EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR

Kupongyűjtő probléma

Bsc Szakdolgozat

Tótok Barbara

Matematika Bsc
Matematikai elemző szakirány

Témavezető:

Csiszár Villő
Valószínűségelméleti és Statisztika tanszék

Budapest
2017

Tartalomjegyzék

1. Bevezetés

A kupongyűjtő probléma a valószínűségszámítás egy klasszikus problémája, ami szorosan kapcsolódik az urna problémákhoz.

A probléma legegyszerűbb és valószínűleg eredeti megfogalmazása a következő: Tegyük fel, hogy N darab kuponból véletlenszerűen visszatevéssel húzunk. Mennyi annak a valószínűsége, hogy több, mint k húzásra van szükségünk ahhoz, hogy megszerezzük mind az N kupont. A kérdés egy érdekes általánosítása, hogy várhatóan hány húzásra lesz szükségünk, hogy megszerezzük mind az N kupont.

A probléma először 1708-ban jelent meg az irodalomban A. De Moivre francia matematikus írásában. 1954-ben H.VonSchelling határozta meg a teljes kollekció összegyűjtésének várható idejét abban az esetben amikor a kuponok valószínűsége nem egyenlő, majd 1960-ban D. J. Newman és L. Shepp meghatározták több kollekció összegyűjtésének várható idejét azonos valószínűségű kuponok esetében.

A problémának a tényleges kollekciók gyűjtésén kívül számos alkalmazása van, különösen villamosmérnöki problémákban. Használható elektronikai és gyorsírótár hibák keresésére, de a biológiában is használják állatfajok számának becslésére. Véletlen nagyságú mintavétel esetében akár betegségek terjedésének vagy bankjegy csere programok várható idejének számítására is használható.

Szakdolgozatomban a várható érték számolásának ismert módszereit fogom bemutatni. Az 2. és 3. fejezetben ezt arra az esetre vizsgálom, amikor egy kollekció összegyűjtése a cél, amiben a különböző típusú kuponok azonos, illetve különböző valószínűséggel fordulnak elő. Az 5. és 6. fejezetben pedig több kollekció összegyűjtését tűzöm ki célul és arra mutatok néhány megoldási lehetőséget. A 7. fejezetben röviden visszatérek az eredeti kérdéshez, hogy fel tudjam használni a képletet a 8. fejezetben, ahol kicsit visszatérve a gyerekkorba egy tényleges kupongyűjtésen keresztül mutatom be, hogy mit jelentenek a fenti képletek számszerűsítve. Zárásként a 9. fejezetben szeretnék megmutatni egyet a kupongyűjtő probléma nem annyira triviális felhasználási lehetőségei közül.

2. Egy kollekció azonos valószínűségekkel

Az alapfeladat: Tegyük fel, hogy van egy kollekció ami N különböző kuponból áll. Az egyszerűség kedvéért ezek $1, 2, ..., N$. Minden csomagban pontosan egy kupon van, amik azonos valószínűséggel fordulnak elő. A kupongyűjtő addig vásárol újabb és újabb csomagokat, amíg teljes lesz a kollekció.

A [6] alapján két módszert fogok bemutatni, amelyekkel kiszámolhatjuk, hogy várhatóan hány csomagot kell vásárolni egy teljes kollekció összegyűjtéséhez.

2.1. Közelítés geometriai eloszlással

Vezessük be a következő jelöléseket: Jelölje X a csomagok számát amit meg kell vennünk, hogy összegyűjtsük a teljes kollekciót és X_i a további szükséges csomagok számát, hogy $(i-1)$ összegyűjtött kupon után egy i. különbözőt gyűjtsünk össze. Így X felírható $X = X_1 + X_2 + ... + X_N$ alakban.
Mivel feltettük, hogy a kuponok azonos valószínűséggel fordulnak elő, így triviálisan $p_1 = 1$ és az i. kupon megtalálása után $\frac{N-i}{N}$ valószínűséggel találunk új kupont. Általánosan minden $i = 1, 2, ..., N$-re X_i független geometriai eloszlású valószínűségi változó $p_i = \frac{N-i+1}{N}$ paraméterrel. Így a várható értékéről tudjuk, hogy $\mathrm{E}(X_i) = \frac{1}{p_i}$.
Ezekből már következik, hogy a teljes kollekció összegyűjtéséhez szükséges csomagok várható értéke:

$$\mathrm{E}(X) = \mathrm{E}(X_1) + \mathrm{E}(X_2) + ... + \mathrm{E}(X_N)$$

$$= 1 + \frac{N}{N-1} + \frac{N}{N-2} + ... + \frac{N}{1} = N \sum_{i=1}^{N} \frac{1}{i} = NH_N, \qquad (1)$$

ahol H_n az n. harmonikus szám.

Ahogy azt a következő táblázatban láthatjuk ezek jelentős mennyiségek lesznek, hiszen már egy 5 elemű kollekciónál is várhatóan kétszer annyi csomagot kell majd megszereznünk, mint a kollekció mérete. Egy 50 elemű kollekciónál pedig már várhatóan 4,5-szer annyi csomagra lesz szükségünk, mint a kollekció mérete.

N	5	10	15	20	25	30	40	50
$\mathrm{E}(X)$	11,43	29,3	49,78	71,96	95,40	119,85	171,14	224,96

Nézzük meg ezt az 2.1. ábrán is. Láthatjuk, hogy a várható érték mennyivel nagyobb ütemben növekszik, mint a kollekció mérete:

1. ábra.

2.2. Közelítés Markov-lánccal

A Markov-lánc egy olyan diszkrét sztochasztikus folyamat amelyben a múltbéli események csak a jelenen keresztül állnak kapcsolatban a jövővel. A kupongyűjtő problémában a jövőt csak az befolyásolja, hogy a jelenre hány különböző kupont tudtunk már összegyűjteni, ezért tudjuk közelíteni Markov-lánc segítségével.

Tegyük fel, hogy bármely időegységben egy csomagot vásárolunk. Jelölje X_i a várakozási időt az i. kuponra, miután $(i-1)$ -et már összegyűjtöttünk és Y_n a különböző kuponok számát n időegység elteltével. A kuponok továbbra is azonos valószínűséggel fordulnak elő, így $p = \frac{1}{N}$. Ekkor Y_n Markov-lánc $S = \{0, 1, ..., N\}$ állapothalmazzal. Mivel Markov-lánc, ábrázolhatjuk irányított gráffal, ahol a csúcsok az állapotok, két csúcsot összekötő él súlya pedig az egyik állapotból a másikba kerülés valószínűsége. Így Y_n átmenetmátrixa a következő:

$$\begin{bmatrix} 0 & 1 & 0 & \dots & \dots & \dots & \dots & 0 \\ 0 & \frac{1}{N} & \frac{N-1}{N} & 0 & \dots & \dots & \dots & 0 \\ 0 & 0 & \frac{2}{N} & \frac{N-2}{N} & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \dots & \frac{N-3}{N} & \frac{3}{N} & 0 & 0 \\ 0 & 0 & \dots & \dots & \dots & \frac{N-2}{N} & \frac{2}{N} & 0 \\ 0 & 0 & \dots & \dots & \dots & \dots & \frac{N-1}{N} & \frac{1}{N} \\ 0 & 0 & \dots & \dots & \dots & \dots & 0 & 1 \end{bmatrix}$$

A kollekció összegyűjtéséhez szükségek csomagok számának várható értéke meg fog egyezni annak a várható idejével amikor Y_n eléri az N állapotot. Tehát a következő egyenletrendszert kell megoldanunk:

$$\begin{cases} k_N = 0 \\ k_i = 1 + \sum_{j \neq N} p_{ij} k_j \quad i \neq N \end{cases}$$

$$\begin{cases} k_0 = 1 + k_1 \\ k_1 = 1 + \frac{1}{N} k_1 + \frac{N-1}{N} k_2 \\ k_2 = 1 + \frac{2}{N} k_2 + \frac{N-2}{N} k_3 \\ \vdots \\ k_{N-3} = 1 + \frac{N-3}{N} k_{N-3} + \frac{3}{N} k_{N-2} \\ k_{N-2} = 1 + \frac{N-2}{N} k_{N-2} + \frac{2}{N} k_{N-1} \\ k_{N-1} = 1 + \frac{N-1}{N} k_{N-1} \\ k_N = 0 \end{cases} \Leftrightarrow \begin{cases} k_0 = 1 + k_1 \\ \frac{N-1}{N} k_1 = 1 + \frac{N-1}{N} k_2 \\ \frac{N-2}{N} k_2 = 1 + \frac{N-2}{N} k_3 \\ \vdots \\ \frac{3}{N} k_{N-3} = 1 + \frac{3}{N} k_{N-2} \\ \frac{2}{N} k_{N-2} = 1 + \frac{2}{N} k_{N-1} \\ \frac{1}{N} k_{N-1} = 1 \\ k_N = 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} k_0 = 1 + k_1 \\ k_1 = \frac{N}{N-1} + k_2 \\ k_2 = \frac{N}{N-2} + k_3 \\ \vdots \\ k_{N-3} = \frac{N}{3} + k_{N-2} \\ k_{N-2} = \frac{N}{2} + k_{N-1} \\ k_{N-1} = N \\ k_N = 0 \end{cases} \Leftrightarrow \begin{cases} k_0 = N + \frac{N}{2} + \frac{N}{3} + \dots + \frac{N}{N-3} + \frac{N}{N-2} + \frac{N}{N-1} + 1 \\ k_1 = N + \frac{N}{2} + \frac{N}{3} + \dots + \frac{N}{N-3} + \frac{N}{N-2} + \frac{N}{N-1} \\ k_2 = N + \frac{N}{2} + \frac{N}{3} + \dots + \frac{N}{N-3} + \frac{N}{N-2} \\ \vdots \\ k_{N-3} = N + \frac{N}{2} + \frac{N}{3} \\ k_{N-2} = N + \frac{N}{2} \\ k_{N-1} = N \\ k_N = 0 \end{cases}$$

Ebből következik, hogy a várakozási idő N kupon összegyűjtéséhez:

$$k_0 = N \sum_{i=1}^{N} \frac{1}{i} = NH_N$$

2.3. Közelítő formula

Kis N-re a fenti formula könnyen és gyorsan számolható. Nagy N-re pedig használhatjuk a következő közelítést:

$$H_N = \sum_{i=1}^{N} \frac{1}{i} = \log(N) + \gamma + \frac{1}{2N} + O\left(\frac{1}{N^2}\right),$$

ahol $\gamma \approx 0,5772156649$ az Euler-Mascheroni állandó. Így nagy N-ekre:

$$\mathrm{E}(X) = N\log(N) + N\gamma + \tfrac{1}{2} + O\big(\tfrac{1}{N}\big),\ N \to \infty$$

2.4. Variancia

X értékére varianciát is számolhatunk. Ezzel azt fogjuk vizsgálni, hogy egy-egy gyűjtés során ténylegesen megvásárolt csomagok száma milyen mértékben szóródik a várhatóan szükséges csomagok számától.

Mivel $X = X_1 + X_2 + \cdots + X_N$, ahol X_i független geometriai eloszlású változók $p_i = \frac{N-i+1}{N}$ paraméterrel és tudjuk, hogy egy p paraméterű geometriai valószínűségi változó szórásnégyzete $\frac{1-p}{p^2}$, ezért X_i varianciáját a következő alakban kapjuk:

$$\mathrm{D}^2(X_i) = \frac{1 - \frac{N-i+1}{N}}{(\frac{N-i+1}{N})^2} = \frac{i-1}{N}\frac{N^2}{(N-i+1)^2} = \frac{N(i-1)}{\big(N-(i-1)\big)^2}$$

Így már könnyen meghatározhatjuk X varianciáját:

$$\mathrm{D}^2(X) = \sum_{i=1}^{N} \frac{N(i-1)}{\big(N-(i-1)\big)^2} = \sum_{i=1}^{N} \frac{N(N-i)}{\big(N-(N-i)\big)^2}$$

$$\sum_{i=1}^{N} \frac{N(N-i)}{i^2} = N^2 \sum_{i=1}^{N} \frac{1}{i^2} - N \sum_{i=1}^{N} \frac{1}{i} = N^2 \sum_{i=1}^{N} \frac{1}{i^2} - NH_N \qquad (2)$$

Mivel tudjuk, hogy

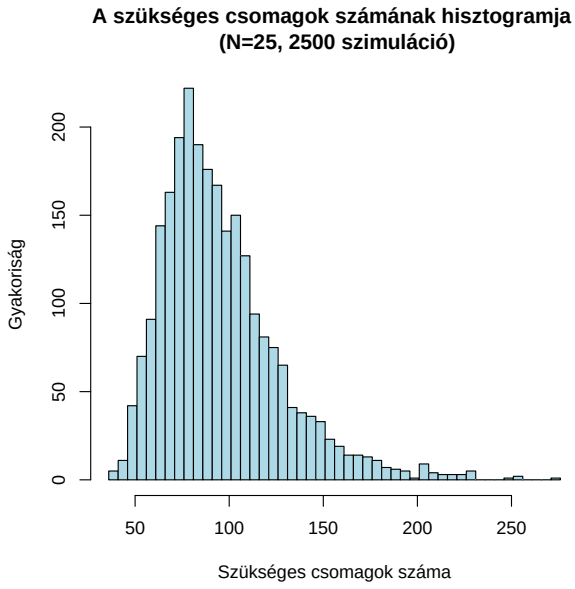
$$\sum_{1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6},$$

ezért a következő módon közelíthetjük a fenti képletet nagy N értékekre:

$$\mathrm{D}^2(X) = \frac{\pi^2}{6}N^2 - N\log(N) - N\gamma - N + O(N) \tag{3}$$

Ez meglehetősen nagy variancia, ezért nagy különbségek lehetnek gyűjtés és gyűjtés között. Nézzünk meg két szimulációt, hogy kicsit pontosabb képet kapjunk a szükséges csomagok számának alakulásáról.

Először az $N = 25$ esetet vizsgáltam. 2500 alkalommal futtattam le a szimulációt, amely során a teljes kollekció összegyűjtéséhez szükséges csomagok számát vizsgáltam. A 2.4. ábrán a szimuláció során kapott értékek hisztogramját láthatjuk. A szimuláció során az értékek átlaga 95,492 volt, ez közel van a $25H_{25} = 95,399$ értékhez. A kapott értékek minimuma 36, míg maximuma 274 volt. Ahogy azt sejtetni lehetett nagy lett a minta terjedelme. A (2) képlettel számolva a szimuláció értékeinek szórása 30,136. Megvizsgálva a mintát azt tapasztaltam, hogy az elemek $72,48\%$-a a $[65, 126]$ intervallumon belül esett, azaz nagy valószínűséggel a szóráson belül marad a teljes kollekció összegyűjtéséhez szükséges csomagok száma. Ezután az $N = 40$

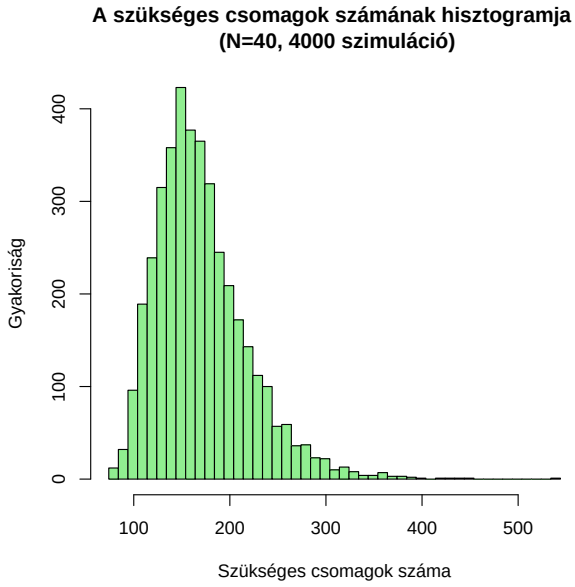


2. ábra.

esetet vizsgáltam meg az előzőhöz hasonlóan. Ebben az esetben 4000 szimulációt hajtottam végre. A kapott értékek hisztogramja a 2.4. ábrán látható. Ekkor az értékek átlaga 171,16 volt ami szintén közel van az $50H_{50} = 171,14$

értékhez. Az értékek minimuma 74, míg maximuma 544 volt és itt is látszik, hogy mennyire nagy a minta terjedelme. A szórás 49,2 volt, viszont itt is azt tapasztaltam, hogy az értékek $72,72\%$-a a szóráson azaz, az $[121, 221]$ intervallumon belül maradt.

Mivel nagyon közeli lett a két eredmény, megvizsgáltam hátha igaz marad más N értékekre is. Kiszámoltam $N = 2, 3, \ldots, 50$ méretű kollekcióra, hogy az 1000 szimulációval kapott értékek hány százaléka marad a szóráson belül és vettem a kapott eredmények átlagát, amely igazolta a sejtést. Tehát várhatóan az esetek 73%-ban a kollekció összegyűjtéséhez szükséges csomagok számának eltérése a várható értéktől a szórás értékén belül marad.



3. ábra.

3. Egy kollekció eltérő valószínűségekkel

Az alapfeladat hasonló, mint az előző esetben. Továbbra is egy N kuponból álló kollekciót szeretnénk összegyűjteni. Csupán annyit változtatunk rajta, hogy a kuponok nem azonos valószínűséggel fordulnak elő. Tehát $\forall i$-re az i kupon előfordulásának valószínűsége $p_i \geq 0$ és $p_1 + p_2 + \cdots + p_N = 1$.

3.1. Közelítés a Maximum-Minimum azonossággal

Legyen X_i a szükséges csomagok száma amit meg kell vásárolnunk, hogy megkapjuk az első i típusú kupont. Így a teljes kollekció összegyűjtéséhez szükséges csomagok száma felírható $X = \max\{X_1, X_2, \ldots, X_N\}$ alakban, ahol X_i geometriai eloszlású változó p_i paraméterrel. Viszont az előző fejezethez képest ezek a valószínűségi változók már nem függetlenek. Mivel $\min\{X_i, X_j\}$ megegyezik a csomagok számával amit szükséges megvásárolnunk, hogy összegyűjtsük az első i vagy j típusú kupont, ezért $i \neq j$-re $\min\{X_i, X_j\}$ egy geometriai eloszlású változó $p_i + p_j$ paraméterrel. Hasonlóan $\min\{X_i, X_j, X_k\}$ megegyezik a szükséges csomagok számával ami az első i, j vagy k kupon valamelyikének összegyűjtéséhez szükséges, ez szintén geometriai eloszlású változó, de $p_i + p_j + p_k$ paraméterrel és így tovább. X várható értékének meghatározásához M. Ferrante, M. Saltalamacchia [6]-ban a Maximum-Minimum azonosságot használja:

$$\max_{i=1,\ldots,N} X_i = \sum_i X_i - \sum_{i<j} \min(X_i, X_j) + \sum_{i<j<k} \min(X_i, X_j, X_k) - \ldots$$

$$\cdots + (-1)^{N+1} \min(X_1, X_2, \ldots, X_N)$$

Ezt az egyenlőséget S.Ross a következő módon bizonyította [5]-ben:
Először is tegyük fel, hogy $X_i \in [0, 1]$ $\forall i$. Később ebből az esetből általánosítunk, de először bizonyítsuk be erre. Ekkor legyen U egy $(0, 1)$ közötti egyenletes eloszlású változó. Minden i értékre jelölje A_i azt az eseményt amikor az U kisebb, mint az X_i, azaz $A_i = \{U < X_i\}$. Mivel tudjuk, hogy ha U kisebb legalább az egyik X_i értéknél, akkor elő fog fordulni legalább egy A_i esemény, ezért:

$$\cup_i A_i = \{U < \max_i X_i\}$$

Így felírhatjuk a következő két egyenlőséget:

$$\mathrm{P}(\cup_i A_i) = \mathrm{P}\big\{U < \max_i X_i\big\} = \max_i X_i$$

$$\mathrm{P}(A_i) = \mathrm{P}(U < X_i) = X_i$$

Ugyanakkor mivel ha U kisebb minden $X_{i_1}, X_{i_2}, \ldots, X_{i_r}$ értéktől, akkor minden $A_{i_1}, A_{i_2}, \ldots, A_{i_r}$ esemény elő fog fordulni, így az események metszete az előzőhöz hasonlóan felírható a következő alakban:

$$A_{i_1} A_{i_2} \ldots A_{i_r} = \{U < \min_{j=1,2,\ldots,r} X_{i_j}\}$$

$$P(A_{i_1} A_{i_2} \ldots A_{i_r}) = \mathrm{P}\big\{U < \min_{j=1,2,\ldots,r} X_{i_j}\big\}$$

Ezek ismeretében felírhatjuk a Szita-formulával az események uniójának valószínűségét:

$$\mathrm{P}(\cup_i A_i) = \sum_i \mathrm{P}(A_i) - \sum_{i<j} \mathrm{P}(A_i A_j) + \sum_{i<j<k} \mathrm{P}(A_i A_j A_k) + \dots$$

$$\dots + (-1)^{N+1} \mathrm{P}(A_1 A_2, \dots, A_N)$$

Ezzel bebizonyítottuk a $X_i \in [0,1]\ \forall i$ esetet. Most tegyük fel, hogy $X_i > 1$ értéket is felvehet, ám továbbra sem lehet negatív. Ekkor vezessünk be egy c konstanst, amelyre teljesül, hogy $X_i < c\ \forall i$. Ekkor a fenti azonosság igaz marad $y_i = \frac{X_i}{c}\ \forall i$ értékekre és a kívánt eredményt a végén c-vel való szorzással kapjuk. Most már csak az az eset maradt amikor X_i értéke akár negatív is lehet. Ekkor vezessünk be egy b konstans értéket, hogy $X_i + b > 0\ \forall i$. Így a korábbi egyenlet a következő módon alakul:

$$\max_i (X_i + b) = \sum_i (X_i + b) - \sum_{i<j} \min(X_i + b, X_j + b) + \dots$$

$$\dots + (-1)^{N+1} \min(X_1 + b, X_2 + b, \dots, X_N + b)$$

Legyen

$$M = \sum_i X_i - \sum_{i<j} \min(X_i, X_j) + \sum_{i<j<k} \min(X_i, X_j, X_k) - \dots$$

$$\dots + (-1)^{N+1} \min(X_1, X_2, \dots, X_N)$$

Így

$$\max_i X_i + b = M + b\left(N - \binom{N}{2} + \dots + (-1)^{N+1}\binom{N}{N}\right),$$

de

$$0 = (1-1)^N = 1 - N + \binom{N}{2} + \dots + (-1)^{N+1}\binom{N}{N}$$

A fenti két egyenletből pedig már következik, hogy $\max_i X_i = M$. $\square$

Most ezek alapján írjuk fel az $\mathrm{E}[X]$-et a képlet segítségével:

$$\mathrm{E}[X] = \mathrm{E}[\max_{i=1,\dots,N} X_i] = \sum_i \mathrm{E}[X_i] - \sum_{i<j} \mathrm{E}[\min(X_i, X_j)] +$$

$$+ \sum_{i<j<k} \mathrm{E}[\min(X_i, X_j, X_k)] - \dots + (-1)^{N+1} \mathrm{E}[\min(X_1, X_2, \dots, X_N)] =$$

$$= \sum_i \frac{1}{p_i} - \sum_{i<j} \frac{1}{p_i + p_j} + \sum_{i<j<k} \frac{1}{p_i + p_j + p_k} - \cdots + (-1)^{N+1} \frac{1}{p_1 + p_2 + \cdots + p_N} \tag{4}$$

Tudjuk, hogy

$$\int_0^{+\infty} e^{-px} dx = -\frac{e^{-px}}{p}\Big|_0^{+\infty} = \lim_{a \to +\infty} -\frac{e^{-pa}}{p} + \frac{e^{-0}}{p} = \frac{1}{p}$$

és ismerjük a következő azonosságot:

$$1 - \prod_{i=1}^{N} (1 - e^{-p_i x}) = \sum_i e^{-p_i x} - \sum_{i<j} e^{-(p_i + p_j)x} + \cdots + (-1)^{N+1} e^{-(p_1 + p_2 + \cdots + p_N)x}$$

Így a szükséges csomagok várható értékét fel tudjuk írni a következő alakban:

$$\mathrm{E}[X] = \int_0^{+\infty} \Big[1 - \prod_{i=1}^{N} (1 - e^{-p_i x})\Big] dx$$

4. Egyenlő- és különböző valószínűségű esetek összehasonlítása

Ebben a fejezetben azt szeretném megmutatni, hogy a különböző valószínűségű esetben, várhatóan mindig több csomagot kell összegyűjtenünk a hiánytalan kollekcióhoz.

Jelöljük $X_{(\frac{1}{N},\ldots,\frac{1}{N})}$-el a csomagok számát amit meg kell vásárolni, ha a kuponok azonos valószínűséggel fordulnak elő a csomagokban és $X_{(p_1,p_2,\ldots,p_N)}$-el a csomagok számát amit meg kell vásárolnunk ha a különböző kuponok $p_1, p_2, \ldots, p_N$ valószínűséggel fordulnak elő. (2) és (4) alapján:

$$E[X_{(\frac{1}{N},\ldots,\frac{1}{N})}] = N \sum_{i=1}^{N} \frac{1}{i}$$

$$E[X_{(p_1,p_2,\ldots,p_N)}] = \sum_i \frac{1}{p_i} - \sum_{i<j} \frac{1}{p_i + p_j} + \cdots + (-1)^{N+1} \frac{1}{p_1 + p_2 + \cdots + p_N}$$

A valószínűségek eloszlása legyen $p = (p_1, p_2, \ldots, p_N)$ és $p_{[j]}$ legyen a j. legnagyobb érték $\{p_1, p_2, \ldots, p_N\}$ közül, ekkor $p_{[1]} \geq p_{[2]} \geq \cdots \geq p_{[N]}$.
Azt mondjuk, hogy $p = (p_1, p_2, \ldots, p_N)$ **majorálja** $q = (q_1, q_2, \ldots, q_N)$-t, ha teljesül, hogy

$$\sum_{i=1}^{k} q_{[i]} \leq \sum_{i=1}^{k} p_{[i]} \qquad \forall\, 1 \leq k \leq (N-1).$$

Jelölés: $q \prec p$
[6]-ban bebizonyították, hogy $(\frac{1}{N}, \ldots, \frac{1}{N}) \prec p$ minden $p = (p_1, p_2, \ldots, p_N)$ eloszlásra. Először is $\frac{1}{N} \leq p_{[1]}$ biztosan teljesül. Ezután indirekten tegyük fel, hogy

$$\exists\, 1 \leq k \leq (N-1)\text{ , hogy} \quad \frac{k}{N} > \sum_{i=1}^{k} p_{[i]}. \tag{5}$$

Mivel $\frac{1}{N} + \cdots + \frac{1}{N} = 1 = p_1 + \cdots + p_N$, ekkor teljesül, hogy

$$\frac{N-k}{N} < \sum_{i=k+1}^{N} p_{[i]}$$

amiből pedig következik, hogy $\exists\, j \in \{k+1, k+2, \ldots, N\}$, hogy $\frac{1}{N} < p_{[j]}$, ami viszont ellentmond a (5) állításnak, amiből következett, hogy $\exists\, l \in \{1, 2, \ldots, k\}$, hogy $\frac{1}{N} > p_{[l]}$.

Azt mondjuk, hogy egy eloszláson definiált $f(p)$ szimmetrikus függvény:

i **Schur konvex**, ha $p \prec q \implies f(p) \leq f(q)$.

ii **Schur konkáv**, ha $p \prec q \implies f(p) \geq f(q)$.

Azt mondjuk, hogy X valószínűségi változó **sztochasztikusan kisebb**, mint Y valószínűségi változó, ha $\mathrm{P}(X > a) \leq \mathrm{P}(Y > a) \quad \forall\, a \in \mathbb{R}$.

T. Nakata [9]-ben a következő tételt bizonyította:

4.1. Tétel. $\mathrm{P}(X_p \leq n)$ *egy Schur konkáv függvénye p-nek.*

4.1.1. Következény. *Ha $p \prec q$, akkor X_p sztochasztikusan kisebb, mint X_q.:*

$$\mathrm{P}(X_p \leq n) \geq \mathrm{P}(X_q \leq n)$$

$$1 - \mathrm{P}(X_p \leq n) = \mathrm{P}(X_p > n) \leq \mathrm{P}(X_q > n) = 1 - \mathrm{P}(X_q \leq n)$$

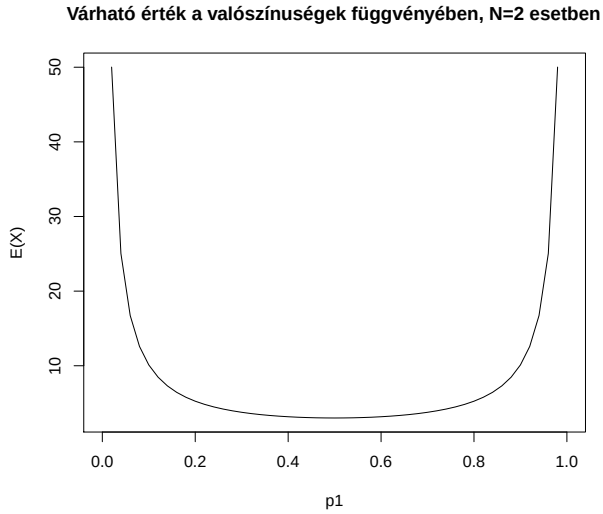
Speciálisan $X_{(\frac{1}{N}, \ldots, \frac{1}{N})}$sztochasztikusan kisebb, mint X_p minden p-re.

4.1.2. Következény. *Ha $p \prec q$, akkor $E(X_p)$ Schur konvex függvénye p-nek.:*

$$E[X_p] = \sum_{n=0}^{\infty} P(X_p > n) \leq \sum_{n=0}^{\infty} P(X_q > n) = E(X_q)$$

Speciálisan $E(X_{(\frac{1}{N},\ldots,\frac{1}{N})}) \leq E(X_{(p_1,p_2,\ldots,p_N)})$ minden p-re.

Az 4. ábrán láthatjuk, hogy az $N = 2$ esetben, hogyan alakul a várhatóan szükséges csomagok száma ha elkezdjük kimozdítani a $p_1 = p_2$ esetből. Az ábrán a legnagyobb ábrázolt érték a $p_1 = 0,02$, $p_2 = 0,98$, a p_1 értéket tovább csökkentve $E(X)$ a végtelenhez tart:



4. ábra.

5. Több kollekció azonos valószínűséggel

Az előző esetekben láthattuk, hogy viszonylag nagy kollekcióknál még az azonos valószínűségű esetben is várhatóan sok kupont kell megvásárolnunk, hogy hiánytalan legyen a kollekciónk. Nyilvánvalóan kevesebb lesz az egy főre jutó kuponok száma, ha többen gyűjtenek össze egy kollekciót, persze ez nem egy ideális eset, hiszen akkor nem lesz mindenkinek sajátja. Ezért most vizsgáljuk azt az esetet, amikor többen gyűjtenek több kollekciót. Legyen m testvér, akik össze szeretnének gyűjteni m teljes kollekciót. Ennek a várható értéke kisebb lesz, mint m-szer az 1 kollekciós eset várható értéke.

5.1. Közelítés Poisson-folyamattal

Lars Holst [8]-ban a Poisson-folyamat segítségével határozta meg a várható értéket.

5.1.1. Beágyazás a Poisson-folyamatba

Tekintsünk egy $\wp$ Poisson-folyamatot 1 intenzitással, $Z_1, Z_2, Z_3, \ldots$ érkezések közötti időkkel. A Z-k legyenek független azonos exponenciális eloszlású valószínűségi változók: $Z_i \sim \mathrm{Exp}(1)\ \forall i = 1, 2, 3, \ldots$. Azt mondjuk, hogy minden érkezéskor egy esemény történik. A $\wp$ folyamat eseményeihez $I_1, I_2, \ldots$ valószínűségi változókat rendelünk. Ezek függetlenek egymástól és az érkezések közötti időktől is, emellett a következő eloszlás jellemző rájuk: $P(I = j) = p_j \quad \forall j = 1, 2, \ldots, N$. A $\wp$ Poisson-folyamat azon eseményei, melyekre $I = j$ teljesül, $\wp_j$ Poisson-folyamatot alkotnak p_j intenzitással. Ekkor a $\wp_1, \wp_2, \ldots, \wp_N$ folyamatok függetlenek egymástól, ezek a Poisson-folyamat ritkításai. Ezzel a módszerrel független multinominális kísérleteket ágyaztunk be folytonos időben.

Azt mondjuk, hogy $\wp_j$ folyamatnak megvan a kvótája, ha legalább m_j esemény előfordult belőle. Jelölje T_j azt az időpontot amíg ez bekövetkezik, erről ismert, hogy gamma eloszlást követ: $T_j \sim \Gamma(m_j, p_j)$. Ekkor a sűrűségfüggvénye:

$$f_{T_j}(t) = \frac{p_j^{m_j} t^{m_j - 1} \mathrm{e}^{-p_j t}}{\Gamma(m_j)} = \frac{p_j^{m_j} t^{m_j - 1} \mathrm{e}^{-p_j t}}{(m_j - 1)!}$$

Jelölje $T_{k:N}$ a várakozási időt mire k folyamat eléri a kvótát. Ez a k. rendezett mintaelem $T_1, T_2, \ldots, T_N$-ből. Továbbá jelölje $W_{k:N}$ az összes bekövetkezett esemény számát $T_{k:N}$ időig, ekkor $W_{k:N}$ független az érkezések közötti időktől és felírható közöttük a következő összefüggés:

$$T_{k:N} = \sum_{\nu=1}^{W_{k:N}} Z_\nu \, .$$

5.1. Tétel. *Legyen $T_j \sim \Gamma(m_j, p_j)\ \forall j = 1, 2, \ldots, N$ és jelölje $T_{1:N}, T_{2:N}, \ldots, T_{N:N}$ a rendezett mintájukat. Független multinominális kísérletekre $p_1, p_2, \ldots, p_N$ valószínűségek és $m_1, m_2 \ldots, m_N$ kvóták mellett, legyen $W_{k:N}$ az események száma, ameddig k folyamat eléri a kvótát. Ekkor $t < \min_j p_j$-re:*

$$E((1-t)^{-W_{k:N}}) = E(\mathrm{e}^{(tT_{k:N})})$$

$$E(W_{k:N}^{[\nu]}) = E(T_{k:N}^{\nu})$$

Bizonyítás: Felhasználva, hogy $W_{k:N}$ és Z-k függetlenek, minden $t < \min_j p_j$-re:

$$E\Big(\exp(tT_{k:N})\Big) = E\bigg(E\Big(\exp\Big(t\sum_{\nu=1}^{W_{k:N}} Z_\nu\Big)\Big|W_{k:N}\Big)\bigg) =$$

$$= E\Big(E\big(\exp(tZ)\big)^{W_{k:N}}\Big) = E\big((1-t)^{-W_{k:N}}\big)$$

Ebből következik, hogy $W_{k:N}$ növekvő faktoriális momentumai teljesítik a következő egyenlőséget:

$$E(W_{k:N}^{[\nu]}) = E\big((W_{k:N})(W_{k:N}+1)\dots(W_{k:N}+\nu-1)\big) = E(T_{k:N}^{\nu})$$

Most vizsgáljunk független multinominális kísérleteket $p_1, p_2, \dots, p_N$ valószínűségekkel. Azt mondjuk, hogy j-nek már megvan a kvótája, ha legalább m_j-szer teljesült. Triviális, hogy a próbálkozások számát a k kvótáig ugyan azzal az eloszlással kapjuk, jelöljük továbbra is $W_{k:N}$-el. Ezzel bizonyítottuk a Tételt. □

Mielőtt folytatnánk még vezessük be az $S_1, S_2, \dots, S_N$ jelölést az egymástól független $\Gamma(m,1)$ eloszlású valószínűségi változókra és $S_{k:N}$, illetve $S_{k:N,m}$ jelölést a k. rendezett mintájukra.

5.1.2. Alkalmazás a kupongyűjtő problémára

Bár a fenti folyamatot különböző valószínűségekkel és kvótákkal vezettük be, most a következő paraméterekkel használjuk: $p_1 = p_2 = \dots = p_N = \frac{1}{N}$ és $m_1 = m_2 = \dots = m_N = m$. Ekkor a megfelelő Poisson folyamatok intenzitása $\frac{1}{N}$ és a T_j várakozási idő, hogy a $\wp_j$ Poisson-folyamatban legalább m esemény előforduljon felírható $T_j = NS_j$ alakban, ahol $S_j \sim \Gamma(m,1)$. Minden Poisson-folyamat a kollekció egy darabjának felel meg, a kvóta pedig az összegyűjtendő teljes kollekciók számának, tehát T_j megfelel annak az időnek, amikor a j. típusú kártyából már legalább m darabbal rendelkezünk. Így az általunk keresett valószínűségi változó $T_{N:N} = NS_{N:N}$. Legyen $W_{N:N}$ az összes húzás száma $T_{N:N}$ időpillanatig. Így a következő egyenletet kapjuk:

$$NS_{N:N} = T_{N:N} = \sum_{\nu=1}^{W_{N:N}} Z_\nu \quad \text{, ahol } Z_i \sim \mathrm{Exp}(1) \;\; \forall i = 1, 2 \dots$$

Felhasználva az 5.1 Tételt $t < \frac{1}{N}$-re:

$$E\big((1-t)^{-W_{N:N}}\big) = E(\mathrm{e}^{tNS_{N:N}})$$

$$E(W_{N:N}^{[\nu]}) = N^{\nu} E(S_{N:N}^{\nu})$$

Tudjuk, hogy $S_{N:N}$ a legnagyobb rendezett mintaelem a $\Gamma(m,1)$ mintában, ezért a $W_{N:N}$ momentumai kifejezhetők $S_{N:N}$ momentumaival. Ha lenne explicit módszer a momentumok számolására már kész is lennénk, így viszont másképp kell kiszámolnunk:

$$\mathrm{E}(W_{N:N}) = NE(S_{N:N}) = N\int_0^{\infty} \mathrm{P}(S_{N:N} > t)dt = N\int_0^{\infty} 1 - \mathrm{P}(S_{N:N} \le t)dt =$$

$$= N\int_0^{\infty} 1 - \mathrm{P}(S_1 \le t, S_2 \le t, \ldots, S_N \le t)dt = N\int_0^{\infty} \left(1 - \left(1 - \sum_{j=0}^{m-1} \frac{t^j \mathrm{e}^{-t}}{j!}\right)^N\right)dt$$

Erre a képletre [1]-ben L. Shepp és D.J. Newmann a következő közelítést bizonyította be:

5.2. Tétel. *m konstans értékre, miközben $N \to \infty$ teljesül a következő egyenlőség:*

$$E(W_{N:N}) = N\Big[\ln(N) + (m-1)\ln\big(\ln(N)\big) + C_m + o(1)\Big]$$

Majd egy évvel később [2] Erdős Pál és Rényi Alfréd bebizonyította, hogy a fenti tételben szereplő $C_m = \gamma - \ln(m-1)!$, ahol $\gamma \approx 0,772156649$ továbbra is az Euler-Mascheroni állandó. Így a várható érték becsülhető a következő kifejezéssel:

$$E(W_{N:N}) = N\Big[\ln(N) + (m-1)\ln\big(\ln(N)\big) + \gamma - \ln(m-1)! + \mathrm{o}(1)\Big] \qquad (6)$$

5.2. Közelítés Markov-lánccal

A gyűjtés módja legyen a következő: Mikor vesznek egy kupont, a legidősebb testvér megnézi, hogy neki már megvan-e a kupon. Ha igen, akkor továbbadja a második legidősebbnek és így tovább.
A kollekció most is N kuponból áll és az 1 kollekciós esethez hasonlóan ezt is felírhatjuk, mint Markov-lánc. Most [6] alapján vizsgáljunk egy alternatív megoldást és tekintsük az $m = 2$ esetet. Legyen X_n $(n \ge 0)$ ez egyes gyerekek kuponjainak száma n csomag megvásárlása után. $\{X_n, n \ge 0\}$ Markov-lánc $S = \{(i,j) : i,j \in \{0,1,\ldots,N\}, i \ge j\}$ állapothalmazon.

Könnyen látszik a $\sum_{i=1}^{n+1} i = \frac{(n+2)(n+1)}{2}$ azonosságból, hogy $|S| = \frac{(N+2)(N+1)}{2}$. Az átmenet mátrix pedig a következőkből adódik:

$$(0,0) \ \to (1,0): \quad p = 1$$

$$(i,j) \ \to \begin{cases} (i,j): & p = \frac{j}{N} \\ (i+1,j): & p = \frac{(N-i)}{N} \\ (i,j+1): & p = \frac{(i-j)}{N} \end{cases} \qquad i \geq j$$

$$(N,N) \ \to (N,N): \quad p = 1$$

Ahhoz, hogy megkapjuk a várható értéket meg kell határoznunk $k_{(0,0)}^{(N,N)}$ értékét. Az 1 kollekciós esethez hasonlóan itt is egy lineáris egyenletrendszert kell megoldanunk, hogy ezt megkapjuk. Nagy N-re ezt csak program segítségével tehetjük meg, ezért most az $N = 4$ eseten keresztül fogom megmutatni a számolást. Az állapothalmaz: $|S| = \frac{4*5}{2} = 10$,

$$S = \{(0,0),(1,0),(1,1),(2,0),(2,1),(2,2),(3,0),(3,1),$$

$$(3,2),(3,3),(4,0),(4,1),(4,2),(4,3),(4,4)\}$$

Az átmenetmátrix pedig:

$$\begin{pmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{1}{4} & \frac{3}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{1}{4} & 0 & \frac{3}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \frac{2}{4} & 0 & \frac{2}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{2}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \frac{2}{4} & 0 & 0 & \frac{2}{4} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{3}{4} & 0 & 0 & \frac{1}{4} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{2}{4} & 0 & 0 & \frac{1}{4} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{2}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{3}{4} & 0 & 0 & 0 & \frac{1}{4} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{2}{4} & \frac{2}{4} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{3}{4} & \frac{1}{4} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}$$

Hogy meghatározzuk a várható értéket meg kell oldanunk a következő egyenletrendszert:

$$\begin{cases} k_{(0,0)} = k_{(1,0)} + 1 \\ k_{(1,0)} = \frac{1}{4}k_{(1,1)} + \frac{3}{4}k_{(2,0)} + 1 \\ k_{(1,1)} = \frac{1}{4}k_{(1,1)} + \frac{3}{4}k_{(2,1)} + 1 \\ k_{(2,0)} = \frac{2}{4}k_{(2,1)} + \frac{2}{4}k_{(3,0)} + 1 \\ k_{(2,1)} = \frac{1}{4}k_{(2,1)} + \frac{1}{4}k_{(2,2)} + \frac{2}{4}k_{(3,1)} + 1 \\ k_{(2,2)} = \frac{2}{4}k_{(2,2)} + \frac{2}{4}k_{(3,2)} + 1 \\ k_{(3,0)} = \frac{3}{4}k_{(3,1)} + \frac{1}{4}k_{(4,0)} + 1 \\ k_{(3,1)} = \frac{1}{4}k_{(3,1)} + \frac{2}{4}k_{(3,2)} + \frac{1}{4}k_{(4,1)} + 1 \\ k_{(3,2)} = \frac{2}{4}k_{(3,2)} + \frac{1}{4}k_{(3,3)} + \frac{1}{4}k_{(4,2)} + 1 \\ k_{(3,3)} = \frac{3}{4}k_{(3,3)} + \frac{1}{4}k_{(4,3)} + 1 \\ k_{(4,0)} = k_{(4,1)} + 1 \\ k_{(4,1)} = \frac{1}{4}k_{(4,1)} + \frac{3}{4}k_{(4,2)} + 1 \\ k_{(4,2)} = \frac{2}{4}k_{(4,2)} + \frac{2}{4}k_{(4,3)} + 1 \\ k_{(4,3)} = \frac{3}{4}k_{(4,3)} + \frac{1}{4}k_{(4,4)} + 1 \\ k_{(4,4)} = 0 \end{cases} \iff \begin{cases} k_{(0,0)} = k_{(1,0)} + 1 \\ k_{(1,0)} = \frac{1}{4}k_{(1,1)} + \frac{3}{4}k_{(2,0)} + 1 \\ \frac{3}{4}k_{(1,1)} = \frac{3}{4}k_{(2,1)} + 1 \\ k_{(2,0)} = \frac{2}{4}k_{(2,1)} + \frac{2}{4}k_{(3,0)} + 1 \\ \frac{3}{4}k_{(2,1)} = \frac{1}{4}k_{(2,2)} + \frac{2}{4}k_{(3,1)} + 1 \\ \frac{2}{4}k_{(2,2)} = \frac{2}{4}k_{(3,2)} + 1 \\ k_{(3,0)} = \frac{3}{4}k_{(3,1)} + \frac{1}{4}k_{(4,0)} + 1 \\ \frac{3}{4}k_{(3,1)} = \frac{2}{4}k_{(3,2)} + \frac{1}{4}k_{(4,1)} + 1 \\ \frac{2}{4}k_{(3,2)} = \frac{1}{4}k_{(3,3)} + \frac{1}{4}k_{(4,2)} + 1 \\ \frac{1}{4}k_{(3,3)} = \frac{1}{4}k_{(4,3)} + 1 \\ k_{(4,0)} = k_{(4,1)} + 1 \\ \frac{3}{4}k_{(4,1)} = \frac{3}{4}k_{(4,2)} + 1 \\ \frac{2}{4}k_{(4,2)} = \frac{2}{4}k_{(4,3)} + 1 \\ \frac{1}{4}k_{(4,3)} = \frac{1}{4}k_{(4,4)} + 1 \\ k_{(4,4)} = 0 \end{cases}$$

$$\iff \begin{cases} k_{(4,4)} = 0 \\ k_{(4,3)} = 4 \\ k_{(4,2)} = 6 \\ k_{(4,1)} \approx 7.33 \\ k_{(4,0)} \approx 8.33 \\ k_{(3,3)} = 8 \\ k_{(3,2)} = 9 \\ k_{(3,1)} \approx 9.78 \\ k_{(3,0)} \approx 10.42 \\ k_{(2,2)} = 11 \\ k_{(2,1)} \approx 11.52 \\ k_{(2,0)} \approx 11.97 \\ k_{(1,1)} \approx 12.85 \\ k_{(1,0)} \approx 13.19 \\ k_{(0,0)} = 14.1887 \end{cases}$$

Tehát ebben az esetben várhatóan 14,2 kupont kell venniük, hogy 2 kollekció is teljes legyen. Négy elemű kollekciónál egy kollekció összegyűjtéséhez várhatóan 8,3 kuponra van szükség. Tehát ha ketten gyűjtenek két kollekciót, akkor 1,2 kuponnal kevesebbet kell vásárolni egy főnek, mintha egyedül gyűjtené. Ez ennél az esetnél nem tűnik jelentősnek, de nagyobb méretű kollekcióknál már sokkal nagyobb ez a különbség:

Várható kuponszám 2 kollekcióra			
Kollekció mérete	Közös gyűjtés	Egyéni gyűjtés	Különbség
4	14.19	16.67	2.48
5	19.04	22.83	3.79
10	46.23	58.58	12.35
15	76.48	99.55	23.07
20	108.7	143.91	35.21
25	142.37	190.8	48.43
30	177.19	239.7	62.51
35	212.97	290.28	77.31

6. Több kollekció eltérő valószínűséggel

A probléma továbbra is az, hogy egy N kuponból álló kollekcióból szeretnénk összegyűjteni m teljes kollekciót. Csupán annyit változtatunk rajta, hogy a kuponok nem azonos valószínűséggel fordulnak elő. Tehát $\forall i$-re az i kupon előfordulásának valószínűsége $p_i \geq 0$ és $p_1 + p_2 + \cdots + p_N = 1$. Mivel ez az eddigieknél nehezebb probléma, nehéz pontos megoldást adni. Ezért ebben a részben csak felső becslést mutatok meg a várható értékre [6]-ból.

6.1. Közelítés a Maximum-Minimum azonossággal

Legyen X_1 a csomagok száma amit meg kell vennünk, hogy legyen m db kuponunk az 1. típusból. Általánosan: legyen X_i a csomagok száma amit meg kell vennünk, hogy legyen m db kuponunk az i. típusból. Így $\mathrm{E}[X] = \mathrm{E}[\max(X_1, X_2, \ldots, X_N)]$ a várható ideje m teljes kollekció összegyűjtésének. A várható érték meghatározásához használjuk a 3.1 fejezetben bizonyított Maximum-Minimum azonosságot:

$$\mathrm{E}[X] = \sum_i \mathrm{E}[X_i] - \sum_{i<j} \mathrm{E}[\min(X_i, X_j)] + \sum_{i<j<k} \mathrm{E}[\min(X_i, X_j, X_k)] - \ldots$$

$$\cdots + (-1)^{N+1}\mathrm{E}[\min(X_1, X_2, \ldots, X_N)]$$

Az X_i most negatív binomiális eloszlást követ (m, p_i) paraméterekkel, ezért $\mathrm{E}[X_i] = \frac{m}{p_i}$. Ezután viszont abba a problémába ütközünk, hogy nem tudjuk meghatározni a következő valószínűségi változók várható értékét:

$$\min_{i<j}\{X_i, X_j\},\ \min_{i<j<k}\{X_i, X_j, X_k\},\ \min_{i<j<k<l}\{X_i, X_j, X_k, X_l\}\ ,\ \ldots$$

Mivel itt kicsit bonyolultabb a helyzet. Ebben az esetben $\min(X_i, X_j)$ értékére nem mondhatjuk, hogy negatív binomiális eloszlást követ $(m, p_i + p_j)$ paraméterrel, ám a következő ötlet mentén, meg tudjuk becsülni az értékét: Könnyen látható, hogy a legszerencsésebb eset ha az i típusú kupon legalább m-szer előfordul mielőtt a j típusúból akár egyet is összegyűjtenénk (vagy fordítva). Ha ez biztosan mindig így alakulna, akkor $\min(X_i, X_j)$ negatív binomiális eloszlást követne $(m, p_i + p_j)$ paraméterrel, így ez megfelelő lesz alsó becslésnek. A másik véglete az esetnek amikor érkezési sorrendjüktől függetlenül i és j típusú kuponból is $(m-1)$ darabot sikerült összegyűjtenünk, mielőtt bármelyikből is megszereznénk az m.-et. Az előzőhöz hasonlóan, ha mindig ilyen módon érkezne az i és j kupon, akkor $\min(X_i, X_j)$ negatív binomiális eloszlást követne $(2(m-1)+1, p_i + p_j)$ paraméterrel. Ez az eset megfelelő felső becslés lesz. Így felírhatjuk a következőket $\forall\ i \neq j$-re:

$$\mathrm{E}[T(i, j; 2)] \leq \mathrm{E}[\min(X_i, X_j)] \leq \mathrm{E}[Z(i, j; 2)],$$

ahol $T(i, j; 2)$ negatív binomiális eloszlást követ $(m, p_i + p_j)$ paraméterrel és $Z(i, j; 2)$ szintén negatív binomiális eloszlást követ $(2m-1, p_i + p_j)$ paraméterrel. Általánosabban $\forall\ 2 \leq k \leq N$-re:

$$\mathrm{E}[T(i_1, i_2, \ldots, i_k; k)] \leq \mathrm{E}[\min(X_{i_1}, X_{i_2}, \ldots, X_{i_k})] \leq \mathrm{E}[Z(i_1, i_2, \ldots, i_k; k)],$$

ahol $T(i_1, i_2, \ldots, i_k; k)$ negatív binomiális eloszlást követ $(m, p_{i_1} + p_{i_2} + \cdots + p_{i_k})$ paraméterrel és $Z(i_1, i_2, \ldots, i_k; k)$ szintén negatív binomiális eloszlást követ $(k(m-1)+1, p_{i_1} + p_{i_2} + \cdots + p_{i_k})$ paraméterrel.
Ezeket felhasználva:

$$\mathrm{E}[X] \leq \sum_i \mathrm{E}[X_i]\ -\sum_{i_1<i_2} \mathrm{E}[T(i_1, i_2; 2)]\ +\sum_{i_1<i_2<i_3} E[Z(i_1, i_2, i_3; 3)]\ -$$

$$-\sum_{i_1<i_2<i_3<i_4} E[T(i_1, i_2, i_3, i_4; 4)] + \cdots + \begin{cases} -\mathrm{E}[T(1, 2, \ldots, N; N)] & \text{, ha } N \text{ páros} \\ \mathrm{E}[Z(1, 2, \ldots, N; N)] & \text{, ha } N \text{ páratlan} \end{cases}$$

$$= \sum_i \frac{m}{p_i} - \sum_{i_1<i_2} \frac{m}{p_{i_1}+p_{i_2}} + \sum_{i_1<i_2<i_3} \frac{3m-2}{p_{i_1}+p_{i_2}+p_{i_3}} - \dots$$

$$\dots - \sum_{\substack{i_1<i_2<\dots<i_l \\ l \text{ páros}}} \frac{m}{p_{i_1}+p_{i_2}+\dots+p_{i_l}} + \dots + \sum_{\substack{i_1<i_2<\dots<i_k \\ k \text{ plan}}} \frac{k(m-1)+1}{p_{i_1}+p_{i_2}+\dots+p_{i_k}} + \dots$$

$$\dots + \begin{cases} -\frac{m}{p_1+p_2+\dots+p_N} & \text{,ha } N \text{ páros} \\ \frac{N(m-1)+1}{p_1+p_2+\dots+p_N} & \text{,ha } N \text{ páratlan} \end{cases}$$

6.1.1. Példa

Vizsgáljuk meg az $N = 3$, $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{3}$, $p_3 = \frac{1}{6}$ példán, hogy különböző m értékekre mennyire van közel a felső becslésünk a szimulált gyűjtések várható értékéhez. A képlet alapján

$$\mathrm{E}(X) \leq \frac{m}{\frac{1}{2}} + \frac{m}{\frac{1}{3}} + \frac{m}{\frac{1}{6}} - \frac{m}{\frac{1}{2}+\frac{1}{3}} - \frac{m}{\frac{1}{2}+\frac{1}{6}} - \frac{m}{\frac{1}{3}+\frac{1}{6}} + \frac{3(m-1)+1}{\frac{1}{2}+\frac{1}{3}+\frac{1}{6}}$$

Sajnos, ahogy a táblázatban láthatjuk m értékének növekedésével a felső becslés jósága csökken.

Várhatóan szükséges csomagok száma m kollekció összegyűjtéséhez									
m	2	3	4	5	6	7	8	9	10
Felső becslés	14,6	22,9	31,2	39,5	47,8	56,1	64,4	72,7	81,0
Szimuláció	13,45	19,3	25,06	31,02	36,96	42,97	48,62	54,95	60,76
Eltérés	1,15	3,6	6,14	8,48	10,84	13,13	15,78	17,75	20,24

7. Egy másik megfogalmazás

A kupongyűjtő probléma egy másik megfogalmazása a következő: Van N különböző kuponunk amit szeretnénk összegyűjteni. Mekkora a valószínűsége, hogy k csomag megszerzése után n különböző kuponunk lesz és várhatóan hány különböző kuponunk lesz ezután?

Kezdjük a kérdés első felével. Jelöljük A_j-vel azokat az eseteket, amikor k csomag egyike sem tartalmazta a j típusú kupont ($j = 1, 2, \dots, N$). Mivel a csomagokban lévő kuponok típusai egymástól függetlenek $\mathrm{P}(A_j) = \left(\frac{N-j}{N}\right)^k$. Jelölje X_k a különböző kuponok számát k csomag megszerzése után. $\mathrm{P}(X_k = n)$

meghatározásához használjuk a Jordán-formulát:

$$\mathrm{P}(X_k = n) = \sum_{i=0}^{N-n} (-1)^i \binom{n+i}{i} \mathrm{S}_{n+i},$$

$$\text{ahol} \quad \mathrm{S}_l = \sum_{1 \le i_1 < \cdots < i_l \le N} \mathrm{P}(\overline{A_{i_1}} \cap \overline{A_{i_2}} \cap \cdots \cap \overline{A_{i_l}})$$

Mivel az A_i események metszetét könnyebben meg tudjuk határozni, számoljunk a következő alakban:

$$\mathrm{P}(X_k = n) = \sum_{i=0}^{n} (-1)^i \binom{N-n+i}{i} \mathrm{S}_{N-n+i},$$

$$\text{ahol} \quad \mathrm{S}_l = \sum_{1 \le i_1 < \cdots < i_l \le N} \mathrm{P}(A_{i_1} \cap \ A_{i_2} \cap \cdots \cap A_{i_l})$$

Felhasználva, hogy $\mathrm{P}(A_{i_s} A_{i_t}) = \left(\frac{N-2}{N}\right)^k \ \ \forall s < t$, $\mathrm{P}(A_{i_s} A_{i_t} A_{i_u}) = \left(\frac{N-3}{N}\right)^k$ $\forall s < t < u \ \ldots$:

$$\mathrm{P}(X_k = n) = \sum_{i=0}^{n} (-1)^i \binom{N-n+i}{i} \binom{N}{N-n+i} \left(\frac{n-i}{N}\right)^k$$

Most térjünk át a kérdés második felére. Definiáljuk $1 \le i \le N$ értékekre a következő függvényt:

$$I_j = \begin{cases} 0 & \text{ha a } j \text{ típusú kupon nem volt benne a } k \text{ csomagban} \\ 1 & \text{egyébként} \end{cases}$$

Ekkor a különböző kuponok száma k csomag megszerzése után:

$$X_k = \sum_{j=1}^{N} I_j$$

Ekkor

$$\mathrm{E}(I_j) = \mathrm{P}(I_j = 1) = \mathrm{P}(\overline{A_j}) = 1 - \left(\frac{N-1}{N}\right)^k$$

Így a különböző kuponok számának várható értéke k csomagban:

$$\mathrm{E}(X_k) = \mathrm{E}\left(\sum_{j=1}^{N} I_j\right) = N\left(1 - \left(\frac{N-1}{N}\right)^k\right)$$

8. Példa a hétköznapokból

Manapság egyre divatosabb, hogy az élelmiszer üzletek kupongyűjtő akciókkal próbálják magukhoz csábítani a vásárlókat. Ilyenkor kártyákat/matricákat lehet gyűjteni néhány hónapon keresztül a következő módon: Általában kétféle csomag elérhető az akció során. Vannak "ingyenes" csomagok, amit olyankor kap a vásárló ha adott összeg felett vásárol és vannak megvásárolható csomagok is, amikkel meg lehet gyorsítani a gyűjtési folyamatot, ezeket a továbbiakban extra csomagoknak fogom nevezni. Ezen kívül persze vannak még gyűjtőalbumok, dobozok és egyéb kiegészítő termékek.

8.1. Motiváció

Én sosem kezdtem bele az ilyen gyűjtésekbe, mert mindig is úgy sejtettem, hogy extra kártyacsomagok vásárlása nélkül nem gyűjthető össze a teljes kollekció. Most szeretném egy kicsit körbejárni ezt a feladatot. Egy korábbi akció alapján szeretném meghatározni, hogy egy kollekcióban vannak-e ritka/gyakori kártyák és ha igen mennyi. Ez alapján meghatározni, hogy várhatóan hány kártyára van szükség a teljes kollekció összegyűjtéséhez. Ezután megnézem, hogy "ideális vásárlással" össze lehetne-e gyűjteni a teljes kollekciót extra kártyacsomagok vásárlása nélkül. Ezután a gyűjthetőség gondolatát továbbvíve megvizsgálom, hogy közös gyűjtéssel mennyit lehet spórolni egy ilyen gyűjtögetés során és, hogy hány embernek kell összedolgoznia ahhoz, hogy extra kártyacsomag vásárlása nélkül mindannyian össze tudják gyűjteni a teljes kollekciót. Végül a gyűjtések kimenetelét vizsgálom meg részletesebben.
Egyenlőre feltesszük, hogy a család nem kap senkitől ajándékba kártyát és nem is cserél senkivel.

8.2. Adatgyűjtés

Kiválasztottam egy korábbi akciót és beléptem egy közösségi oldalon létrejött csoportba ahol ezeket a kártyákat cserélgették. Innen írtam ki 50 ember arra vonatkozó adatát, hogy milyen kártyát keres és milyet kínál cserére. Ez alapján csináltam egy táblázatot. 0 értéket kaptak a keresett kártyák, 2 értéket kaptak a cserére kínált kártyák és 1 értéket kapott a kollekció többi darabja. A gyakoriság alapján minden kártyára kaptam egy előfordulási valószínűséget. Ekkor az egyes típusok valószínűségei 0,021 ás 0,034 között alakultak, melyek közel esnek az azonos valószínűségű esethez, ahol 0,0278 az egyes kár-

tyák valószínűsége. Bár ez a rossz minta miatt is adódhat most feltételezem, hogy azonos valószínűséggel fordultak elő a kuponok.

8.3. A kiválasztott akció adatai

Így az általam választott akció adatai a következők: A kollekció 36 kártyából állt, aminek összegyűjtésére 56 napig volt lehetőségük a vásárlóknak. Összesen 7.600.000 darab ingyenes kártyacsomagot és 1.400.000 darab extra kártyacsomagot gyártottak. Ingyenes csomagokat 4.000 Ft-onként kapott a vásárló, az extra csomagokat pedig 100 Ft-ért lehetett megvásárolni. A különböző kuponok azonos valószínűséggel fordultak elő.

8.4. A várhatóan szükséges kártyák száma

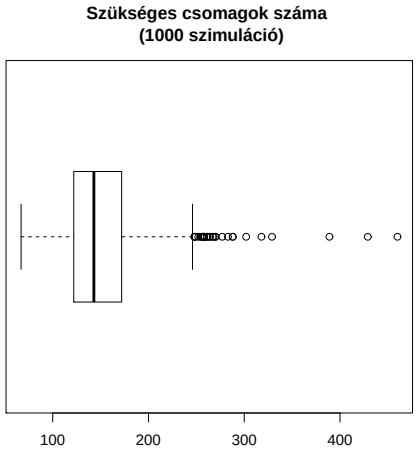
Az (1) képlet alapján 150,28 kártyára lesz szükségünk egy ilyen kollekció összegyűjtéséhez, amihez R-es szimuláció alapján kapott 150,88 várható kuponok száma közel van. Ezt az eredményt 1000 gyűjtés szimulációjából kaptam. Ezek közül a leggyorsabb gyűjtéshez 67 kupon vásárlására volt szükség, míg a leglassabb gyűjtéshez 460 kupont kellett szerezni, tehát láthatjuk, hogy elég nagy a szórása a valószínűségi változónak, így extrém esetek is létrejöhetnek.

Nézzük meg a értékek 8.4.(box-plot) ábráját. A medián értéke 143, így várhatóan az esetek felében nem is lesz szükségünk a várhatóan szükséges kártyákra. Az interkvartilis terjedelme 50, a legtöbb eset 122 és 172 érték között marad. Mind a 30 kieső érték a jobb oldalon van, de ez mindössze az esetek 3%-a.

A 6. ábra hisztogramja is ezt tükrözi. Kimagaslóan sok érték esik a (127,147) intervallumba és a legtöbb érték körülötte csoportosul így a hisztogram bal felére csak néhány szélsőséges érték kerül.

8.5. Össze tudjuk gyűjteni?

A KSH 2015-ös adatai szerint a háztartások egy főre jutó havi fogyasztási kiadása élelmiszerekre és alkoholmentes italokra 19885 Ft. Én most négy fős, két gyermekes családok kiadásával fogok számolni, mert az általam választott promóció elsősorban a gyermekeket célozta. Tehát az havi 79540 Ft-ot jelent. Az egyszerűség kedvéért számoljunk 80000Ft-al.



5. ábra.

8.5.1. Ideális költés

Tegyük fel, hogy az akció időtartama alatt a család odafigyel, hogy minden ilyen jellegű kiadását az akciót hirdető üzletben költse el és a vásárlás végösszege közel legyen a 4000 Ft valamilyen többszöröséhez, mert akkor kapják a legtöbb kártyát. A promóció nagyjából 2 hónapig tartott, tehát az ideális költés mellett összesen 40 kártyát tudnak megszerezni ingyenesen. Így, ha össze szeretnék gyűjteni a teljes kollekciót, várhatóan még 11028 Ft-ot el kell költeniük az extra kártyacsomagokra.

8.6. Várhatóan hány különböző kártyánk lesz?

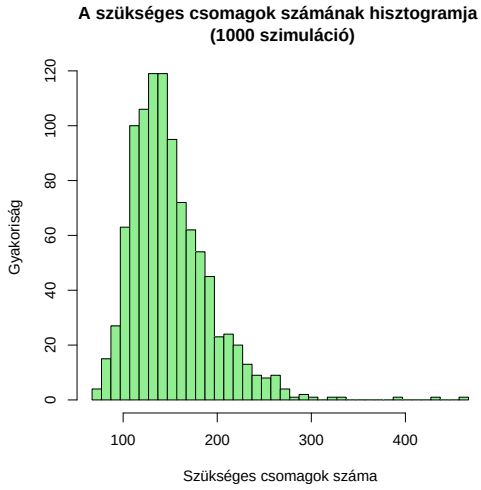
Tegyük fel, hogy nem vagyunk hajlandóak pénzt költeni az extra csomagokra. Ekkor az akció végén az ismert képlet segítségével könnyen számolható, hogy várhatóan még 11,67 kártyánk hiányozni fog a kollekcióból:

$$36\left(1-\left(1-\frac{1}{36}\right)^{40}\right)=24,33$$

8.7. Hogyan gyűjtsük mégis össze?

8.7.1. Extra csomag vásárlása

Mint ahogy azt korábban már kiszámoltuk a legoptimálisabb vásárlással is még várhatóan 11028 Ft-ot kell extra csomagokra költenünk. Ez heti 3 vá-



6. ábra.

sárlásnál azt jelenti, hogy 4-5 csomagot kell vennünk alkalmanként.

8.7.2. Gyűjtsük többen

Láthattuk a 5. fejezetben, hogy ismerősökkel közösen gyűjtve a kuponokat, biztosan kevesebb kupont kell vásárolnunk. De vajon hány családnak kell összefogni ehhez? A korábbi (6) képlet segítségével:

$$E(W_{36:36}) = 36 * \Big[\ln(36) + (m-1)\ln\big(\ln(36)\big) + \gamma - \ln(m-1)! + o(1)\Big]$$

Ezzel a képlettel számolva a 1. táblázatban láthatjuk, hogy hány kártyát kell szereznünk ha egyedül és ha közösen gyűjtjük a kártyákat $m = 2, 3, \ldots$ családdal. Összesen 7 családnak kell közösen gyűjtenie a kártyákat a fenti ideális módon ha nem szeretnének költeni extra kártyákra. Ezt már épp elég nagy ahhoz, hogy nehéz legyen megoldani a szomszédokkal és rokonokkal. Viszont 4 családnak már könnyebb összefogni és ha heti háromszori vásárlás mellett minden alkalommal csupán 1,03 extra csomagot vásárolunk, akkor vásárlásonként ezzel a 103 Ft extra költséggel várhatóan össze tudnak gyűjteni 4 teljes kollekciót.

8.8. Milyen kimenetelre számíthatunk?

Egy ilyen gyűjtögetésben persze sok múlik a szerencsén, de nézzünk meg néhány esetet, hogy hogyan alakult a gyűjteményünk. Egy szimulációval 1000-szer lefuttattam a gyűjtési folyamatot és megszámoltam, hogy az egyes

Gyűjtők száma (fő)	Szükséges kupon/fő (db)	Ráfordítandó összeg/fő (Ft)
1	150,28	11028
2	115,87	7587
3	84,24	4424
4	64,78	2478
5	51,03	1103
6	40,53	53
7	32,09	0

1. táblázat.

típusú kártyákból mennyit gyűjtöttünk a teljes kollekció megszerzéséig.

A 2. táblázatban három esetben szeretném megmutatni, hogy hogyan alakult az egyes típusok száma: a legkedvezőbb esetben, egy átlagos(várható) esetben és a legszerencsétlenebbül hosszasan gyűjtött esetben:

Kártya	Min.	Max.	Átl.	Kártya	Min.	Max.	Átl.
1	2	12	6	19	2	10	1
2	4	10	5	20	2	10	2
3	2	10	4	21	3	8	5
4	1	12	2	22	2	8	2
5	2	11	6	23	1	7	5
6	1	13	1	24	2	11	1
7	3	8	2	25	1	8	3
8	1	9	2	26	1	11	4
9	5	13	6	27	2	1	5
10	1	14	4	28	1	13	7
11	1	14	4	29	1	12	6
12	1	20	5	30	3	9	6
13	2	9	2	31	1	11	2
14	1	7	4	32	2	11	3
15	2	11	4	33	2	14	4
16	2	9	3	34	2	11	5
17	2	12	3	35	1	9	6
18	2	10	8	36	4	13	4

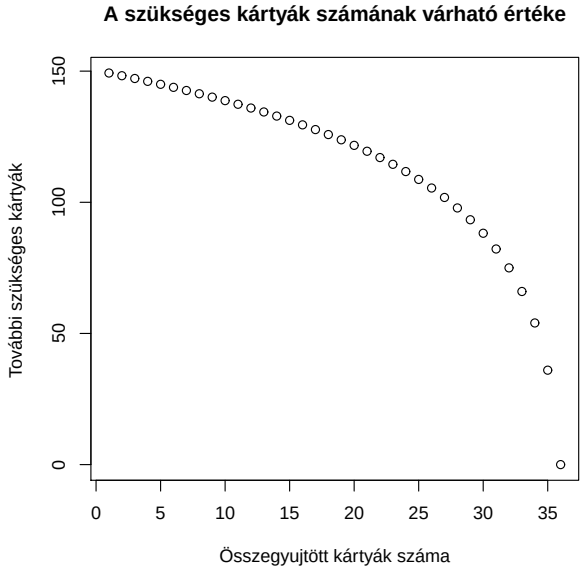
2. táblázat.

A **Min. esetben** 14 db olyan kártya van ami csak egyszer fordul elő a kollekciónkban, ami a kártyák fele, tehát tényleg elég szerencsésen gyűjtögettünk. Ezt tovább erősíti a tény, hogy a további kártyákból(amikből legalább 2-t gyűjtöttünk) átlagosan 2,45 kártyát szereztünk és maximum 5-öt. De még ebben az esetben is 68 csomagra volt szükségünk, ami a kollekció méretének közel kétszerese.

A **Átlagos esetben** Már csak 3 olyan kártya volt ami egyetlen egyszer fordul elő. Ez már sokkal kevesebb a "szerencsés" esethez képest. Itt a további kártyákból átlagosan 4,2 db is összegyűlt, de még mindig csak maximum 8 db kártyánk volt egy típusból. Ehhez viszont már 142 csomagra volt szükségünk, ami a kollekció méretének közel négyszerese.

A **Max. eset** egy nagyon szerencsétlen gyűjtésből származó eset. Ebben csupán egy olyan kártyánk volt ami csak egyszer fordul elő a kollekcióban. Mivel ezt nehezen sikerült összegyűjtenünk, közben a többi típusból sokat felhalmoztunk. Hogy egészen pontosak legyünk a további kártyákból átlagosan 10,8 darab volt típusonként és volt olyan is amiből 20 darabot is összegyűjtöttünk. Ebben az esetben 381 csomagot szereztünk be mire teljes lett a kollekciónk.

Könnyen látszik, hogy a fenti táblázatban a sorok közötti legnagyobb különbség, hogy hány darab 1-es tartalmaznak, azaz, a kártyák közül, hány olyan típus van amiből pontosan egy darabbal rendelkezünk. Így felmerül a kérdés, hogy vajon mennyit tudunk spórolni a szükséges csomagokon ha szerzünk valakitől vagy csere útján néhány olyan kártyát ami még hiányzik a kollekciónkból? Ha cserével vagy ismerősöktől meg tudunk szerezni 4 ilyen kártyát, akkor várhatóan meg tudunk spórolni 75 csomagot. Ha 5 ilyen kártyát is tudunk szerezni, akkor várhatóan további 7,2 csomagot spórolhatunk meg. Viszont nem feltétlenül szerencsés minél több kártya sorsát a szerencsére bízni, hiszem folyamatosan csökken a további megspórolt kártyák száma. A 8.8. ábrán láthatjuk, hogy x különböző kártya összegyűjtése után még várhatóan hány további csomagra lesz szükségünk a teljes kollekció összegyűjtéséhez. Ahogy csökken a pontok közötti függőleges távolság egyre kevesebb csomagot tudunk megspórolni egy újabb kártya kihagyásával. Így nagyjából 32 különböző kártyáig lehet érdemes gyűjtögetni és ezután szerencsét próbálni, ekkor várhatóan 75,28 csomag kártyát kell megszereznünk mielőtt cserélgetésbe fognánk. Ez a szám már nincs is annyira messze attól az esettől amikor 4 család közösen gyűjt. Így ezzel a módszerrel csupán kicsit több anyagi ráfordítással és szerencsével várhatóan szintén össze tudjuk gyűjteni a teljes kollekciót.



7. ábra.

9. Alkalmazás az onkológiában

A következőkben a rákos megbetegedés egy többfokozatú szimmetrikus modelljében fogom megmutatni a kupongyűjtő probléma megjelenését.

9.1. A rák alapjai

A sejtosztódás vagy sejtburjánzás egy fiziológiai folyamat. Normális esetben a sejtek burjánzásának és elhalásának egyensúlya szigorúan szabályozottan megy végbe, hogy a szervek és szövetek integritása megmaradjon. A rákbetegségek közös jellemzője, hogy megszakítják ezt a rendezett folyamatot, amivel szabályozatlanná válik a sejtszaporulat. Ezt a kontrollálatlan növekedést olyan DNS-hibák, genetikai mutációk okozzák, melyek a sejtciklus szabályozásában vesznek részt. Általában több ilyen mutációra van szükség a daganat kialakulásához. Néhány ilyen hibát kemikáliák, fizikai hatások okoznak, mások öröklődnek vagy éppen spontán jelennek meg, azaz genetikus és környezeti tényezők együttesen vezethetnek eltorzult növekedési szabályozáshoz. Az ilyen szabályozatlan és gyakran igen gyors sejtburjánzás jóindulatú vagy rosszindulatú daganat (rák) kialakulásához vezethet.

9.2. Modell

Louis Anthony Cox Jr. [7] irodalomban a következő modellt vizsgálta. A sejtvonalunk K különböző transzformációból álló készletből fokozatosan felhalmoz különböző transzformációkat, mint például a tesi öröklődő mutációt, DNS-hibát, stb... A transzformációk véletlenszerűen tűnnek fel függetlenül az időtől, így megfeleltethetők Poisson folyamatoknak megfelelő intenzitással és λ időegységenkénti átlagos előfordulással. (Viszont így az olyan transzformációk, melyek a sokkal kisebb előfordulási rátájuk miatt nem szabályozhatók ilyen módon figyelmen kívül maradhatnak.) Tegyük fel, hogy ha egy K transzformációból bármelyik feltűnik, akkor az tartós és visszafordíthatatlan változást okoz. Emellett tudjuk, hogy ha egy speciális K transzformáció többször is feltűnik, akkor az már nem gyorsítja tovább a folyamatot, tehát az, hogy például ez a speciális eset egyszer vagy tízszer tűnik fel az nem számít. Tudjuk még, hogy egy sejtvonal élettartama véges, jelöljük ezt T idővel. Így ezen ismeretek birtokában a modellünk úgy fog kinézni, hogy ha minden K különböző transzformáció feltűnik a T idő előtt, akkor a sejtvonal rosszindulatúvá válik.

9.3. Megoldás

Az egész modell motivációja tehát, hogy bármelyik K transzformáció felbukkanhat elsőként anélkül, hogy jelentős hatása lenne, de a felbukkanása után (mivel egy felbukkanás visszafordíthatatlan) annak a valószínűsége, hogy a következőnek felbukkanó transzformáció hatása is jelentéktelen lesz $\frac{K-1}{K}$-ra csökken. A második transzformáció felbukkanása után $\frac{K-2}{K}$-ra, aztán így tovább egészen addig amíg ez a valószínűség $\frac{1}{K}$-ra csökken. Jelölje n^* a sejt rosszindulatúvá válásának idejét, azaz az előforduló transzformációk száma mire minden K transzformáció legalább egyszer előfordult. Ebben a megfogalmazásban ez megfelel egy K elemű kollekció összegyűjtésének várható idejével. Azaz, (1) képlettel kapjuk, hogy $\mathrm{E}(n^*) = K\ln(K)+K\gamma+\frac{1}{2}+O\big(\frac{1}{K}\big)$, ahol $\gamma \approx 0,57721$ az Euler-Mascheroni állandó. Ezzel még nem végeztünk, mert a kupongyűjtő problémával ellentétben itt a szükséges transzformációk száma mellett lényeges vagy talán még lényegesebb is, hogy ez a transzformáció mennyiség mikorra fordul elő. Jelölje t^* azt az időpontot amikorra minden K transzformáció legalább egyszer előfordult. Mivel minden transzformáció egymástól függetlenül érkezik, egységesen λ arányban, ezért $K\lambda$ lesz annak az aránya, hogy épp melyik transzformáció érkezik az egységnyi időintervallumban. Így felírhatjuk a következő egyenlőséget:

$$\mathrm{E}(t^*) = \frac{\mathrm{E}(n^*)}{K\lambda} = \frac{1}{\lambda}\Big[\ln(K) + \gamma + \frac{1}{2K} + O\big(\frac{1}{K^2}\big)\Big]$$

Motwani és Raghavan (1995) bebizonyította, hogy az n^* eloszlásnak éles átmenete van $\mathrm{E}(n^*)$ körül. Ez a megoldás fixen tartja n^*-ot. Tudjuk, hogy a transzformációk megjelenése felírható Poisson folyamatként, ezért t^* Gamma eloszlást követ $\frac{n^*}{K\lambda}$ várható értékkel és $\frac{n^*}{K^2\lambda^2}$ varianciával. Mivel n^* fix, a szórás aránya a várható értékhez megegyezik $(n^*)^{-\frac{1}{2}} \approx \big(K(\ln(K)+\gamma)\big)^{-\frac{1}{2}}$. Ez az érték 0-hoz tart miközben K tart a végtelenhez.

9.4. Eredmény

Tehát a fenti egyszerűsített modell alapján a modellben van egy éles átmeneti idő: $T^* \approx \frac{1}{\lambda}(\ln K + \gamma)$, ami a első rosszindulatú sejt megjelenésének várható időpontja. Hogyha a sejtvonal elhalásának T időpontja nagyobb, mint ez az érték, akkor a rosszindulatú sejt megjelenésére szinte biztosan számítani lehet. Ellenkező esetben viszont nagyon valószínűtlen a megjelenése.
A valós modellek persze ettől bonyolultabbak, mert a szervezetünkben a rákos sejtek kialakulását segítő gének mellett a kialakulást megakadályozó gének is vannak, amit a modell nem vesz figyelembe. Viszont így is tanulságos az eredmény, hogy komplex sztochasztikus rendszerek olykor felírhatók átmenetekkel és valószínűségi törvényekkel annak ellenére is, hogy a rendszer részletei ismeretlenek.

Hivatkozások

[1] Donald J. Newman, Lawrence Shepp, *The double dixie cup problem.* American Mathematical Monthly, 67:58-61, 1960.

[2] Erdős P.,Rényi A *On a classical problem of probability theory.* Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei, 6: 215–220, 1961.

[3] Craig A. Tracy, *The coupon problem*, Tananyag
`www.math.ucdavis.edu/∼tracy/courses/math135A/UsefullCourseMaterial/couponProblem.pdf`

[4] Központi statisztikai hivatal, *A háztartások fogyasztása*, 2015
`https://www.ksh.hu/docs/hun/xftp/stattukor/haztfogy/haztfogy1512.pdf`

[5] S. Ross, *A first course in probability,*. 9th Edition, Pearson, 2012

[6] M. Ferrante, M. Saltalamacchia, *The Coupon Collector's Problem.* MATerials MATemàtics Volum 2014, treball no. 2, 35 pp. ISSN: 1887-1097

[7] Louis Anthony Cox Jr., *Risk Analysis of Complex and Uncertain Systems.*

[8] Lars Holst, *On Birthday, Collectors', Occupancy and Other Classical Urn Problems.* Internat. Statist. Rev., 54:15-27, 1986.

[9] T. Nakata, *Coupon collector's problem with unlike probabilities.* Preprint, 2008.