Col·leccions de cromos

Emili Pecharroman Ravella

Grau d'Estadística Aplicada
Universitat Autònoma de Barcelona

Tutor: Xavier Bardina Simorra

26 de gener de 2017

Índex

Resum

En aquest treball estudiarem alguns dels diferents aspectes dins de les col·leccions de qualsevol tipus. Val a dir, que d'aquí en endavant suposarem que estem treballant amb col·leccions de cromos, com a fi de poder explicar els procediments i resultats de manera més senzilla i clara.

La principal incògnita que ens pararem a estudiar és la quantitat de cromos en mitjana que hem de comprar per a completar una col·lecció de N cromos. Així, presentarem diferents mètodes per al càlcul d'aquesta esperança, i, segons ens anem endinsant en el món de les col·leccions, ens interessarem en calcular aquesta esperança sota diferents hipòtesis, com per exemple, que les probabilitats d'obtindre els cromos siguin diferents, és a dir, que hi hagin alguns cromos dins de la col·lecció que siguin més dificils d'aconseguir que uns altres, que és el que a priori, sembla més realista. També calcularem aquesta esperança suposant que hi han vàries persones fent la mateixa col·lecció i, fins i tot, quants cromos de més hauríem de comprar per a completar una segona col·lecció després d'haver-ne completat una, guardant-ne els repetits.

Paraules clau: Col·leccions de cromos, Cadenes de Markov, Adrenalyn, Simulació.

Resumen

En este trabajo estudiaremos algunos de los diferentes aspectos dentro de las colecciones de cualquier tipo. Cabe decir, que, de aquí en adelante supondremos que estamos trabajando con colecciones de cromos, con el fin de poder explicar los procedimientos y resultados de manera más sencilla y clara.

La principal incógnita que nos pararemos a estudiar es la cantidad media de cromos que hemos de comprar para completar una colección de N cromos. De tal modo, presentaremos diferentes métodos para el cálculo de esta esperanza, y según nos vayamos adentrando en el mundo de las colecciones, nos interesaremos en calcular esta esperanza bajo diferentes hipótesis, como por ejemplo, que las probabilidades de obtener los cromos sean diferentes, es decir, que hayan algunos cromos dentro de la colección que sean más difíciles de obtener que otros, lo que, a priori, parece más realista. También calcularemos esta esperanza suponiendo que hay varias personas haciendo la misma colección e, inclusive, cuantos cromos de más deberíamos comprar para completar una segunda colección después de haber completado la primera guardando los repetidos.

Palabras clave: Colecciones de cromos, Cadenas de Markov, Adrenalyn, Simulación.

Abstract

In this dissertation we will study some of the different aspects within the collections of any type. It should be said that, from now on, we will assume that we are working with collections of cards, in order to be able to explain the procedures and results in a simpler and clearer way.

The main and first unknown point that we will study will be the average amount of cards that we have to buy to complete a collection of N cards. Thus, we will present different methods for the calculation of this mean, once we have estimated this mean, we will be interested in calculating the same mean under different hypotheses, for example, assuming the chances of obtaining the cards are different, in other words, there are some cards in the collection that are more difficult to obtain than others, this assumption, theoretically, seems to be more realistic. We will also calculate this average by assuming that there are several people making the same collection, and even, how many cards should we buy to complete a second collection after completing the first one by keeping the repeated ones.

Keywords: Card collections, Markov chains, Adrenalyn, Simulation.

1. Introducció

En aquest treball presentarem un problema clàssic en probabilitat combinatòria, com és el problema del col·leccionista.

Qui no s'ha preguntat alguna vegada si fer una col·lecció concreta resultarà molt complicat?

Si ens parem a pensar, cada cop que comprem aleatòriament un cromo és com si juguèssim a una loteria, la qual estaria condicionada segons les probabilitats de sortir de cada cromo, i també estaria condicionada per la quantitat de cromos no repetits que tenim respecte de la quantital total de cromos diferents que hi ha. Per exemple, si totes les probabilitats fossin iguals i tinguèssim 10 cromos diferents, la primera loteria que jugaríem encertaríem segur per obtindre un de diferent, la segona costaria una mica més ja que servirien només 9 dels 10 cromos, i així fins a completar-la.

Llavors sembla lògic pensar que si quan juguem a la loteria sabem que el més probable és no guanyar, quan fem una col·lecció ens podem fer una idea de que un cop tenim una quantitat elevada de cromos diferents respecte del total de cromos diferents, obtenir un altre que ens interessi serà realment complicat, més encara si els cromos que ens quedessin tinguessin menys probabilitat de sortir.

A l'hora de completar una col·lecció amb N cromos diferents, poden surgir-nos algunes preguntes.

- Quants cromos hauré de comprar/obtenir en mitjana per completar la col·lecció?
- Si es fa una única col·lecció entre i persones, quants cromos s'hauràn de comprar per a tenir completa una sola col·lecció?
- Si considerem les probabilitats d'aparició dels cromos desproporcionals, com varien els temps per completar una col·lecció?

Un cop haguem donat resposta a aquestes qüestions, ens centrarem en un cas concret i actual. I veurem empíricament quant es trigaria a fer la col·lecció.

2. Compra unitària i probabilitats iguals

En aquest apartat ens interessarem per la quantitat de cromos en mitjana que s'han de comprar per a completar una col·lecció de N cromos diferents.

Com a premisa suposarem que la probabilitat d'obtenir cada cromo és la mateixa, és a dir: $p = \frac{1}{n}$.

2.1. Distribució geomètrica

En aquest cas utilitzem la distribució geomètrica ja que quan comprem un cromo poden passar només dues situacions: que el cromo obtingut sigui un cromo que ja tinguem, o el que és el mateix, un cromo repetit (fracàs) o que el cromo obtingut encara no el tinguem, és a dir un cromo faltant (èxit).

Si definim X com el nombre de proves independents que hem de realitzar fins a obtindre un cromo diferent (èxit), tenim:

$$P\{X = m\} = pq^{(m-1)} = p(1-p)^{(m-1)}$$

Com sabem l'esperança d'una variable aleatòria amb distribució geomètrica és $E(X) = \frac{1}{p}$ ja que:

$$E(X) = \sum_{m=1}^{\infty} mp(1-p)^{(m-1)} = \frac{p}{(1-(1-p))^2} = \frac{1}{p}$$

Que l'esperança d'una geomètrica sigui justament $1/p$, ens servirà pel cas d'arribada unitària on tots els cromos tenen la mateixa probabilitat de sortir.

Si ara denotem X com el total de cromos diferents que podem obtindre, podem dir que X és igual al conjunt de tenir tots els cromos X_i. Dit d'una altra manera:

$X = (X_1, X_2, \ldots, X_n)$ suposant que cada X_i és el número de cromos a comprar per a pasar de tenir $X_{(i-1)}$ a tenir X_i.

Per tant, hem vist sota aquesta definició, X_i són variables aleatòries que seguixen una distribució geomètrica, on l'exit és aconseguir un cromo que encara no tinguem a la col·lecció.

Quan tenim $(i-1)$ cromos, la probabilitat d'obtrindre un cromo que encara no tinguem és:

$$\frac{FALTANTS}{TOTALS} = \frac{N-(i-1)}{N}$$

Completar la col·lecció X implica, per tant, obtenir 1 èxit en cada X_i. Per tant el temps d'espera per completar la col·lecció serà l'esperança de la variable aleatòria X.

$$E(X) = E(X_1 + X_2 + \cdots + X_N) = E(X_1) + E(X_2) + \cdots + E(X_N) = \frac{1}{p_1} + \frac{1}{p_2} + \cdots + \frac{1}{p_N} =$$

$$= \frac{N}{N} + \frac{N}{N-1} + \frac{N}{N-2} + \cdots + \frac{N}{2} + N = N \sum_{i=1}^{N} \frac{1}{i}$$

Exemple:

Una col·lecció de $N = 10$ cromos diferents amb una probabilitat igual de sortir cada cromo. $p_i = \frac{N-(i-1)}{N}$

Utilitzarem de forma genèrica l'esperança de la geomètrica per estimar quants cromos calen comprar per a tenir-ne de $(i-1)$ cromos diferents a i cromos diferents.

$E[X_1] = 1/p_1 = N/N = 10/10 = 1$

$E[X_2] = 1/p_2 = N/(N-1) = 10/9$

$E[X_3] = 1/p_3 = N/(N-2) = 10/8$

$E[X_4] = 1/p_4 = N/(N-3) = 10/7$

$E[X_5] = 1/p_5 = N/(N-4) = 10/6$

$E[X_6] = 1/p_6 = N/(N-5) = 10/5$

$E[X_7] = 1/p_7 = N/(N-6) = 10/4$

$E[X_8] = 1/p_8 = N/(N-7) = 10/3$

$E[X_9] = 1/p_9 = N/(N-8) = 10/2$

$E[X_{10}] = 1/p_{10} = N/(N-9) = 10$

Per tant, si volem saber quants cromos calen en mitjana per a completar la col·lecció de 10 cromos, podem fer la suma de totes les esperances de les geomètriques.

$$E[X] = 1 + 10/9 + 10/8 + 10/7 + 10/6 + 10/5 + 10/4 + 10/3 + 10/2 + 10 = 29.2896$$

2.2. Funció d'R

De manera genèrica podem utilitzar la següent funció d'R per calcular les esperances de completar una col·lecció amb probabilitats iguals i arribada unitària depenent de n.

Primer definim la quantitat de cromos diferents que té la col·lecció. En la funció hem agafat $n = 10$.

```
n=10
```

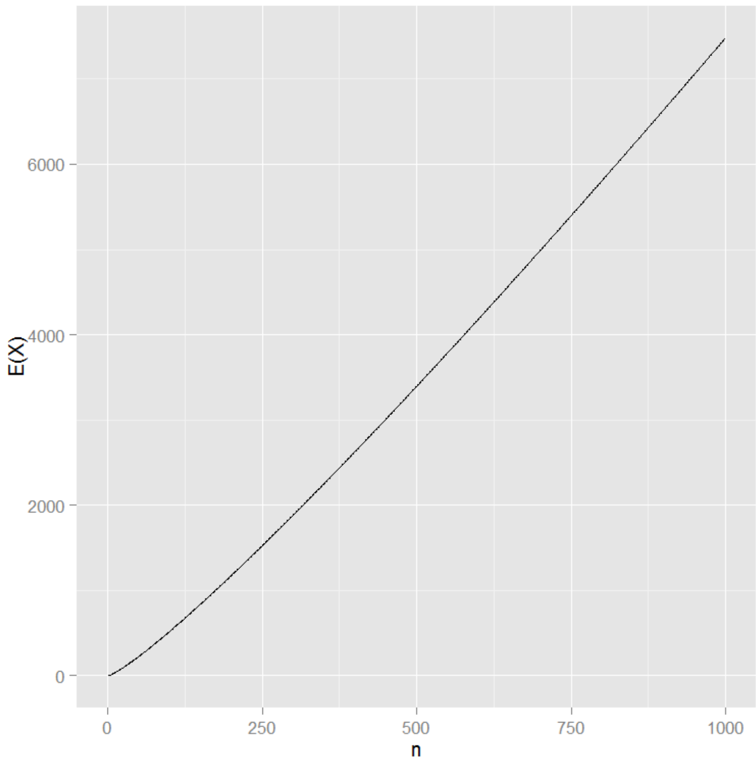
Un cop definit n, fem de manera iterativa els càlculs anteriors:

```
for (i in 1:n){
x=x+(n/i)
}
```

Finalment tenim guardat a la variable x la mitjana de cromos que haurem de comprar per a completar la col·lecció. En aquest cas $x = 29.29$ cromos.

A continuació observem una taula amb altres exemples d'esperances d'X i un gràfic en funció de les diferents N.

$\boldsymbol{N}$	10	20	50	100	1000
$\boldsymbol{E(X)}$	29.29	71.95	224.96	518.74	7485.47



Gràfic 1. E(X) en funció de n, probabilitat igual, arribada unitària

2.3. Cadenes de Markov

Ara emprant les cadenes de Markov, tractarem de resoldre la probabilitat de completar una col·lecció de N cromos amb les mateixes probabilitats de sortir, de manera equivalent a l'apartat anterior.

Suposem que cada cromo arriba en un temps t, de manera discreta.

X_i tornarà a ser el temps d'espera per obtindre l'ièssim cromo diferent tenint-ne (i-1) de diferents.

Definim Y_n com el nombre de cromos diferents obtinguts havent-ne comprat n. La probabilitat d'obtindre cada cromo és $1/N$.

La variable Y_n és una cadena de Markov amb estats de 0 fins a N, un total de $N+1$ estats.

La matriu de transició en un pas és la següent:

$$P = \begin{pmatrix} 0 & 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \frac{1}{N} & \frac{N-1}{N} & 0 & \cdots & \cdots & 0 \\ \vdots & 0 & \frac{2}{N} & \frac{N-2}{N} & 0 & \cdots & 0 \\ \vdots & \vdots & 0 & 0 & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \frac{N-2}{N} & \frac{2}{N} & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & \frac{N-1}{N} & \frac{1}{N} \\ 0 & \cdots & \cdots & \cdots & 0 & 0 & 1 \end{pmatrix}$$

Ara, com a l'apartat anterior, volem calcular el temps per a completar la col·lecció desde zero. És a dir, passar de l'estat 0 a l'estat N en w passos. On w seràn el total de cromos que comprarem per obtenir la col·lecció sencera.

Per definició sabem que en les cadenes de Markov, el temps d'espera per arribar per primer cop a j partint de l'estat i és:

$$\mu_{ij} = 1 + \sum_{k \neq j} p_{ik}\, \mu_{kj}$$

p_{ik} és la probabilitat de passar de l'estat i a l'estat k en un sol pas.

Llavors, volem calcular μ_{0N} que és el temps en passar de tenir 0 cromos diferents a tenir-ne N de diferents. Definim $t_i = \mu_{iN}$, que serà el temps d'espera per a completar la col·lecció partint de i cromos diferents.

Volem arribar a calcular $t_0 = \mu_{0N}$

$$\begin{cases} t_N = 0 \\ t_i = 1 + \sum_{j \neq N} p_{ij}\, t_j, \qquad i \neq N \end{cases}$$

$$\begin{cases} t_N = 0 \\ t_{N-1} = 1 + \dfrac{N-1}{N}\, t_{N-1} \\ t_{N-2} = 1 + \dfrac{N-2}{N}\, t_{N-2} + \dfrac{2}{N}\, t_{N-1} \\ t_{N-3} = 1 + \dfrac{N-3}{N}\, t_{N-3} + \dfrac{3}{N}\, t_{N-2} \\ \vdots \\ t_1 = 1 + \dfrac{1}{N}\, t_1 + \dfrac{N-1}{N}\, t_2 \\ t_0 = 1 + \dfrac{0}{N}\, t_0 + \dfrac{N}{N}\, t_1 = 1 + t_1 \end{cases}$$

$$\begin{cases} t_N = 0 \\ t_{N-1} = N \\ t_{N-2} = N + \dfrac{N}{2} \\ t_{N-3} = N + \dfrac{N}{2} + \dfrac{N}{3} \\ \vdots \\ t_1 = N + \dfrac{N}{2} + \dfrac{N}{3} + \cdots + \dfrac{N}{N-2} + \dfrac{N}{N-1} \\ t_0 = N + \dfrac{N}{2} + \dfrac{N}{3} + \cdots + \dfrac{N}{N-2} + \dfrac{N}{N-1} + 1 \end{cases}$$

Finalment obtenim el temps per a completar la col·lecció desde que comencem fins que l'acabem sencera.

$$\mu_{0N} = t_0 = N \sum_{i=1}^{N} \frac{1}{i}$$

Exemple:

Una col·lecció de $N = 10$ cromos diferents amb una probabilitat de igual de que surti cada cromo.

La matriu de transició en un sol pas és:

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.1 & 0.9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.2 & 0.8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.3 & 0.7 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.4 & 0.6 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.6 & 0.4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.7 & 0.3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.8 & 0.2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.9 & 0.1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\mu_{0N} = 10 + \frac{10}{2} + \frac{10}{3} + \frac{10}{4} + \frac{10}{5} + \frac{10}{6} + \frac{10}{7} + \frac{10}{8} + \frac{10}{9} + 1 = 29.2896$$

És la mateix resultat que havíem obtingut al apartat anterior, per tant concluim que el temps de espera per a completar una col·lecció de N cromos és:

$$E(X) = N \sum_{i=1}^{N} \frac{1}{i}$$

2.4. Aproximació

Als apartats anteriors hem vist com la fórmula que ens calcula la quantitat de cromos que hem d'obtenir per acabar la col·lecció és:

$$E(X) = N\sum_{i=1}^{N}\frac{1}{i}$$

Com podem observar, aquesta fórmula inclou una sèrie harmònica.

$$\sum_{i=1}^{N}\frac{1}{i} = 1 + \frac{1}{2} + \cdots + \frac{1}{N}$$

Quan volem fer càlculs per n grans, els càlculs que requereix la sèrie harmònica fan que el procediment sigui lent.

En canvi, podríem trobar alguna altra fórmula que aproximi bé $E(X)$ i que sigui més eficient a l'hora de calcular-la per n grans.

Aproximarem la sèrie harmònica per:

$$\sum_{i=1}^{N}\frac{1}{i} = \ln(N) + \gamma + \frac{1}{2N}$$

on $\gamma = \pm 0.5772156649015328606065120$9 és la constant de Euler-Mascheroni.

$$\gamma = \lim_{n\to\infty}\left(1 + \frac{1}{2} + \cdots + \frac{1}{n} - \ln(n)\right) = 0.5772156649015328606065120 9\ldots$$

Per tant utilitzant aquesta funció podem aproximar per n grans com:

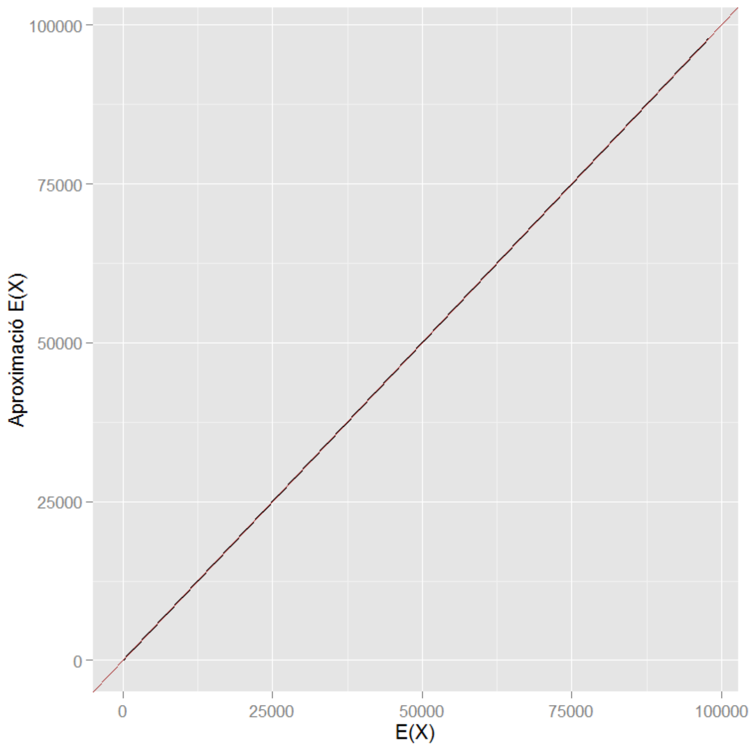
$$\widehat{E(X)} = N\cdot\left(\ln(N) + \gamma + \frac{1}{2N}\right)$$

A la taula següent s'observen els errors en l'estimació de la funció utilitzada, respecte al valor real.

N	5	10	20	100	1000	10000
$E(X)$	11.417	29.290	71.955	518.738	7485.471	97876.06
$\widehat{E(X)}$	11.433	29.298	71.959	518.739	7485.471	97876.06
$\widehat{E(X)} - E(X)$	0.01660	0.00833	0.00417	0.00083	0.00008	0.00001

Error en l'aproximació de E(X)

Veiem com a mesura que augmenta la n l'error disminueix, per tant és una bona aproximació.



Gràfic 2. Comparació entre $E(X)$ i $\widehat{E(X)}$

3. Compra unitària i probabilitats desiguals

Després d'haver-nos parat a estudiar les col·leccions amb probabilitats iguals, ens interessarem ara en els casos més reals, que són les col·leccions amb probabilitats desiguals. Estudiarem una altra vegada la quantitat de cromos en mitjana que s'han de comprar per a completar una col·lecció de N cromos diferents.

Com a premisa suposarem que la probabilitat d'obtenir un cromo i és $p_i > 0$, amb $p_1 + p_2 + \cdots + p_N = 1$.

3.1. Fórmula Maximums-Minimums Identity

X_i, serà el nombre aleatori de cromos que hem de comprar fins a tenir una unitat del cromo i. Llavors per a completar la col·lecció sencera haurem de comprar el màxim en tot X_i, per tant definirem la variable X com $X = \max\{X_1, \dots, X_N\}$, que definit en esperances resulta:

$$E(X) = E(\max\{X_i, \dots, X_N\}) = E(max_i X_i)$$

La fórmula de "Maximums-minimums Identity" serveix per calcular el màxim de qualsevol successió de variables aleatòries positives i finites. És la següent:

$$max_i X_i = \sum_i X_i - \sum_{i<j} \min(X_i, X_j) + \sum_{i<j<k} \min(X_i, X_j, X_k) - \cdots + (-1)^{(N+1)} \min(X_1, \dots, X_N)$$

Ens interessa trobar $E(X)$.

Llavors, aplicant les igualtats anteriors, tenim:

$$E(X) = E(max_i X_i) =$$
$$= \sum_i E(X_i) - \sum_{i<j} \mathrm{E}(\min(X_i, X_j)) + \sum_{i<j<k} \mathrm{E}(\min(X_i, X_j, X_k)) - \cdots$$
$$+ (-1)^{(N+1)} \mathrm{E}(\min(X_1, \ldots, X_N))$$

Sabem també que les variables X_i segueixen distribucions geomètriques de paràmetre p_i, però no són idènticament distribuides.

Deduim que:

$$E(X) = \sum_i \frac{1}{p_i} - \sum_{i<j} \left(\frac{1}{p_i + p_j}\right) + \sum_{i<j<k} \left(\frac{1}{p_i + p_j + p_k}\right) - \cdots + (-1)^{(N+1)} \left(\frac{1}{p_1 + \cdots + p_N}\right)$$

Aquesta fórmula és la que usarem per a calcular l'esperança de completar una col·lecció de N cromos amb probabilitats p_i.

Exemple:

- $N = 2, p_1 = \frac{1}{3};\ p_2 = \frac{2}{3}$
  $$E\left(X_{\left(\frac{1}{3},\frac{2}{3}\right)}\right) = \frac{1}{p_1} + \frac{1}{p_2} - \frac{1}{p_1 + p_2} = 3 + \frac{3}{2} - 1 = \frac{7}{2} = 3.5$$
- $N = 3, p_1 = \frac{1}{10};\ p_2 = \frac{2}{10}; p_3 = \frac{7}{10}$
  $$E\left(X_{\left(\frac{1}{10},\frac{2}{10},\frac{7}{10}\right)}\right) = \frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3} - \frac{1}{p_1 + p_2} - \frac{1}{p_1 + p_3} - \frac{1}{p_2 + p_3} + \frac{1}{p_1 + p_2 + p_3} =$$
  $$= 10 + 5 + \frac{10}{7} - \frac{10}{3} - \frac{10}{8} - \frac{10}{9} + 1 \approx 11.73$$

3.2. Funció d'R

De manera genèrica podem fer una funció d'R per la qual, segons un vector de probabilitat p, ens calculi l'esperança de completar una col·lecció de N cromos, on N, seria la longitut del vector p. Evidentment, la funció també calcula les probabilitats del apartat de probabilitats iguals sempre que el vector que introduim tingui les mateixa probabilitat per a tots els cromos.

Comprovarem que la funció fa els càlculs de maner correcta emprant els exemples anteriorment utilitzats. En aquest cas calcularem $E\left(X_{\left(\frac{1}{10},\frac{2}{10},\frac{7}{10}\right)}\right)$

```
p=c(1/10,2/10,7/10)
```

Comprovem que p efectivament és un vector de probabilitat ja que la suma de les seves components és 1.

```
sum(p) == 1
```

Utilitzarem la variable s per anar sumant iteriativament.

```
s=0
```

El bucle que usarem és farà un total de "j" vegades, on "j" és la quantitat de cromos que hi ha, o equivalentment, la longitut del vector p. Per a cada "j" trobarà el nombre de combinacions del vector "p" agafats de "j" en "j", posteriorment mitjançant un altre bucle, aplicarem la forma genèrica de la funció Maximum-Minimum Identity per a cada combinació trobada: $(-1)^{(j+1)}\left(\frac{1}{q[k,]}\right)$.

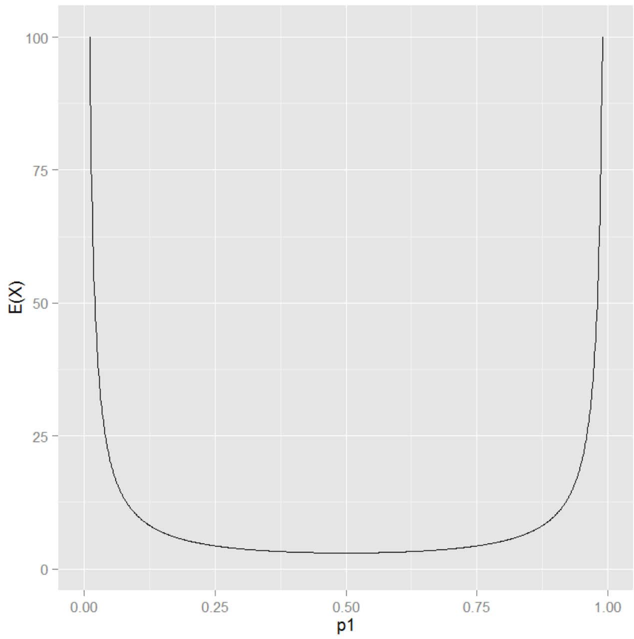
El resultat d'aplicar la fórmula per a cada combinació per a cada j, s'anirà acumulant en la variable s, la qual, quan la funció hagi acabat serà la $E(X)$ que buscàvem.

```
for (j in 1:length(p)){

q=CombSet(p,j)

        for (k in 1:dim(q)[1]){
        s=s+((-1)^(j+1))*sum(1/sum(q[k,]))
        }
}
```

Finalment tenim $s = 11.73$, el mateix resultat que a exemple anterior.

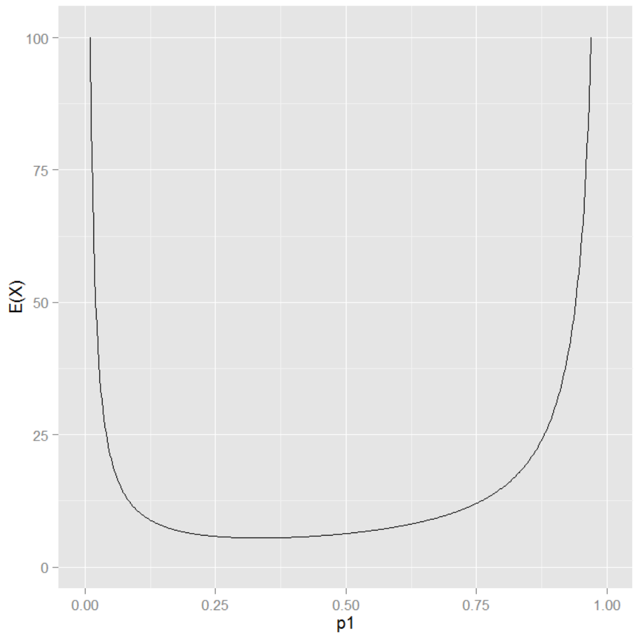
Si el vector p fós un vector de probabilitats iguals com per exemple, una col·lecció de $n = 10$ amb probabilitats d'obtenir cada cromo $p = 1/10$. Tenim que s, i per tant $E(X) = 29.29$. Que coincideix amb la probabilitat trobada al apartat de probabilitats iguals.

A continuació presentem alguns gràfics interessants obtinguts:



Gràfic 3. $E(X)$ en funció de p_1 amb $N = 2$.

Podem veure com l'esperança d'X es comporta com una exponencial en els extrems del gràfic, això ens indica precissament que, com més propera a 0 sigui la probabilitat de sortir d'un dels dos cromos, l'esperança tendirà cada cop més a infinit. El mínim s'assoleix quan els dos cromos tenen la mateixa probabilitat de sortir.



Gràfic 4. $E(X)$ en funció de p_1, per $N = 3$ amb $p_2 = p_3$.

Com estem fixant $p_2 = p_3$, això vol dir que si p_1 creix en un 1%, p2 i p3 es reduriràn en 0.5%, ja que els dos valen sempre (1-p3)/2. Seguint aquest criteri podem observar exactament el mateix que en el gràfic anterior, tenint en compte que el creixement de la funció tendeix més ràpid a infinit per la detra que per l'esquerra pel motiu que acabem d'explicar.

4. Cas concret. Col·lecció cromos Adrenalyn.

4.1. Funció d'R.

En aquest apartat estudiarem un cas concret mitjançant simulacions.

La col·lecció escollida es tracta d'una col·lecció de Panini anomenada Adrenalyn. Consisteix en 559 cromos de jugadors de fútbol de primera divisió dels anys 2013-2014.

A través de la pàgina web on es venen cromos, hem pogut deduir que les probabilitats de que apareguin els cromos són desiguals, ja que alguns es venen més cars que altres. En concret, la col·lecció es divideix en 4:

- CARDS BASE+CAMPEÓN+ACTUALIZ.+NUEVO FICH.+PLUS ENTRENADOR: **0.50€ (N=447)**
- ÍDOLOS+NUEVOS ÍDOLOS: **1.20€ (N=46)**
- PORTERAZOS+DIAMANTES+NUEVOS DIAM: **1.60€ (N=22)**
- BALÓN DE ORO+CARD INVENCIBLE+JUGONES+(NUEVOS) SÚPER CRACKS: **3€ (N=44)**

Els cromos estàn numerats de 1 a 559, i no estàn ordenats, és a dir, els cromos van intercalats segons els preus. Amb la següent funció d'R creem la nostra llista de ID pels cromos:

```
id=c(1:559)
```

Amb la següent fució d'R calculem les probabilitats de que surti cada cromo dels 559. Com els cromos no estàn en ordre, el vector on guardarem les probabilitats, el fem de manera que tingui en compte tots els canvis de preu segons passem del cromo i al cromo $(i+1)$.

```
p=1/(447*1/0.5+46*1/1.2+1/1.6*22+1/3*44)
p1=p*1/0.5
p2=p*1/1.2
p3=p*1/1.6
p4=p*1/3
```

```
precio=c(rep(0.5,360),rep(1.6,20),rep(1.2,40),rep(3,37),rep(0.5,56),rep(1.6,2),rep(1.2,6),rep(3,
7),rep(0.5,31))
```

```
prob=c(rep(p1,360),rep(p3,20),rep(p2,40),rep(p4,37),rep(p1,56),rep(p3,2),rep(p2,6),rep(p4,7),
rep(p1,31))
```

El programa permet escollir quantes persones fan la col·lecció a l'hora, en aquest cas possarem que hi han 3 persones fent la col·lecció a la vegada.

```
num.pers=3
```

El programa fa els càlculs suposant que totes les persones completin una mateixa col·lecció.

Per a que el programa funcioni hem d'inicialitzar les variables que utilitzarem.

```
xyz.tiempo.total.length=NULL
xyz.tiempo.mitad.length=NULL
xyz.tiempo.total.2colecion.length=NULL
xyz.tiempo.mitad.2colecion.length=NULL
m2=NULL
```

Un cop inicialitzades, podem començar amb la funció. La variable "k" serà el nombre de vegades que es farà la simulació.

```
for (k in 1:100){
```

Amb la següent instrucció realitzem una mostra aleatòria de la col·lecció, tenint en compte les probabilitats calculades anteriorment. Fixem-nos que la grandària de la mostra està sent multiplicada per "num.pers", el número de persones que fan la col·lecció.

```
x2<-sample(id,20000*num.pers, prob=prob,replace=TRUE)
```

Creem la matriu "m1" on cada fila serà un individu que fa una col·lecció, això s'aconsegueix dividint la mostra aleatòria feta en el pas anterior entre el número de persones que fan la col·lecció.

```
m1=matrix(0,num.pers,(length(x2)/num.pers))
k1=1
k2=(length(x2)/num.pers)

for (i in 1:num.pers){
        m1[i,]=x2[k1:k2]
        k1=k1+(length(x2)/num.pers)
        k2=k2+(length(x2)/num.pers)
}
```

Seguidament, ara ens interessarem en els cromos totals que hem de comprar per a completar la col·lecció sencera (559 cromos diferents), i els cromos totals que hem de comprar per a completar el 50% de la mateixa (280 cromos diferents). Per aquest fi, usarem les variables internes "j" i "j2". Les quals serviran per indexar els vectors, i aniràn incrementant-se fins a cumplir els requeriments demanats. En aquest cas "j" deixarà d'incrementar en el moment que ja haguem obtingut 280 cromos diferents (meitat de la col·lecció) i "j2" deixarà d'incrementar-se quan la tinguem sencera (559 cromos diferents). Així, per exemple, sabrem que en "j2" passos (és a dir, cromos totals comprats), s'ha completat la col·lecció.

```
j=1
j2=1
```

El bucle on anirem incrementant "j" i "j2" és farà com a màxim tants cops com columnes tinguem a la matriu "m1", ja que recordem que cada fila correspon a un individu.

```
for (i in 1:dim(m1)[2]){
```

Si agafant el vector desde 1 fins a "j" encara no s'han agafat 280 cromos diferents, la variable "j" s'incrementa en 1 unitat.

```
if (length(unique(c(m1[,c(1:j)]))) < length(id)/2){
        j=j+1
        j2=j
}
```

La variable "j2" comença valent "j" un cop s'hagi completat el 50% de la col·lecció, i segueix la mateixa metodologia. Si de 1 a "j2" no s'ha completat encara la col·lecció sencera, s'incrementa "j2" en 1 unitat.

```
if ((length(unique(c(m1[,c(1:j2)]))) >= length(id)/2) & (length(unique(c(m1[,c(1:j2)]))) <
        length(id))){
        j2=j2+1
}

}
```

Guardem els valors de "j" i "j2" en les següents variables, les quals guardaràn "k" valors de "j" i de "j2". Ja que "k" és el número de simulacions que es faran.

```
xyz.tiempo.total.length[k]=j2
xyz.tiempo.mitad.length[k]=j
```

També ens interessem per la quantitat de cromos que haurem de comprar per a completar una segona col·lecció, després d'haver-ne completat la primera, és a dir, partirem amb una gran quantitat de cromos repetits.

El vector "v" gardarà tots els cromos que hi han entre 1 i "j2" (completar la col·lecció).

```
v=c(m1[,c(1:j2)])
```

El que farem serà restar de "v" una unitat de cada cromo, per a representar que ja s'ha completat una col·lecció i a la vegada mantenim els repetits. La següent funció, treballa de manera que mentre la "w" sigui 0, per a cada cromo troba la primera vegada que surt i el borra, en aquest moment la variable "w" passa de valer 0 a valer 1 i, per tant, canvia de cromo i torna a començar el bucle.

```
for (j in 1:length(id)){
        w=0

        for (i in 1:length(v)){
                if (v[i]==id[j] & w==0){
                        v=v[-i]
                        w=1
                }
        }

}
```

A partir d'aqui farem el procediment equivalent a "j" i "j2", però en aquest cas, en comptes de treballar amb els vectors de 1 a "j" o de 1 a "j2", el que farem serà treballar, amb el vector "v" seguit del vector començant en "j2 + 1" i acabant en "j3" o "j4".

La variable "j3" serà la quantitat total de cromos per a completar la 2a col·lecció. La variable "j4" serà la quantitat total de cromos per a completar la meitat 2a col·lecció. Si al acabar la funció restem "j3" menys "j2" tindrem exactament el número de cromos que haurem de comprar per a completar la segona col·lecció despres de completar la primera.

```
j4=j2+1
j3=j2+1

for (i in (j2+1):dim(m1)[2]){
```

La variable "j4" valdrà "j2" la majoria dels casos, ja que casi sempre s'haurà completat la meitat de la col·lecció només amb els repetits.

```
if (length(unique(c(v))>length(id)/2)) {
        j4=j2
}

if (length(unique(c(v,m1[,c((j2+1):j4)]))) < length(id)/2){
        j4=j4+1
        j3=j4
}

if ((length(unique(c(v,m1[,c((j2+1):j3)]))) >= length(id)/2) &
(length(unique(c(v,m1[,c((j2+1):j3)]))) < length(id))){
        j3=j3+1
}

}

xyz.tiempo.total.2colecion.length[k]=j3
xyz.tiempo.mitad.2colecion.length[k]=j4
```

Finalment creem la matriu "m2", on guardarem per a cada "k" alguns resultats interesants de la simulació.

```
m3=matrix(0,j2,4)
m3=rbind(m3,m2)
m2=m3
```

La matriu "m2" tindrà tantes files com la suma de totes les "j2", i per a cada fila calcularà:

- A la primer columna: "Preu total per a completar la col·lecció comprant els que falten".
- A la segona columna: "Quants cromos diferents falten per completar la col·lecció".
- A la tercera columna: "Quants cromos totals falten per completar la col·lecció".
- A la quarta columna: "Quants cromos s'han comprat fins al moment".

```
for (i in 1:j2){
        h=unique(sort(c(m1[,c(1:i)])))
        m2[i,1]=sum(precio[-c(h)])
        m2[i,2]=length(id)-length(unique(c(m1[,c(1:i)])))
        m2[i,3]=(j2-i)
        m2[i,4]=i*num.pers
}
```

I aquí tanquem el bucle que es repetirà "k" vegades

```
}
```

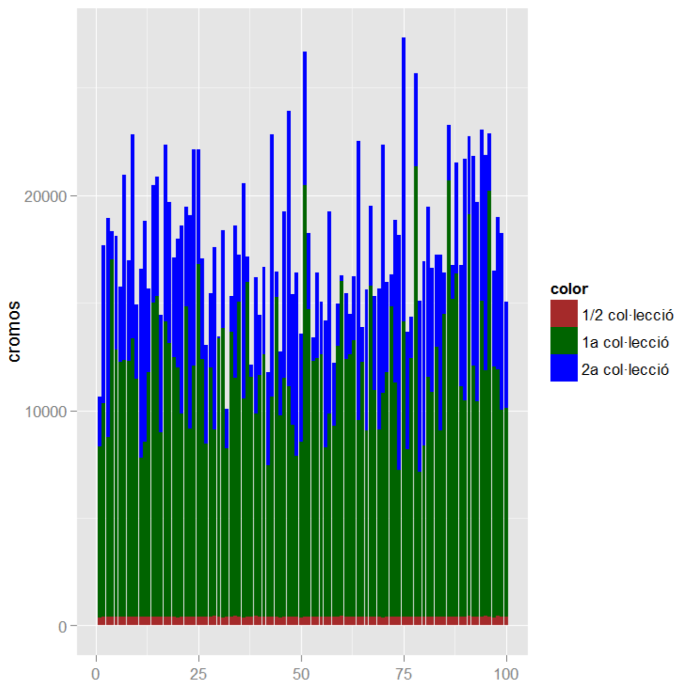
4.2. Resultats d'aplicar la funció amb diferents paràmetres.

A continuació tenim diferents gràfics i resultats després d'executar la funció, detalladament explicada a l'apartat anterior, en simulacions de 100 col·leccions, amb diferents paràmetres:

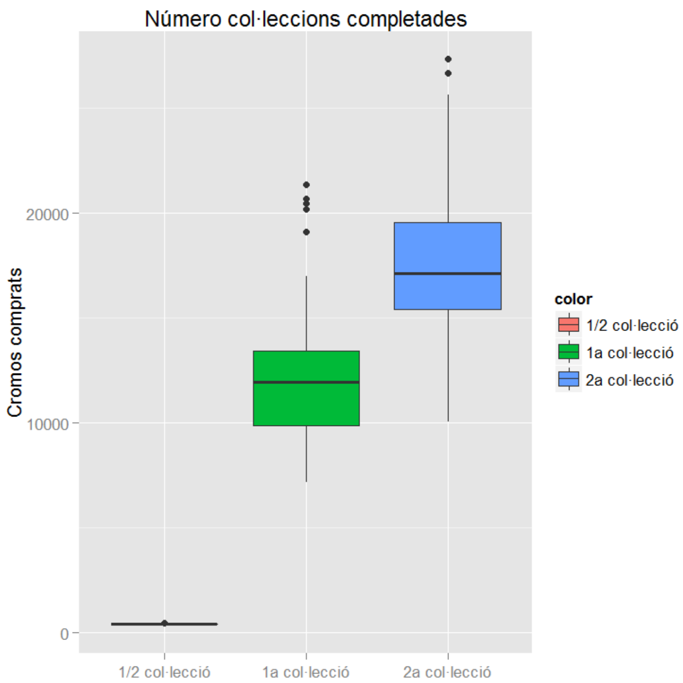
- **k=100; num.pers=1; prob=desiguals**

Els següents resultats provenen d'una simulació per k=100, amb una única persona realitzant la col·lecció a la vegada i amb les probabilitats calculades a partir del preu dels cromos.

```
k=100
num.pers=1
prob=c(rep(p1,360), rep(p3,20), rep(p2,40), rep(p4,37), rep(p1,56), rep(p3,2),
rep(p2,6), rep(p4,7),rep(p1,31))
```



Gràfic 5. Simulacions de 100 col·leccions fins a ser completades.



Gràfic 6. Boxplot de les $E(X)$ per a les diferents col·leccions simulades.

Podem comprovar que les esperances no són gens proporcionals. Mitja col·lecció es completa comprant 400 cromos aproximadament.

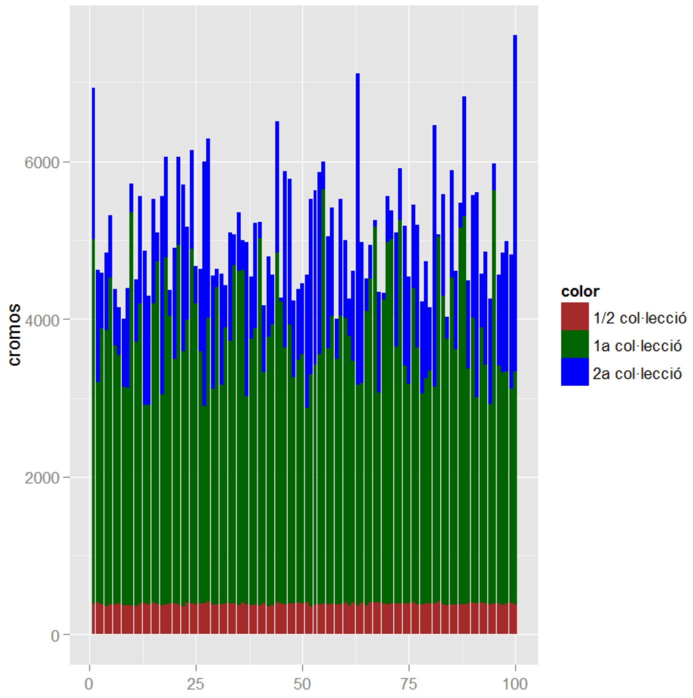
Una col·lecció sencera requereix bastants més, concretament 12000 cromos aproximadament (30 cops més). Això vol dir que si 30 persones completessin una única col·lecció a la vegada, comprant cadascuna 400 cromos, probablement aconseguirien entre tots completar-ne una.

Per últim, per a completar una segona col·lecció després d'haver-ne completat la primera, calen en total uns 17700 cromos aproximadament. És a dir, comprar 5700 cromos més (en mitjana) després de completar la primera col·lecció.

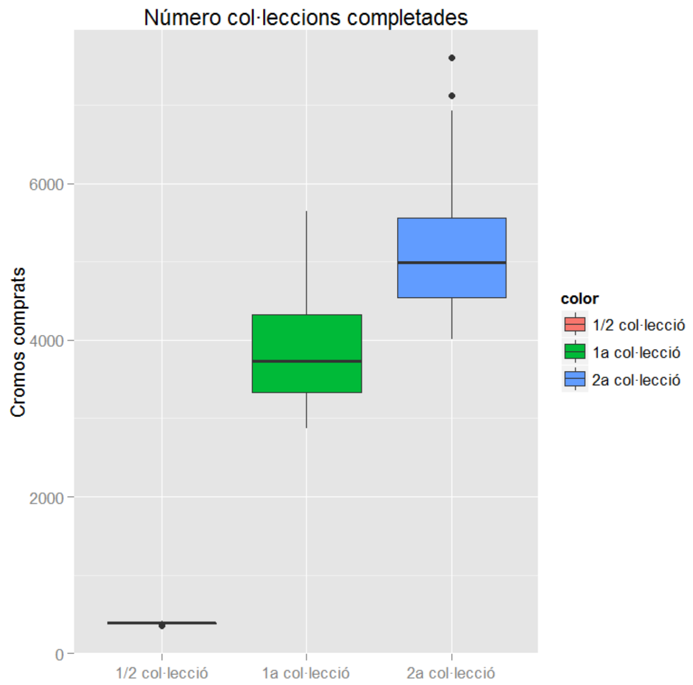
- **k=100; num.pers=1; prob=iguals**

Els següents resultats provenen d'una simulació per k=100, amb una única persona realitzant la col·lecció a la vegada i amb les probabilitats iguals.

```
k=100
num.pers=1
prob=rep(1/length(id),length(id))
```



Gràfic 7. Simulacions de 100 col·leccions fins a ser completades.



Gràfic 8. Boxplot de les $E(X)$ per a les diferents col·leccions simulades.

Amb probabilitats iguals, les esperances per a completar les col·leccions es redueixen bastant, de fet quan les probabilitats són les mateixes les esperances per a completar les col·leccions són les mínimes.

Per a completar mitja col·lecció hem de comprar 390 cromos aproximadament en mitjana. Per a completar una col·lecció hem de comprar 3890 cromos aproximadament en mitjana. Per a completar la segona col·lecció hem de comprar 5100 cromos en mitjana.

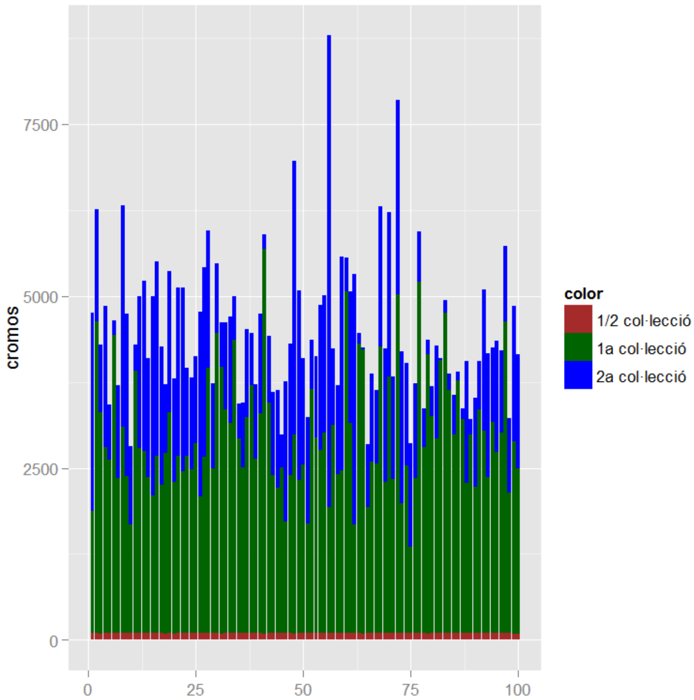
Comparant amb l'apartat anterior, veiem que ara per a completar la col·lecció, el nombre de cromos a comprar es redueix bastant, concreament es redueix en 3 vegades. A més, en aquest cas, la quantitat de cromos a comprar per a completar una col·lecció és 10 vegades els cromos a comprar per a completar-ne mitja (3 vegades inferior que a l'apartat anterior, on eren 30). El que vol dir que només amb 10 persones i, a més, cada persona comprant 10 cromos menys que a l'apartat anterior, podríem completar una col·lecció sencera.

La quantitat de cromos a comprar per a completar la segona col·lecció també és 3 vegades inferior que a l'apartat anterior.

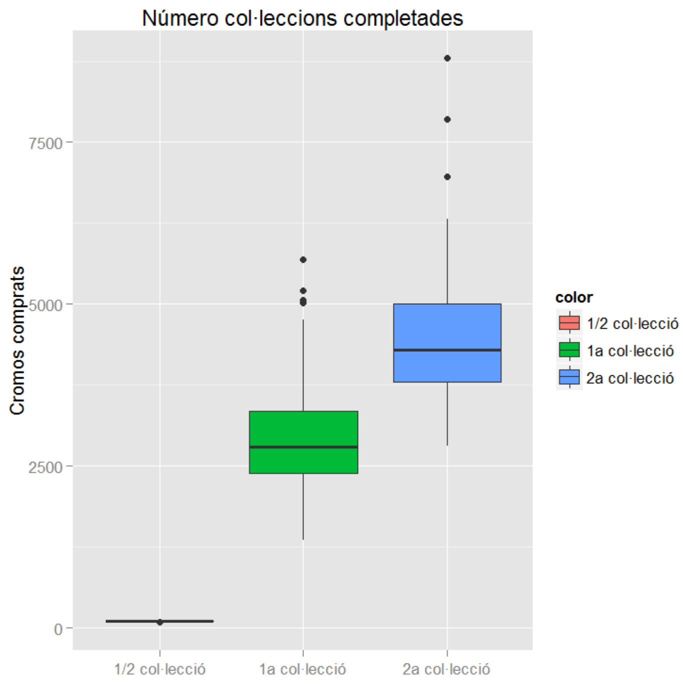
- **k=100; num.pers=4; prob=desiguals**

Els següents resultats provenen d'una simulació per k=100, amb 4 persones realitzant la col·lecció a la vegada i amb les probabilitats calculades a partir del preu dels cromos.

```
k=100
num.pers=4
prob=rep(1/length(id),length(id))
```



Gràfic 9. Simulacions de 100 col·leccions fins a ser completades.



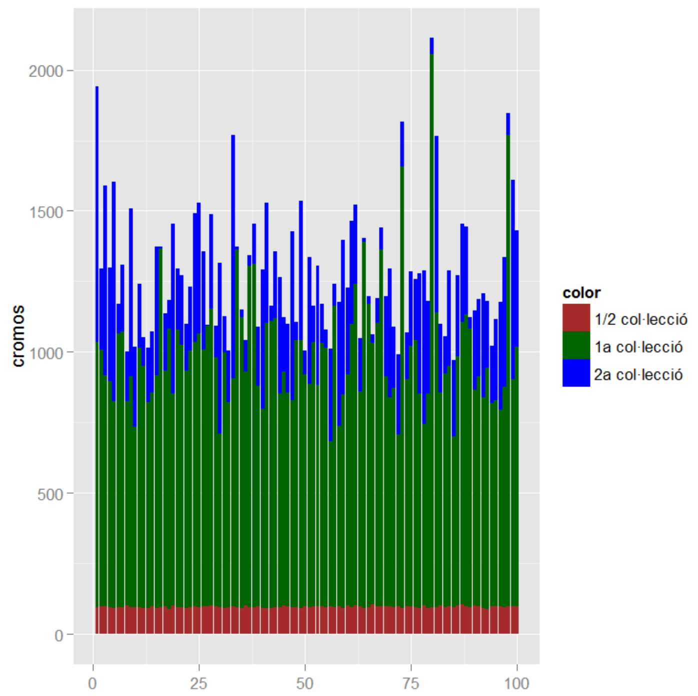
Gràfic 10. Boxplot de les $E(X)$ per a les diferents col·leccions simulades.

Podem compravar com els resultats obtinguts són exactament equivalents als resultats per a una única persona fent la col·lecció amb les probabilitats agafades a partir del preu dels cromos, però en aquest cas els cromos que haurem de comprar són la quarta part, ja que la col·lecció és fa entre 4.

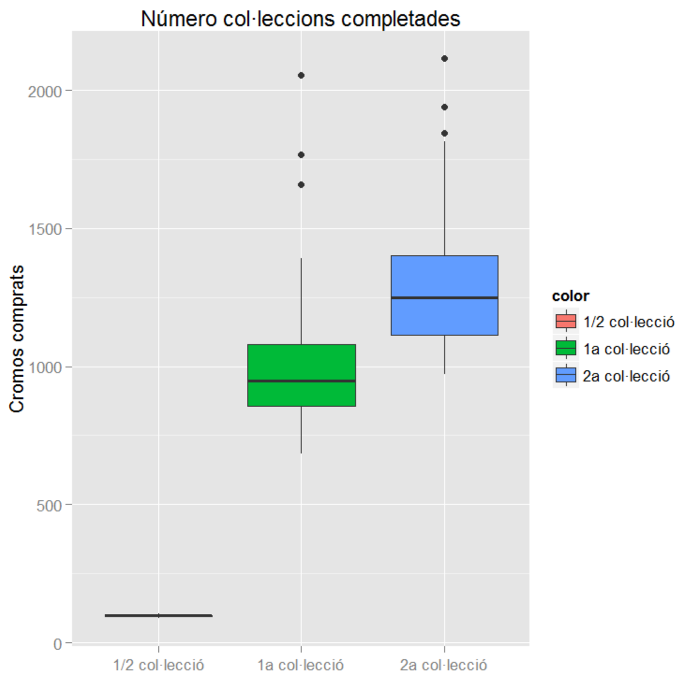
- **k=100; num.pers=4; prob=iguals**

Els següents resultats provenen d'una simulació per k=100, amb 4 persones realitzant la col·lecció a la vegada i amb les probabilitats iguals.

```
k=100
num.pers=4
prob=rep(1/length(id),length(id))
```



Gràfic 11. Simulacions de 100 col·leccions fins a ser completades.



Gràfic 12. Boxplot de les $E(X)$ per a les diferents col·leccions simulades.

Podem compravar com els resultats obtinguts són exactament equivalents als resultats per a una única persona fent la col·lecció amb les mateixes probabilitats d'obtenir cada cromo, però en aquest cas els cromos que haurem de comprar són la quarta part, ja que la col·lecció és fa entre 4.

Taula resum:

A continuació indiquem una taula resum, amb els resultats obtinguts i explicats anteriorment.

Probabilitats	**Número persones**	$E(X)$ **mitja col·lecció**	$E(X)$ **1a col·lecció**	$E(X)$ **2a col·lecció**
Vector prob*	1	408.26	12087.15	17731.67
Iguals	1	387.48	3887.71	5103.95
Vector prob*	4	102.85	3003.07	4500.04
Iguals	4	97.25	998.2	1282.05

**prob és el vector de probabilitats calculat a partir del preu específic dels cromos*

5. Conclusions

La quantitat de cromos que hem de comprar per a completar una col·lecció varia molt en funció de diferents paràmetres.

El paràmetre que més afecta a aquesta quantitat de cromos a comprar són les probabilitats que té de sortir cada cromo. Com més complicat sigui d'aconseguir un cromo concret, més cromos haurem de comprar. I, de forma lògica, el mínim de cromos per a completar la col·lecció es troba en el cas que tots els cromos tinguin la mateixa probabilitat de sortir.

Si fem una col·lecció entre i persones, la quantitat de cromos totals que hem de comprar entre les i persones és la mateixa, però la quantitat que hem de comprar individualment és la ièssima part del total de cromos comprats. A la vida real l'ideal seria que tots compartíssim els cromos per així haver-ne de comprar menys.

Pel nostre cas concret de la col·lecció Adrenalyn 2013-2014, podem veure que si volguèssim fer la col·lecció sencera nossaltres sols, hauríem de comprar un total de 12000 cromos aproximadament en mitjana. Sabem que els cromos es compren a les botigues en sobres d'1€ on a cada sobre hi han 6 cromos. Per tant, per a completar la col·lecció hauríem de comprar en mitjana 2000 sobres, el que es tradueix en 2000€ per a completar la col·lecció comprant tots els cromos a través de sobres.

A més, els cromos es poden comprar també a través de la seva pàgina web, com hem explicat, hi han alguns més cars que d'altres. Si volguèssim completar la col·lecció comprant-los tots a través de la pàgina web ens costaria exactament 445.9€, 4 vegades i mitja més barat que comprant els cromos en sobres.

No obstant mitjançant la pàgina web només deixen comprar una quantitat limitada de cromos, concretament:

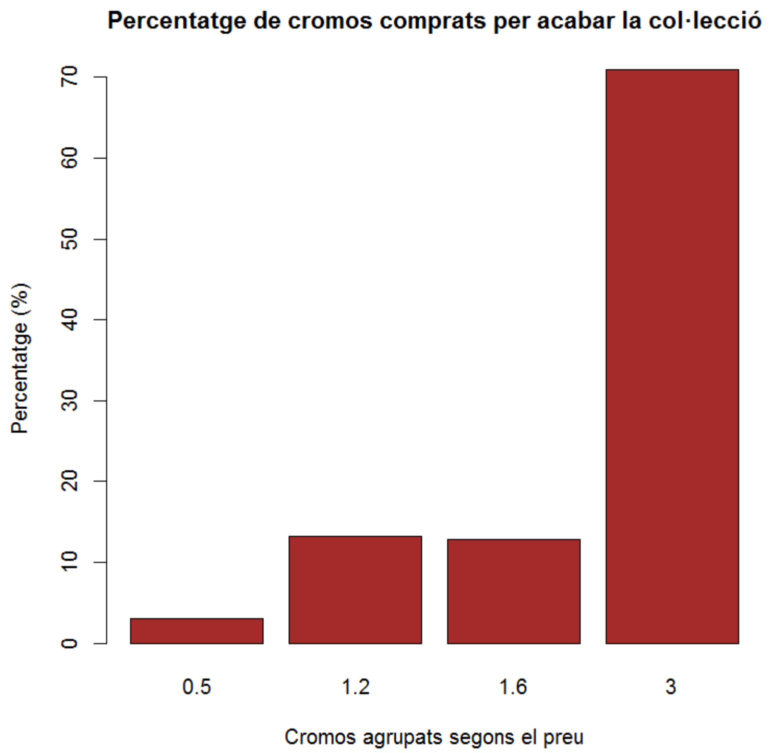
- 66 cromos del total de 447 cromos de 0.5€.
- 18 cromos del total de 46 cromos de 1.20€.
- 9 cromos del total de 22 cromos de 1.60€.
- 12 cromos del total de 44 cromos de 3€.

Un total de 105 cromos. El que suposa una despessa de 105€, si compréssim el màxim, sense comptar les despeses d'enviament.

Hem creat un programa derivat i específic per calcular els diners aproximats que gastaríem en completar una col·lecció si decidíssim comprar els cromos faltants un cop tenim com a mínim: 381 cromos diferents de 0.5€, 28 cromos diferents de 1.20€, 13 cromos diferents de 1.60€ i 32 cromos diferents de 3€. És a dir, un cop tenim el mínim de cromos per a poder completar la col·lecció comprant el màxim que ens deixen comprar a la pàgina. La funció d'R la podem trobar a l'annex final com a "Funció final. Simulació cas real. Comprant cromos".

Els resultats obtinguts, en una simulació de 100 col·lecions, són:

El preu total que gastaríem comprant sobres i els cromos faltants conjuntament és de 650€ en mitjana. En mitjana, hauríem de comprar de manera aleatòria 3650 cromos, 608 sobres aproximadament i hauríem de comprar mitjançant la pàgina web un total de 17 cromos en mitjana. D'aquest 17 cromos, estan distribuits segons el següent gràfic, en el qual es pot observar que la majoria dels que demanaríem són dels cromos més cars i per tant més complicats d'obtenir.



Gràfic 13. Percentatge de cromos de cada tipus que hauríem de comprar per a completar una col·lecció comprant els cromos que falten

6. Bibliografia

[1] M. Ferrante, N. Frigo, *A note in the coupon-collector's problem with multiple arrivals and the random sampling.*

[2] M. Ferrante, N. Frigo, *On the expected number of different records in a random sample.*

[3] Ross, Sheldon, *A First Course in Probability.*

[4] Xavier Bardina, *Rècords: Quina és la probabilitat d'obtenir-ne? Quan apareixen? Quins valors prenen?*

[5] Armengol Gassull, *Sumes harmòniques.*

Dades obtingudes de la pàgina web: http://collectibles.panini.es/

- *http://collectibles.panini.es/store/col_esp_es/002733cesid-it-football-es-2013-14-tcg-liga-figurina-idolos.html*
- *http://collectibles.panini.es/store/col_esp_es/002733cesbb-it-football-es-2013-14-tcg-liga-figurina-base.html*
- *http://collectibles.panini.es/store/col_esp_es/002733cesbi-it-football-es-2013-14-tcg-liga-figurina-balon-invincible.html*
- *http://collectibles.panini.es/store/col_esp_es/002733cesjs-it-football-es-2013-14-tcg-liga-figurina-jugones-supercrak.html*
- *http://collectibles.panini.es/store/col_esp_es/002733cespd-it-football-es-2013-14-tcg-liga-figurina-porterazos-diamantes.html*

7. Annex

Paquets necessaris per a poder executar el codi R.

```
install.packages("ggplot2")
install.packages("combinat")
```

Probabilitats iguals, arribada unitària.

```
n=NULL
a=NULL
for (n in 1:1000){
        x=0
        for (i in 1:n){
                x=x+(n/i)
        }
        a[n]=x
}

qplot(1:n,a,geom="line",xlab="n", ylab="E(X)")
```

Cadenes de markov arribada unitaria probs iguals.

```
n=10
b=matrix(0,n+1,n+1)
colnames(b)=c(0:n)
rownames(b)=c(0:n)
a=0

for (i in 1:n){
        fila=c(rep(0,a),a/n,(n-a)/n,rep(0,n-(a+1)))
        b[i,]=fila
        a=a+1
}

b[n+1,]=c(rep(0,n),1)

b #Matriu de transició de tenir i cromos a tenir-ne i+1 de diferents
```

Càlcul de t_{0N}.

```
n=100
t=c(rep(0,n+1))
res=0
```

```
for (i in 1:n){
	res=res+n/i
	t[i+1]=res
}

t0=t[n+1]
```

Càlcul de t_{0N} per a totes les mides mostrals de 1 fins a n.

```
n=100
j=0
t0=c(rep(0,n))

for (x in 0:n){
	t=c(rep(0,x+1))
	res=0
	for (i in 1:x){
		res=res+x/i
		t[i+1]=res
	}

	t0[j]=t[x+1]
	j=j+1

}

t0
```

Aproximació d'$E(X)$.

COMPARACIÓ ENTRE LAS 2 FUNCIONS QUE APROXIMEN E(X) quan n tendeix a infinit

```
#Funció a=n*sumatorio(1/i); i=(1..n)

#La funció que aproxima a la funció a serà b=n*log(n)

a=NULL
b=NULL
i=NULL

for (n in 1:10000){
	i=c(1:n)
	a[n]=n*sum(1/i)
	b[n]=n*(log(n)+(1/(2*n))+0.57721566490153286060651209)
}

qplot(a,b,xlab="E(X)",ylab="Aproximació E(X)",geom="line")+ geom_abline(col="brown")
```

Probabilitats desiguals, arribada unitària. Serveix per a qualsevol llargària del vector p, és a dir, per a qualsevol N, però recordem que p sempre ha de sumar 1.

```
p=c(1/3,1/3,1/3)
s=0

for (j in 1:length(p)){
	q=CombSet(p,j)
	for (k in 1:dim(q)[1]){
		s=s+((-1)^(j+1))*sum(1/sum(q[k,]))
	}
}

s
```

Funció final. Simulació cas real. Serveix per a qualsevol vector "prob", però recordem que sempre ha de sumar 1. Serveix també per a qualsevol "num.pers".

```
p=1/(447*1/0.5+46*1/1.2+1/1.6*22+1/3*44)
p1=p*1/0.5
p2=p*1/1.2
p3=p*1/1.6
p4=p*1/3

prob=c(rep(p1,360),rep(p3,20),rep(p2,40),rep(p4,37),rep(p1,56),rep(p3,2),rep(p2,6),rep(p4,7),
rep(p1,31))
precio=c(rep(0.5,360),rep(1.6,20),rep(1.2,40),rep(3,37),rep(0.5,56),rep(1.6,2),rep(1.2,6),rep(3,
7),rep(0.5,31))
id=c(1:559)

xyz.tiempo.total.length=NULL
xyz.tiempo.mitad.length=NULL
xyz.tiempo.total.2colecion.length=NULL
xyz.tiempo.mitad.2colecion.length=NULL
m2=NULL

num.pers=1 #numero de persones que fan una mateixa col·lecció a la vegada

for (k in 1:100){
	x2<-sample(id,40000*num.pers, prob=prob,replace=TRUE)
	m1=matrix(0,num.pers,(length(x2)/num.pers))

	k1=1
	k2=(length(x2)/num.pers)
```

#Creem matriu m1 on cada fila és un individu que fa la col·lecció

```
for (i in 1:num.pers){
	m1[i,]=x2[k1:k2]
	k1=k1+(length(x2)/num.pers)
	k2=k2+(length(x2)/num.pers)
}
```

#del cromo 1 al cromo j ó j2

```
j=1
j2=1

for (i in 1:dim(m1)[2]){
	if (length(unique(c(m1[,c(1:j)]))) < length(id)/2){
		j=j+1
		j2=j
	}
	if ((length(unique(c(m1[,c(1:j2)]))) >= length(id)/2) &
	(length(unique(c(m1[,c(1:j2)]))) < length(id))){
		j2=j2+1
	}
}

xyz.tiempo.total.length[k]=j2
xyz.tiempo.mitad.length[k]=j
```

#Completar una segona colecció després de la primera

```
v=c(m1[,c(1:j2)])
for (j in 1:length(id)){
	w=0
	for (i in 1:length(v)){
		if (v[i]==id[j] & w==0){
		v=v[-i]
		w=1
		}
	}
}
```

#v es el vector al cual en j2 le hemos quitado 1 cromo de cada

```
j4=j2+1
j3=j2+1

for (i in j2+1:dim(m1)[2]){
	if (length(unique(c(v))>length(id)/2)) {
```

```
                j4=j2
                }

                if (length(unique(c(v,m1[,c((j2+1):j4)]))) < length(id)/2){
                        j4=j4+1
                        j3=j4
                }

                if ((length(unique(c(v,m1[,c((j2+1):j3)]))) >= length(id)/2) &
                (length(unique(c(v,m1[,c((j2+1):j3)]))) < length(id))){
                        j3=j3+1
                }

        }

        xyz.tiempo.total.2colecion.length[k]=j3
        xyz.tiempo.mitad.2colecion.length[k]=j4

#Preu fins a acabar la col·lecció si compréssim els faltants 1 per 1

        m3=matrix(0,j2,4)
        m3=rbind(m3,m2)
        m2=m3

        for (i in 1:j2){
                h=unique(sort(c(m1[,c(1:i)])))
                #Preu total comprant els que falten per completar la col·leccio:
                m2[i,1]=sum(precio[-c(h)])
                #Quants cromos diferents queden per completar la col·lecció:
                m2[i,2]=length(id)-length(unique(c(m1[,c(1:i)])))
                #Quants cromos totals falten per a completar la col·leccio:
                m2[i,3]=(j2-i)
                #Quants cromos s'han comprat fins al temps i:
                m2[i,4]=i*num.pers
        }

        print(k)

}

head(m2)

mean(xyz.tiempo.total.length) #E(X) 1a col·lecció
mean(xyz.tiempo.mitad.length) #E(X) 1/2 col·lecció
mean(xyz.tiempo.total.2colecion.length) #E(X) 2a col·lecció
mean(xyz.tiempo.mitad.2colecion.length) #E(X) 1/2 de la 2a col·lecció
xyz.total.repetidos = (xyz.tiempo.total.length * num.pers)-length(id) #Cromos repetits en la 1a
col·lecció
```

```
#Gràfics
dummy=rep(c(1:100),3)
color=factor(c(rep(1,100),rep(2,100),rep(3,100)),labels=c("1/2 col·lecció", "1a col·lecció","2a
col·lecció"))
cromos=c(xyz.tiempo.mitad.length, (xyz.tiempo.total.length-xyz.tiempo.mitad.length),
(xyz.tiempo.total.2colecion.length-xyz.tiempo.total.length))
cromos2=c(xyz.tiempo.mitad.length,xyz.tiempo.total.length,xyz.tiempo.total.2colecion.length)

g1=data.frame(color,dummy,cromos2,cromos)

ggplot(g1, aes(x=dummy, y=cromos, fill=color)) + geom_bar(stat="identity") +
scale_fill_manual(values=c("brown","darkgreen","blue"))

qplot(color, cromos2, data=g1, geom=c("boxplot"), fill=color, main="Número col·leccions
completades", xlab="", ylab="Cromos comprats")
```

Funció final. Simulació cas real. Comprant cromos. La variable "j.cromprant105" recull la quantitat de cromos a comprar fins a arribar al punt on podem completar la col·lecció comprant els faltants.

```
p=1/(447*1/0.5+46*1/1.2+1/1.6*22+1/3*44)
p1=p*1/0.5
p2=p*1/1.2
p3=p*1/1.6
p4=p*1/3

#Agrupant per preus
precio=c(rep(0.5,360+56+31),rep(1.2,40+6),rep(1.6,20+2),rep(3,37+7))
prob=c(rep(p1,360+56+31),rep(p2,40+6),rep(p3,20+2),rep(p4,37+7))

#Intervals diferents probabilitats de cromos:
#0.5: 1-447
#1.2: 448-493
#1.6: 494-515
#3:   516-559

j.comprant105=NULL
preu.comprant=NULL
cromosf.comprant=NULL
cromos.comprats=NULL
sobres.comprats=NULL
diners.gastats.completar=NULL
l=NULL
l2=NULL
num.pers=1 #numero de persones que fan una mateixa col·lecció a la vegada
```

```
for (k in 1:100){
	j.comprant105=0
	x2<-sample(id,40000*num.pers, prob=prob,replace=TRUE)
	m1=matrix(0,num.pers,(length(x2)/num.pers))
	k1=1
	k2=(length(x2)/num.pers)
	#Creem matriu m1 on cada fila és un individu que fa la col·lecció
	for (i in 1:num.pers){
		m1[i,]=x2[k1:k2]
		k1=k1+(length(x2)/num.pers)
		k2=k2+(length(x2)/num.pers)
	}
	#del cromo 1 al cromo j ó j2
	j=1
	j2=1
	for (i in 1:dim(m1)[2]){
		if (j.comprant105==0){
			if (length(unique(c(m1[,c(1:j)]))) < length(id)/2){
			j=j+1
			j2=j
				if (j.comprant105==0 & sum(sort(unique(c(m1[,c(1:j2)])))<448)
				>= (447-66) & sum(sort(unique(c(m1[,c(1:j2)])))<494 &
				sort(unique(c(m1[,c(1:j2)])))>447) >= (46-18) &
				sum(sort(unique(c(m1[,c(1:j2)])))<516 &
				sort(unique(c(m1[,c(1:j2)])))>493) >= (22-9) &
				sum(sort(unique(c(m1[,c(1:j2)])))<560 &
				sort(unique(c(m1[,c(1:j2)])))>515) >= (44-12)){
					j.comprant105=j2
				}
			}
			if ((length(unique(c(m1[,c(1:j2)]))) >= length(id)/2) &
			(length(unique(c(m1[,c(1:j2)]))) < length(id))){
				j2=j2+1
				if (j.comprant105==0 & sum(sort(unique(c(m1[,c(1:j2)])))<448)
				>= (447-66) & sum(sort(unique(c(m1[,c(1:j2)])))<494 &
				sort(unique(c(m1[,c(1:j2)])))>447) >= (46-18) &
				sum(sort(unique(c(m1[,c(1:j2)])))<516 &
				sort(unique(c(m1[,c(1:j2)])))>493) >= (22-9) &
				sum(sort(unique(c(m1[,c(1:j2)])))<560 &
				sort(unique(c(m1[,c(1:j2)])))>515) >= (44-12)){
					j.comprant105=j2
				}
			}
		}
	}
```

```
        l=(sort(unique(m1[,c(1:j.comprant105)])))
        l2=c(l2,precio[-l])
        preu.comprant[k]=sum(precio[-l])
        cromosf.comprant[k]=559-length(sort(unique(m1[,c(1:j.comprant105)])))
        cromos.comprats[k]=j.comprant105
        sobres.comprats[k]=j.comprant105/6
        diners.gastats.completar[k]=(j.comprant105/6)+sum(precio[-l])

        print(k)
}

data.frame(cromosf.comprant, preu.comprant, cromos.comprats, sobres.comprats,
diners.gastats.completar)

round(prop.table(table(l2))*100,2)
```