Probabilités du programme de 3^{e} Documents d'accompagnement des programmes de collège

André PRESSIAT

Maître de Conférences

de mathématiques à l'IUFM d'Orléans

DIDIREM - Université Paris 7

1. La nouveauté dans les programmes de 3e : les probabilités

Le programme de la classe de troisième a pour objectif de permettre :

• dans la partie « organisation et gestion de données, fonctions » :

- d'approcher la notion de fonction ;
- d'acquérir une première connaissance des fonctions linéaires et affines et de synthétiser le travail conduit sur la proportionnalité dans les classes antérieures ;
- de poursuivre la mise en place de paramètres (de position et de dispersion) d'une série statistique et d'envisager ainsi la notion de résumé statistique ;
- ***de mettre en pratique sur des exemples simples la notion de probabilité*** **;**

Extrait du bandeau relatif au titre

1. Organisation et gestion de données, fonctions.

Pour les séries statistiques, l'étude des paramètres de position est poursuivie : médiane et quartiles. Une première approche de la dispersion est envisagée. L'éducation mathématique rejoint ici l'éducation du citoyen : prendre l'habitude de s'interroger sur la signification des nombres utilisés, sur l'information apportée par un résumé statistique.

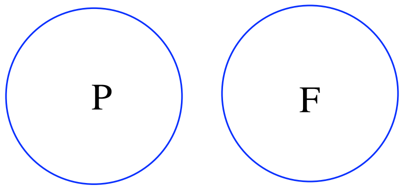
De même, c’est pour permettre au citoyen d’aborder l’incertitude et le hasard dans une perspective rationnelle que sont introduits les premiers éléments relatifs à la notion de probabilité.

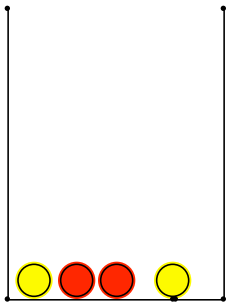
Contenus	Compétences	Exemples d'activité, commentaires
1.4 Notion de probabilité [Thèmes de convergence]	Comprendre et utiliser des notions élémentaires à propos des probabilités dans des contextes familiers d'expérimentation.	La notion de probabilité est abordée à partir de situations familières concernant des instruments produisant du hasard (pièces de monnaie, dés, roues de loteries, urnes).
		Certaines de ces situations (lancer d'un objet comme une punaise par exemple) permettent de rencontrer des cas pour lesquelles les probabilités ne sont pas définies à partir de considérations intuitives de symétrie ou de comparaison mais sont approximativement évaluées par les fréquences observées expérimentalement (approche fréquentiste des probabilités).

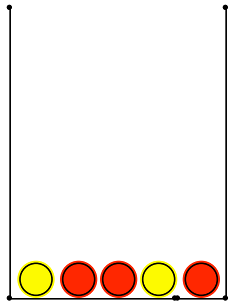
Exemples d'activité, commentaires
La notion de probabilité est utilisée pour traiter des situations de la vie courante pouvant être modélisées simplement à l'aide de ces instruments. Les situations étudiées concernent les expériences aléatoires *à une ou à deux épreuves*.

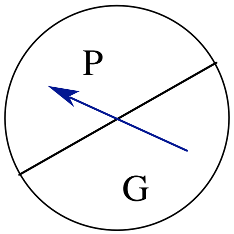
Il y a aura un document d'accompagnement portant spécialement sur les probabilités et leur enseignement en 3e …

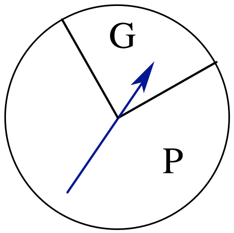
Les situations familières concernant les instruments produisant du hasard

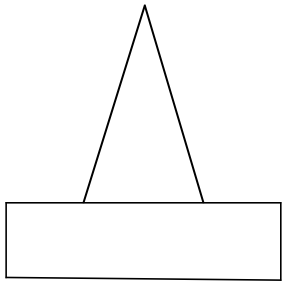




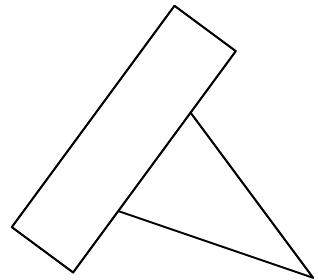








G



P

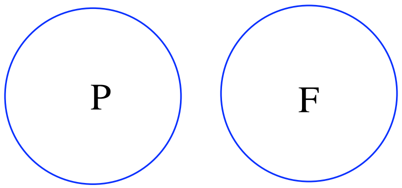
En lançant un grand nombre de fois la punaise, on obtient une suite de G et de P. Pour un petit nombre d'expériences, cette suite ne semble suivre aucune loi ; mais le résultat global laisse apparaître une régularité dans la fréquence de sortie de P et de G.

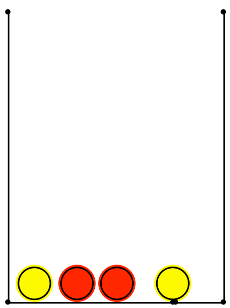
Au début, la fréquence (relative) du nombre de G varie très fortement. Mais à la longue, elle tend à se stabiliser autour d'une valeur p [qui vaut à peu près 5/6].

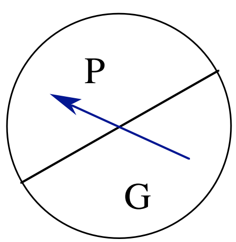
C'est pour traduire ce fait empirique que l'on dit que la probabilité d'obtenir G est p.

Pour le lancer de la punaise, on ne peut approcher cette probabilité que par l'expérimentation.

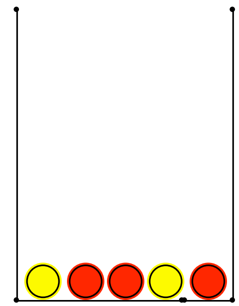
En revanche, pour les jeux évoqués précédemment, on peut obtenir la probabilité d'un résultat (d'une issue) par des considérations de symétrie ou de comparaison.



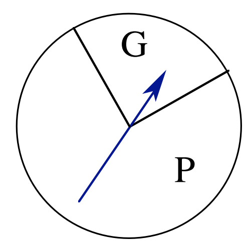




Pour chacun des jeux, chacun des deux résultats possibles ont une chance sur 2 de se produire ; ils ont la même probabilité : 1/2

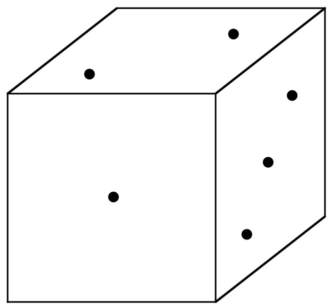


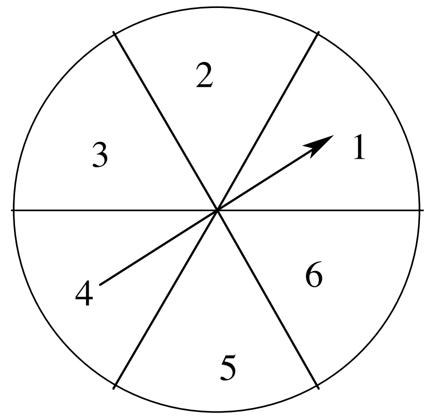
La probabilité d'obtenir une boule jaune est 2/5.
On a 3 chances sur 5 d'obtenir une boule rouge.



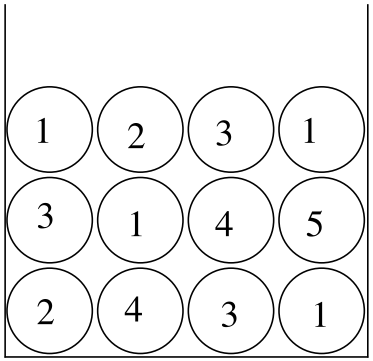
La probabilité de gagner est 1/4, …

Autres exemples :





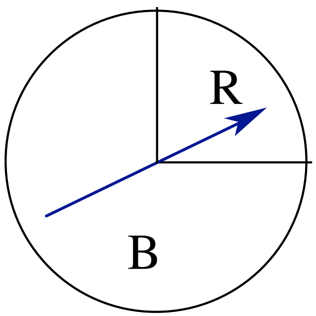
Chaque résultat a la même probabilité : 1/6.

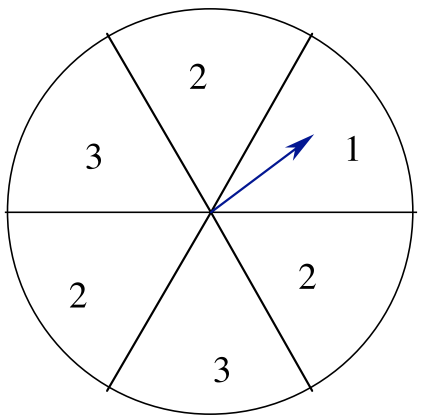


Les résultats 1, 2, 3, 4 et 5 ont respectivement comme probabilités :
1/3, 1/6, 1/4, 1/6 et 1/12.

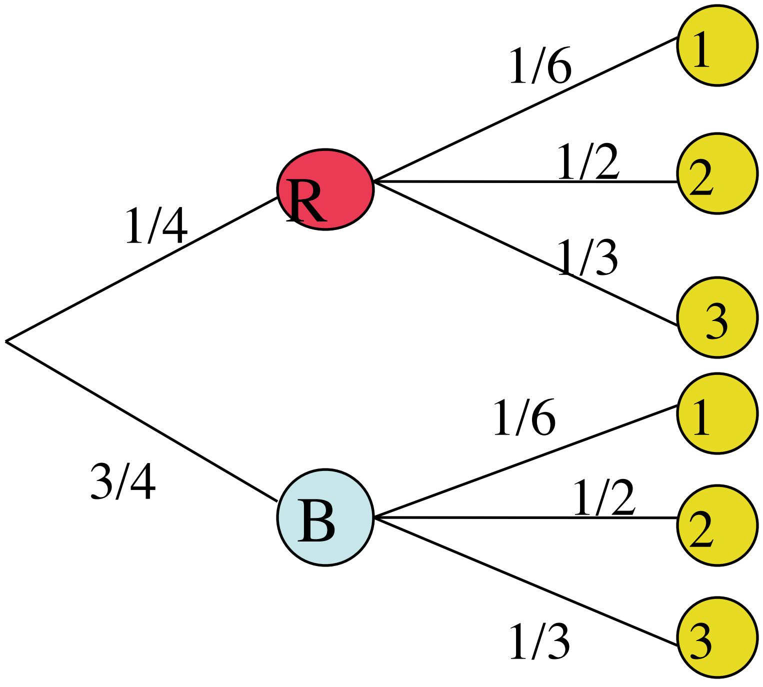
La probabilité d'obtenir un résultat pair est 1/6 + 1/6, c'est-à-dire 1/3.

Expériences à deux épreuves





Les résultats possibles sont (R, 1), (R, 2), (R, 3), (B, 1), (B, 2), (B, 3).



Chacun de ces résultats est représenté dans l'arbre ci-contre par une branche (ou chemin).

Comment évaluer la probabilité de chacun d'eux ?

Imaginons que l'on reproduise 120 (ou *N*) fois l'expérience.
1/4 de ces expériences suivront la branche vers R, et parmi celles-ci 1/6 iront vers 1. Donc il y en aura :

$$\frac{1}{6}\times\frac{1}{4}\times 120\left(\text{ou }\frac{1}{6}\times\frac{1}{4}\times N\right)$$

$$\text{soit } 5\left(\text{ou }\frac{1}{24}\times N\right)$$

La fréquence (relative) du résultat (R, 1) est donc 5/120 (ou 1/24).

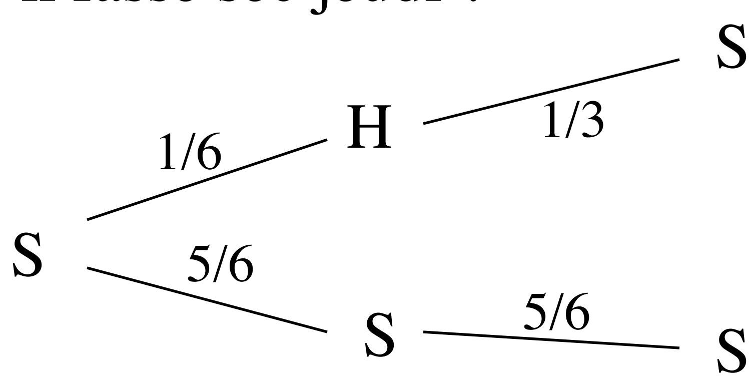
Ceci conduit à admettre que, de manière générale, la probabilité "d'un chemin" est égale au produit des probabilités "rencontrées le long de ce chemin".

On peut traiter avec ces représentations en arbres les questions relatives à deux tirages successifs dans une urne, avec remise ou sans remise.

C'est ce modèle de traitement qu'il conviendra de savoir réinvestir dans une situation simple de la vie courante.

E x e m p l e :
S'il fait sec un jour (S), alors il fera encore sec le lendemain avec la probabilité 5/6. S'il fait humide (H), alors il fera humide le lendemain avec la probabilité 2/3.

Aujourd'hui, nous sommes mardi et il fait sec. Quelle est la probabilité qu'il fasse sec jeudi ?



2. Documents d'accompagnement

Déjà parus :

Articulation école - collège

Ainsi que les ***projets*** de document d'accompagnement suivants

La proportionnalité (Juillet 2005)

Organisation et gestion de données (août 2005)

Du numérique au littéral (5 avril 2006)

Les nombres au collège (été 2006)

Vont bientôt paraître, les projets de documents suivants :

Le calcul numérique au collège

Grandeurs et mesures

Le document “Organisation et gestion de données” a été amendé.

Afin de mieux distinguer ce qui relève de la gestion de données d’une part, de la démarche statistique d’autre part…

… cette différence étant souvent occultée par le fait que, dans les deux cas, on utilise les mêmes modes de représentation, et en particulier les mêmes types de graphiques.

Les mêmes exemples (relatifs à la pêche dans les Côtes d’Armor) ont été conservés pour ce qui concerne la gestion de données.

En revanche, le passage à la statistique est maintenant illustré à partir d’un autre exemple, celui de la direction et de la vitesse des vents sur un plan d’eau accueillant une école de voile.

On part de questions :

Quels sont les vents dominants sur le plan d’eau en juillet - aôut ?

Quel est le nombre de jours où il est impossible de faire sortir les stagiaires pour des raisons de sécurité (où la vitesse du vent dépasse un certain seuil) ?

Un vent de vitesse supérieure à 12 nœuds est-il un vent “fort” sur ce plan d’eau ?

Ou

Combien de chances a-t-on d’avoir un vent supérieur à 12 nœuds ?

A partir d’une telle situation, on illustre les notions de population, de caractère qualitatif ou quantitatif.

date	direction observée	vitesse mesurée en nœud	Beaufort
09/07/2005	W	4,3	2
10/07/2005	NW	5,9	2
11/07/2005	NW	7	3
12/07/2005	NE	9,8	3
13/07/2005	N	2,8	1
14/07/2005	NE	1,1	1
15/07/2005	NE	6,2	2
16/07/2005	NE	11	4
17/07/2005	NW	5,5	2
18/07/2005	W	7,3	3
19/07/2005	W	7,5	3
20/07/2005	NW	8	3
21/07/2005	N	15,6	4
22/07/2005	N	7,4	3
23/07/2005	S	4,2	2
24/07/2005	SW	11,7	4
25/07/2005	W	6,8	2
26/07/2005	SW	0,5	0
27/07/2005	S	2,8	1
28/07/2005	SW	3	1
29/07/2005	NW	6,7	2
30/07/2005	W	6	2
31/07/2005	W	6,5	2
01/08/2005	N	9,2	3
02/08/2005	NW	5,5	2
03/08/2005	N	15,2	4
04/08/2005	NW	4,8	2
05/08/2005	W	7	3
06/08/2005	N	10,7	3
07/08/2005	NE	13,2	4
08/08/2005	NE	4,3	2
09/08/2005	NE	5,1	2
10/08/2005	E	4,5	2
11/08/2005	N	8,6	3
12/08/2005	NW	12,1	4
13/08/2005	W	6,5	2
14/08/2005	NW	18,1	5
15/08/2005	NE	8	3
16/08/2005	NE	10,1	3
17/08/2005	E	6,4	2
18/08/2005	W	4,1	2
19/08/2005	NW	20,4	5
20/08/2005	N	13,2	4
21/08/2005	N	12,3	4
22/08/2005	NW	2,5	1
23/08/2005	N	11,2	4
24/08/2005	SW	9,6	3
25/08/2005	SW	11,8	4
26/08/2005	W	9,7	3
27/08/2005	W	2,2	1
28/08/2005	NE	4,8	2

On part d’un relevé de données réelles.

La “population” est constituée des jours d’observation.

Caractère qualitatif : direction du vent

Caractère quantitatif : sa vitesse, en nœuds.

Le traitement statistique commence au moment où on s’affranchit des individus, pour s’intéresser à la façon dont ils se distribuent vis à vis du caractère observé.

Travail souvent déjà fait par l’énoncé ou par le professeur, et donc peu “visible” pour les élèves.

Questions génératrices de la statistique

Pour justifier le fait de s'intéresser aux "statistiques descriptives", on évoque souvent la trop grande complexité des informations recueillies, et la nécessité d'en fabriquer un résumé. Mais à quoi un tel résumé peut-il servir ?

- Un animal, disons un taureau, pèse 520 kg. Peut-on dire qu'il est gros ? Ou du moins qu'il est gros pour son âge ?

Pour répondre, il faut remonter jusqu'au domaine même de la statistique : en regardant cet animal comme un *individu* d'une certaine *population* (parente effective), au sein de laquelle on essaiera de le situer du point de vue d'un *caractère* déterminé, son poids. Cela conduira à s'intéresser à un échantillon de cette population, puis à la *série statistique* des poids des individus de cet échantillon.

On cherche à situer le poids observé par rapport à ceux de l'échantillon : par exemple, on peut comparer le poids avec un résumé numérique de la série : ainsi arrive-t-on au *secteur d'études* (Statistique descriptive). Si la *moyenne* est de 536 kg, on ne pourra pas dire que l'animal est gros. Mais est-il proche de la moyenne ? Est-il "dans la moyenne" ? Ou bien est-il "anormalement petit" ? On est ainsi amené à faire entrer en jeu la *dispersion*.

Exemple adapté, tiré de Chevallard, 2002 (Voir Bibliographie)

On peut aussi s'intéresser à la *distribution* (des effectifs, des fréquences) de cette "variable". Sa rareté relative est alors la proportion de taureaux moins lourds que lui. Cette proportion peut s'exprimer en pourcentage. Par exemple, notre taureau peut être gros au seuil de p % (dans l'échantillon, p % des taureaux sont moins lourds que lui).

Guy Brousseau met en valeur cette idée fondamentale : représenter la distance d'un objet à une collection d'objets par la "proportion" des objets de la collection plus éloignés que lui (plus rares). Une puce qui n'a qu'un millième des puces plus grandes qu'elle est plus grande qu'un éléphant donné, si 30% des éléphants sont plus grands que lui.

Cette idée peut être utilisée pour introduire la médiane en 3^e^ (et en 2^e^) …

Considérons une population Ω, un caractère X défini sur Ω à valeurs dans une partie V de $\mathbf{R}$. Soit a un élément de V.
Les individus ω tels que $X(\omega) > a$ sont-ils fréquents ?
Autrement dit : “Avoir un X supérieur à a, est-ce avoir un “gros” X ?”

Pour répondre à cette question, il est intéressant de connaître une valeur v (si elle existe) telle qu’à peu près 50% des individus ω de Ω vérifient $X(\omega) < v$ tandis qu’à peu près 50% vérifient $X(\omega) > v$.

L’introduction de la médiane d’une série est alors motivée.

On peut alors situer a par rapport à la médiane …
Un individu dont le X est a est situé dans la moitié supérieure (ou inférieure) de la population.

On peut affiner, ce qui justifie l’introduction des … quartiles …

Retour sur “On s’affranchit des individus”

Individus	Valeurs
ω_1	x_1 (= v_1)
ω_2	x_2 (= v_2)
ω_3	x_3 (= v_3)
ω_4	x_4 (= v_4)
ω_5	x_5 (= v_3)
…	…
ω_N	x_N (= v_k)

Lorsqu’on réduit ce tableau au N-uplet $(x_1, \ldots, x_N)$, c’est-à-dire lorsqu’on le réduit à la série statistique correspondante, on perd l’identité des individus ω_i (on ne sait plus qui est au juste ω_i). On ne retient que la valeur x_i que prend sur lui le caractère X : on ne s’intéresse pas aux individus en tant que tels, mais à la distribution des valeurs de X sur Ω.

La **distribution des observations** se déduit du tableau des données ponctuelles, dont elle est "l'inverse" :

Valeurs	Échantillons
v_1	$\{\omega_1\}$
v_2	$\{\omega_2\}$
v_3	$\{\omega_3, \omega_5\}$
v_4	$\{\omega_4\}$
…	…
v_p	$\{\omega_l, \ldots\}$

La **distribution des effectifs** (ou des fréquences absolues), n_j ($1 \leq j \leq p$), qui se déduit également de la série statistique, fait perdre le rang dans la série statistique de chacune des valeurs obtenues lors du recueil des données.

Valeurs	Effectifs
v_1	n_1 (= 1)
v_2	n_2 (= 1)
v_3	n_3 (= 2)
v_4	n_4 (= 1)
…	…
v_p	n_p (= …)

3. Les documents et leur exploitation en classe

Ils ne sont pas rédigés “par niveau”.

Pourquoi ?

Que leur manque-t-il ?

Que manque-t-il à certains d’entre eux ?

Quelles sont les évolutions souhaitées ?

Eléments de bibliographie :

Brousseau G., 2005, *Situations fondamentales et processus génétiques de la statistique*, in Balises en didactique des mathématiques, cours de la 12e école d’été de didactique des mathématiques, coordonnés par A. Mercier et C. Margolinas.

Chevallard Y., 2002, *Organiser l’étude : écologie et régulation*, in Actes de la 11e école d’été de didactique des mathématiques, coordonnés par J.-L. Dorier, M. Artigue, M. Artaud, R. Berthelot, R. Floris.

Dupuis C. et Rousset-Bert S., 1998, *De l’influence des représentations disponibles sur la résolution de problèmes élémentaires de probabilité et sur l’acquisition du concept d’indépendance,* Annales de didactique et de sciences cognitives, Vol. 6, ULP IREM de Strasbourg.

Engel A., 1975, L’enseignement des probabilités et de la statistique, Cedic,

Parzysz B., 1993, *Des statistiques aux probabilités : exploitons les arbres*, Repères IREM n° 10.

Pressiat A., 2005, *De l’épistémologie de la statistique vers sa didactique dans l’enseignement secondaire*, in Balises en didactique des mathématiques, cours de la 12e école d’été de didactique des mathématiques, coordonnés par A. Mercier et C. Margolinas.