Le problème du Calife

M.Gouy G.Huvent A. Ladureau

26 mars 2002

A Bagdad la magnifique, en l'an 840, le Calife *Al-Mamun* est un homme comblé. Son harem[a] compte vingt épouses toutes plus belles les unes que les autres. Et là est le problème du Calife. Comment choisir l'épouse du jour[b]. Cruel dilemme, que l'héritier de la dynastie abbasside ne sait résoudre. Le Calife se souvient alors qu'il compte parmi ses protégés un Mathématicien, un certain *Mohammed ibn Musa al Khawarizmi*, dont le génie est reconnu dans tous le califat. Convoqué, ce dernier propose une solution qui enthousiasme le Calife :

-Je sais, dit-il, que l'un des eunuques est un artiste exceptionnel. Qu'il reproduise les visages des femmes du Calife sur des images. Chaque soir, le Calife tirera d'une urne, au hasard[c], l'une de ces images puis la remettra. Il laissera ainsi la main de Dieu choisir pour lui.

Mais, se demande le Calife, combien de jour me faudra-t-il pour voir toutes mes femmes ? *Mohammed ibn Musa al Khawarizmi* ne sait répondre.

Douze siècles plus tard, on se propose de trouver une réponse à cette question. Plus précisément quel est en moyenne le nombre de jours nécessaire au Calife pour qu'il rencontre, au moins une fois, chacune de ses vingt épouses.

[a] Toute ressemblance avec des personnages ayant existé est délibérée.
[b] Pour la chérir et l'aimer, cela va sans dire.
[c] Az-zahr, les dés.

1 Simulation du phénomène sur calculatrice ou ordinateur

On se place provisoirement dans le cas où il n'y a que deux épouses. On désire donc faire une simulation de ce que l'on dénomme "un tirage aléatoire avec remise".

1.1 Algorithme

Pour étudier ce problème, on a besoin d'illustrer la possession des deux images. On note celles-ci A et B. On simule alors à la main le tirage aléatoire des images :

Jour numéro	Image obtenue	Images déjà vues
1	B	A : non B : oui
2	B	A : non B : oui
3	A	A : oui B : oui

Il faut alors trouver un moyen simple pour traduire sur la calculatrice le fait que l'on a déjà tiré ou pas les images A et B.
Cela peut se faire tout simplement par une liste de deux termes avec pour convention par exemple : 0 signifie on possède l'image, 1 on ne l'a pas.
Dans le cas précèdent on aurait alors :

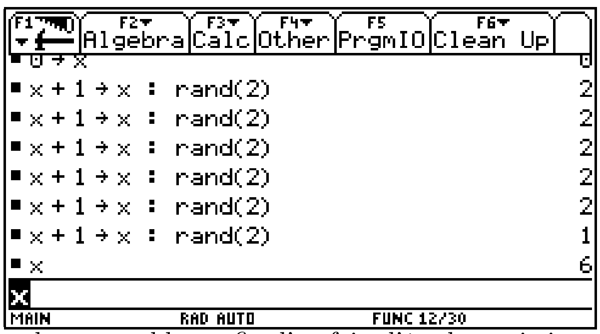
Achat numéro	Image obtenue	Etat de la liste L1
1	B	{1 ;0}
2	B	{1 ;0}
3	A	{0 ;0}

La possession des deux images se traduira donc par le fait que la somme des deux termes de la liste est nulle.
Pour une étude statistique ultérieure, on peut également constituer une liste contenant le nombre de jours écoulés pour que chaque épouse ait été rencontrée au moins une fois. Dans le cas de deux épouses, quelques essais montrent qu'il faut en général moins de 20 jours.
On note alors X et C les variables comptant respectivement le nombre de jours pour une simulation et le cumul sur les N simulations demandées (ainsi C/N fournit une estimation de la réponse à donner au calife, i.e. une estimation du nombre moyen de jours à attendre pour que chaque épouse ait été choisie au moins une fois).

1.2 Une première programmation à la main

Pour le cas d'une unique simulation, on prend l'exemple d'une TI-83 :

Tapez :
`0` → `X` EXE
`X + 1` → `X : randInt(1, 2)` EXE
EXE
etc jusqu'à l'obtention du 1 et du 2.
Pour connaître le nombre de jours nécessaires, il suffit de demander la valeur de `X` (`X` EXE)



On peut alors répéter la simulation et consigner les résultats dans un tableau afin d'en faire l'étude statistique.

1.3 Où l'on automatise la simulation

L'algorithme suivant simule le phénomène étudié dans le cas de plusieurs essais.

```
Entrer le nombre d'essais N
Mettre 0 dans C
Création de la liste L2 de dimension 20 contenant vingt zéros
Pour I allant de 1 à N avec un pas de 1
    Mettre 0 dans X
    Mettre {1,1} dans L1
    Tant que somme des éléments de L1 est non nulle
        Tirer au sort un nombre A entier entre un et deux
        Mettre 0 dans L1(A) : X augmente de 1
    Fin du tant que
L2(X) augmente de 1
C augmente de X
Fin du Pour
Afficher C/N
```

Cet algorithme donne alors les programmes suivants sur calculatrices :

TI-89	TI-83	CASIO
collect()	PromptN	"N = "? → N
Prgm	seq(0, i, I, 20, 1) → L_2	seq(0, I, 1, 20, 1) → List 2
Prompt n	0 → c	0 → c
Delvar l1, l2, i, c, a, x	For I, 1, N, 1	For 1 → I to N Step1
seq(0, i, 1, 20, 1) → l2	seq(I, I, 1, 2, 1) → L_1	seq(I, I, 1, 2, 1) → List 1
0 → c	0 → X	0 → X
For i, 1, n, 1	While sum(L_1) ≠ 0	While sum (List 1)≠ 0
seq(i, i, 1, 2, 1) → l1	RandInt(1, 2) → A	Int (2 ∗ Ran#) → a
0 → x	0 → L_1(A)	0 → List 1[A]
While sum(l1) ≠ 0	X + 1 → X	X + 1 → X
rand(2) → a	End	WhileEnd
0 → l1[a]	L_2(X) + 1 → L_2(X)	List 2[X] + 1 → List 2[X]
x + 1 → x	C + X → X	C + X → X
EndWhile	End	Next
l2[x] + 1 → l2[x]	Disp C/N	C/N
c + x → c		
EndFor		
Disp c/n		
EndPrgm		

Exemples de résultats obtenus (à remplir par le lecteur)

Essai avec N =										
Fréquence observée										

Exercice 1 *Comment modifier le programme précédent pour traiter le cas de* 3 *épouses puis de manière plus générale de* m *épouses ?*

Exercice 2 *Quelle estimation pouvez-vous donner du nombre moyen de jours passés pour que chaque épouse soit vue, au moins une fois, dans les cas suivants ?*

m =										
Estimation										

Exercice 3 *Dans le cas de* $m = 6$, *construire le diagramme statistique faisant apparaître, pour chaque nombre* x *de* 1 *à* 60, *le nombre de fois où il ne fallait pas plus de* n *jours pour que chaque épouse soit rencontrée au moins une fois.*

2 A la recherche d'une première démonstration

On suppose que toutes les images ont la même probabilité d'être choisies. On note X la variable aléatoire égale au nombre de jours à attendre pour avoir vu chaque épouse au moins une fois.

2.1 Le cas de deux images.

Il est clair que X peut prendre toute valeur entière supérieure ou égale à deux.
Soit k un entier strictement supérieur à 1, $X = k$ signifie que les $k-1$ premiers achats ont amené la même image. On en déduit donc que :
$P(X = k) =$ nombre de choix de la première image $\times \left(\frac{1}{2}\right)^{k-1} \times \frac{1}{2} = 2 \times \frac{1}{2}^{k} = \left(\frac{1}{2}\right)^{k-1}$

Remarque 4 *On pourra vérifier que :*

- *La somme des probabilités églale* 1.
- $P(X > 20)$ *est bien faible.*

Le calcul de $\mathrm{E}(X)$ s'avère impossible pour un élève de lycée. On a en effet

$$\mathrm{E}(X) = 2 \times \left(\frac{1}{2}\right) + 3 \times \left(\frac{1}{2}\right)^{2} + \cdots + k \times \left(\frac{1}{2}\right)^{k-1} + \cdots = \sum_{k=2}^{+\infty} k \times \left(\frac{1}{2}\right)^{k-1}$$

Posons $f(x) = \sum_{k \geq 0} x^{k}$, cette série converge sur $]-1,1[$ vers $\dfrac{1}{1-x}$.
On a $f'(x) = \sum_{k \geq 1} k \times x^{k-1}$, cette série convergeant sur $]-1,1[$ vers $\dfrac{1}{(1-x)^2}$
On en déduit que :

$$\mathrm{E}(X) = f\left(\frac{1}{2}\right) - 1 = 3$$

En fait, on peut raisonner autrement et rendre ce calcul abordable pour un lycéen. Il suffit de connaître les suites géométriques.

On pose $f_n(x) = \sum_{k=1}^{n} kx^{k-1}$

On a :

$$\begin{array}{rcccccccccccc}
\sum_{k=1}^{n} kx^{k-1} & = 1 & + & 2x & + & 3x^2 & + & 4x^3 & + & \cdots & + & nx^{n-1} \\
 & = 1 & + & x & + & x^2 & + & x^3 & + & \cdots & + & x^{n-1} \\
 & & + & x & + & x^2 & + & x^3 & + & \cdots & + & x^{n-1} \\
 & & & & + & x^2 & + & x^3 & + & \cdots & + & x^{n-1} \\
 & & & & & & & & & \vdots & & \\
 & & & & & & & & & & + & x^{n-1}
\end{array}$$

Ce qui donne en ajoutant par lignes,

$$\begin{aligned}
\sum_{k=1}^{n} kx^{k-1} = \quad & 1 + x + x^2 + x^3 + \cdots + x^{n-1} \\
& +x + x^2 + x^3 + \cdots + x^{n-1} \\
& +x^2 + x^3 + \cdots + x^{n-1} \\
& \vdots \\
& +x^{n-1} \\
= \quad & \frac{1-x^n}{1-x} + x\frac{1-x^{n-1}}{1-x} + x^2\frac{1-x^{n-2}}{1-x} + ... + x^{n-2}\frac{1-x^2}{1-x} + x^{n-1}\frac{1-x^1}{1-x} \\
= \quad & \left(1 + x + x^2 + x^3 + \cdots + x^{n-1}\right) \times \frac{1}{1-x} - n\frac{x^n}{1-x} \\
= \quad & \frac{1-x^n}{1-x} \times \frac{1}{1-x} - n \times \frac{x^n}{1-x}
\end{aligned}$$

Soit

$$\sum_{k=1}^{n} kx^{k-1} = \frac{1-x^n}{(1-x)^2} - \frac{nx^n}{1-x} \tag{1}$$

En remarquant que pour tout x de $]-1,1[$,on a :

$$\lim_{n\to+\infty} \left(\frac{1-x^n}{1-x}\right) = \frac{1}{1-x}$$

$$\lim_{n\to+\infty} n \times x^n = 0$$

On en déduit que :

$$\lim_{n\to+\infty} \sum_{k=1}^{n} kx^{k-1} = \frac{1}{(1-x)^2}$$

Soit :

$$\sum_{k\geq 1} kx^{k-1} = \frac{1}{(1-x)^2}$$

Remarque 5 *On peut donner une preuve plus rapide de l'égalité* (1) *.*
En remarquant que $\left(\sum_{k=0}^{n} x^k\right)' = \sum_{k=1}^{n} kx^{k-1}$ *soit* $\sum_{k=1}^{n} kx^{k-1} = \left(\frac{1-x^{n+1}}{1-x}\right)' = \frac{1-x^n}{(1-x)^2} - \frac{nx^n}{1-x}$. *La méthode exposée ci-dessus est en fait bien utile, on la retrouvera ultérieurement.*

2.2 Qu'en est-il pour une collection de trois images ?

Première idée : Il est clair que X peut prendre toute valeur entière supérieure ou égale à trois.
Soit k un entier supérieur ou égal à trois, on calcule $P(X = k)$
On note A,B et C les trois images, on a :

$$P(X = k) = 6 \times P(\text{obtenir en exactement } k \text{ tirages les } 3 \text{ images A,B et C dans cet ordre}) = 6P_1$$

Calcul de P_1 : On note, pour tout entier j compris entre 1 et k, U_j la variable aléatoire égale au résultat du j-ième tirage.
Le premier tirage donne l'image A.
Soit i, le plus petit entier compris entre 2 et $k-1$, pour lequel U_i vaut B

L'évènement $E=$ ≪obtenir en exactement k tirages les 3 images A, B et C dans cet ordre≫ est la réunion des évènements de la forme :

Tirage 1	2	3	4	$\cdots$	$i-1$	i	$i+1$	$\cdots$	$\cdots$	$k-1$	k
A	A	A	A	$\cdots$	A	B	A ou B	A ou B	A ou B	A ou B	C

où i varie de 2 à $k-1$.

La probabilité d'un tel évènement est :

$$p_i = \left(\frac{1}{3}\right)^{i-1} \times \frac{1}{3} \times \left(\frac{2}{3}\right)^{k-i-1} \times \frac{1}{3}$$

D'où

$$P_1 = \sum_{i=2}^{K-1} \left(\frac{1}{3}\right)^{i} \times \left(\frac{2}{3}\right)^{K-(i+1)} \times \frac{1}{3}$$

$$\begin{aligned} P(X = K) &= 6 \times \sum_{i=2}^{K-1} \left(\frac{1}{3}\right)^{i} \times \left(\frac{2}{3}\right)^{K-(i+1)} \times \frac{1}{3} \\ &= 6 \times \sum_{i=2}^{K-1} \frac{2^{K-(i+1)}}{3^K} = \left(\frac{2}{3}\right)^{K-1} - 2 \times \left(\frac{1}{3}\right)^{K-1} \end{aligned}$$

On en déduit donc que :

$$\begin{aligned} \mathrm{E}(X) &= \sum_{k\geq 3} k \times P(X = k) = \sum_{k\geq 3} k \times \left(\frac{2}{3}\right)^{k-1} - 2\sum_{k\geq 3} k \times \left(\frac{1}{3}\right)^{k-1} \\ &= f^{\theta}\left(\frac{2}{3}\right) - 1 - 2 \times \frac{2}{3} - 2 \times \left(f^{\theta}\left(\frac{1}{3}\right) - 1 - 2 \times \frac{1}{3}\right) \end{aligned}$$

soit

$$\mathrm{E}(X) = 9 - 1 - \frac{4}{3} - 2\left(\frac{9}{4} - 1 - \frac{2}{3}\right) = \frac{20}{3} - \frac{7}{6} = \frac{33}{6}$$

Ce calcul, bien que simple, se révèle compliqué à généraliser.
On va donc proposer une seconde approche du problème.

Seconde idée : On suppose que chaque image a un numéro de un à trois. On note T_i le numéro d'apparition de l'image numéro i. T_i suit alors une loi géométrique où la probabilité de succès est $\frac{1}{3}$.
D'où : $\forall\ k \in \mathbb{N}^*\ p(T_i = k) = \left(\frac{2}{3}\right)^{k-1} \times \frac{1}{3}$.
On a

$$X = \max(T_1, T_2, T_3)$$

Soit K un entier supérieur ou égal à trois, on a

$$\begin{aligned} X > K &\Longleftrightarrow \max(T_1, T_2, T_3) > K \\ &\Longleftrightarrow T_1 > K \text{ ou } T_2 > K \text{ ou } T_3 > K \end{aligned}$$

Or

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &\quad - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &\quad + P(A \cap B \cap C) \end{aligned}$$

Mais pour tout i compris entre 1 et 3, on a

$$P(T_i > K) = \left(\frac{2}{3}\right)^K$$

pour tout couple (i, j) et $i < j$,

$$P(T_i > K, T_j > K) = \left(\frac{1}{3}\right)^K$$

et

$$P(T_1 > K, T_2 > K, T_3 > K) = 0$$

d'où

$$P(X > K) = 3 \times \left(\left(\frac{2}{3}\right)^K - \left(\frac{1}{3}\right)^K\right)$$

On en tire donc que

$$\begin{aligned} P(X = K) &= P(X > K-1) - P(X > K) \\ &= 3 \times \left(\left(\frac{2}{3}\right)^{K-1} - \left(\frac{1}{3}\right)^{K-1}\right) - 3 \times \left(\left(\frac{2}{3}\right)^K - \left(\frac{1}{3}\right)^K\right) \\ &= \left(\frac{2}{3}\right)^{K-1} - 2 \times \left(\frac{1}{3}\right)^{K-1} \end{aligned}$$

Quitte à calculer $P(X=K)$ d'une autre manière, autant poursuivre et proposer une autre démonstration pour le calcul de $\mathrm{E}(X)$[1].
On a

$$\begin{aligned}
\mathrm{E}(X) &= \textstyle\sum_{K\geq 3} K\times P(X=K)\\
&= \begin{array}{llllllllll} & P(X=3) & + & P(X=3) & + & P(X=3) & & & & \\ + & P(X=4) & + & P(X=4) & + & P(X=4) & + & P(X=4) & & \\ + & P(X=5) & + & P(X=5) & + & P(X=5) & + & P(X=5) & + & P(X=5) \\ + & \cdots & & & & & & & & \end{array}\\
&= \overbrace{P(X>2)} + \overbrace{P(X>2)} + \overbrace{\sum_{K=2}^{+\infty} P(X>K)}
\end{aligned}$$

d'où

$$\mathrm{E}(X) = 2\times\left(3\times\frac{4}{9}-3\times\frac{1}{9}\right)+3\sum_{K=2}^{+\infty}\left(\left(\frac{2}{3}\right)^K-\left(\frac{1}{3}\right)^K\right)=\frac{11}{2}=5,5$$

L'avantage de ces deux méthodes est qu'elles se généralisent bien. Il suffit de connaître (ou de démontrer) la formule de Poincare[2] (cf annexe), dite aussi formule du crible.

Une autre approche de la première idée : Soit k un entier compris entre 1 et 3, on note Y_k la variable aléatoire égale au nombre de jour écoulés pour obtenir k images différentes. On se convainc que :

- $Y_1 = 1$
- $Z_k = Y_k - Y_{k-1}$ représente le nombre de jour à attendre pour voir la $k-$ième épouse lorsqu'on en a déjà vu $k-1$ différentes.
- $X = Y_3 = (Y_3 - Y_2) + (Y_2 - Y_1) + Y_1$

Z_2 représente le nombre de nombre de jour à attendre pour tirer de l'urne la deuxième image lorsqu'on en a déjà vu une. Comme il y trois images, à chaque essai, on a la probabilité $\frac{1}{3}$ de tomber sur l'image déjà vue et la probabilité $\frac{2}{3}$ de tomber sur une image différente. Z_2 suit donc une loi géométrique de paramètre $\frac{2}{3}$.

Z_3 représente le nombre de jour à attendre pour obtenir la troisième image lorsqu'on en déjà vu deux. Comme il y trois images, à chaque essai, on a la probabilité $\frac{2}{3}$ de tomber sur l'une des images vue et la probabilité $\frac{1}{3}$ de tomber sur la dernière image. Z_3 suit donc une loi géométrique de paramètre $\frac{1}{3}$.
Intuitivement, on peut affirmer que Y_1, Z_2 et Z_3 sont indépendantes (inutile pour l'espérance mais indispensable pour la variance). On en déduit alors que :

$$\begin{aligned}
\mathrm{E}(X) &= \mathrm{E}(Y_1)+\mathrm{E}(Z_2)+\mathrm{E}(Z_3)\\
&= 1+\frac{1}{\frac{1}{3}}+\frac{1}{\frac{2}{3}}\\
&= \frac{11}{2}
\end{aligned}$$

[1] Pour le calcul de $\mathrm{E}(X)$, on retrouve l'idée développée page 5.
[2] Jules Henri Poincaré $(1854-1912)$ mathématicien, physicien, astronome et philosophe français.

Exercice 6 *(On reprend la seconde idée)*

1. *Montrer la formule de Poincaré.*
2. *Faire les calculs pour le cas* $n = 4$.
3. *Comparer les résultats avec les simulations faites au A.*
4. *Calculer dans le cas général* $P(X > K)$, *en déduire alors* $P(X = K)$

3 Pour aller plus loin...

On se propose dans cette dernière partie de traiter le problème posé par le Calife dans le cas général.

3.1 Un cas élémentaire : le jeu de pile ou face.

On traite tout d'abord le cas d'un jeu de pile ou face. Afin de compliquer un peu, on suppose que la probabilité d'apparition de la première face est p, celle d'apparition de l'autre face est alors $q = 1 - p$.
La probabilité d'apparition de la première face après k lancers est $q^{k-1}p$ (pour $k \geq 1$ et 0 si $k = 0$!).
Cette simple remarque va nous permettre de résoudre le problème. Avant cela on donne quelques rappels sur les fonctions génératrices de probabilités.

3.1.1 Fonction génératrice de probabilité

Soit X une variable aléatoire à valeurs entières, la fonction génératrice de probabilité est la fonction $G_X(z)$ définie par

$$G_X(z) = \sum_{k=0}^{+\infty} P(X = k)\, z^k$$

Cette fonction est définie comme la somme d'un série entière à coefficients positifs. Puisque $\sum_{k=1}^{+\infty} P(X = k) = 1$, cette série converge normalement sur $[-1, 1]$ et

$$G_X(1) = 1$$

Exemple 7 *Dans le cas du jeu de pile ou face, soit X la variable aléatoire égale au premier rang à partir duquel apparaît la face pile. On a vu que si $k \geq 1$, $P(X = k) = q^{k-1}p$ (X suit une loi géométrique de paramètre p). Ainsi*

$$G_X(z) = \sum_{k=1}^{+\infty} q^{k-1}pz^k = \frac{pz}{1 - qz}$$

3.1.2 Calcul de l'espérance et de la variance

Cette fonction permet un calcul rapide de l'espérance de X et de sa variance. En effet

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{k=0}^{+\infty} kP(X = k) = G'_X(1) \\
\mathrm{E}\left(X^2\right) &= \sum_{k=0}^{+\infty} k^2P(X = k) = G''_X(1) - G'_X(1) \\
V(X) &= G''_X(1) + G'_X(1) - G'_X(1)^2
\end{aligned}$$

Exemple 8 *Dans le cas du jeu de pile ou face,* $\mathrm{E}(X) = \dfrac{d}{dz}\left(\dfrac{pz}{1-qz}\right)_{z=1} = \dfrac{1}{p}$ *et* $V(X) = \dfrac{q}{p^2}$.
Résultat classique pour la loi géométrique.

3.2 Le problème du Calife

On peut maintenant aborder le problème annoncé, ou plus généralement le problème suivant :

3.2.1 Enoncé du problème et calcul de la fonction génératrice de probabilité

Problème 9 *On jette un dé à m faces*[3]*, on note X_k le premier rang à partir duquel on a vu k faces différentes. Quelle est l'espérance de X_k, quelle est la probabilité de $X_k = n$ où n est donné ?*

Pour répondre à la première question (la deuxième est un peu plus difficile), on introduit les fonctions

$$f_k(z) = \sum_{i=0}^{+\infty} P(X_k = i)\, z^i$$

qui sont les fonctions génératrices de probabilité de la variable X_k.

La clef du problème repose dans la remarque suivante :

Ayant vu k faces distinctes, tout se passe comme si l'on joue à pile ou face avec la probabilité $q_k = \dfrac{k}{m}$ de ne pas avoir de nouvelle face.

Ainsi

$$P(X_{k+1} = i) = \sum_{j=0}^{i-1} P(X_k = j)\, q_k^{i-j-1} p_k \text{ où } q_k = 1 - p_k \qquad (E_1)$$

Cette égalité traduit le fait que pour voir, pour la première fois au i-ième tirage, $k+1$ faces, il faut avoir vu k faces (pour la première fois) à un tirage précédent, puis n'avoir pas vu de nouvelles faces avant le i-ième tirage. Si on pose $a_0 = 0$, $a_i = p_k^{i-1} q_k$, l'égalité (E_1) s'écrit

$$P(X_{k+1} = i) = \sum_{j=0}^{i} P(X_k = j)\, a_{i-j}$$

On reconnaît dans cette égalité le produit de deux séries entières, ainsi

$$\begin{aligned} f_{k+1}(z) &= f_k(z) \times \sum_{i=1}^{+\infty} q_k^{i-1} p_k z^i \\ &= f_k(z) \times \frac{p_k z}{1 - q_k z} \\ &= f_k(z) \times \frac{(m-k)\, z}{m - kz} \end{aligned}$$

Avec $f_0(z) = 1$, on obtient[4]

$$f_k(z) = \prod_{i=0}^{k-1} \frac{(m-i)\, z}{m - iz}$$

[3] Est-il posible de construire un dé à m faces équiprobables. Il s'agit d'un autre problème...
On peut énoncer le problème avec une urne dans laquelle on tire, avec remise, des boules numérotées

[4] Il est clair que $f_1(z) = z$, donc il faut $f_0(z) = 1$

Remarque 10 *On peut trouver plus rapidement cette fonction. En effet, si* $Y_1, .., Y_k$ *sont des variables aléatoires indépendantes, et si* $X_k = Y_1 + ... + Y_k$*, alors la fonction génératrice de* X_k *est égale au produit des fonctions génératrices des* Y_i*. Dans notre exemple, on prend* $Y_1 = X_1 (= 1)$, Y_2 *est égal au nombre de lancer supplémentaire pour voir (pour la première fois) une deuxième face, ainsi* $X_2 = Y_1 + Y_2$*. On définit de même* $Y_3, .., Y_n$*. Ainsi* Y_k *est le nombre de lancers qui séparent la première apparition de* $k-1$ *faces et la première apparition de k faces. Il est clair que les* Y_i *sont indépendants. Quand à la loi qui suit* Y_k*, il s'agit d'une loi de pile ou face, i.e. une loi géométrique de paramètre* p_k*, c'est l'idée développée page* 8.

Calcul de l'espérance La fonction f_k s'écrit comme un produit de fonctions qui valent 1 en $z = 1$ (ce sont aussi des fonctions génératrices de probabilité). Si on pose $\phi_i(z) = \frac{(m-i)z}{m-iz}$ alors

$$f'_k(z) = \sum_{i=0}^{k-1} \left(\prod_{\substack{j=0 \\ j \neq i}}^{k-1} \phi_j(z) \right) \phi'_i(z)$$

et par conséquent

$$\begin{aligned} f'_k(1) &= \mathrm{E}(X_k) = \sum_{i=0}^{k-1} \phi'_i(1) \\ &= \sum_{i=0}^{k-1} \frac{m}{m-i} \\ &= m \left(\frac{1}{m} + \frac{1}{m-1} + ... + \frac{1}{m-k+1} \right) \\ &= m(H_m - H_{m-k}) \end{aligned}$$

où

$$\begin{aligned} H_m &= 1 + \frac{1}{2} + \frac{1}{3} + ... + \frac{1}{m} \text{ est le } m\text{-ième nombre harmonique} \\ H_0 &= 1 \end{aligned}$$

Remarque 11 *On constate que* $\mathrm{E}(X_1) = 1$ *(ouf!) et que* $\mathrm{E}(X_m) = mH_m$.

Exemple 12 *Pour :*

Un dé à $m = 2$ *faces. L'espérance de* X_2 *est* $\mathrm{E}(X_2) = 3$

Pour $m = 3$, $\mathrm{E}(X_3) = \frac{11}{2}$

Enfin pour $m = 6$, $\mathrm{E}(X_6) = \dfrac{147}{10} = 14,7$

Exemple 13 *Pour le problème du Calife, avec un harem comptant vingt épouses,* $\mathrm{E}(X_{20}) \simeq 71,95$*. Il faut attendre en moyenne* 72 *jours.*

On va maintenant chercher la probabilité de l'événement $X_k = i$ i.e la probabilité d'avoir vu, pour la première fois, k faces différentes au i-ième jet du dé. Le problème semble simple, on connaît la fonction génératrice de probabilité, il suffit de chercher son développement en séries entières. Pour cela, on a besoin d'étudier une famille de nombres analogues aux coefficients du binôme.

3.2.2 Vade mecum sur les nombres de Stirling (de seconde espèce)

On donne ici les principaux résultats sur les nombres de Stirling[5]. En fait seul un de ces résultats est utile à notre propos.

Définition, relation de récurrence

Définition 14 *Soit n et k deux entiers positifs, on appelle nombre de Stirling de seconde espèce l'entier $\left\{ {n \atop k} \right\}$ égal au nombre de partitions en k parties d'un ensemble à n éléments*[6]*.*

Définition 15 *Par convention, on pose* $\left\{ {0 \atop 0} \right\} = 1$

Théorème 16 (Relation de récurrence pour les nombres de Stirling)

$$\forall n > 0,\ \left\{ {n \atop k} \right\} = k\left\{ {n-1 \atop k} \right\} + \left\{ {n-1 \atop k-1} \right\}$$

Preuve. *Il suffit de considérer parmi les partitions de $\{1, ..., n\}$ en k parties celles qui contiennent le singleton $\{n\}$ (au nombre de $\left\{ {n-1 \atop k-1} \right\}$) et les autres. Ces dernières sont construites à partir d'une partition en k parties de $\{1, ..., n-1\}$ (au nombre de $\left\{ {n-1 \atop k} \right\}$) à laquelle on ajoute, à une des parties, l'élément n.* ■

Proposition 17 *On a*

$$\begin{aligned} \left\{ {n \atop 0} \right\} &= 0 \\ \left\{ {n \atop 1} \right\} &= \left\{ {n \atop n} \right\} = 1 \\ \left\{ {0 \atop k} \right\} &= 0 \ si\ k > 0 \end{aligned}$$

$$\begin{aligned} \left\{ {n \atop 2} \right\} &= 2^{n-1} - 1 \ si\ n > 0 \\ \left\{ {n \atop n-1} \right\} &= C_n^2 = \frac{n(n-1)}{2} \\ \left\{ {n \atop k} \right\} &= 0 \ si\ k > n \end{aligned}$$

La proposition suivante donne une forme close pour les nombres de Stirling. Il est évident que cette expression n'est pas utilisable pour leur calcul. On préférera utiliser un tableau semblable au triangle de Pascal, rempli grâce à la relation de récurrence sur les nombres de Stirling.

---

[5] James Stirling (1692 – 1770). Mathématicien écossais surtout connu pour le développement asymptotique à cinq termes de $\ln(n!)$ qu'il établit en 1730. L'équivalent classique $n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$ qui s'en déduit est donné la même année par De Moivre.

[6] La notation est analogue à celle utilisée pour le coefficient du binôme $\binom{n}{p} = C_n^p$.

Proposition 18 *On a*

$$\left\{ {n \atop k} \right\} = \frac{1}{k!} \sum_{i=0}^{k} (-1)^{k-i} \, i^n C_k^i$$

Le lecteur désireux d'en savoir plus pourra consulter le livre de Graham, Knuth et Patashnik [1] et l'article de Kaeser [2] cités dans la bibliographie.

Fonction génératrice des nombres de Stirling On introduit pour $k > 0$, les séries entières

$$S_k(z) = \sum_{n=0}^{+\infty} \left\{ {n \atop k} \right\} z^n$$

De la relation de récurrence sur les nombres de Stirling, on déduit

$$S_k(z) = zS_{k-1}(z) + kzS_k(z)$$

ce qui fournit

$$S_k(z) = \frac{z}{1-kz} S_{k-1}(z)$$

et ainsi

$$\forall k > 0, \; S_k(z) = \sum_{n=0}^{+\infty} \left\{ {n \atop k} \right\} z^n = \frac{z^k}{(1-z)(1-2z)(1-3z)\ldots.(1-kz)}$$

On peut maintenant établir le résultat attendu.

3.2.3 Calcul de $P(X_k = n)$

On revient maintenant au problème de dé. On a vu que

$$\begin{aligned}
f_k(z) &= \prod_{i=0}^{k-1} \frac{(m-i)z}{m-iz} \\
&= \prod_{i=0}^{k-1} (m-i) \times \prod_{i=0}^{k-1} \frac{z}{m-iz} \\
&= \frac{m!}{(m-k)!} \left(\frac{z}{m}\right)^k \prod_{i=0}^{k-1} \frac{1}{1-i\frac{z}{m}} \\
&= \frac{m!}{(m-k)!} \left(\frac{z}{m}\right) S_{k-1}\left(\frac{z}{m}\right) \\
&= \frac{m!}{(m-k)!} \left(\frac{z}{m}\right) \sum_{n=0}^{+\infty} \left\{ {n \atop k-1} \right\} \left(\frac{z}{m}\right)^n \\
&= \sum_{n=0}^{+\infty} \frac{m!}{m^{n+1}(m-k)!} \left\{ {n \atop k-1} \right\} z^{n+1}
\end{aligned}$$

Ce qui permet d'affirmer que

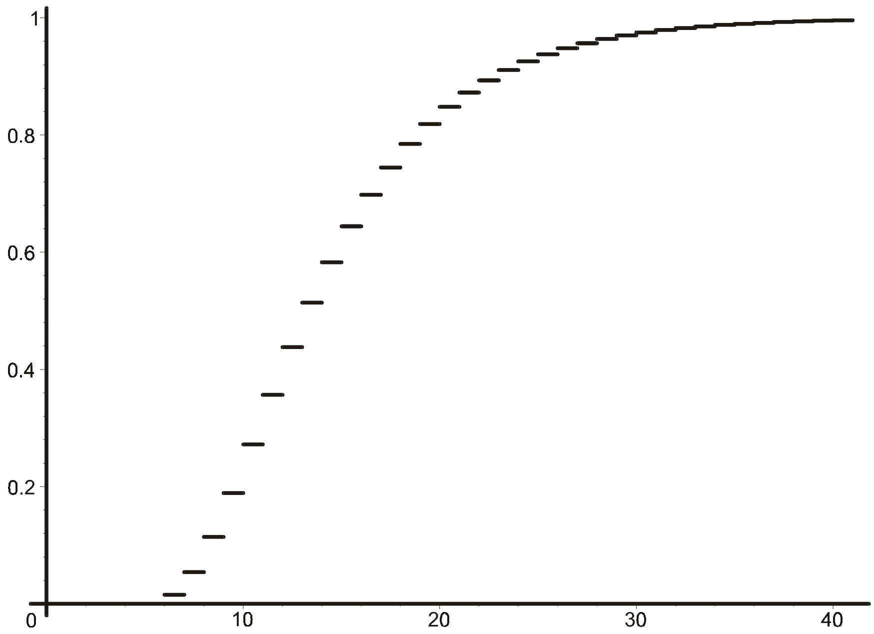
$$\boxed{\forall k > 0, \; \forall n > 0, \; P(X_k = n) = \frac{m!}{m^n (m-k)!} \left\{ {n-1 \atop k-1} \right\}} \qquad (2)$$

Exemple 19 *Dans le cas d'un dé à six faces, la probabilité de voir pour la première fois les six faces au énième lancer est*

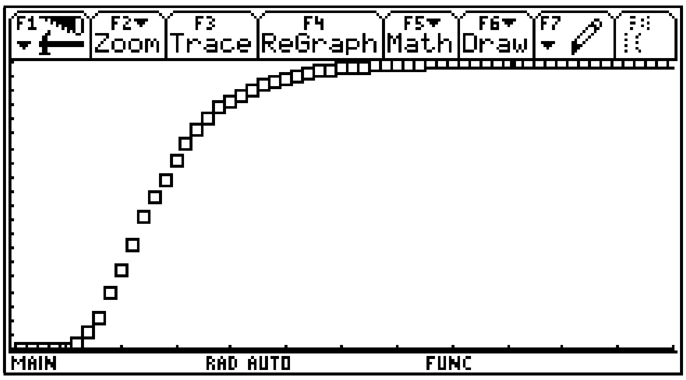
$$\textit{Pour } n \geq 6,\ P(X_6 = n) = \frac{720}{6^n}\left(\frac{1}{24} - \frac{2^n}{24} + \frac{3^n}{36} - \frac{4^n}{96} + \frac{5^n}{600}\right)$$

Remarque 20 *La formule* (2) *est bien cohérente.*
Si $k = 1$,*on doit avoir* $P(X_1 = 1) = 1$ *et* $P(X_1 = n) = 0$ *si* $n \geq 2$ *(au premier lancer, on voit une face et ceci pour la première fois). C'est bien le cas avec la convention* $\left\{ \begin{matrix} 0 \\ 0 \end{matrix} \right\} = 1.$

Remarque 21 *Voici la fonction génératrice de la probabilité de* X_6 *lorsque* $m = 6$ *(c'est l'exemple du dé à* 6 *faces)*



On pourra comparer avec la courbe suivante obtenue expérimentalement.



4 Conclusion

L'histoire du Calife est certes amusante[7] mais semble difficile à exposer à nos élèves sans susciter émois et agitations. On propose donc une autre formulation :

Une image "cadeau" vous est offerte dans chaque paquet de votre chocolat préféré. Il y a vingt images différentes. Vous avez décidé de collectionner toutes les images. Combien de paquets allez-vous acheter en moyenne[8] ?

Ce problème est a priori différent, mais, en première approche, si l'on considère que le nombre de paquets mis en circulation est grand, la modélisation proposée dans cet article rend assez bien compte de la réalité. Il reste ensuite à en faire une critique et à proposer d'autres simulations plus réalistes. Mais cela est une autre histoire.

[7] Et à notre avis, elle distrait davantage les hommes que les femmes qui y verront, avec raison, de nombreux points "*politiquement incorrects*"

[8] Et combien de crises de foies ?

Fiche élève

<u>Problème :</u> Votre marque préférée de chocolat vous propose de collectionner A images. Combien en moyenne devrez-vous acheter de plaques de chocolat pour obtenir la collection complète.

Partie A

1. Sur votre calculatrice, quelle(s) instruction(s) vous permettent de choisir au hasard un entier compris entre 1 et A.

2. On pose dans un premier temps $A = 2$

   (a) Simuler, sans programme, l'achat de plaques de chocolat jusqu'à obtention des deux images.

   (b) Comment peut-on s'y prendre pour ne pas avoir à compter, à la main le nombre de plaques achetées ?

   (c) Faire quelques essais et compléter le tableau suivant :

Essai	1	2	3	4	5	6	7	8	9
Nb tirages									

$\boxed{\text{Partie B}}$

On suppose désormais que $A = 6$. La méthode précédente n'étant guère pratique, on se propose d'écrire un programme simulant l'expérience.

1. Simulation d'une partie

   (a) Que doit-on stocker comme informations (on pourra se référer à la méthode utilisée à la question 2 partie A).

   (b) On décide de stocker les images obtenues dans une liste $L1$ de dimension 6. Par convention, on mettra 0 au ième rang de cette liste si l'image est obtenue et 1 dans le cas contraire. Comment peut-on traduire qu'on a obtenu la collection complète ?

2. Simulation de plusieurs parties

   (a) Que doit-on stocker comme informations ?

3. Compléter l'algorithme suivant.

```
Enter le nombre d'essais ....
Mettre 0 dans ....
Pour I allant de 1 à ... avec un pas de 1
Mettre 0 dans ......
Mettre { , , , , , } dans L1
Tant que somme des éléments de L1 est ....................
Tirer au sort un nombre A entier entre .....................
Mettre dans L1(........ ) : X ...................
Fin du tant que
C augmente de ............
Fin du Pour
Afficher ............
```

4. Compléter le programme correspondant à votre machine.

TI-89	TI-83	CASIO

`collect()`
`Prgm`
`Prompt`
`Delvar`
 	
`For i, 1, ..., 1`
`seq(i, i, 1, ..., 1)` $\rightarrow$ `l1`
 	
`While sum(l1).....`
 	
`EndWhile`
 	
`EndFor`
`Disp`
`EndPrgm`
	`Prompt ....`
	
	`For I, 1, ...., 1`
	`seq(I, I, 1, ..., 1)` $\rightarrow$ L_1
	
	`While sum(`L_1`)`
	
	`End`
	
	`End`
		`".... = "?` $\rightarrow$ `...`
		
		`For 1` $\rightarrow$ `I to ... Step1`
		`seq(I, I, 1, ...., 1)` $\rightarrow$ `List 1`
		
		`While sum (List 1)`
		
		`WhileEnd`
		
		`Next`

5. Compléter le tableau ci-dessous :

Essai	1	2	3	4	5	6	7	8	9
Nb tirages									

$\boxed{\text{Partie C}}$

Pour aller plus loin

1. Comment modifier le programme précédent pour travailler avec A images, A étant choisi par l'utilisateur.

2. Le nombre moyen de plaques à acheter est-il proportionnel au nombre d'images de la collection ?

3. A partir d'une série de 500 essais, on désirerait tracer à l'aide de la calculatrice l'ensemble des points M_k d'abscisse k et d'ordonnée fréquence de l'évènement "il a fallu acheter k plaques pour avoir la collection complète".
   (a) Quelle(s) information(s) supplémentaire(s) doit-on recueillir.

   (b) Au vu des résultats obtenus dans la partie B, dans quel intervalle peut-on considérer que k varie ?

   (c) Modifier votre programme pour tenir compte de ce changement.

   (d) Effectuer 500 essais, puis faire apparaître le diagramme demandé. quelles remarques celui-ci vous inspire-t-il ?

Annexe

Proposition 22 (Formule de Poincaré) *Soient* $(A_1, \ldots, A_n)$ *une famille de* n *événements alors*

$$\begin{aligned}
P\left(\bigcup_{i=1}^{n} A_i\right) &= P(A_1) + P(A_2) + \ldots + P(A_n) \\
&- \left(P(A_1 \cap A_2) + P(A_1 \cap A_3) + \ldots\right) \\
&\vdots \\
&+ (-1)^{k+1} \sum_{\substack{I \subset \{1,\ldots,n\} \\ \mathrm{card}(I)=k}} P\left(\bigcap_{i \in I} A_i\right) \\
&\vdots \\
&+ (-1)^{n+1} P(A_1 \cap A_2 \ldots \cap A_n)
\end{aligned}$$

Références

[1] R.L.Graham, D.E. Knuth, O. Patashnik, *Mathématiques Concrètes-Fondations pour l'informatique,* International Thomson Publishing France, Paris, 1998

[2] P. Kaeser, *Dénombrement des schémas de rimes*, in Quadrature n°40.

[3] C. Bouzitat, G. Pages, *Pour quelques images de plus,* in Quadrature n°25.