

SZUPERPONÁLT VÉLETLEN JELZETEKET ALKALMAZÓ DOKUMENTÁCIÓS RENDSZEREK MATEMATIKAI KÉRDÉSEIRŐL

OROSZ GÁBOR¹⁾ és TAKÁCS LAJOS

Bevezetés

A tudományos és szakirodalom nagy mértékű növekedése az irodalomkutatás új eljárásainak kifejlesztésére vezetett. Az irodalom feltárásának s e feltárás alapján nyújtandó adatszolgáltatásnak — közismert néven: dokumentációnak — korábbi egyéni és manuális jellegű módszerei helyett mind szélesebb körben az intézményes és mechanizált megoldások kerülnek alkalmazásra. A feltárási munkát helyi, országos vagy nemzetközi intézmények szakértői végzik, akik a dokumentumok (könyvek, tanulmányok, kutatási jelentések, műszaki-, szabadalmi- és szabvány-leírások, prospektusok stb.) tartalmáról egyrészt ismertető kivonatot készítenek, másrészt kiemelik abból a tudományos vagy gyakorlati használat szempontjából tekintetbe vehető adatokat, tudományos fogalmakat, szakkifejezéseket, jellemzőket. Eme analizáló munka eredményeként minden dokumentumról úgynevezett dokumentációs felvétel készül, amely a dokumentum bibliográfiai adatait, tartalmi kivonatát s a kiemelt adatokat tartalmazza. Az így elkészített tételek összessége dokumentációs gyűjteményt képez, mely hatalmas adattömegével a tudományos kutatás nélkülözhetetlen forrása. A gyűjtemény sorolási rendje a kiemelt adatok nyilvántartási módján alapszik, s így egy meghatározott témára anyagot keresni szintén ennek alapján lehet.

A fejlődés másik tényezője az adatszolgáltatás mechanizálása. A dokumentációs gyűjteményekben való keresés gyors és hibamentes lefolytatására különleges berendezéseket, úgynevezett dokumentációs szelektorokat szerkesztettek. E berendezések a gyűjtemény felvétel-tömegéből az adott témára reference-ként használható tételeket a vonatkozó adatok leérzékelése alapján automatikusan válogatják ki. A szelektoroknak részint a gyűjtemény terjedelméhez méretezeten, részint az alkalmazott technikai megoldások különbözősége folytán igen sokféle típusa használatos a manuális lyukkártyás szerkezetektől a Hollerith-rendszerű elektromos gépeken, a fotocellás filmszelektorokon keresztül egészen a digitális számológépekkel rokon elektronikus berendezésekig.

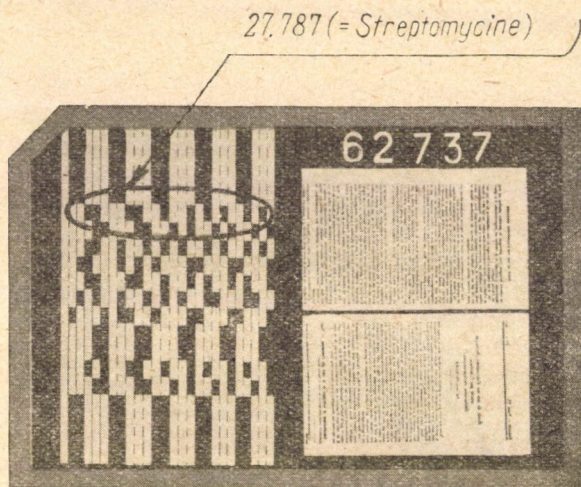
A szelektorok a dokumentumokból kiemelt adatokat az emberek közötti szokásos — írás- vagy hangbeli — közlési eljárások útján felvenni,

¹⁾ A budapesti Egyetemi Könyvtár munkatársa.

illetve ezek alapján feldolgozni a technikai fejlettség mai fokán nem képesek. Ezért a szelektorok leérzékelő szerkezete számára az adatokat megfelelő jelzésrendszer szerint leképezetten (átkódolva) kell az alkalmazott felvívó közegre (kártyára, filmre, magnetofon-szalagra) rávinni. A dokumentációs adatfeldolgozás és adatszolgáltatás munkamenete tehát jelzetelési eljárást — át- és visszakódolást — foglal magában.

A dokumentációs jelzésrendszereknek rendkívül sokféle típusa alakult ki. A jelzetek megválasztásának módja szerint három csoportba sorolhatók:

a) *Szisztematikus rendszerek.* Jelzeteik valamely tudományos-tároló rendszer alapján vannak szerkesztve. Ilyen például a kétsoros peremlyukasztású kártyákra szerkesztett metallurgiai dokumentációs rendszer, mely



1.a ábra

A Samain-féle Filmorex-szelektor egy filmlapja. Egysoros almezők 5 számjegyű exkluzív jelzetekkel

az ASM—SLA Metallurgical Classification Systemen²⁾ alapul. Vizsgálatuk tulajdonképpen a vonatkozó tudományos-tároló rendszer vizsgálatába megy át, tehát osztályozás- és tudományelméleti problémát képez.

b) *Jelentéselemekkel dolgozó rendszerek,* melyek a dokumentumban szereplő adatok, meghatározások szóképből leválasztott gyököket alkalmaznak jelzetekként. Ezt az eljárást alkalmazza a Perry-csoport által a kémiai dokumentáció céljaira szerkesztett jelzésrendszer, melynek jelzetei Hollerith-kártyákra tehetők, s feldolgozásukra a külön e célra szerkesztett IBM X794 Electronic Data Searching Machine szolgál. E rendszerek tanulmányozása elsősorban a jelentéstan körébe tartozik.

c) *Véletlenszerűen választott jelzeteket alkalmazó eljárások.* Ezeknél a jelzetek numerikus vagy alfabetikus, avagy numerikus és alfabetikus kom-

²⁾ ASM—SLA Metallurgical Literature Classification. Prepared by a Joint Committee of the American Society for Metals and the Special Libraries Association. Cleveland, 1950.

binatorikus formációk, melyeket teljesen tetszésszerű választással rendelünk a leképezendő fogalmakhoz, természetesen azzal a megkötéssel, hogy minden fogalomnak saját egyedi jelzete legyen. E rendszereknél részint a jelzetek struktúrája, részint véletlenszerű megválasztása s bejelölési módja folytán bizonyos matematikai problémák lépnek fel. Az alábbiakban éppen ezeket fogjuk behatóbb vizsgálat tárgyává tenni.

[illegible]

1:b ábra

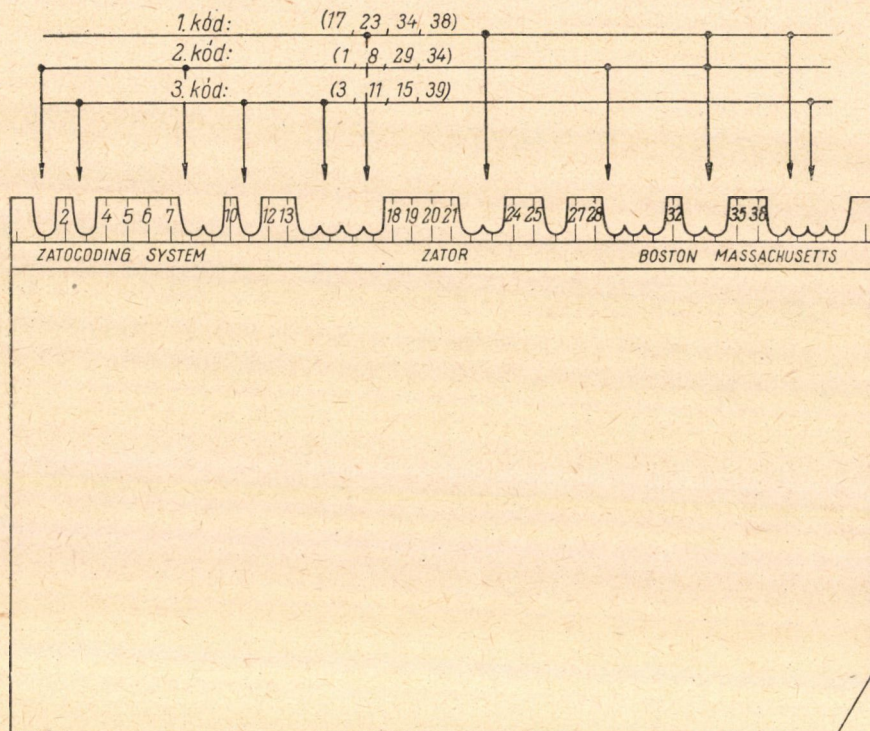
A Microfilm Rapid Selector jelmezejének részlete 7 számjegyű exkluzív jelzetek felvételére szolgáló 2 soros almező 5 jelhelyes mezőszelvényekkel

	A	B	C	D	E
0	◆	◆			
1	◆		◆		
2	◆			◆	
3	◆				◆
4		◆	◆		
5		◆		◆	
6		◆			◆
7			◆	◆	
8			◆		◆
9				◆	◆

1.c ábra

A Microfilm Rapid Selector kódrendszerének kulcsa

A véletlenszerű rendszerek jelzeteit a jelmezőbe, illetve ennek elhatárolt részeibe, az almezőkbe, kétféleképpen jelölhetjük be. Vagy úgy, hogy az almezők mindegyikébe az éppen soronkövetkező jelzet kerül. Így bármelyik jelzet akármelyik almezőbe juthat, de ha az almező egy jelzetet már felvett, többet nem jelölhetünk bele. (Lásd: 1. a, b, c ábrák.) Ezek az úgynevezett esetleges elhelyezésű exkluzív jelzetek. Vagy pedig úgy történhetik a bejelölés, hogy a jelzeteket a jelmezők, illetve az almezők meghatározott pozícióihoz rögzítjük, de ugyanakkor a jelmezőben egymáson keresztül-kasul futtatva több jelzetet is bejelölünk. (Lásd: 2. ábra.) Ezek a rögzített elhelyezésű szuperponált jelzetek.



2. ábra

Dokumentációs-kártya az elektrovibrátoros Zator-szelektorhoz.
 Línearis szuperpozíciós jelmező kombinációs jelzetekkel

Mielőtt tovább mennénk, közbevetőleg közöljük a következőkben használt jelöléseinket:

- n a jelmező pozícióinak száma
- p a jelmező almezőinek száma
- v a jelzet elemeinek száma
- m egy almező pozícióinak száma
- q a szuperponált rendszerekben az egyes almezőkbe eső jelzetrészek pozícióinak száma
- N egy jelmezőbe bejelölhető jelzetek száma
- W a jelzésrendszer összes jelzeteinek száma, a leképezhető fogalmak összessége.

Az exkluzív csoportba tartozó jelzésrendszerek közös jellemzője, hogy egy jelzetelem leképezéséhez a jelmezőnek egynél több pozíciója szükséges;

a jelmező kongruens almezőkre osztott;

minden almező egyetlen teljes jelzet felvételére szolgál, tehát ezeknél

$$N = p.$$

Egyedi jellemzőjük éppen jelzetképeiknek struktúrája. A jelzetek leképezése ugyanis lehet:

Biner: ekkor a tízes számrendszerben megadott jelzet biner számokra áttett formában van bejelölve. A jelzet biner képe az egymás után következő almező-pozíciók jelzeteltségének, illetve jelzeteletlenségének rendjéből adódik. Csak numerikus jelzetek leképezésére alkalmazható. A rendszer kapacitása:

$$W = 2^m - 1.$$

Szemibiner: az almezők v számú kongruens szelvényre vannak osztva. A szelvényekbe az egyes jelzetelemek biner formában kerülnek be, s e bejelölésekből az almezőben a teljes jelzet szemibiner képe tevődik össze. Az almező-szelvények terjedelme numerikus jelzetelem esetén 4, alfabetikus jelzetelem esetén 5 pozíció, mivel a 0–9, illetve 1–26 számok biner ekvivalenseinek leképezéséhez ennyi pozíció szükséges. Ennélfogva

numerikus jelzetek esetében	$m = 4v$ és $W = 10^v$
alfabetikus jelzetek esetében	$m = 5v$ és $W = 26^v$.

Kombinatorikus: az almezők itt is v számú szelvényre osztottak. A jelzetelemek képei a szelvények pozícióiból alkotott egyazon osztályba tartozó ismétlés nélküli kombinációk. A jelzetek viszont a jelzetelemek ismétléses variációi. Numerikus jelzetelemek számára 5 pozíciós szelvény szükséges, amelyben 2–2 pozíció jelölésével képezhetők le a számjegyek. Alfabetikus jelzetelemeknek 7 pozícióból álló szelvények kellenek, mert 7 az a legkisebb szám, amelynek annyi azonos osztályú (3-adosztályú) kombinációja képezhető, amennyi az alfabetum jegyeinek számát felülről legjobban közelíti. Az almezők $m = 5v$, illetve $m = 7v$ terjedelműek. A jelzés-rendszerek kapacitása ugyanakkora, mint a szemibiner leképezésű rendszereké.

A szuperpozíciós rendszereket az előbb tárgyaltakkal szemben éppen az jellemzi, hogy:

minden jelzetelem leképezése egyetlen pozíció igénybevételével történik;

a jelzetelemek és a vonatkozó pozíciók közötti kapcsolat szükségszerű, ennél fogva a jelzet képe a rendre következő tételek (kártyák, filmkockák) jelmezőiben mindig ugyanazon pozíciókból áll;

a jelmező lehet osztatlan vagy almezőkre osztott — ez utóbbi esetben az egyes almezőkbe nem a teljes jelzetkép, hanem annak csupán vonatkozó része esik;

a jelzetelemek tetszés szerinti megválasztása és a jelzetek szuperponálódása folytán egyes pozíciókon többszörös jelzeteltség állhat elő.

A jelzet szerkezete szerint a szuperponált rendszer lehet:

Kombinációs: a jelmező egyetlen összefüggő egész, s ennek pozícióiból képezett v -edosztályú ismétlés nélküli kombinációk a jelzetek, vagyis

$$W = \binom{n}{v}.$$

Variáció: A jelmező v számú kongruens almezőre van osztva, s a jelzetek az egyenként m pozícióból álló almezők egy-egy pozíciójából képezett v -adosztályú ismétléses variációk. Így

$$n = mv, \quad W = m^v.$$

Variáció—kombináció: A jelmező p számú kongruens almezőre osztott. A jelzetek az egymás után következő almezők pozícióiból választott q számosságú csoportokból tevődnek össze, vagyis a p számú almező m számosságú pozícióiból képezett q -adosztályú ismétlés nélküli kombinációk p -adosztályú ismétléses variációi. Tehát

$$n = pm, \quad v = pq, \quad W = \binom{m}{q}^p.$$

Mind az exkluzív, mind a szuperpozíciós típusnak megvannak a maga kétségtelen előnyei, amelyek használatukat adott esetben indokolttá teszik, de vannak hátrányai is.

Exkluzív jelzetek alkalmazása esetén szelektáláskor kizárólag azokat a tételeket kapjuk, amelyek a szelektálás szempontjait képező jelzeteket ténylegesen tartalmazzák, vagyis, amelyek a szóbanforgó problémára valóban szolgáltatnak reference-anyagot. Hátrányuk, hogy a jelzetek elhelyezésének esetlegessége következtében komplikáltabb leérzékelő mechanizmus szükséges, mert a kívánt jelzetet a jelmezőben meg kell keresni. Ha pedig több jelzetnek szimultán leérzékelése szükséges, akkor megfelelő memória-berendezésről kell gondoskodni a korábban befutott leérzékelések tárolására az utolsó beérkeztéig. Oly szelektoroknál, amelyekben ilyen különleges technikai berendezés nem alkalmazható, mint például a peremlyukasztású kártyákkal dolgozó egyszerű kézi berendezéseknél, az exkluzív típusú jelzetek nem is használhatók.

A szuperponált jelzetek megtalálása éppen rögzített elhelyezésük következtében a vonatkozó pozíciók leérzékelése folytán közvetlenül adódik. A leérzékelő mechanizmus a szelektálás alapjául szolgáló jelzetek pozícióira eleve beállítható, s így találat esetén egy viszonylag egyszerű kiváltószervezet útján hajtható végre a kívánt tétel kiemelése a többi közül. Memóriaszerkezetre nincs szükség, mert az összes vonatkozó pozíciók leérzékelése egyidejűleg következik be. Másik előnyük, hogy az igénybe veendő jelmező az előbbi típus jelmezőihez képest lényegesen kisebb terjedelmű, ami szintén fontos körülmény, tekintettel a szelektorok adatfelvevő közegének terjedelmét determináló technikai szempontokra. Viszont hátrányt jelent az a tény, hogy a jelzetek szuperpozíciója következtében más jelzeteknek a keresett jelzet pozícióival azonos pozícióiból a keresett jelzet fiktív mása imitálódik, s ennek következtében a szelektálás eredményeként kapott tételek között fölös, a vizsgált témára anyagot nem szolgáltatató tételek is előfordulhatnak. Ezeket kiválogatni, azaz a szelektálás eredményét hibamentessé tenni már csak emberi munkával lehet.

Az alábbi táblázatban két konkrét esetre közöljük az exkluzív és szuperponált rendszerek különböző fajtáinak jelmező terjedelmét:

$N = 12$	$W \approx 100\,000$			$W \approx 1\,000\,000$		
	n	r	m	n	r	m
<i>Szuperponált rendszerek</i>						
Kombinációs	70	4	—	72	4	—
Kombinációs	—	—	—	85	5	—
Kombinációs—variációs	70	4	35	92	4	46
Variációs	72	4	18	90	5	18
<i>Exkluzív rendszerek</i>						
Biner	204	—	17	240	—	20
Szemibiner—numerikus	240	5	20	288	6	24
Szemibiner—alfabetikus	240	4	20	300	5	25
Kombinatorikus (numerikus)	300	5	25	360	6	30
Kombinatorikus (alfabetikus)	336	4	28	420	5	35

A fent tárgyalt kérdéseknek az utóbbi években igen gazdag irodalma fejlődött ki. Elsősorban R. S. CASEY és J. W. PERRY [1] könyvét említjük meg, amelyben C. S. WISE írt egy fejezetet a szuperponált kódrendszerek matematikai kérdéseiről (276—302. oldalak). Az említett könyv a kérdés irodalmának kiterjedt összefoglalását tartalmazza. A szuperpozíciós rendszerek matematikai vizsgálatával C. N. MOOERS [2], [3], [4] munkáiban foglalkozik. Ezen munkák többnyire közelítő megoldásokat tartalmaznak. Pontos megoldást a [4] munka tartalmaz, amelyben módszert ad a jelzetelt pozíciók száma eloszlásának meghatározására. Eljárása rekurzív képleteken alapszik, amelyeket Boole-algebra segítségével nyer. A felmerülő problémák egzakt és közelítő megoldását a szerzők [5] dolgozatukban adták meg, további általánosításokat pedig [6] tartalmaz. Jelen dolgozat eredményei azonosak az [5] dolgozatban korábban közölt eredményekkel.

A következőkben rátérünk a szuperpozíciós rendszerek speciális matematikai problémáinak vizsgálatára.

1. §. A szuperpozíciós rendszerek matematikai problémái

Mint már a bevezetésben említettük, tegyük fel, hogy a dokumentációt n pozíciót tartalmazó lapokon végezzük. Minden egyes dokumentumról egy lapot készítünk, amely N jelzetet (kódot) tartalmaz. Minden egyes jelzetnek megfelelően a lapon r pozíciót kilyukasztunk.³⁾ Feltesszük, hogy a jelmező p számú kongruens almezőre van osztva, amelyek mindegyike m pozícióból áll. Tehát $n = mp$. Továbbá feltesszük, hogy az egyes kódoknak minden egyes almezőben q pozíció felel meg, azaz $r = pq$. Tekintsük a 3. ábrát, amely $n = 36$, $p = 2$, $r = 6$, $m = 18$ adatokkal jellemzett lap skémáját tünteti fel, és $N = 3$ kód bejelölését tartalmazza, amelyek a következők: (2, 7, 14, 8, 12, 18), (4, 14, 18, 1, 12, 15), (7, 9, 14, 3, 6, 14).

³⁾ A továbbiakban egyszerűség kedvéért mindig „lapokat” és „lyukasztást” mondunk. Megállapításaink természetesen általánosak, és a „lapok” alatt a szelektor adatfelvevő közegének egysége, „lyukasztáson” pedig a jelmező elem (pozíció) jelzeteltsége értendő.

1		3		5	6		8	
10	11	12	13		15	16	17	
	2		4	5		7		9
10	11		13			16	17	

3. ábra

Kombinációs—variációs módszer

A fent vázolt eljárás a kombinációs—variációs módszer. Ha speciálisan $p = 1$, akkor a kombinációs módszert nyerjük, míg ha $q = 1$, akkor a variációt. Így elég a kombinációs—variációs módszerrel foglalkozni, abból speciálisan kiadódik a másik két módszer.

Az egyes fogalmakhoz tartozó kódokat véletlenszerűen állapítjuk meg, mégpedig a lehetséges

$$W = \binom{m}{q}^p$$

számú kód mindegyike ugyanazon valószínűséggel jöhet tekintetbe. A hozzárendelést végezhetjük a következőképpen: Az összes lehetséges W számú kódnak megfelelő lapocskákat készítünk, és azokat elhelyezzük egy urnában. Minden egyes új fogalomhoz úgy rendelünk hozzá kódot, hogy az urnából *viSSzatevés nélkül* húzunk egy lapocskát. Természetesen ismétlődő fogalmakhoz mindig ugyanolyan kódot rendelünk. A fenti modell felhasználásával nyert eredmények pontosak. Ha $W \gg N$, akkor jó közelítéssel alkalmazhatunk egy, az előzőhöz hasonló modellt, csupán feltesszük, hogy az urnából a húzásokat *viSSzatevéssel* végezzük. Ekkor persze előfordulhat, hogy ugyanazon kódot különböző fogalmakhoz rendeljük hozzá, de ennek igen kicsiny a valószínűsége, és az eredményekben alig jut kifejezésre ennek a közelítésnek a hatása.

Bevezetjük a következő jelöléseket: Legyen η az egy lapon lyukasztott pozíciók száma. A fenti feltevésből következik, hogy η valószínűségi változó. Ennek eloszlása, várható értéke, szórása és momentumainak meghatározása

bír érdekességgel. Legyen $\mathbf{P}\{\eta = k\} = P_k$ és $\mathbf{M}\left\{\binom{\eta}{s}\right\} = B_s$, úgy $\mathbf{M}\{\eta\} = B_1$

és $\mathbf{D}^2\{\eta\} = 2B_2 + B_1 - B_1^2$. Továbbá jelölje χ valószínűségi változó egy tetszőlegesen választott pozíció esetén azt, hogy az hányszor van jelzetelve. Legyen $\mathbf{P}\{\chi = j\} = p_j$. A dokumentációs szelekció szempontjából igen fontos annak a valószínűségét megállapítani, hogy s kód szerinti szelekció alkalmával szelektálva legyen egy olyan lap, amelyen nincs mind az s kód bejelölve. Jelölje ennek a valószínűségét R_s . Ez szolgáltatja a nem kívánatos lapok kiválasztásának valószínűségét. Ehhez viszont szükséges ama valószínűség ismerete, hogy egy lapon meghatározott μ pozíció lyukasztva legyen. Jelölje ezt Q_μ .

A következőkben a fenti valószínűségek és várható értékek meghatározásával fogunk foglalkozni. Megadjuk ezek explicit alakját a pontos és a közelítő esetben egyaránt. Az általános tárgyalás előtt a $p = 1$ -nek megfelelő speciális esettel foglalkozunk, mivel ez gyakorlatilag a legfontosabb eset és elméletileg a legegyszerűbben tárgyalható. Ezután a legáltalánosabb kombinációs—variációs módszert tárgyaljuk, amelyből speciálisan adódik a kombinációs és variációs módszer. A probléma tovább általánosítható arra az esetre is, midőn a jel-mezőkben az egyes almezőkbe eső pozíciók száma és a kód elemeinek az egyes almezőkbe eső része különböző. Ezzel jelenleg nem foglalkozunk. A szerzők [6] dolgozata vizsgálja ezt az esetet.

A felvetett kérdések megoldására JORDAN KÁROLY [7] általános valószínűségi tételét használjuk fel.

2. §. Az általános valószínűségi tétel

Tekintsünk egy véletlen kísérletet, amelynek eredményét az $A_1, A_2, \dots, A_n$ események bekövetkezése szempontjából vizsgáljuk. Jelölje ξ valószínűségi változó az $A_1, A_2, \dots, A_n$ események közül bekövetkezők számát. A ξ valószínűségi változó a $0, 1, 2, \dots, n$ értékeket veheti fel. Vezessük be továbbá a következő jelöléseket: $\mathbf{P}\{\xi = k\} = P_k(A_1, A_2, \dots, A_n)$ és

$$\mathbf{M}\left\{\binom{\xi}{k}\right\} = B_k \quad (k = 0, 1, 2, \dots, n).$$

JORDAN KÁROLY 1927-ben megjelent [7] dolgozatában közölte az úgynevezett általános valószínűségi tételt, amely a fenti jelölésben a következő:

$$(1) \quad P_k(A_1, A_2, \dots, A_n) = \sum_{j=k}^n (-1)^{j-k} \binom{j}{k} S_j,$$

ahol $S_0 = 1$, és $j = 1, 2, \dots, n$ esetén

$$(2) \quad S_j = \sum_{(i_1, i_2, \dots, i_j)} \mathbf{P}\{A_{i_1} A_{i_2} \dots A_{i_j}\},$$

ahol az összegezés kiterjesztendő az $(1, 2, \dots, n)$ számokból alkotható valamennyi j -edosztályú $(i_1, i_2, \dots, i_j)$ ismétlés nélküli kombinációra.

A fenti tételeknek fontos speciális esetét kapjuk, ha feltesszük, hogy az $A_1, A_2, \dots, A_n$ események ekvivalens események. Ezen azt értjük, hogy bárhogyan is választunk ki az $A_1, A_2, \dots, A_n$ események közül j számú különbözőt ($j = 1, 2, \dots, n$), ezen események együttes bekövetkezésének a valószínűsége független a kiválasztás módjától és csupán a j számtól függ, azaz fennállnak a

$$\mathbf{P}\{A_{i_1} A_{i_2} \dots A_{i_j}\} = \mathbf{P}\{A_1 A_2 \dots A_j\}, \quad 1 \leq i_1 < i_2 < \dots < i_j \leq n, \\ j = 1, 2, \dots, n$$

összefüggések. Vezessük be ekkor a következő jelölést: $\pi_0 = 1$, és $j = 1, 2, \dots, n$ esetén

$$\pi_j = \mathbf{P}\{A_1 A_2 \dots A_j\}.$$

Ekkor speciálisan ekvivalens eseményekre (1) és (2) a következő egyszerűbb alakot ölti:

$$(3) \quad \begin{aligned} P_k(A_1, A_2, \dots, A_n) &= \sum_{j=k}^n (-1)^{j-k} \binom{j}{k} \binom{n}{j} \pi_j = \\ &= \binom{n}{k} \sum_{j=k}^n (-1)^{j-k} \binom{n-k}{n-j} \pi_j \end{aligned}$$

és

$$(4) \quad S_j = \binom{n}{j} \pi_j.$$

Megjegyezzük, hogy a ξ változó várható értéke:

$$(5) \quad \mathbf{M}\{\xi\} = S_1,$$

szórásnégyzete:

$$(6) \quad \mathbf{D}^2\{\xi\} = 2S_2 + S_1 - S_1^2$$

és s -edik momentuma:

$$(7) \quad \mathbf{M}\{\xi^s\} = \sum_{j=1}^s j! \mathfrak{S}_s^j S_j,$$

ahol $\mathfrak{S}_s^j$ a másodfajú Stirling-számokat jelöli.

A fentiekkel kapcsolatban utalunk még az egyik szerző [8] dolgozatára.

3. §. Az osztatlan jelmező esete

Tekintsük most a $p = 1$ speciális esetet, vagyis azt az esetet, midőn az n pozícióval rendelkező lapokra N számú v elemből álló kódot jelölünk be.

Ekkor a lehetséges kódok száma, a szókinsz: $W = \binom{n}{v}$.

Jelölje most $A_1, A_2, \dots, A_n$ rendre azt az eseményt, hogy a vizsgált lapon az első, a második, ..., az n -edik pozíció nincs lyukasztva. Mint könnyen belátható, a pontos és a közelítő modell esetén egyaránt érvényes, hogy az $A_1, A_2, \dots, A_n$ események ekvivalens események, és annak a valószínűsége, hogy közülük j számú előfordul, a pontos modell esetén:

$$(8) \quad \mathbf{P}\{A_1 A_2 \dots A_j\} = \frac{\binom{n-j}{v}}{\binom{n}{v}},$$

a közelítő modell esetén pedig :

$$(9) \quad \mathbf{P}\{A_1 A_2 \dots A_j\} = \left[\frac{\binom{n-j}{v}}{\binom{n}{v}} \right]^N.$$

Legyen a rövidség kedvéért $\mathbf{P}\{A_1 A_2 \dots A_j\} = \pi_j$. A fenti képletek könnyen adódnak a visszatevés nélküli és a visszatevéses húzások alapján.

1°. *A lyukasztott pozíciók számának eloszlása.* Most nyilvánvalóan felírhatjuk, hogy $P_k = \mathbf{P}\{\eta = k\} = P_{n-k}(A_1, A_2, \dots, A_n)$. Ugyanis a lapon akkor lesz k pozíció lyukasztva, ha az $A_1, A_2, \dots, A_n$ események közül pontosan k nem fordul elő, vagyis $n - k$ fordul elő. Mivel (3) szerint ekvivalens eseményekre

$$P_{n-k}(A_1, A_2, \dots, A_n) = \binom{n}{k} \sum_{j=0}^k (-1)^{k-j} \binom{k}{j} \pi_{n-j},$$

tehát a pontos modell esetén :

$$(10) \quad P_k = \frac{\binom{n}{k}}{\binom{n}{v}} \sum_{j=0}^k (-1)^{k-j} \binom{k}{j} \binom{j}{v} \binom{j}{N},$$

a közelítő modell esetén pedig :

$$(11) \quad P_k = \frac{\binom{n}{k}}{\left[\binom{n}{v}\right]^N} \sum_{j=0}^k (-1)^{k-j} \binom{k}{j} \left[\binom{j}{v}\right]^N.$$

Speciálisan mindkét modellre :

$$\mathbf{M}\{\eta\} = n(1 - \pi_1)$$

és

$$\mathbf{D}^2\{\eta\} = n^2(\pi_2 - \pi_1^2) + n(\pi_1 - \pi_2).$$

2°. *A pozíciók lyukasztási multiplicitása.* χ -vel jelöljük azt a változót, amely azt mutatja, hogy egy kiválasztott pozíció hányszorosan van lyukasztva. A $\mathbf{P}\{\chi = j\} = p_j$ valószínűsége könnyen felírható a pontos modell esetén, hogy

$$(12) \quad p_j = \frac{\binom{n-1}{v-1} \binom{n-1}{N-j}}{\binom{n}{N}} ,$$

a közelítő modell esetén pedig :

$$(13) \quad p_j = \binom{N}{j} \left(\frac{v}{n}\right)^j \left(1 - \frac{v}{n}\right)^{N-j} .$$

Ugyanis az $\binom{n}{v}$ lehetséges kód között $\binom{n-1}{v-1}$ olyan van, amely egy kiválasztott pozíciót tartalmaz, és $\binom{n-1}{v}$ olyan van, amely azt nem tartalmazza. Az első esetben visszatevés nélkül történik a húzás, míg a másodikban visszatevéssel és $\binom{n-1}{v-1} : \binom{n}{v} = \frac{v}{n}$ a kedvező kód húzásának a valószínűsége.

Mindkét esetben

$$\mathbf{M}\{\chi\} = N \frac{v}{n} .$$

A pontos esetben :

$$\mathbf{D}^2\{\chi\} = N \frac{v}{n} \left(1 - \frac{v}{n}\right) \left(1 - \frac{N-1}{\binom{n}{v} - 1}\right) ,$$

a közelítő esetben pedig :

$$\mathbf{D}^2\{\chi\} = N \frac{v}{n} \left(1 - \frac{v}{n}\right) .$$

3°. Az *extra kártyák kiválasztásának a valószínűsége*. Tegyük fel, hogy egy vagy több kód szerint szelektálást végzünk. Ekkor kiválasztódnak mindazon kártyák, amelyeken ez a kód vagy ezek a kódok ténylegesen be vannak jelölve, de esetleg kiválasztódnak olyan kártyák is, amelyeken a speciális kód vagy kódok nem szerepelnek, de a többi kód szuperpozíciója révén kiadódnak. Kérdés, mennyi a valószínűsége annak, hogy egy olyan kártya kiválasztódik, amelyen a szelektálás alapjául szolgáló kód vagy kódok nincsenek bejelölve. Jelölje ezt a valószínűséget s kód szerinti szelektálás esetén R_s ($s = 1, 2, \dots, N$). Ezt a valószínűséget a következő valószínűség segítségével határozhatjuk meg. Jelölje Q_μ annak a valószínűségét, hogy egy kártyán kiválasztott μ pozíció lyukasztva van. Az $A_1, A_2, \dots, A_n$ események ekvivalenciájából következik, hogy ez a valószínűség :

$$Q_\mu = \mathbf{P}\{\bar{A}_1 \bar{A}_2 \dots \bar{A}_\mu\}$$

Ez pedig JORDAN KÁROLY (3) tételének alkalmazásával a következőképpen fejezhető ki a π_j mennyiségek segítségével:

$$Q_\mu = \sum_{j=0}^{\mu} (-1)^j \binom{\mu}{j} \pi_j.$$

Ugyanis most annak a valószínűségéről van szó, hogy az $A_1, A_2, \dots, A_n$ események közül egy sem fordul elő, azaz a $P_0(A_1, A_2, \dots, A_n)$ valószínűségről. Most a pontos modell esetén

$$(14) \quad Q_\mu = \frac{1}{\binom{n}{\nu}} \sum_{j=0}^{\mu} (-1)^j \binom{\mu}{j} \binom{n-j}{\nu} \binom{n}{N},$$

a közelítő modell esetén pedig:

$$(15) \quad Q_\mu = \frac{1}{\left[\binom{n}{\nu}\right]^N} \sum_{j=0}^{\mu} (-1)^j \binom{\mu}{j} \left[\binom{n-j}{\nu}\right]^N.$$

Speciálisan annak a valószínűsége, hogy egy kártyán egy kijelölt kód pozíciói lyukasztva vannak, Q_ν : ugyanis egy kód ν számú elemet tartalmaz. Most felírható a pontos modell esetén, hogy

$$Q_\nu = \frac{N}{\binom{n}{\nu}} + \left(1 - \frac{N}{\binom{n}{\nu}}\right) R_1,$$

a közelítő modell esetén pedig:

$$Q_\nu = \left[1 - \left(1 - \frac{1}{\binom{n}{\nu}}\right)^N\right] + \left(1 - \frac{1}{\binom{n}{\nu}}\right)^N R_1.$$

A fenti képletek a teljes valószínűség tétele alapján adódnak. Ugyanis Q_ν annak a valószínűsége, hogy egy kiválasztott kódnak megfelelő pozíciók ki vannak lyukasztva a kártyán, és ez két egymást kizáró módon jöhet létre: az illető kód vagy tényleg szerepel a kártyán — aminek a valószínűsége a pontos modell esetén $N/\binom{n}{\nu}$, a közelítő modell esetén pedig $1 - \left(1 - 1/\binom{n}{\nu}\right)^N$

és ekkor 1 valószínűséggel ki vannak lyukasztva a megfelelő pozíciók; vagy az illető kód nem szerepel — aminek a valószínűsége a pontos modell esetén

$1 - N / \binom{n}{v}$, a közelítő modellnél pedig $\left(1 - 1 / \binom{n}{v}\right)^N$, de a többiek szuperpozíciója révén kiadódik, aminek a valószínűsége R_1 .

Így tehát egy kód szerinti szelektáláskor annak a valószínűsége, hogy egy olyan kártyát is kiválasztunk, amely nem kívánatos, a pontos modell esetén

$$(16) \quad R_1 = \frac{Q_v - N / \binom{n}{v}}{1 - N / \binom{n}{v}},$$

a közelítő modell esetén pedig

$$(17) \quad R_1 = \frac{Q_v - \left[1 - \left(1 - \frac{1}{\binom{n}{v}}\right)^N\right]}{\left(1 - \frac{1}{\binom{n}{v}}\right)^N}.$$

Ha egyidejűleg több kód szerint végezzük a szelektálást, úgy először össze kell számolni, hogy ezen kódok szuperpozíciója összesen hány különböző elemet tartalmaz. Tegyük fel, hogy s kód szerint végezzük a szelektálást, és az összes különböző elemek száma μ . Ekkor Q_μ szolgáltatja annak valószínűségét, hogy az s kód szuperpozíciójának megfelelő μ pozíció a kártyán ki van lyukasztva. Most a fentihez hasonló okoskodással felírható a pontos modell esetén:

$$Q_\mu = \frac{\binom{\binom{n}{v} - s}{N - s}}{\binom{\binom{n}{v}}{N}} + \left[1 - \frac{\binom{\binom{n}{v} - s}{N - s}}{\binom{\binom{n}{v}}{N}}\right] R_s,$$

a közelítő modell esetén pedig:

$$Q_\mu = \sum_{j=0}^s (-1)^j \binom{s}{j} \left[1 - \frac{j}{\binom{n}{v}}\right]^N + \left\{1 - \sum_{j=0}^s (-1)^j \binom{s}{j} \left[1 - \frac{j}{\binom{n}{v}}\right]^N\right\} R_s.$$

A pontos modell esetén ugyanis

$$\frac{\binom{\binom{n}{v} - s}{N - s}}{\binom{\binom{n}{v}}{N}}$$

annak a valószínűsége, hogy az N kód között mind az s megjelölt kód szerepel, míg a közelítő modell esetén JORDAN KÁROLY (3) tétele alapján könnyen nyerjük, hogy

$$\sum_{j=0}^s (-1)^j \binom{s}{j} \left[1 - \frac{j}{\binom{n}{v}} \right]^N$$

a kívánt valószínűség. Így tehát a pontos modell esetében:

$$(18) \quad R_s = \frac{Q_\mu - \left(\binom{n}{v} - s \right) / \left(\binom{n}{N} \right)}{1 - \left(\binom{n}{v} - s \right) / \left(\binom{n}{N} \right)},$$

a közelítő modellnél pedig:

$$(19) \quad R_s = \frac{Q_\mu - \sum_{j=0}^s (-1)^j \binom{s}{j} \left[1 - \frac{j}{\binom{n}{v}} \right]^N}{1 - \sum_{j=0}^s (-1)^j \binom{s}{j} \left[1 - \frac{j}{\binom{n}{v}} \right]^N}.$$

Megjegyezzük, hogy ama valószínűségek, amelyek azt fejezik ki, hogy az N kód közül kiválasztott s kód be van jelölve a kártyán, nem bírnak különösebb fontossággal. Ugyanis szelektáláskor mindazon kártyák, amelyek a megjelölt kódok szerepelnek, minden esetben kiesnek. A fenti valószínűségek csak tájékoztató adatot szolgáltatnak a kódok előfordulására, ha az egyes fogalmak egyenlően valószínűen fordulnak elő. A nem kívánatos kártyák számának megbecslésére azonban jól alkalmazható a fenti R_s valószínűség. Ha összesen S kártyánk van, úgy s kód szerinti szelektáláskor a nem kívánatos kártyák várható száma SR_s .

Az eddig megadott exakt formulák birtokában lehetőség van arra, hogy szuperpozíciós kódokkal dolgozó dokumentációs rendszerek tervezésénél az eddig alkalmazott közelítő formulák helyett pontosabbakat alkalmazzunk, és így valóban optimális viszonyokat érjünk el.

4. §. Az osztott jelmező esete

Tekintsük most a bevezetésben említett általános esetet, midőn a jelmező p számú kongruens almezőre van osztva. Ekkor a lehetséges kódok száma, a szókinsz: $W = \binom{m}{q}^p$. A kódok választása a pontos modellnél vissza-

tevés nélküli húzással, a közelítő modellnél pedig visszatevéses húzással történik.

Jelölje rendre $A_1, A_2, \dots, A_n$ azt az eseményt, hogy a vizsgált kártyán az első, a második, ..., az n -edik pozíció nincs lyukasztva. Az osztott jelmezű esetében az $A_1, A_2, \dots, A_n$ események már nem lesznek ekvivalens események, de érvényes lesz, hogy ugyanazon almezőre vonatkozó események egymásközt ekvivalensek. Ha most meg akarjuk állapítani j számú esemény együttes előfordulásának a valószínűségét, akkor ez attól fog függni, hogy az egyes események közül a különböző almezőkre hány vonatkozik. Így a pontos modell esetén:

$$\mathbf{P}\{A_{i_1} A_{i_2} \dots A_{i_j}\} = \frac{\binom{m-j_1}{q} \binom{m-j_2}{q} \dots \binom{m-j_p}{q}}{\binom{m}{q}^p},$$

ha az $A_{i_1}, A_{i_2}, \dots, A_{i_j}$ események közül j_1 számú az első, j_2 számú a második, ... és j_p számú a p -edik almezőre vonatkozik. A fenti valószínűség a közelítő modell esetén:

$$\mathbf{P}\{A_{i_1} A_{i_2} \dots A_{i_j}\} = \left[\frac{\binom{m-j_1}{q} \binom{m-j_2}{q} \dots \binom{m-j_p}{q}}{\binom{m}{q}^p} \right]^N.$$

I^o A lyukasztott pozíciók számának eloszlása. Most is, miként az osztatlan jelmezű esetében, felírható: $P_k = \mathbf{P}\{\eta = k\} = P_{n-k}(A_1, A_2, \dots, A_n)$, és ezen utóbbi valószínűséget JORDAN KÁROLY (1) képlete szolgáltatja, amely szerint

$$(20) \quad P_k = \sum_{j=0}^k (-1)^{k-j} \binom{n-j}{n-k} B_{n-j},$$

ahol a B_j mennyiségek most a $\xi = n - \eta$ változó binomiális momentumait jelölik. (2) szerint a pontos modell esetén fennáll:

$$(21) \quad B_j = \sum_{j_1+j_2+\dots+j_p=j} \binom{m}{j_1} \binom{m}{j_2} \dots \binom{m}{j_p} \frac{\binom{m-j_1}{q} \binom{m-j_2}{q} \dots \binom{m-j_p}{q}}{\binom{m}{q}^p},$$

a közelítő modell esetén pedig :

$$(22) \quad B_j = \sum_{j_1+j_2+\dots+j_p=j} \binom{m}{j_1} \binom{m}{j_2} \dots \binom{m}{j_p} \left[\frac{\binom{m-j_1}{q} \binom{m-j_2}{q} \dots \binom{m-j_p}{q}}{\binom{m}{q}^p} \right]^N.$$

Most η várható értéke :

$$\mathbf{M}\{\eta\} = n - B_1$$

és szórásnégyzete :

$$\mathbf{D}^2\{\eta\} = 2 B_2 + B_1 - B_1^2,$$

ahol a pontos modellre

$$B_1 = n \frac{\binom{m-1}{q} \binom{m}{q}^{p-1}}{\binom{m}{q}^p}.$$

és

$$B_2 = \frac{p \binom{m}{2} \left(\frac{\binom{m-2}{q} \binom{m}{q}^{p-1}}{N} \right) + m^2 \binom{p}{2} \left(\frac{\binom{m-1}{q}^2 \binom{m}{q}^{p-2}}{N} \right)}{\binom{m}{q}^p},$$

míg a közelítő modellre

$$B_1 = n \left(1 - \frac{q}{m} \right)^N$$

és

$$B_2 = \frac{n(m-1)}{2} \left[\left(1 - \frac{q}{m} \right) \left(1 - \frac{q}{m-1} \right) \right]^N + \frac{n(n-m)}{2} \left(1 - \frac{q}{m} \right)^{2N}.$$

2°. *A pozíciók lyukasztási multiplicitása.* Jelölje ismét χ egy kiválasztott pozíció lyukasztásának multiplicitását, és legyen $p_j = \mathbf{P}\{\chi = j\}$. A pontos modell esetén felírható, hogy

$$(23) \quad p_j = \frac{\binom{m-1}{q-1} \binom{m}{q}^{p-1}}{j} \frac{\binom{m-1}{q} \binom{m}{q}^{p-1}}{N-j} \frac{1}{\binom{m}{q}^p},$$

a közelítő modell esetén pedig

$$(24) \quad p_j = \binom{N}{j} \left(\frac{q}{m}\right)^j \left(1 - \frac{q}{m}\right)^{N-j}.$$

Most mindkét modellre :

$$\mathbf{M}\{\chi\} = N \frac{q}{m},$$

és a pontos modellre :

$$\mathbf{D}^2\{\chi\} = N \frac{q}{m} \left(1 - \frac{q}{m}\right) \left(1 - \frac{N-1}{\binom{m}{q}^p - 1}\right),$$

a közelítő modellre pedig :

$$\mathbf{D}^2\{\chi\} = N \frac{q}{m} \left(1 - \frac{q}{m}\right).$$

3°. Az *extra kártyák kiválasztásának a valószínűsége*. Az osztatlan jellemző esetéhez hasonlóan meg kell állapítani először a Q_v valószínűséget, vagyis annak a valószínűségét, hogy a szelektálás alapjául szolgáló kód elemeinek megfelelő pozíciók ki vannak lyukasztva. Ezt a valószínűséget szintén JORDAN KÁROLY általános valószínűségi tétele alapján írhatjuk fel. A pontos modell esetén az eredmény :

$$(25) \quad Q_v = \sum_{j=0}^v (-1)^j \sum_{j_1+j_2+\dots+j_p=j} \binom{q}{j_1} \binom{q}{j_2} \dots \binom{q}{j_p} \frac{\binom{m-j_1}{q} \binom{m-j_2}{q} \dots \binom{m-j_p}{q}}{\binom{m}{q}^p - N},$$

a közelítő modell esetén pedig :

$$(26) \quad Q_v = \sum_{j=0}^v (-1)^j \sum_{j_1+j_2+\dots+j_p=j} \binom{q}{j_1} \binom{q}{j_2} \dots \binom{q}{j_p} \left\{ \frac{\binom{m-j_1}{q} \binom{m-j_2}{q} \dots \binom{m-j_p}{q}}{\binom{m}{q}^p} \right\}^N.$$

Jelölje T_1 annak a valószínűségét, hogy egy kiszemelt kód szerepel a kártyán. Erre könnyen beláthatóan fennáll a pontos modell esetén

$$(27) \quad T_1 = \frac{\binom{m}{q}^p - 1}{\binom{m}{q}^p - N} = \frac{N}{\binom{m}{q}^p}.$$

a közelítő modell esetén pedig:

$$(28) \quad T_1 = 1 - \left(1 - \frac{1}{\binom{m}{q}^p} \right)^N.$$

Végül az R_1 valószínűség a teljes valószínűség tétele alapján felírható

$$Q_v = T_1 + (1 - T_1) R_1$$

összefüggésből határozható meg, ahonnan

$$(29) \quad R_1 = \frac{Q_v - T_1}{1 - T_1}.$$

Ha a szelektálás egyidejűleg s kód szerint történik, úgy össze kell számolni, hogy ezen s kód szuperpozíciója összesen hány különböző elemet tartalmaz. Legyen ez μ . Ezután meg kell nézni, hogy az egyes almezőkbe ezen elemekből hány jut. Legyenek ezek számai rendre $\mu_1, \mu_2, \dots, \mu_p$. Most annak a valószínűsége, hogy az s kiszemelt kód szuperpozíciójának megfelelő $\mu = \mu_1 + \mu_2 + \dots + \mu_p$ pozíció a kártyán lyukasztva van, JORDAN KÁROLY (3) tétele alapján a pontos modellre:

$$(30) \quad Q_\mu = \sum_{j=0}^{\mu} (-1)^j \sum_{j_1+j_2+\dots+j_p=j} \binom{\mu_1}{j_1} \binom{\mu_2}{j_2} \dots \binom{\mu_p}{j_p} \frac{\binom{m-j_1}{q} \binom{m-j_2}{q} \dots \binom{m-j_p}{q}}{\binom{\binom{m}{q}^p}{N}},$$

a közelítő modellre pedig:

$$(31) \quad Q_\mu = \sum_{j=0}^{\mu} (-1)^j \sum_{j_1+j_2+\dots+j_p=j} \binom{\mu_1}{j_1} \binom{\mu_2}{j_2} \dots \binom{\mu_p}{j_p} \left\{ \frac{\binom{m-j_1}{q} \binom{m-j_2}{q} \dots \binom{m-j_p}{q}}{\binom{m}{q}^p} \right\}^N.$$

Jelölje T_s annak a valószínűségét, hogy s kiszemelt kód mindegyike szerepel a kártyán. A pontos modellre könnyen felírható, hogy

$$(32) \quad T_s = \frac{\binom{\binom{m}{q}^p - s}{N - s}}{\binom{\binom{m}{q}^p}{N}},$$

a közelítő modellre pedig JORDAN KÁROLY (3) tétele alapján

$$(33) \quad T_s = \sum_{j=0}^s (-1)^j \binom{s}{j} \left[1 - \frac{j}{\binom{m}{q}} \right]^N.$$

Végül a teljes valószínűség tétele szerint fennáll:

$$(34) \quad Q_\mu = T_s + (1 - T_s) R_s,$$

és innen

$$R_s = \frac{Q_\mu - T_s}{1 - T_s}$$

szolgáltatja annak a valószínűségét, hogy egy olyan kártyát kiválasszunk, amelyen a kiszemelt s kód mindegyike nincs bejelölve.

(Beérkezett: 1956. augusztus 15.)

IRODALOM

- [1] R. S. CASEY—J. W. PERRY: *Punched cards, their applications to Science and Industry*. Reinhold, New York 1951.
- [2] C. N. MOOERS: *Application of random codes to the gathering of statistical information*. Zator Company, Boston, 1949.
- [3] C. N. MOOERS: *Zatocoding for punched cards*. Zator Company, Boston, 1950.
- [4] C. N. MOOERS: *The exact distribution of the number of positions marked in a Zatocoding field*. Zator Technical Bulletin No. 73., Boston (1952)
- [5] G. OROSZ—L. TAKÁCS: „Über die Codesysteme mit zufallsartig gewählten superponierten Koden.” *Dokumentation* 3 (1956) 9—22.
- [6] G. OROSZ—L. TAKÁCS: „Some probability problems concerning the marking of codes into the superimposition field.” *Journal of Documentation* (sajtó alatt).
- [7] JORDAN K.: „A valószínűségszámítás alapfogalmai” („Les fondements du calcul des probabilités”). *Matematikai és Fizikai Lapok* 34 (1927) 109—136.
- [8] TAKÁCS L.: „Az általános valószínűségi tételről.” *A Magyar Tudományos Akadémia III. (Matematikai és Fizikai) Osztályának Közleményei* 5 (1955) 467—476.

О МАТЕМАТИЧЕСКИХ ВОПРОСАХ ДОКУМЕНТАЦИОННЫХ СИСТЕМ С НАЛАГАЮЩИМИСЯ СЛУЧАЙНЫМИ КОДАМИ

G. OROSZ и L. TAKÁCS

Резюме

Введение знакомит с различными документационными системами. § 1. знакомит с математическими проблемами случайных налагающихся систем. Относящиеся сюда более ранние исследования содержатся в работах K. S. WISE [1] (стр. 272—302), C. N. MOOERS ([2], [3], [4]) авторов ([5], [6]). Работы [1], [2], [3] содержат приближенные решения. Точным решением проблемы занимается работа [4], но она не содержит явных формул, а дает лишь рекурсивный метод для их вычисления. Явное решение возникших вопросов в общем случае дается в предыдущих работах авторов [5] и [6]. Настоящая работа содержит результаты работы [5].

Документация со случайными кодами производится следующим образом. О каждом документе изготавливается перфокарта, на которой отмечается N типичных для данного документа понятий. Предположим, что данная перфокарта содержит n позиций (n возможностей перфорации), каждому понятию соотносится некоторый код с r элементами и код обозначается на перфокарте так, что соответствующие коду элементы перфорируются. Мы говорим о *неделимом поле знаков*, если элементы кода могут любым образом выбираться среди n позиций. Тогда число возможных кодов $W = \binom{n}{r}$. Мы говорим о *делимом поле знаков*, если поле знаков разбито на p одинаковых подполей, состоящих из m позиций, и элементы кода содержат по q позиций из каждого подполя. Тогда $n = mp$ и $r = pq$. Число всевозможных кодов в этом случае $W = \binom{m}{q}^p$. Предположим, что к отдельным понятиям соответствующий код соотносится случайным выбором из всех возможных кодов. Нужно различать два случая: *при точном рассмотрении* предполагается, что к новому понятию коды соотносятся выбором без возвращения, в то время как *при приближенном рассмотрении* считается, что соотношение кодов производится выбором с возвращением.

При сделанных выше предположениях определяется 1° Число перфорированных позиций на перфокарте, распределение η , т. е. вероятности $P\{\eta = k\} = P_k$, математическое ожидание $M\{\eta\}$, дисперсия $D^2\{\eta\}$ и моменты $M\{\eta^s\}$. 2° Пусть переменное χ обозначает, сколько раз перфорирована выбранная на перфокарте позиция. Определяем распределение кратности перфорации χ , т. е. вероятности $P\{\chi = j\} = p_j$, математическое ожидание $M\{\chi\}$ и дисперсию $D^2\{\chi\}$. 3° Наконец, определяем $R_s (s = 1, 2, \dots, N)$ вероятность того, что одновременно проводя селектацию по s кодам, выберем такую перфокарту, на которой не отмечены все s кодов, служащие основой для селектации. Коротко говоря, s вероятность исключительных перфокарт.

Упомянутые выше вероятности определяются следующим образом: вводятся события $A_1, A_2, \dots, A_n$, обозначающие соответственно тот факт, что на перфокарте позиции $1, 2, \dots, n$ не перфорированы, и к этим событиям применяется общая теорема К. ЖОРДАН [7]. (§ 2.)

В дальнейшем перечисляются относящиеся к вероятностям P_k, p_k, R_1 и R_s формулы для следующих четырех случаев: а) неделимое поле знаков, точная модель, б) неделимое поле знаков, приближенная модель, в) делимое поле знаков, точная модель, г) делимое поле знаков, приближенная модель (§§ 3—4).

Относительно P_k : а) (10), б) (11), в) (20), где t определяется (21), г) (20), где B_j дается (22).

Относительно p_j : а) (12), б) (13), в) (23), г) (24).

Относительно R_1 : а) (16), где Q_r дается (14), б) (17), где Q_r определяется (15), в) (29), где Q_r дается (25) и T_1 (27), г) (29), где Q_r определяется (26), а T_1 (28).

Относительно R_s : а) (18), где Q_μ дается (14), б) (19), где Q_μ определяется (15), в) (34), где Q_μ дается (30), а T_s (32), г) (34), где определяется Q_μ (31) и T_s (33).

Зная эти формулы, можно вместо применявшихся до сих пор приближенных формул применять точные формулы при планировании документационных систем и создать таким образом лучшие условия.

ÜBER GEWISSE MATHEMATISCHE PROBLEME DER DOKUMENTATIONSSYSTEME MIT ZUFALLSARTIG GEWÄHLTEN SUPERPONIERTEN KODEN

G. OROSZ und L. TAKÁCS

Zusammenfassung

In der Einführung werden die verschiedenen Dokumentationssysteme besprochen. §. 1. behandelt die mathematischen Probleme der zufallsartigen Superpositionssysteme. Diesbezügliche frühere Untersuchungen sind in den Arbeiten von C. S. WISE ([1], pp. 276—302), C. N. MOOERS ([2], [3], [4]) und denjenigen der Verfasser ([5], [6]) enthalten. Die Arbeiten [1], [2], [3] enthalten nur angenäherte Lösungen. Die exakte Lösung des Problems wurde in der Arbeit [4] mitgeteilt; sie enthält aber keine explizite Formel, nur eine rekursive Methode zur Berechnung derjenigen. Die

expliziten Lösungen der wichtigsten Fragen sind für den allgemeinen Fall in den früheren Arbeiten ([5], [6]) der Verfasser enthalten. Die gegenwärtige Arbeit reproduziert die Ergebnisse von [5].

Die Dokumentation durch zufallsartige Kodern geschieht folgendermassen. Von jedem Dokument wird eine Unterlage verfertigt, auf welche N Begriffe, die das betreffende Dokument charakterisieren, eingezeichnet werden. Es sei angenommen, dass die gegebene Karte n Positionen (n Durchlochungsmöglichkeiten) enthält, zu jedem einzelnen Begriffe ein Kod von v Elementen angeordnet wird, und die Kode auf der Unterlage derart eingezeichnet wird, dass die dem Elementen der Kode entsprechenden Positionen ausgelocht werden. Man spricht von einem *ungeteilten Kodefeld* wenn die Elemente der Kode aus den n Positionen beliebig ausgewählt werden können. Dann

wird die Anzahl der möglichen Kode $W = \binom{n}{v}$ sein. Man spricht von einem *geteilten Kodefeld*, wenn das Kodefeld auf p gleiche Teilfelde mit je m Positionen geteilt ist und die Elemente der Kodern q Positionen aus jedem Teilfelde enthalten. Dann wird $n = mp$ und $v = pq$ sein. Die Anzahl aller möglichen Kodern wird dann $W = \binom{m}{q}^p$ sein. Es wird

angenommen, dass die passende Kode durch zufallsartige Ziehen aus allen Kodern zu den einzelnen Begriffen zugeordnet werden. Es werden zwei Fälle unterschieden: Bei der *exakten Behandlung* nimmt man an, dass die Kodern durch Ziehen ohne Zurückwerfen zu den neuen Begriffen zugeordnet werden; bei der *angenäherten Behandlung* wird es angenommen, dass die Wahl der Kodern durch Ziehen mit Zurückwerfen geschieht.

Neben den obigen Voraussetzungen wird folgendes bestimmt: 1° Die Verteilung der Anzahl η der ausgelochten Positionen auf einer Karte, d. h. die Wahrscheinlichkeiten $P\{\eta = k\} = P_k$, der Erwartungswert $M\{\eta\}$, die Streuungsquadrat $D^2\{\eta\}$ und die Momente $M\{\eta^s\}$. 2° Es bezeichne die Veränderliche χ wie vielmal eine ausgelesene Position auf einer Karte gelocht wurde. Die Verteilung der Multiplizität der Auslochung χ , d. h. die Wahrscheinlichkeiten $P\{\chi = j\} = p_j$, der Erwartungswert $M\{\chi\}$ und der Streuungsquadrat $D^2\{\chi\}$ werden berechnet. 3° Endlich wird R_s ($s = 1, 2, \dots, N$) berechnet, R_s ist die Wahrscheinlichkeit, dass — bei einer Selektierung nach s gleichzeitigen Kodern — eine Karte ausgewählt wird, auf welcher nicht alle s Kodern, die den Grund für die Selektierung bieten, eingezeichnet sind. Kurz: R_s ist die Wahrscheinlichkeit für extra Karten.

Die obigen Wahrscheinlichkeiten werden folgendermassen bestimmt: Es seien die Ereignisse $A_1, A_2, \dots, A_n$ eingeführt, welche der Reihe nach bezeichnen, dass die 1, 2, ..., n -te Position auf einer Karte nicht ausgelocht wurde. Auf diese Ereignisse wird dann der allgemeine Wahrscheinlichkeitssatz von KARL JORDAN [7] angewandt (§. 2.).

Im darauffolgenden Teil werden Formeln bezüglich der Wahrscheinlichkeiten P_k, p_j, R_1 und R_s für die folgenden vier Fälle aufgeführt: a) Ungeteiltes Kodefeld, exaktes Modell, b) ungeteiltes Kodefeld, angenähertes Modell, c) geteilte Kodefeld, exaktes Modell, d) geteiltes Kodefeld, angenähertes Modell (§. 3.—4.).

Bezüglich P_k : a) (10), b) (11), c) (20), wobei B_j durch (21) definiert wird, d) (20), wobei B_j durch (22) definiert wird.

Bezüglich p_j a) (12), b) (13), c) (23), d) (24).

Bezüglich R_1 a) (16), Q_v wird durch (14) definiert, b) (17), Q_v wird durch (15) definiert, c) (29) wo Q_v und T_1 werden durch (25), bzw. (27) bestimmt, d) (29) wobei Q_v und T_1 durch (26) bzw. (28) bestimmt wird.

Bezüglich R_s a) (18) wobei Q_μ durch (14) gegeben wird, b) (19) wobei Q_μ durch (15) bestimmt wird, c) (34), wobei Q_μ und T_s durch (30) bzw. (32) gegeben werden, d) (34), wobei Q_μ und T_s durch (31) bzw. (33) bestimmt werden.

Das Kenntnis der obigen Formeln gibt Möglichkeit exakte Formeln anstatt der bisher angewandten Näherungsformeln beim Planen von Dokumentationssystemen anzuwenden und dadurch bessere Verhältnisse zu schaffen.