_{i=k+1}^{n+1} \binom{M+1}{i} \binom{N-M}{n+1-i}}{\binom{N+1}{n+1}}. \tag{7}$$

Ez az azonosság azonban ugyanazt mondja ki, mint az (5) összefüggés. Tehát a (3)-at a Bayes-tétel alkalmazásával és az (5) egyenlet bizonyításával is igazolhatjuk, a (4)-et és a (6)-ot a Bayes-tétel pótolja.

Ha az *Oderfeld* által követett utat követtük volna, akkor még a (7)-et is aritmetikai úton kellett volna bizonyítanunk. Ez az út igen bonyolultnak látszik.

A következő pontokban látni fogjuk, hogy fentebbi bizonyításunknak az is előnye, hogy könnyen kiterjeszthető.

### 3. *Polihipergeometrikus eloszlás*

Be fogjuk bizonyítani, hogy ha az egyszerű alternatíva helyett általában $s+1$ osztályozási lehetőség van, a dualizmus elve kiterjeszthető a következő alakban:

$$\mathbf{P}\left(\prod_{r=1}^{s}\left(\sum_{i=1}^{r}\zeta_i \leq \sum_{i=1}^{r}\frac{M_i}{N}\right)\,\Bigg|\,\prod_{r=1}^{s}\left[\varkappa_{rnN}=k_r\right)\right)= \tag{8}$$
$$=\mathbf{P}\left(\prod_{r=1}^{s}\left(\sum_{i=1}^{r}\varkappa_{i,n+s,\,N+s}\geq\sum_{i=1}^{r}k_i+r\right)\,\Bigg|\,\prod_{r=1}^{s}\left(\zeta_r=\frac{M_r+1}{N+s}\right)\right).$$

Itt $\zeta_i$ jelenti a tételben az $i$-edik osztályba tartozó elemek részarányát, $\varkappa_{inN}$ jelenti az $N$ elemű tételből vett $n$ elemű mintában az $i$-edik osztályba tartozó elemek számát. A produktumjel azt jelenti, hogy az utána álló esemény a futó index mindegyik értékére egyidejűleg fennáll.

A (8) a priori Bose—Einstein-eloszlás esetén érvényes, amikor tehát a tétel bármely $s+1$ osztályba való eloszlása egyenlően valószínű.

Most egy $N+s$ elemű sorszámozott tételt gondolunk el s ebből kiveszszük a $\xi_1, \xi_2, \ldots \xi_{n+s}$ $(n+s)$-elemű mintát. Legyen

$$\eta_1<\eta_2<\cdots<\eta_{n+s}$$

a $\xi_1, \xi_2 \ldots \xi_{n+s}$-ből,

$$\eta^*_1<\eta^*_2<\cdots<\eta^*_s$$

a $\xi_1, \xi_2, \ldots \xi_s$-ből alakított rendezett minta. Felírható, hogy

$$\mathbf{P}\left(\prod_{r=1}^{s}\left(\eta_{\sum_{i=1}^{r}k_i+r}\leq\sum_{i=1}^{r}M_i+r\right)\,\Bigg|\,\prod_{r=1}^{s}\left(\eta_{\sum_{i=1}^{r}k_i+r}=\eta^*_r\right)\right)= \tag{9}$$
$$=\mathbf{P}\left(\prod_{r=1}^{s}\left(\eta_{\sum_{i=1}^{r}k_i+r}\leq\sum_{i=1}^{r}M_i+r\right)\right).$$

A (9)-et hasonló módon láthatjuk be, mint a (4)-et. Az $n+s$ elemű minta első $s$ eleme nincs kitüntetve, az $\eta^*_1, \ldots \eta^*_s$-ot képezhetnénk a minta bármelyik $s$ eleméből, a (9) baloldala nem változik, tehát nem változik meg akkor sem, ha a feltételt elhagyjuk.

Továbbá belátható a

$$\mathbf{P}\left(\prod_{r=1}^{s}\left(\eta_{\sum\limits_{i=1}^{r} k_i+r} \leq \sum_{i=1}^{r} M_i + r\right)\right) =$$

(10)

$$= \mathbf{P}\left(\prod_{r=1}^{s}\left(\sum_{i=1}^{r} \varkappa_{i,n+s,N+s} \geq \sum_{i=1}^{r} k_i + r\right) \,\middle|\, \prod_{r=1}^{s}\left(\zeta_r = \frac{M_r+1}{N+s}\right)\right)$$

összefüggés érvényessége is, teljesen hasonló módon, mint az (5)-é, csak most az $N + s$ elemű tételnek azokat az elemeit tekintjük $r$-edik tulajdonságúaknak, amelyeknek sorszáma

$$\geq \sum_{i=1}^{r-1} M_i + r - 1 \quad \text{de} \quad \leq \sum_{i=1}^{r} M_i + r.$$

Végül a (6)-hoz hasonlóan belátható a

$$\mathbf{P}\left(\prod_{r=1}^{s}\left(\eta_r^* \leq \sum_{i=1}^{r} M_i + r\right) \,\middle|\, \prod_{r=1}^{s}\left(\eta_{\sum\limits_{i=1}^{r} k_i+r} = \eta_r^*\right)\right) =$$

(11)

$$= \mathbf{P}\left(\prod_{r=1}^{s}\left(\sum_{i=1}^{r} \zeta_i \leq \sum_{i=1}^{r} \frac{M_i}{N}\right) \,\middle|\, \prod_{r=1}^{s} [\varkappa_{rnN} = k_r]\right)$$

összefüggés is. Ha az $N + s$ elemű tételből kivesszük egyelőre a $\xi_1, \xi_2, \ldots \xi_s$ $s$ elemű mintát, a megmaradó elemek egy $N$ elemű tételt alkotnak. Ebben $r$-edik tulajdonságúnak tekintjük azt a $\eta_r^* - \eta_{r-1}^* - 1$ elemet (itt $\eta_0^* = 0$ és $\eta_{s+1}^* = N + s + 1$ értendő), amelynek eredeti sorszáma $> \eta_{r-1}^*$, de $< \eta_r^*$. Ezáltal $N$ elemű tételünk eloszlása Bose—Einstein-féle lesz, vagyis az a priori feltevéssel megegyező. Ugyanúgy tehát, mint előbb az (5)-nél, arra jutunk, hogy ilyen a priori feltevés mellett a (11) áll.

A (9), (10) és (11)-ből következik a (8).

### 4. Polinomiális eloszlás

Ha $N$ végtelenhez tart, a (8) átmegy a

$$\mathbf{P}\left(\prod_{r=1}^{s}\left(\sum_{i=1}^{r} \zeta_i \leq \sum_{i=1}^{r} p_i\right) \,\middle|\, \prod_{r=1}^{s} [\varkappa_{rn} = k_r]\right) =$$

(12)

$$= \mathbf{P}\left(\prod_{r=1}^{s}\left(\sum_{i=1}^{r} \varkappa_{i,n+s} \geq \sum_{i=1}^{r} k_i + r\right) \,\middle|\, \prod_{r=1}^{s} [\zeta_r = p_r]\right)$$

képletbe. Itt $\varkappa_{rn}$ jelenti az $n$ elemű mintában talált $r$-edik tulajdonságú elemek számát. A (12) akkor áll, ha a tétel a priori eloszlása a Bose—Einstein-eloszlásnak megfelelő folytonos eloszlás, az ú. n. intervallum-eloszlás, azaz minden lehetséges $s + 1$ osztályba való eloszlás egyforma valószínűség-sűrűségű.

Ebben a pontban a (12)-t közvetlenül, határátmenet nélkül is be fogjuk bizonyítani.

Most legyenek $\xi_1, \dots \xi_{n+s}$ a $(0, 1)$ intervallumban egyenletes eloszlású független valószínűségi változók, $\eta_i$ és $\eta_i^*$ értelmezése legyen ugyanaz, mint az előző pontban ; itt az $\eta$-k, ill. $\eta^*$-ok között egyenlők is lehetnek, de csak 0 valószínűséggel. Felírhatjuk ismét a

$$\mathbf{P}\left(\prod_{r=1}^{s}\left(\eta_{\sum\limits_{i=1}^{r} k_i+r} \leq \sum_{i=1}^{r} p_i\right) \,\middle|\, \prod_{r=1}^{s}\left(\eta_{\sum\limits_{i=1}^{r} k_i+r} = \eta_r^*\right)\right) = \tag{13}$$

$$= \mathbf{P}\left(\prod_{r=1}^{s}\left(\eta_{\sum\limits_{i=1}^{r} k_i+r} \leq \sum_{i=1}^{r} p_i\right)\right)$$

összefüggést. A (13) baloldala most sem változik meg, ha $\eta_1^*, \dots \eta_s^*$-et a $\xi_1, \dots \xi_{n+s}$ sorozat tetszőleges $s$ eleméből képezzük. Ha különböző kombinációkat veszünk, az így adódó különböző feltételek egymást — 0 valószínűségű kivételtől eltekintve — kizárják, tehát a valószínűség nem változik, ha a feltételt elhagyjuk.

Bebizonyíthatjuk továbbá, hogy

$$\mathbf{P}\left(\prod_{r=1}^{s}\left(\sum_{i=1}^{r} \varkappa_{i,n+s} \geq \sum_{i=1}^{r} k_i + r\right) \,\middle|\, \prod_{r=1}^{s} [\zeta_r = p_r]\right) = \tag{14}$$

$$\mathbf{P}\left(\prod_{r=1}^{s}\left(\eta_{\sum\limits_{i=1}^{r} k_i+r} < \sum_{i=1}^{r} p_i\right)\right).$$

Annak a valószínűsége ugyanis, hogy bármelyik $\xi_i$ $x$ és $x + p_r$ közé essék, $p_r$, tehát ugyanaz, mint annak valószínűsége, hogy az $n + s$ elemű minta bármely eleme $r$-edik tulajdonságú legyen, ha az $r$-edik tulajdonsághoz tartozó valószínűség $p_r$. Az viszont, hogy a $\xi_i$-k közül a nagyságrendben $\sum\limits_{i=1}^{r} k_i + r$-edik 0 és $\sum\limits_{i=1}^{r} p_i$ közé essék, azt jelenti, hogy legalább $\sum\limits_{i=1}^{r} k_i + r$ esik 0 és $\sum\limits_{i=1}^{r} p_i$ közé s ennek ugyanaz a valószínűsége, mint hogy a felírt feltétel mellett az $n + s$ elemű mintában legalább $\sum\limits_{i=1}^{r} k_i + r$ első $r$ tulajdonsághoz tartozó elemet találtunk. Következtetésünk igaz marad akkor is, ha az $r$ különböző értékeire felírt események egyidejű fennállását vesszük tekintetbe. Így tehát a (14) fennáll. Fennáll továbbá, hogy

$$\mathbf{P}\left(\prod_{r=1}^{s}\left(\sum_{i=1}^{r}\zeta_i<\sum_{i=1}^{r}p_i\right)\Bigg|\prod_{r=1}^{s}[\varkappa_{rn}=k_r]\right)=$$

$$-\mathbf{P}\left(\prod_{r=1}^{s}\left(\eta_r^*<\sum_{i=1}^{r}p_i\right)\Bigg|\prod_{i=1}^{s}\left(\eta_{\sum_{i=1}^{r}k_i+r}=\eta_r^*\right)\right). \qquad (15)$$

Az $\eta_1^*, \eta_2^*, \ldots \eta_s^*$-gal ugyanis definiálhatunk egy olyan tételt, amelynek eloszlása intervallum-eloszlás úgy, hogy az $r$-edik tulajdonsághoz tartozó valószínűségnek $(\eta_r^*-\eta_{r-1}^*)$-ot tekintjük (itt $\eta_0^*=0$ és $\eta_{s+1}^*=1$). Ebben az esetben mind a feltételben lévő, mind a feltétel előtt írt események a (15) bal- és jobboldalán rendre megegyeznek egymással, tehát a (15) igaz. A (12) tehát a (13), (14) és (15)-ből következik.

Abban a speciális esetben, amikor $s=1$, eredményünk egy új bizonyítást ad az (1)-re.

## 5. *Poisson-eloszlás*

Az eddigiekben láttuk azt, hogy a dualitási elv mind a pontos hipergeometrikus eloszlás, mind a közelítő binomiális eloszlás feltételezése esetén fennáll. Ismeretes azonban, hogy a hipergeometrikus eloszlás bizonyos esetekben Poisson-eloszlással is megközelíthető. Vessük fel tehát azt a kérdést, hogy változatlanul fennáll-e és közvetlenül igazolható-e a dualitási elv, ha feltételezzük azt, hogy a minta selejtes darabjainak száma Poisson-eloszlást követ? Természetesen a határátmenet arra vezet, hogy érvényes, ezért a közvetlen igazolás lehetőségét vizsgáljuk.

Megjegyezzük, hogy a gyakorlatban előfordulhat olyan eset, amikor a selejtszám eloszlásául a Poisson-eloszlást nem a binomiális eloszlás megközelítéseként tételezzük fel, pl. mikor szövetből veszünk bizonyos hosszúságú mintát.

Legyen egy bizonyos méretű mintában $\varkappa$ a hibák száma, $\lambda$ a hibák számának várható értéke ugyanolyan méretű mintában. Akkor a dualitás elvének erre az esetre felírt alakja

$$\mathbf{P}(\lambda<l\,|\,\varkappa=k)=\mathbf{P}(\varkappa\geq k+1\,|\,\lambda=l). \qquad (16)$$

Ha $n$-et végtelenhez közelítjük úgy, hogy $np=l$ legyen, akkor az (1) átmegy a (16)-ba. A (16) azonban közvetlenül is igazolható. Segédtételünk felhasználása helyett most a jelen esetben egyszerűbb analitikus utat választjuk. A Bayes-formula szerint

$$f(l|k)\,dl=\frac{\dfrac{l^k}{k!}e^{-l}\,dl}{\displaystyle\int_0^\infty\frac{x^k}{k!}e^{-x}\,dx}=\frac{l^k}{(k+1)!}e^{-l}\,dl \qquad (17)$$

s ez egyszersmind egyenlő annak a valószínűségével, hogy a $(k+1)$-edik hiba $l$ és $l+dl$ között van. Így tehát a (17)-et integrálva a (16)-ot kapjuk, mivel az, hogy a $(k+1)$-ik hiba 0 és $l$ közé esik, ugyanazt jelenti, mint hogy 0 és $l$ között legalább $k+1$ hiba van. Ezzel tehát a dualitás elvét arra az esetre is igazoltuk, ha a hibák száma Poisson-eloszlást követ.

## 6. *Nem egyenletes a priori eloszlás*

Ha az a priori eloszlás folytonos, sűrűségfüggvénye $f(p)$, akkor a Bayes-tétel alkalmazásával:

$$\mathbf{P}(\zeta < p | \varkappa_n = k) = \frac{\int_0^p f(x)\, x^k (1-x)^{n-k} dx}{\int_0^1 f(x)\, x^k (1-x)^{n-k} dx}.$$

(Itt és a következőkben adott tételből kivett minta selejteloszlását binomiálisnak tekintjük.)

Ha az a priori eloszlás diszkrét, a $w_1, w_2, \ldots$ értékeket $p_1, p_2, \ldots$ valószínűséggel veszi fel, akkor

$$\mathbf{P}(\zeta < p | \varkappa_n = k) = \frac{\sum\limits_{w_i < p} c_i w_i^k (1-w_i)^{n-k}}{\sum\limits_{i=1}^{\infty} c_i w_i^k (1-w_i)^{n-k}}.$$

Kérdés, hogy mely a priori eloszlások esetére terjeszthető ki a dualitási elv?

Erre a kérdésre itt általánosságban választ adni nem tudunk. Ismeretes azonban *Oderfeld* eredményéből [1], hogy ha az a priori eloszlás béta-eloszlás, tehát ha

$$f(x) = \frac{x^r (1-x)^{s-r}}{B(r+1, s-r+1)},$$

akkor alkalmazható a dualitási elv, éspedig ekkor

$$\mathbf{P}(\zeta < p | \varkappa_n = k) = \mathbf{P}(\varkappa_{n+s+1} \geq k+r+1 | \zeta = p). \tag{18}$$

Megjegyezzük, hogy a béta-eloszlás érdekes kapcsolatban van az intervallum-eloszlással: az előbbi az utóbbiból úgy származtatható, hogy az első $r$ tulajdonságú elemet selejtesnek, a többit jónak tekintjük.

Ennek megfelelően, ha a 4. pontban közölt bizonyítást úgy módosítjuk, hogy a produktum-jelet elhagyjuk és az $r$-nek határozott értéket tulajdonítunk, a (18)-ra jutunk.

Az alábbiakban még egy típusú a priori eloszlást adunk meg, amely mellett fennáll a dualitási elv egy alakja. Legyen az a priori eloszlás béta-eloszlások keveréke, tehát sűrűségfüggvénye legyen

$$f(x) = \sum_{i=1}^{\infty} c_i \frac{x^{r_i}(1-x)^{s_i - r_i}}{B(r_i + 1, s_i - r_i + 1)}.$$

Ebben az esetben kapjuk, hogy:

$$\mathbf{P}(\zeta < p \mid \varkappa_n = k) = \tag{19}$$

$$= \frac{\displaystyle\sum_{i=1}^{\infty} \frac{c_i (s_i + 1)\binom{s_i}{r_i}}{(s_i + n + 1)\binom{s_i + n}{r_i + k}} \mathbf{P}(\varkappa_{n+s_i+1} \geq k + r_i + 1 \mid \xi = p)}{\displaystyle\sum_{i=1}^{\infty} \frac{c_i (s_i + 1)\binom{s_i}{r_i}}{(s_i + n + 1)\binom{s_i + n}{r_i + k}}}.$$

Ha az összetevők száma kicsiny, a (19) képlet alkalmas a számításra, bár nem annyira, mint egyszerű béta-eloszlás esetén, a kész prospektív paraméterek felhasználása itt nem lehetséges.

## IRODALOM


[1] *G. Castelnuovo:* Calcolo delle probabilità, Milano, 1919.

[2] *E. C. Molina:* Bayes Theorem: An Expository Presentation, Bell System Technical Journal, 273. o. (1931)

[3] *J. Oderfeld:* Statystyczny odbiór towarów klasyfikowanych wedlug alternatywy, Warszawa (1950).

[4] *J. Oderfeld:* On the dual aspect of sampling plans. Colloquium Mathematicum II. 2. 89. o. (1951)

[5] *L. E. Simon:* An Engineers' Manual of Statistical Methods, New-York, 184. o. (1941).

[6] *H. Steinhaus:* Quality control by sampling. Colloquium Mathematicum, II. 2. 98. o. (1951).


## О ПРИНЦИПЕ ДВОЙСТВЕННОСТИ ПО ОТНОШЕНИЮ К ВЕРОЯТНОСНТЫМ ГРАНИЦАМ ПРОПОРЦИИ БРАКА, ОПРЕДЕЛЯЕМЫМ ПО БАЙЕСУ


К. Шаркади



Резюме

Статья присоединяется к статьям Штейнхауза [6] и Одерфельда [3], [4]. В упомянутых статьях авторы критикуют ту точку зрения, согласно которой метод, основанный на постулате Байеса –, должен быть исключен из современных методов математической статистики. Указывается что применение постулата Байеса во многих случаях приводит в основном к тем же результатам, как метод доверительных границ.

Принцип двойственности изложенный в упомянутых статях Одерфелда являетстя результатом подобного типа, относительно исследования брака. В данной статье автор обобщает упомянутый принцип двойственности на случаи гипергеометрического, полигипергеометрического, полиномиального распределении и распределения Пуассона.

# ON THE RULE OF DUALISM CONCERNING THE BAYES' PROBABILITY LIMITS OF THE FRACTION DEFECTIVE

K. SARKADI

*Summary*

This paper departs from the publications of Steinhaus and Oderfeld [6],[3],[4], in which they criticise the view that the method based on Bayes' rule must be excluded from up-to-date mathematical-statistical methods and point out that in many cases Bayes' rule gives the same result as the method of confidence limits. Concerning acceptance inspection, a statement of the same character is also expressed in the »rule of dualism« dealt with in Oderfeld's papers. Author generalizes this »rule of dualism« for the cases of hypergeometric, polyhypergeometric, polynomial and Poisson distributions.