

Différencier une suite aléatoire d'une autre qui ne l'est pas

Jean François Kentzel(*)

Première partie : comptage des inversions

(Récit détaillé de deux séances en classe de première)

Description des activités

Première séance

Je raconte à mes élèves qu'un d'entre eux à qui on a demandé de lancer 50 fois une pièce en notant la suite des résultats successifs PFPF... décide pour gagner du temps d'inventer une liste de P et de F. Son subterfuge risque-t-il d'être découvert ? Comment doit-il s'y prendre pour ne pas être pris ?

Cette question intervenant au cours du temps consacré au calcul des probabilités, les élèves (de première STG) conviennent rapidement qu'il faut obtenir des effectifs de pile et face raisonnables, c'est à dire proches de 25, ce que le tableur confirme.

Je leur annonce alors que je connais un moyen de déceler « assez souvent » une telle escroquerie.

Je leur demande, sur une demi-page où ils écrivent leur nom, de donner deux listes de longueur 50 PFPF..., l'une étant obtenue avec la touche simulation de leur calculatrice, l'autre étant inventée. Ils écrivent ces deux listes dans l'ordre qu'ils veulent et ils notent soigneuse-

Commentaires

Comme beaucoup d'enseignants je demande chaque année à mes élèves de seconde ou de première de lancer 50 ou 100 fois un ou deux dés ou une pièce afin de les sensibiliser au phénomène de la fluctuation d'échantillonnage. La question posée ici leur semble naturelle et intéressante.

La loi binomiale régissant le nombre de pile, ici $B(50, 1/2)$, n'est pas au programme de leur classe.

Je ne sais pas trop comment l'affaire va se passer et surtout je tiens à ce qu'on soit d'accord qu'il ne peut exister de méthode infaillible, ne serait-ce que parce que chaque suite a exactement la même probabilité d'apparition.

L'objectif pour chaque élève est alors d'essayer de « cacher » la fausse liste, étant entendu que les feuilles vont être ramassées puis redistribuées (dans le désordre) et qu'il faudra alors, à l'aide du moyen que je vais donner, essayer de retrouver laquelle des deux listes est inventée.

(*) Jkentzel@ac-toulouse.fr. Une version plus détaillée de cet article est sur le site du lycée Pardailhan à Auch (32) où j'enseigne, <http://pardailhan.ent.midipyrenees.fr/>, site désigné par PARD dans ce qui suit. (*)

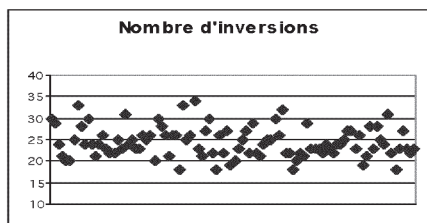
ment (en dehors de la feuille) laquelle des deux (liste 1 ou liste 2) est simulée et laquelle est inventée.

Puisque de toutes façons ils comptent les pile et face je leur demande d'inscrire ces nombres sur leur feuille.

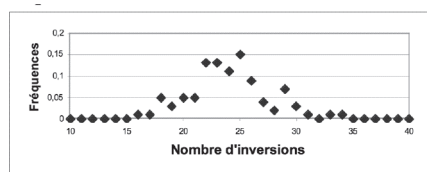
Deuxième séance

Avant de redistribuer les feuilles, je leur annonce d'entrée de jeu qu'on va compter les « inversions », c'est à dire les sous suites du type PF ou FP.

On simule avec un tableur 100 séries de 50 lancers et on regarde ce qu'est « en moyenne » le nombre d'inversions.



Pour confirmer l'impression donnée par la représentation ci-dessus, on calcule les fréquences :



Je redistribue les feuilles dans le désordre et chaque élève compte alors les inversions de chacune des deux séries.

Chaque élève écrit sur la feuille qu'il a : « la série inventée est la liste (1 ou 2) car... » et il signe.

Je ramasse et redistribue chaque copie à son auteur.

La série inventée a-t-elle été repérée ?

On obtient alors des feuilles où n'apparaissent que des suites dont les effectifs de pile sont assez proches de 25 (et parfois égaux).

J'ai lu [1] que beaucoup de gens créent un nombre excessif d'inversions pour « faire plus aléatoire » ; je veux savoir si c'est vrai.

Bien sûr, je ne dis pas aux élèves que le nombre d'inversions suit la loi $B(49;1/2)$!

En simulant le lancer dans la colonne A (formule : $=ENT(ALEA()*2)$ dans la cellule A1), on compte les inversions dans la colonne B.

La formule : $=A1+A2$ dans la cellule B2, proposée par un élève, est moins naturelle que $=A2-A1$, mais plus pratique car il suffit d'écrire : $=NB.SI(B2 :B50 ; « =1 »$ en B51 pour compter.

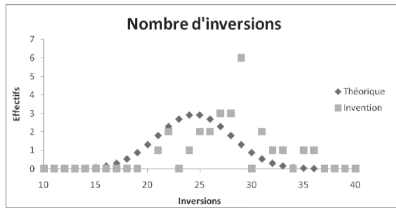
On écrit 10, 11...40 en B52 :AF52 et la formule $=NB.SI(\$B51:\$GR51;B52)/100$ en B53 (recopiée jusqu'à AF53).

Je deviens nerveux car on a investi du temps dans cette activité alors qu'en regardant quelques copies, il me semble probable que je vais à la catastrophe (un score proche de 50 % de réussite, voire inférieur à 50 % !).

Certains élèves sont contraints d'écrire qu'ils répondent au hasard car ils ont deux nombres d'inversions très proches.

Oui pour 18 des 26 élèves ayant participé à l'expérience (69 % de réussite).

Pour qu'on comprenne mieux ce qui s'est passé, on fait un tableau du nombre d'inversions des séries inventées.



En remplissant le début de ce tableau, je crois que les élèves ne suivent pas car très peu lèvent la main pour annoncer le nombre d'inversions de la série qu'ils ont inventée...

Dans le graphique ci-contre les nombres d'inversions pour la série intitulée « théorique » sont en fait calculés à partir des probabilités et on obtient évidemment une plus belle courbe de Gauss que ci-dessus.

Deuxième partie : autres comptages

Ce qui suit a été fait avec des élèves de terminale S mais peut être fait en classe de première ou dans une autre classe de terminale.

Définition : soit k un entier compris entre 1 et n . On appelle **k -suite** toute suite de k résultats consécutifs identiques.

Par exemple, la suite 100101111 contient neuf 1-suites, quatre 2-suites, deux 3-suites et une 4-suite.

On va s'intéresser aux k -suites pour obtenir d'autres façons de différencier les deux suites.

C'est un prolongement naturel de ce qui précède : compter les piles et les faces, c'est s'intéresser aux 1-suites ; compter les inversions c'est, en creux, compter les 2-suites.

1. Longueur de la plus longue k -suite

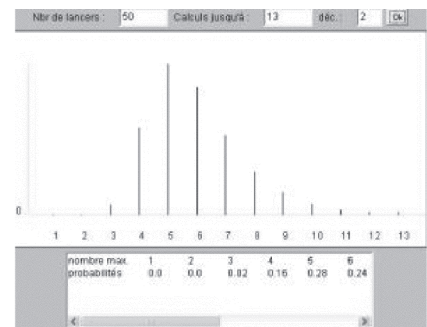
On note L la variable aléatoire « Longueur de la plus longue k -suite ».

Dans l'exemple de la définition (avec $n = 9$), $L = 4$.

On trouve la loi de cette variable aléatoire, pour n'importe quelle valeur de n , sur l'excellent site de Jean-Paul Quelen [3].

Son diagramme est donné ci-contre pour $n = 50$.

Les calculs donnant les valeurs exactes intervenant dans la loi, sont expliqués sur une page de ce site liée à celle du diagramme.



2. Nombre de k -suites

k étant un naturel non nul, on note Y_k la variable aléatoire « Nombre de k -suites ».

Dans l'exemple de la définition (avec $n = 9$), $Y_1 = 9$, $Y_2 = 4$, $Y_3 = 2$, $Y_4 = 1$.

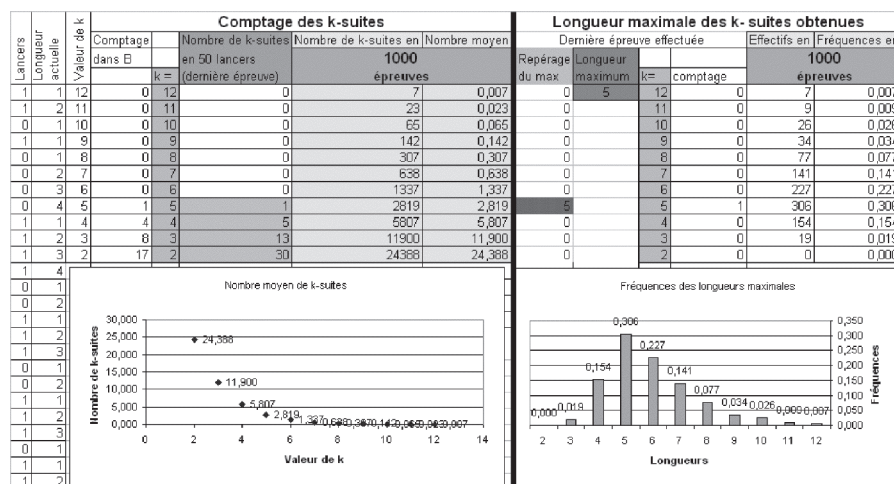
La loi de Y_k est un peu compliquée mais on obtient facilement son espérance (pour $n = 50$) :

Valeur de k	1	2	3	4	5	6	7	8	9	10	11	12	13
Nombre moyen de k -suites	50	24,5	12	5,9	2,9	1,4	0,7	0,3	0,2	0,1	0,04	0,02	0,01

3. Utilisation de ces nouveaux comptages

Ces comptages sont lisibles sur la page de simulation qui suit, disponible sur le site PARD.

On obtient rapidement des valeurs proches de celles des deux tableaux précédents si le tableau utilisé réagit bien quand on appuie en continu sur la touche F9 (recalcul).



Après avoir présenté ces nouveaux comptages dont chacun donne un critère pour différencier les deux suites, on redistribue (presque aléatoirement) aux élèves les feuilles contenant les listes et on termine en construisant le tableau suivant dont ils doivent remplir les trois premières lignes, chaque case contenant un des signes : 1, 2, ?.

La ligne « Moyenne » permet parfois (voir 4° ci-après) de repérer immédiatement, pour chacun des critères, la liste la plus « suspecte ».

	Nombre de 2-suites	Nombre de 3-suites	Nombre de 4-suites	Nombre de 5-suites	Nombre de 6-suites	Nombre de 7-suites	Longueur maximum	Conclusion finale
Moyenne théorique	24,5	12	6	3	1,4	0,7	5,8	
Liste 1								xxxxxx
Liste 2								xxxxxx
Conclusion : la liste inventée est								
Conclusion : bonne ou mauvaise ?								

Chaque élève récepteur de listes va ensuite demander (si ce n'est pas déjà fait) à l'élève émetteur de ses listes quelle était la liste fautive : il l'écrit sur sa feuille et peut ensuite remplir la dernière ligne de son tableau.

Remarque : déclarer qu'une série est plus vraisemblable qu'une autre en utilisant seulement la proximité des deux valeurs observées avec la moyenne d'une variable aléatoire X alors qu'on ne sait rien de la loi de X n'est a priori pas sérieux. Mais les courbes ci-après montrent que le procédé utilisé fonctionne en pratique.

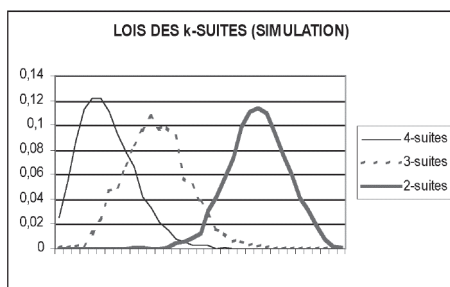
4 Cas litigieux

	Nombre de 2-suites	Nombre de 3-suites
Moyenne	24,5	12
Liste 1	22	<u>7</u>
Liste 2	<u>17</u>	11
Conclusion : la liste inventée est	la 2	la 1

Les élèves n'ont que rarement eu à gérer des conflits du type ci-contre (où les valeurs suspectes sont en gras), car les sept variables étudiées sont largement dépendantes (dans la série inventée, il y a souvent plus d'inversions et, du coup, moins de k -suites pour la plupart des k valant au moins 2).

Les valeurs « suspectes », sont 7 pour les 3-suites (liste 1) et 17 pour les 2-suites (liste 2).

Or un coup d'œil sur les courbes ci-contre montre que la valeur 7 pour un nombre de 3-suites est moins marginale que la valeur 17 pour une 2-suite⁽¹⁾.



On déclare donc finalement que la liste fautive (inventée) est la 2 car elle est moins vraisemblable.

(1) Cette constatation visuelle (évidente avec cet exemple qui a été choisi pour ça) peut être précisée par les valeurs numériques figurant sur la page de tableau.

Troisième partie : commentaires

1) En remplaçant 50 lancers par 100 lancers pour cette activité, on obtient évidemment pour la suite aléatoire des observations plus proches des probabilités attendues.

2) Il faut demander aux élèves d'inventer une liste avant d'en copier une sur la machine afin d'éviter les recopiations intempestifs !

3) Il faut se fixer un vocabulaire dont on ne sort plus pour désigner les listes : la vraie (vraiment aléatoire) et la fausse (inventée par le faussaire). Éviter de parler de la « bonne liste » !

4) Je me méfie un peu des générateurs de suites pseudo-aléatoires des calculatrices⁽²⁾, en tout cas pour un n petit comme 50, car j'ai lu [4] que ces générateurs ne sont testés que sur de longues plages. Si vous êtes très méfiant, vous pouvez remplacer la suite pseudo-aléatoire par le résultat de lancers effectifs ; je n'ai pas fait ce choix afin de donner moins de travail aux élèves.

5) **Les élèves ont ils fait des mathématiques ?** Question essentielle ! Est ce qu'on n'inflige pas aux élèves un exposé de calcul des probabilités un peu trop difficile pour eux (critique inverse si on n'utilise que des simulations) avant de leur demander d'effectuer des comptages qui seraient faisables en classe de sixième ? Je pense que non car avec cette activité on atteint deux objectifs : montrer la puissance du calcul des probabilités et présenter un exemple de « démarche statistique », en opposition à la démarche probabiliste (théorique). En classe de terminale cela peut être intéressant quelque temps avant de traiter l'adéquation à une loi équirépartie.

6) La bonne surprise que j'ai eue en regardant les copies de mes élèves est que tous ont « l'esprit logique » ; ils ne m'en donnent pas la preuve tous les jours ! Je n'ai trouvé que peu de fautes de raisonnement. **Pour être sûr que tout a été compris, il faut demander aux élèves un texte justifiant leur conclusion finale.**

7) Il me semble inutile de dire aux élèves qu'ils se livrent à des test statistiques, même si l'expression « test » apparaît naturellement. Il s'agit seulement de donner une vision de ce qu'est une **démarche statistique** en rendant naturelle cette notion de test.

8) Il demeure que l'exercice proposé est critiquable car on ne peut pas toujours conclure de façon satisfaisante et il peut en ressortir une grande perplexité... Conseil pratique : ne pas faire cette activité une semaine avant le bac ! Elle est, dans certains cas, un peu déstabilisante, soit parce qu'on ne sait pas quoi conclure, soit parce qu'un raisonnement correct aboutit à une conclusion fausse. Une réponse est qu'on peut prudemment (mais à mon avis c'est dommage) se limiter à deux critères.

9) Si on a l'occasion d'emmener ses élèves devant des ordinateurs, on peut informatiser complètement les activités. On peut alors pour le même prix prendre n valant 100 ou 200 au lieu de 50. « Même prix » seulement pour les élèves, car pour l'enseignant il faut modifier les pages de tableur concernant les nombres de k -suites : elles sont faites pour $n = 50$. En revanche la simulation de la loi de la variable aléatoire L figure sur le site de Jean Paul Quelen pour toute valeur de n .

(2) En revanche ceux des tableurs semblent corrects comme on peut le vérifier avec la simulation présentée ci-dessus.

Bibliographie succincte :

[1] DELAHAYE Jean-Paul, *Complexités* (Belin, octobre 2006). Voir le chapitre 5 : Les dés pipés du cerveau.

[2] *Document d'accompagnement des programmes de mathématiques de Seconde* (octobre 2000), fiche « Faites vos jeux » page 37. Disponible sur www.cndp.fr.

[3] QUELEN Jean-Paul <http://perso.orange.fr/jpq/index.htm>. Splendide. À découvrir d'urgence si vous ne le connaissez pas !

[4] Commission inter-IREM Statistiques et probabilités, *Statistique au lycée* (brochures APMEP 156 et 167). Voir notamment dans le volume 1 : « Quelques questions à propos des tables et des générateurs aléatoires » de Bernard PARZYSZ.

[5] VOGEL Nicole, *Peut-on imiter le hasard ?* in Bulletin APMEP n° 451 (mars-avril 2004). Une étude des k -suites à l'aide des graphes probabilistes. Disponible sur <http://www.apmep.asso.fr/spip.php?article2655>

Une bibliographie plus complète, des photocopiés (modifiables) destinés aux élèves et divers compléments figurent sur le site PARD (taper : Rubriques des disciplines/Mathématiques/Documents destinés aux enseignants).
