La loi des séries … pédantisme des médias ou réalité ?!

1. Objectifs :

Apporter une réponse aux élèves intrigués par l'expression « la loi des séries », entendue fréquemment dans les médias et qui est plus une incantation qu'une réalité mathématique.

De quoi s'agit-il ? 2 accidents aériens la même semaine et hop ! « la loi des séries » ; 3 séismes en 15 jours et hop ! …

Y aurait-il un fantasmatique lien de causalité ? L'indépendance, cette notion difficile de probabilité a bien du mal à être assimilée.

Y aurait-il illusion de résultats en séries ? Mais là pas besoin de loi, il suffit de calculer les probabilités ou en seconde de simuler.

2. Enoncé :

Simuler le lancer de 100 pièces de monnaie et compter les séries de 5 fois « Face ».
Utiliser les fonctions SI, TROUVE et ESTNUM.

3. De quoi s'agit-il ?:

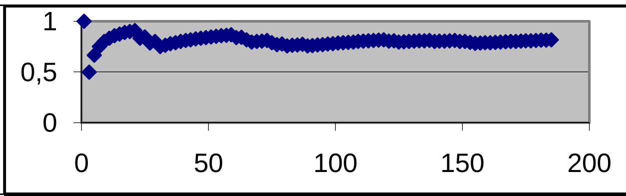
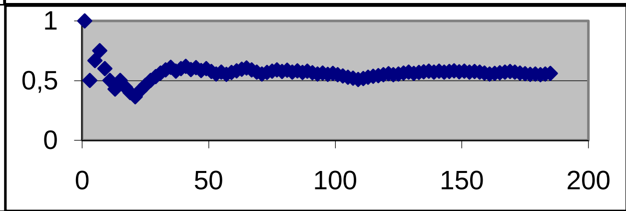
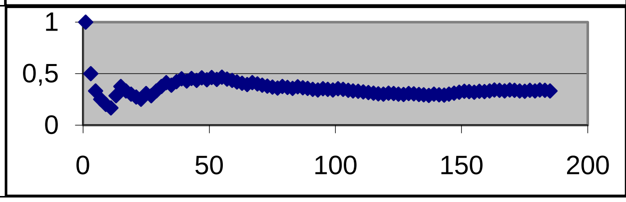
En général à la première séance de simulation en Seconde, je donne comme travail (!) à la maison de jeter 100 fois une pièce monnaie. A chaque fois il y a des petits malins qui, par flemme après quelques lancers, trouvent plus rapide (ou moins bruyant) de remplir à la main la série de PFFPF… .
Assez probable de les démasquer ; surtout quand n'apparaît aucune suite de FFFF ou plus (théoriquement 97,25% des tirages) ; et même quasiment 100% si on cherche FFFF ou PPPP.
Essayer de générer « humainement » du hasard est difficile, car on répugne à se lancer dans des grandes séries répétitives, et pourtant …

Finalement, l'accumulation de coïncidence est banale et ne relève d'aucune loi autre que celle du hasard. Sans doute des réminiscences de vieux dictons « un malheur n'arrive jamais seul », « jamais 2 sans 3 » (comme disait peut être Peugeot ?).

<u>4. Autre activité :</u> on demande à une classe de 23 élèves la probabilité pour que deux d'entre eux au moins aient la même date de naissance, les réponses tournent en général autour d'une chance sur cinq. Et pourtant, la bonne réponse est une chance sur deux.

<u>Eléments de réponse :</u>

FFFPPPPPPPFPPP**FFFF**FPPFPFPPPPPFFPFPPPFFPFFPFPFFPPPFPPPPFPFPPFFPFFFFPPFFPFFPPFFPPFFFFFFFFFFPFFPPFPPFPFPPP	
1	*la formule utilisée est* =SI(ESTNUM(TROUVE("FFFFF";*cellule*));1;0)

L'apparition de FFFFF dans une série de 100 lancers de pièces est plutôt banale !	
L'apparition de FFFFFF de pièces aussi.	
Et même FFFFFFFF arrive aussi de temps en temps.	

Conclusion : ce qui serait « rare » ce serait les séries extrêmement alternées du type PFPF.

<u>Compléments :</u>

- les calculs exacts de ces probabilités
- l'anniversaire
- la série noire d'août 2005 et la loi de Poisson (un document fort intéressant de l'Université de Rouen à consulter sur le web http://www.univ-rouen.fr/LMRS/Persopage/Janvresse/Publi/series_noires.pdf)

Détails 1:

Les calculs exacts de la probabilité d'obtenir une série de 4 « Face » pour la première fois au rang i, puis lors de 100 lancers, en faisant la somme, la probabilité de voir apparaître FFFF ou plus.

Pour i entre 1 et 3, $p(X=i)=0$; $p(X=4)=\frac{1}{2^4}$ le seul cas étant FFFF ;

$p(X=5)=\frac{1}{2}\frac{1}{2^4}=\frac{1}{2^5}$ le seul cas étant PFFFF ; $p(X=6)=\frac{1}{2^5}$ les seuls cas étant ?PFFFF ;

$p(X=7)=\frac{1}{2^5}$ les seuls cas étant ??PFFFF ; $p(X=8)=\frac{1}{2^5}$ les seuls cas étant ???PFFFF ;

Ensuite, il faut tenir compte que dans les premiers lancers, FFFF n'est pas apparu.

Pour i entre 6 et 100 , $p(X=i)=(1-p(X\leq i-4-1)).\frac{1}{2^5}$ or, $p(X\leq k)=p(X\leq k-1)+p(X=k)$

d'où une récurrence et une somme …

Un peu d'aide de Maple pour effectuer cette somme :

restart: n:=100: N:=4: for i from 1 to N-1 do p(i):=0 od:
> p(N):=1/2^N:
*> for i from N+1 to 2*N do p(i):= 1/2^(N+1) od:*

*> for k from 2*N+1 to n*
*> do p(k):= (1- sum(p(z), z=1..k-N-1))*1/2^(N+1);*
> od:
> sum(p(z),z=1..100); $\frac{1233062807221151952231797979607}{1267650600228229401496703205376}\simeq 0{,}9727150423$

De la même façon, pour N=5 on obtient environ 81%, pour N=6 environ 55% et même pour N=8 environ 17% ce qui n'est pas négligeable.

Détails 2:

Les calculs pour trouver les apparitions de FFFF ou PPPP, donc une série d'au moins 4 dans 100 lancers, peuvent être faits en comptant les i-uplets dans lesquels il n'y a pas ce type de série.

i° élément	suite sans FFFF ni PPPP ou plus	nombre de i-uplets	nombre de suites contenant PPPP ou FFFF	ces suites sont:	probabilité d'avoir une telle suite sans FFFF ni PPPP
1	2	2			0
2	4	4			0
3	8	8			0
4	14	16	2	PPPP, FFFF	0,125
5	26	32	6	PPPPP, PPPPF, FFFFP, FFFFF, PFFFF, FPPPP	0,1875
6	48	64	16		0,25
7	88	128	40		0,3125
8	162	256	94		0,3671875

Comment les compter ?

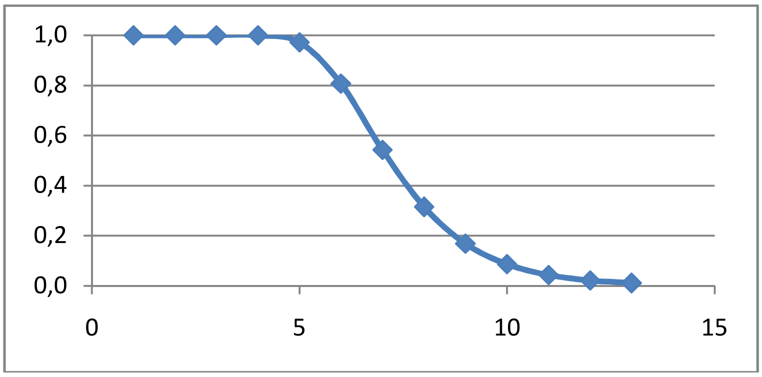
Appelons *pas-glop* ! un i-uplet n'ayant pas de répétitions FFFF ni PPPP

$n_1 = 2$: F et P sont *pas-glop* ! ; $n_2 = 4$; $n_3 = 8$

- pour le rang i≥4, les i-uplets *pas-glop* ! sont :
  ceux *pas-glop* !, de rang i-1 auquel ont adjoint un i° tirage différent du i-1° ; il y en a ainsi n_{i-1} *pas-glop* ! qui se terminent par PF ou FP
- ceux qui se terminent de la même manière par PFF ou FPP quand les i-2-uplets sont *pas-glop* ! ; il y en a ainsi n_{i-2} *pas-glop* !
- ceux enfin, qui se terminent de la même manière par PFFF ou FPPP quand les i-3-uplets sont *pas-glop* ! ; il y en a ainsi n_{i-3}.

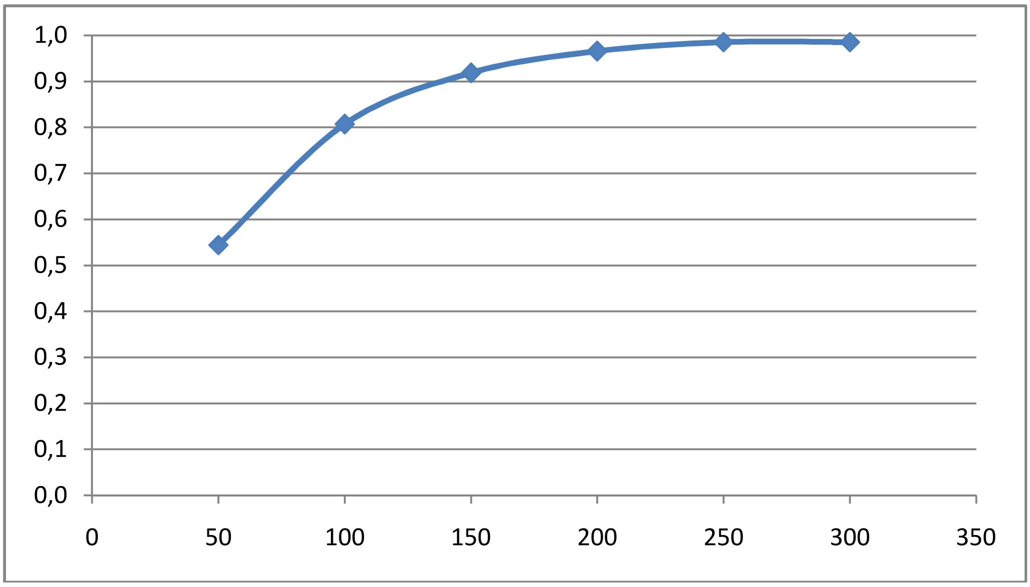
Au total, $n_i = n_{i-3} + n_{i-2} + n_{i-1}$, ce qui permet de calculer par récurrence n_{100}

La probabilité recherchée d'apparition de FFFF ou PPPP au moins lors de 100 lancers est $1 - \frac{n_{100}}{2^{100}} \simeq 0{,}9997153847$.



Ce sont les probabilités respectives d'apparitions de séries de 4,5,6 … lors de 100 lancers.

Il est intéressant aussi d'observer la convergence vers 1 de l'apparition de 6F ou 6P consécutifs lorsque le nombre de lancers augmente.



<u>Curiosités:</u>

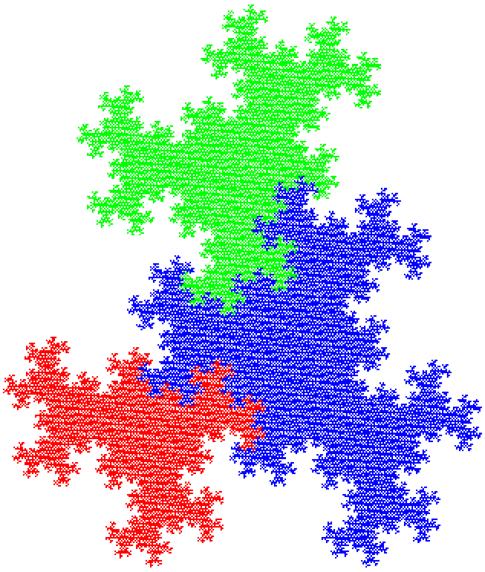
La récurrence précédente pour 4, $\begin{cases} n_1 = 2 \\ n_2 = 4 \\ n_3 = 8 \\ n_i = n_{i-1} + n_{i-2} + n_{i-3} \end{cases}$

est une extension de suite de Fibonacci.

On peut même nommer suite de « Tri-bonacci » la suite
$\begin{cases} n_1 = 0 \\ n_2 = 0 \\ n_3 = 1 \\ n_i = n_{i-1} + n_{i-2} + n_{i-3} \end{cases}$

qui amène au nombre d'argent
$\dfrac{\sqrt[3]{19+3\sqrt{33}}+\sqrt[3]{19-3\sqrt{33}}+1}{3}$

au problème du découpage d'un saucisson en tranches de 1,2 ou 3



et à la fractale de Rauzy.

De la même manière, les autres récurrences sont à relier aux suites de k-bonacci.

Détails 3:

On peut aborder aussi le problème de la manière suivante :
la probabilité qu'au rang i>4, on rencontre FFFF ou PPPP (ou plus) se calcule en tenant compte des cas où c'était déjà ainsi au rang i-1, des cas où on vient de sortir exactement le 4°F ou le 4°P.

La suite récurrente est alors : $\begin{cases} p_1 = 0 \\ p_2 = 0 \\ p_3 = 0 \\ p_4 = \dfrac{2}{2^4} \\ \\ p_i = p_{i-1} + \dfrac{1}{2^4}(1 - p_{i-4}),\ pour\ i > 4 \end{cases}$

Un petit calcul avec Maple :

```
> restart;
> for i from 1 to 3 do f(i):=0 od:   f(4):=2/2^4:
> for i from 5 to 100 do f(i):=f(i-1)+1/2^4*(1-f(i-4)) od:
> f(100);
```

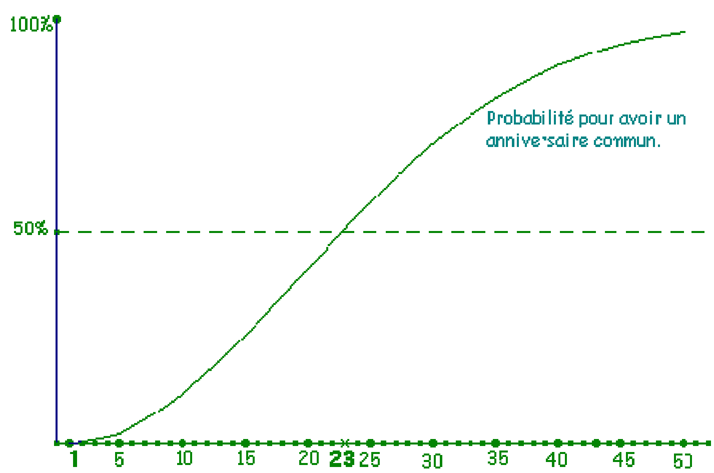
$$\frac{633644903733299599847137445049}{633825300114114700748351602688}$$

```
> evalf(");
```

.9997153847

Deux élèves ont le même anniversaire dans une classe de 23 :

$$1 - \frac{\underbrace{365 \times 364 \times ....}_{23}}{365^{23}} \simeq 50,72\%$$



La loi des séries noires

É. Janvresse et T. de la Rue
Laboratoire de mathématiques Raphaël Salem*
CNRS, université de Rouen

La succession d'accidents d'avions en août 2005 a donné lieu à de nombreux débats sur la sécurité aérienne. Une telle série noire était-elle improbable ? L'analyse statistique des accidents aériens permet d'affirmer le contraire.

Le 2 août 2005, un avion d'Air France sort de piste lors de son atterrissage à Toronto et s'enflamme. Le 6 août, un avion de Tuninter tombe en mer à proximité de Palerme. Le 14 août, un avion d'Hélios Airways percute une montagne près d'Athènes. Le 16 août, un avion de West-Carribean s'écrase au Vénézuéla. Et, le 23 août, un avion de Tans s'écrase en Amazonie.

L'été dernier, cette série noire a suscité l'inquiétude générale : était-elle liée à l'augmentation du trafic aérien ? Trahissait-elle une soudaine hausse du risque d'accidents s'expliquant par le relâchement des contrôles et de la maintenance ? Ou encore par le vieillissement de la flotte aérienne ?

On peut trouver quelques éléments de réponse dans l'analyse statistique des accidents aériens. Comment procède-t-on ? D'abord, on considère que les accidents surviennent indépendamment les uns des autres. Ensuite, on suppose que la fréquence à laquelle surviennent ces accidents est constante. La question que l'on se pose alors est : une série de 5 accidents sur une période de 22 jours peut-elle être le simple fruit du hasard, ou doit-elle être expliquée par d'autres raisons ?

Méthode de calcul

Entre 1995 et 2004, on a recensé en moyenne un accident tous les dix jours (367 en 10 ans)[1]. Nous n'avons retenu que les vols transportant au moins 30 passagers (les plus médiatisés), soit à peu près deux vols sur cinq. Ainsi, le nombre quotidien d'accidents impliquant ce type de vols vaut en moyenne $1/10 \times 2/5 = 0,04$. Avec ces données, comment calculer le nombre total d'accidents auquel il faut statistiquement s'attendre sur une période donnée ? Pour le savoir, on fait appel à ce que l'on nomme la *loi de Poisson*, dite encore *loi*

* http ://www.univ-rouen.fr/LMRS/
[1] www.airdisaster.com

1

des événements rares. Celle-ci, introduite par Siméon Denis Poisson en 1837, permet de décrire la somme de multiples phénomènes de faible probabilité et indépendants les uns des autres.

La loi de Poisson est appropriée pour étudier les accidents aériens car le nombre d'avions qui décollent chaque jour est très élevé (20 000), la probabilité d'avoir un accident est infime pour chaque vol ($0,04/20\,000 = 1/500\,000$), et les accidents surviennent indépendamment les uns des autres. La loi de Poisson décrivant le nombre d'accidents sur une période donnée est caractérisée par un paramètre, appelé *intensité*, correspondant au nombre *moyen* d'accidents sur cette période.

Lorsqu'une variable aléatoire N suit une loi de Poisson d'intensité λ, la probabilité que N prenne une certaine valeur entière $k \geq 0$ est donnée par

$$\text{Proba}(N = k) \ = \ \exp(-\lambda)\frac{\lambda^k}{k!},$$

où $k!$ (« factorielle k ») désigne le produit de tous les entiers entre 1 et k si $k \geq 1$, et $0! = 1$.

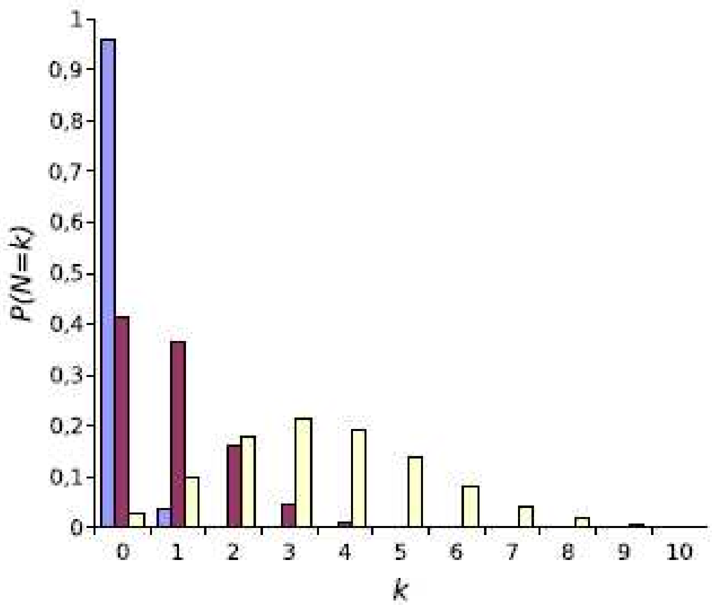


FIG. 1 – Distribution de la loi de Poisson pour les intensités $\lambda = 0,04$ (bleu), $\lambda = 0,88$ (rouge) et $\lambda = 3,6$ (jaune).

Si on s'intéresse au nombre d'accidents sur une période de 22 jours, l'intensité, qui est le nombre moyen d'accidents sur une telle durée, vaut $\lambda = 0,04 \times 22 = 0,88$. La probabilité qu'il soit inférieur ou égal à 4 vaut

2

$$
\begin{aligned}
&\mathrm{Proba}(N \leq 4)\\
&= \mathrm{Proba}(N=0) + \mathrm{Proba}(N=1) + \cdots + \mathrm{Proba}(N=4)\\
&= \exp(-0,88)\left(1 + 0,88^1 + \frac{0,88^2}{2\times 1} + \frac{0,88^3}{3\times 2\times 1} + \frac{0,88^4}{4\times 3\times 2\times 1}\right)\\
&\simeq 0,9978,
\end{aligned}
$$

soit environ $1 - 2/1000$. Ainsi, la probabilité d'une série d'au moins 5 accidents sur cette période est d'environ deux pour mille, ce qui paraît peu probable.

Pour savoir s'il faut remettre en cause les hypothèses de notre modèle, nous devons calculer la probabilité qu'une série de 5 accidents survienne sur une fenêtre quelconque de 22 jours au cours d'une année entière. Ce calcul est difficile car les fenêtres à considérer se chevauchent : le nombre d'accidents survenant entre le 1er et le 22 janvier est bien sûr corrélé avec le nombre d'accidents survenant entre le 2 et le 23 janvier !

Dans un premier temps, nous pouvons simplifier le problème en découpant l'année en 16 périodes disjointes de 22 jours (on néglige les 13 jours restant). L'absence de série noire sur une année complète impliquerait qu'il y ait, au plus, 4 accidents par période. Sur chacune d'elles, ceci arrive avec une probabilité d'environ $1 - 2/1\,000$. Comme elles sont disjointes, les nombres d'accidents par période sont indépendants et donc la probabilité qu'aucune ne comporte de série noire est de l'ordre de $(998/1\,000)^{16} \simeq 0,97$. Ainsi, la probabilité que l'une des 16 périodes comporte une série noire vaut environ $1 - 0,97 = 3\%$. C'est assez faible, mais ce calcul est très approximatif puisque nous ne tenons pas compte de toutes les périodes possibles de 22 jours consécutifs. On peut s'attendre à ce que la probabilité qui nous intéresse soit sensiblement plus élevée.

Une chance sur dix

Comment mener le calcul lorsque l'on considère des périodes qui ne sont pas indépendantes ? Ce type de problème est particulièrement difficile et a donné lieu à de nombreux travaux depuis les années 1960. Les probabilistes ont nommé « statistiques de balayage » le nombre maximum d'événements survenant dans une fenêtre de taille fixée prenant toutes les positions possibles dans une région donnée. Ces statistiques permettent de détecter et d'analyser une succession d'événements rapprochés, afin de déterminer si elle est le fruit du hasard ou si elle est, au contraire, atypique. Par exemple, l'observation localisée d'un nombre élevé de cas de cancers conduit à s'interroger sur l'existence d'une cause commune (habitudes alimentaires, pollution industrielle...). Elles servent aussi en génétique pour le repérage de motifs exceptionnels dans l'ADN.

Des formules exactes ont été obtenues dans les années 1970 pour exprimer ces statistiques, mais elles sont complexes à mettre en œuvre. Aussi, on fait appel à des approximations plus faciles à manipuler et fournissant des résultats

3

convenables[2]. Ainsi la probabilité pour que se produise une succession de 5 accidents en 22 jours sur une année est d'environ 0, 11. En d'autres termes, il y a chaque année plus d'une chance sur dix d'en observer une. Cette probabilité est bien plus importante que les 3% obtenus en ne prenant en compte que des périodes disjointes.

Si nous avons l'impression qu'une série rapprochée d'accidents témoigne d'un changement dans les conditions de sécurité, c'est que nous sommes habitués à raisonner « en moyenne », alors que le hasard, lui, ne répartit pas les événements de manière uniforme. Et même avec un faible niveau de risque, une succession aussi rapprochée d'accidents arrive avec une probabilité non négligeable. Notons que notre modèle ne tient pas compte des variations saisonnières du trafic. Or, si on les prenait en compte, on verrait augmenter la probabilité de voir des accidents regroupés, et ce, même si le risque individuel d'accidents demeurait inchangé. En effet, les accidents se concentreraient davantage sur les périodes où le trafic est intense.

[2]Scan statistics and applications. Édité par J. Glaz et N. Balakrishnan. Birkhäuser Boston, 1999.

4