Statistiques et probabilité :

Au collège :

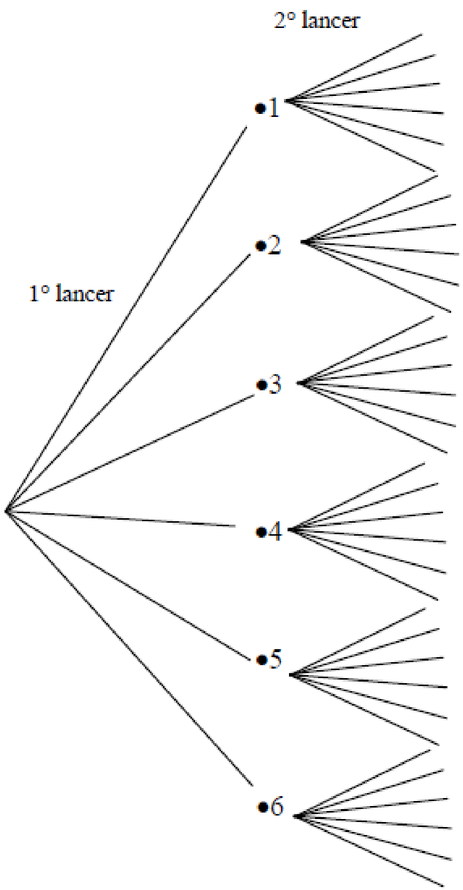
	Classe de sixième	Classe de cinquième	Classe de quatrième	Classe de troisième
Organisation et gestion de données	Organiser des données en choisissant un mode de représentation adapté. Lire et compléter une graduation sur une demi-droite graduée. Lire et interpréter des informations à partir d'une représentation graphique.	Repérage sur une droite graduée et dans le plan. Classes, effectifs, fréquences. Tableaux de données : lectures, interprétation, élaboration, représentations graphiques.	Moyenne pondérée.	Caractéristiques de position : médiane, quartiles. Approche de caractéristiques de dispersion : étendue. Notion de probabilité

En 3ème

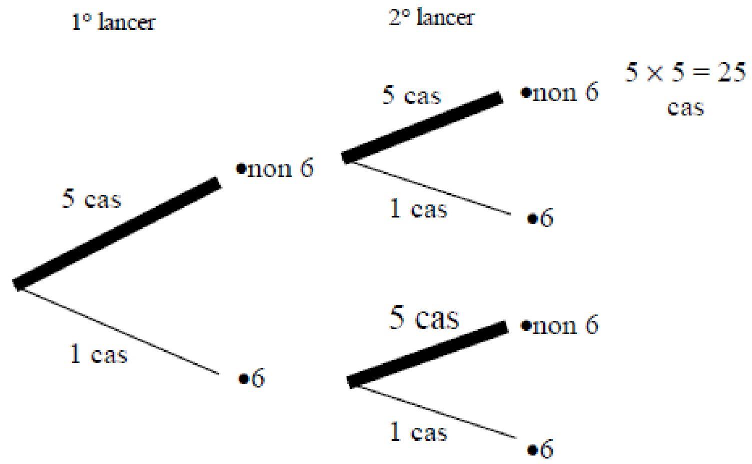
1.4. Notion de probabilité [Thèmes de convergence]	- Comprendre et utiliser des notions élémentaires de probabilité. - Calculer des probabilités dans des contextes familiers.	La notion de probabilité est abordée à partir d'expérimentations qui permettent d'observer les fréquences des issues dans des situations familières (pièces de monnaie, dés, roues de loteries, urnes, etc.). La notion de probabilité est utilisée pour modéliser des situations simples de la vie courante. Les situations étudiées concernent les expériences aléatoires à une ou *à deux épreuves*.

- Continuité dans les apprentissages
- Apprentissage progressif des arbres pondérés : attention, cette notion n’est formellement abordée qu’en classe de 1ere.

Un exemple pour introduire les règles d'utilisation d’un arbre pondéré : deux lancers d’un dé supposé bien équilibré



On dénombre toutes les issues qui sont équiprobables avec un arbre (36 chemins). Puis, si on s'intéresse seulement à l'apparition du numéro 6, on propose de simplifier l'arbre en regroupant les branches qui conduisent à un numéro différent de 6. On calcule alors grâce au modèle d'équiprobabilité la probabilité des différentes issues.



[1] Ce texte est en partie issu des travaux menés par le groupe chargé de la rédaction du document ressources « Statistiques - probabilités ».

En seconde (probabilités)

CONTENUS	CAPACITÉS ATTENDUES	COMMENTAIRES
Probabilité sur un ensemble fini		
Probabilité d'un événement.	• Déterminer la probabilité d'événements dans des situations d'équiprobabilité. • Utiliser des modèles définis à partir de fréquences observées.	La probabilité d'un événement est définie comme la somme des probabilités des événements élémentaires qui le constituent. Pour les calculs de probabilités, on utilise des arbres, des diagrammes ou des tableaux.
Réunion et intersection de deux événements, formule : $p(A \cup B) + p(A \cap B) = p(A) + p(B)$.	• Connaître et exploiter cette formule.	

En seconde (statistiques)

CONTENUS	CAPACITÉS ATTENDUES
Statistique descriptive, analyse de données Caractéristiques de position et de dispersion • médiane, quartiles ; • moyenne.	• Utiliser un logiciel (par exemple, un tableur) ou une calculatrice pour étudier une série statistique. • Passer des effectifs aux fréquences, calculer les caractéristiques d'une série définie par effectifs ou fréquences. • Calculer des effectifs cumulés, des fréquences cumulées. • Représenter une série statistique graphiquement (nuage de points, histogramme, courbe des fréquences cumulées).
Échantillonnage Notion d'échantillon. Intervalle de fluctuation d'une fréquence au seuil de 95%*. Réalisation d'une simulation.	• Concevoir, mettre en œuvre et exploiter des simulations de situations concrètes à l'aide du tableur ou d'une calculatrice. • Exploiter et faire une analyse critique d'un résultat d'échantillonnage.

Notion d’échantillon

Dans le sens commun des sondages, un échantillon est un sous ensemble obtenu par prélèvement aléatoire dans un population.

Deux types d'échantillons :

- Échantillons constitués *sans remise*
- Échantillons constitués *avec remise*

Le programme de Seconde 2009 ne retient que le 2ème type d'échantillons :

"*Un échantillon est constitué des résultats de n répétitions indépendantes de la même expérience*".

L'échantillonnage est l'étude des distributions de fréquences de variables définies sur l’ensemble des échantillons (proportion, moyenne, variance…).

Un exemple

On considère une population de 4 enfants :

Adeline, Benjamin, Clara et David, d'âges respectifs 12, 13, 14 et 15 ans et on s'intéresse aux enfants de plus de 14 ans et demi. Il y en a une proportion p = 1/4 dans la population-mère.

On constitue (avec remise) des échantillons de taille 3.

On peut ainsi constituer 4^3=64 échantillons.

Proportion d'enfants de plus de 14 ans et demi dans les échantillons

A	A	A	0
A	A	B	0
A	A	C	0
A	A	D	1/3
A	B	A	0
A	B	B	0
A	B	C	0
A	B	D	1/3
A	C	A	0
A	C	B	0
A	C	C	0
A	C	D	1/3
A	D	A	1/3
A	D	B	1/3
A	D	C	1/3
A	D	D	2/3

B	A	A	0
B	A	B	0
B	A	C	0
B	A	D	1/3
B	B	A	0
B	B	B	0
B	B	C	0
B	B	D	1/3
B	C	A	0
B	C	B	0
B	C	C	0
B	C	D	1/3
B	D	A	1/3
B	D	B	1/3
B	D	C	1/3
B	D	D	2/3

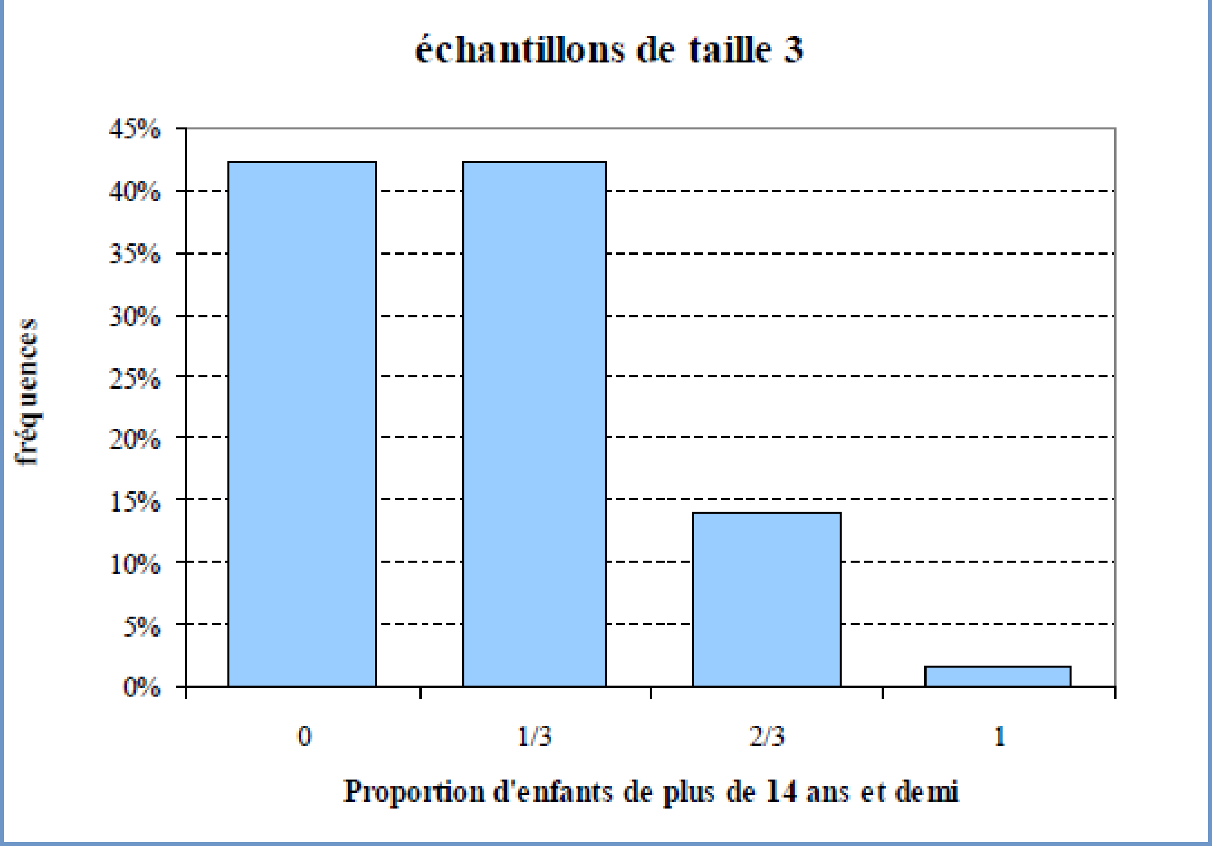
C	A	A	0
C	A	B	0
C	A	C	0
C	A	D	1/3
C	B	A	0
C	B	B	0
C	B	C	0
C	B	D	1/3
C	C	A	0
C	C	B	0
C	C	C	0
C	C	D	1/3
C	D	A	1/3
C	D	B	1/3
C	D	C	1/3
C	D	D	2/3

D	A	A	1/3
D	A	B	1/3
D	A	C	1/3
D	A	D	2/3
D	B	A	1/3
D	B	B	1/3
D	B	C	1/3
D	B	D	2/3
D	C	A	1/3
D	C	B	1/3
D	C	C	1/3
D	C	D	2/3
D	D	A	2/3
D	D	B	2/3
D	D	C	2/3
D	D	D	1

Parmi les 64 échantillons :

Proportion d'enfants de plus de 14 ans et demi dans l'échantillon	Nombre d'échantillons	Fréquences d'échantillons
0/3	27	42,2 %
1/3	27	42,2 %
2/3	9	14 %
3/3	1	1,6%

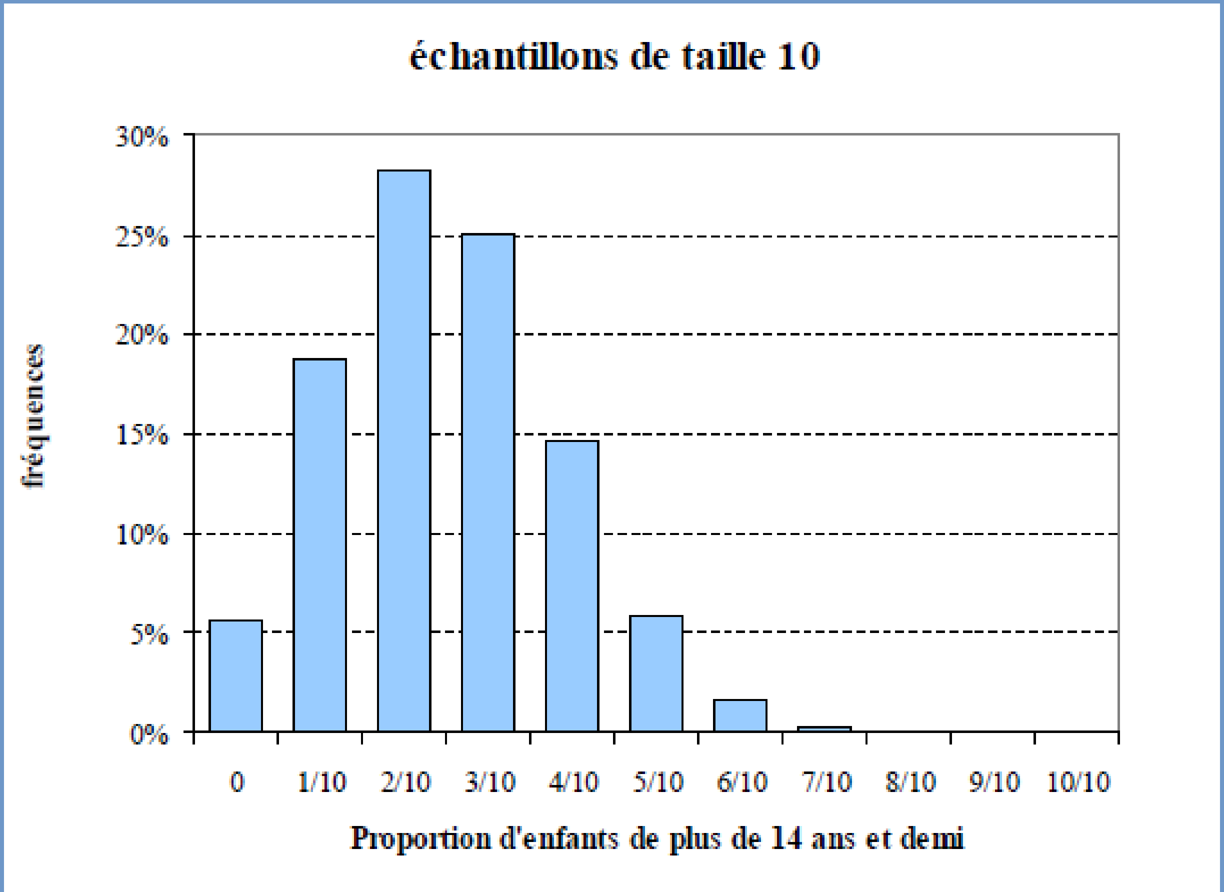
Distribution d'effectifs et de fréquences de la proportion d'enfants de plus de 14 ans et demi dans les échantillons

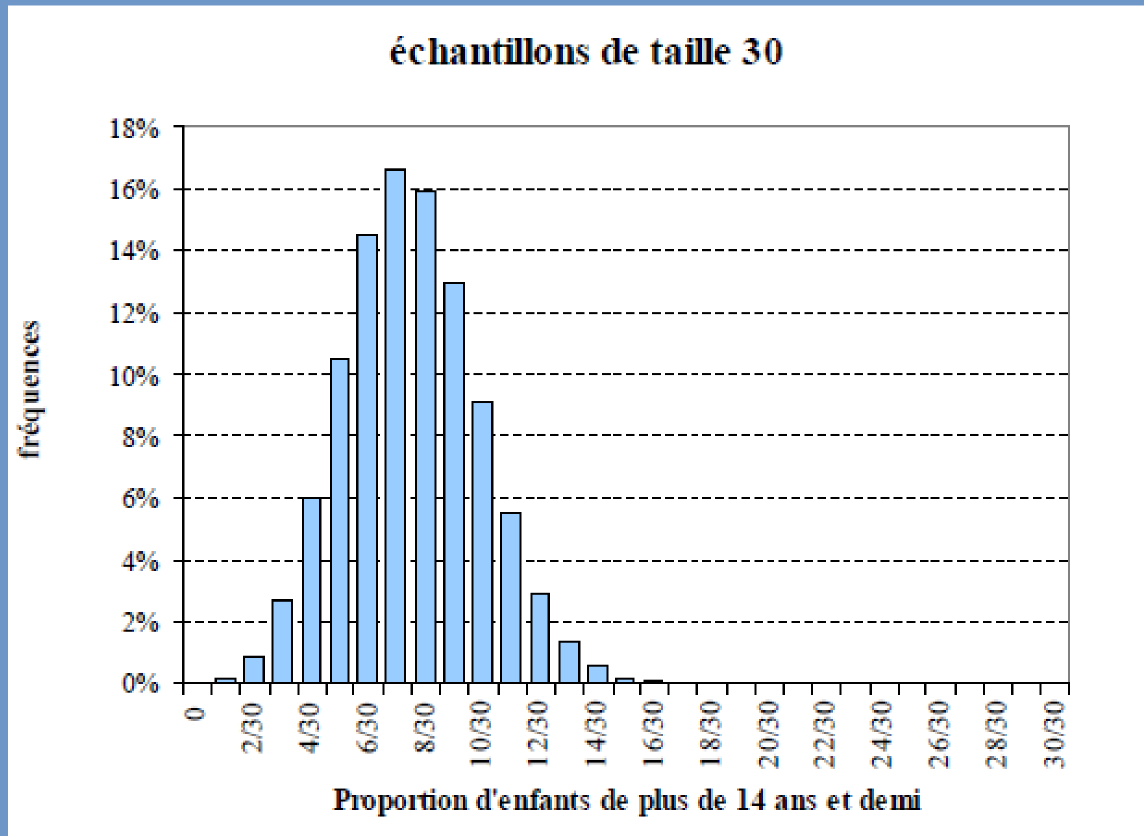

échantillons de taille 3
fréquences
45%
40%
35%
30%
25%
20%
15%
10%
5%
0%
0
1/3
2/3
1
Proportion d'enfants de plus de 14 ans et demi

Échantillons de taille 10 issus d'une population contenant 1/4 d'enfants de plus de 14 ans et demi :

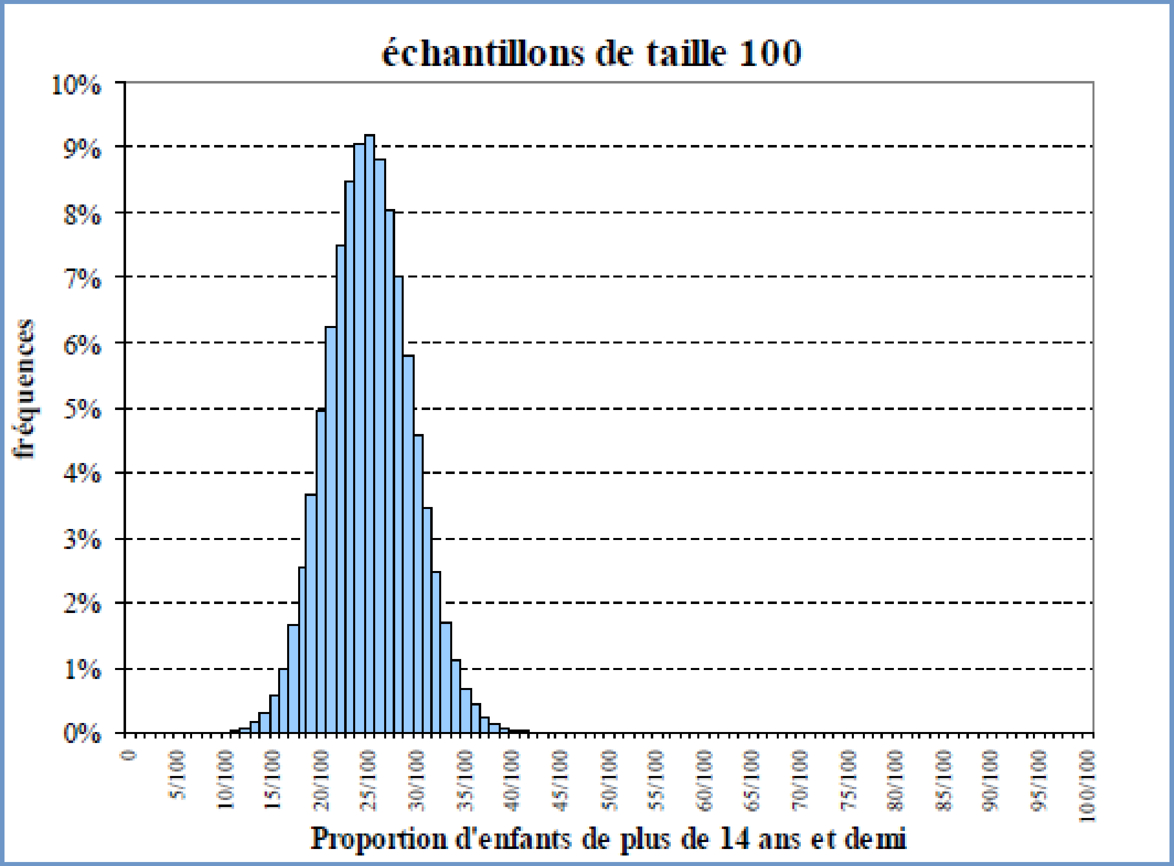
Échantillons avec une proportion de	Fréquences d'échantillons
0/10	5,6%
1/10	18,8%
2/10	28,2%
3/10	25,0%
4/10	14,6%
5/10	5,8%
d'enfants de plus de 14 ans et demi	

Échantillons avec une proportion de	Fréquences d'échantillons
6/10	1,6%
7/10	0,3%
8/10	0,0%
9/10	0,0%
10/10	0,0%
d'enfants de plus de 14 ans et demi	


échantillons de taille 10
fréquences
30%
25%
20%
15%
10%
5%
0%
0
1/10
2/10
3/10
4/10
5/10
6/10
7/10
8/10
9/10
10/10
Proportion d'enfants de plus de 14 ans et demi



échantillons de taille 30
18%
16%
14%
12%
10%
8%
6%
4%
2%
0%
fréquences
0
2/30
4/30
6/30
8/30
10/30
12/30
14/30
16/30
18/30
20/30
22/30
24/30
26/30
28/30
30/30
Proportion d'enfants de plus de 14 ans et demi

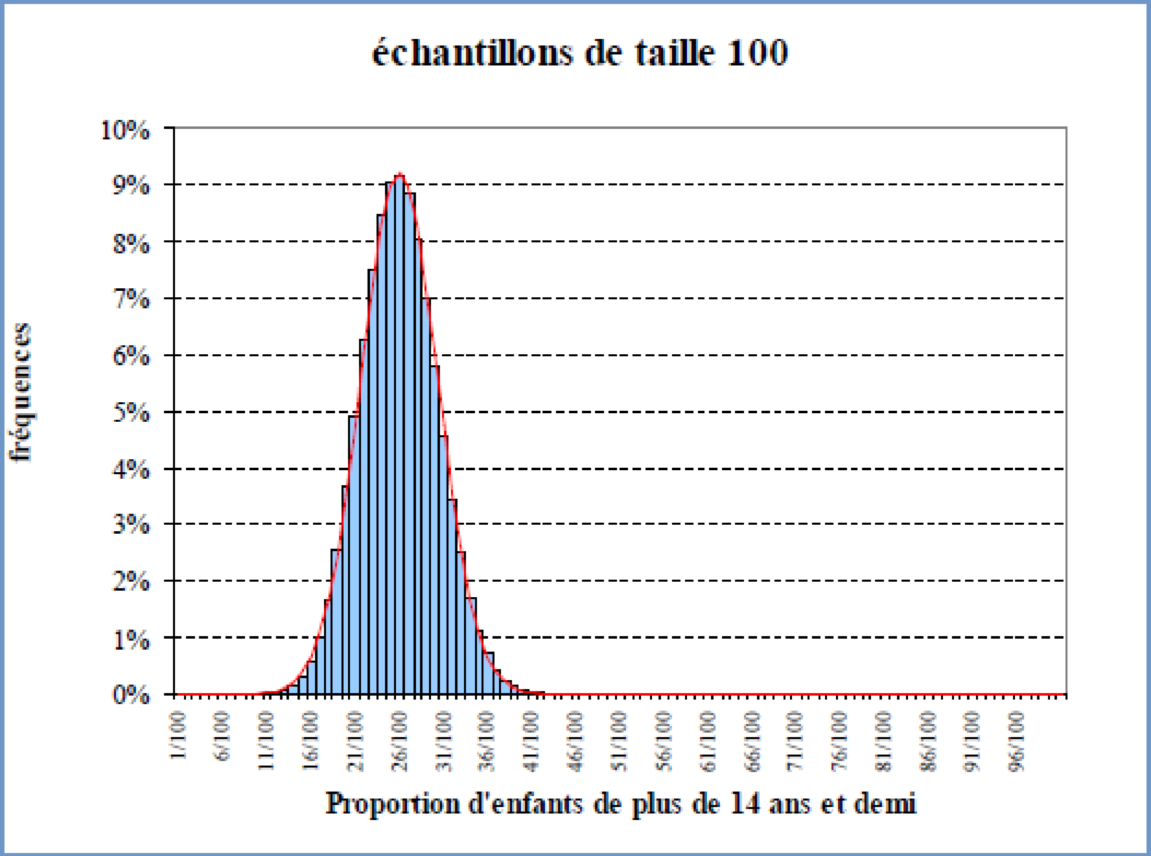

échantillons de taille 100
fréquences
10%
9%
8%
7%
6%
5%
4%
3%
2%
1%
0%
0
5/100
10/100
15/100
20/100
25/100
30/100
35/100
40/100
45/100
50/100
55/100
60/100
65/100
70/100
75/100
80/100
85/100
90/100
95/100
100/100
Proportion d'enfants de plus de 14 ans et demi

Résultat :

Quand n augmente, les proportions observées sont de plus en plus souvent proches de la proportion du caractère dans la population-mère.

C'est la loi faible des grands nombres.

Lorsque n est grand la distribution de fréquence de la proportion d'échantillonnage s'approche d'une « distribution en cloche »


échantillons de taille 100
fréquences
10%
9%
8%
7%
6%
5%
4%
3%
2%
1%
0%
1/100
6/100
11/100
16/100
21/100
26/100
31/100
36/100
41/100
46/100
51/100
56/100
61/100
66/100
71/100
76/100
81/100
86/100
91/100
96/100
Proportion d'enfants de plus de 14 ans et demi

Définition :

L’intervalle de fluctuation d’une fréquence ou proportion à 95%, pour des échantillons de taille n, est l’intervalle :

– d'amplitude minimale,

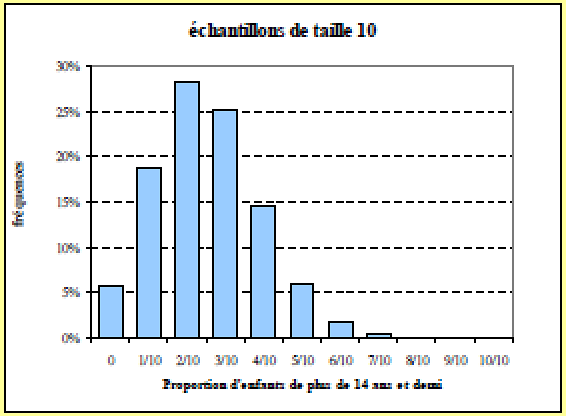
– centré autour de p, proportion du caractère dans la population,

– contenant la proportion observée sur un échantillon aléatoire de taille n, avec une probabilité au moins égale à 0,95.

Échantillons de taille 10 issus d'une population contenant 1/4 d'enfants de plus de 14 ans et demi :

p = 25 %

0/10	5,6 %
1/10	18,8 %
2/10	28,2 %
3/10	25,0 %
4/10	14,6 %
5/10	5,8 %
6/10	1,6 %
7/30	0,3 %
8/10	0,0 %
9/10	0,0 %
10/10	0,0 %



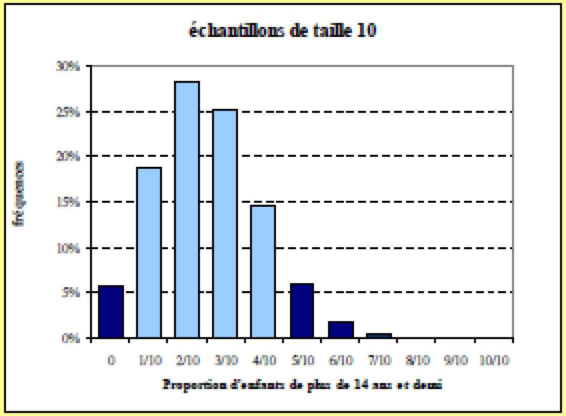
Distribution des fréquences

Échantillons de taille 10 issus d'une population contenant 1/4 d'enfants de plus de 14 ans et demi :

p = 25 %

0/10	5,6 %
1/10	18,8 %
2/10	28,2 %
3/10	25,0 %
4/10	14,6 %
5/10	5,8 %
6/10	1,6 %
7/30	0,3 %
8/10	0,0 %
9/10	0,0 %
10/10	0,0 %

86,6 %



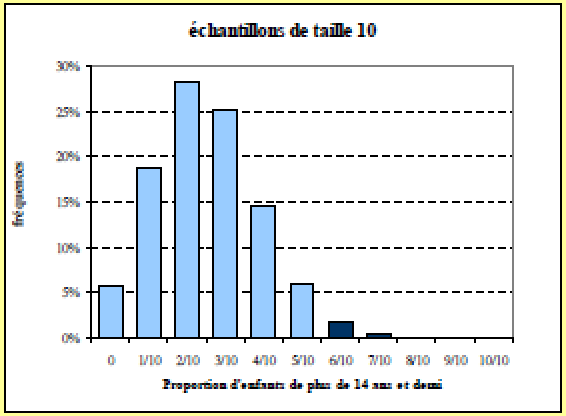
Distribution des fréquences

Échantillons de taille 10 issus d'une population contenant 1/4 d'enfants de plus de 14 ans et demi :

p = 25 %

0/10	5,6 %
1/10	18,8 %
2/10	28,2 %
3/10	25,0 %
4/10	14,6 %
5/10	5,8 %
6/10	1,6 %
7/30	0,3 %
8/10	0,0 %
9/10	0,0 %
10/10	0,0 %

98 %



Distribution des fréquences

L'intervalle de fluctuation est [0 ; 0,5].

Échantillons de taille 30 issus d'une population contenant 1/4 d'enfants de plus de 14 ans et demi :

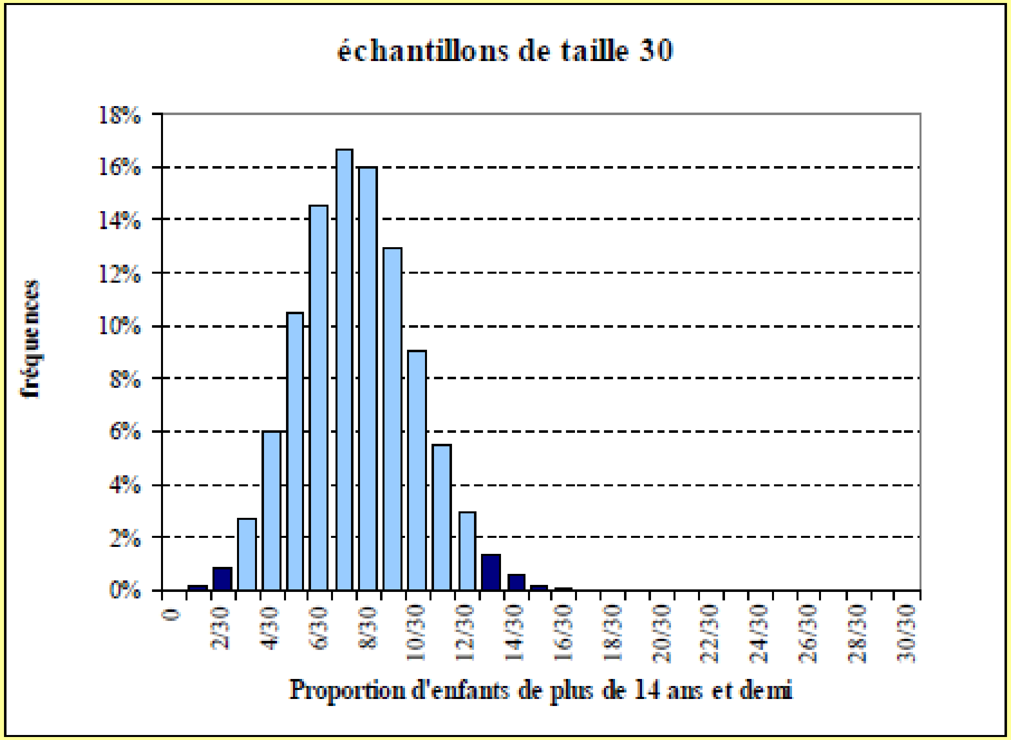
0/30	0,0%
1/30	0,2%
2/30	0,9%
3/30	2,7%
4/30	6,0%
5/30	10,5%
6/30	14,5%
7/30	16,6%
8/30	15,9%
9/30	13,0%
10/30	9,1%

11/30	5,5%
12/30	2,9%
13/30	1,3%
14/30	0,5%
15/30	0,2%
16/30	0,1%
17/30	0,0%
18/30	0,0%
19/30	0,0%
20/30	0,0%

21/30	0,0%
22/30	0,0%
23/30	0,0%
24/30	0,0%
25/30	0,0%
26/30	0,0%
27/30	0,0%
28/30	0,0%
29/30	0,0%
30/30	0,0%

96,7 %

p = 25 %



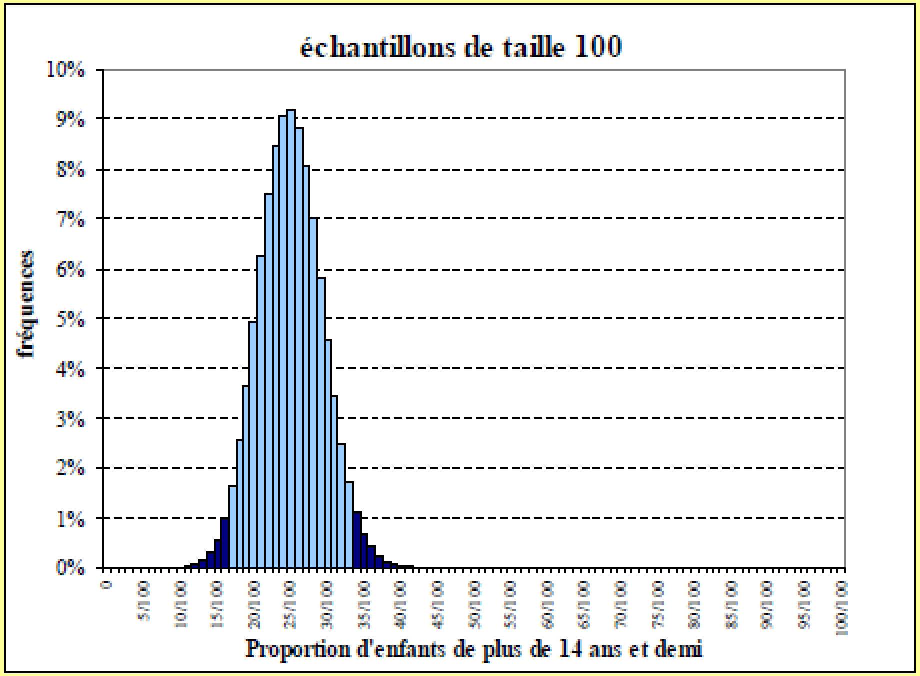
L'intervalle de fluctuation est [0,1 ; 0,4].

Échantillons de taille 100 issus d'une population contenant 1/4 d'enfants de plus de 14 ans et demi :

95,1 %

25 %

11/100	0,0%	21/100	6,3%	31/100	3,4%
12/100	0,1%	22/100	7,5%	32/100	2,5%
13/100	0,1%	23/100	8,5%	33/100	1,7%
14/100	0,3%	24/100	9,1%	34/100	1,1%
15/100	0,6%	25/100	9,2%	35/100	0,7%
16/100	1,0%	26/100	8,8%	36/100	0,4%
17/100	1,7%	27/30	8,1%	37/100	0,2%
18/100	2,5%	28/30	7,0%	38/100	0,1%
19/100	3,7%	29/100	5,8%	39/100	0,1%
20/100	4,9%	30/100	4,6%	40/100	0,0%



L'intervalle de fluctuation est [0,17 ; 0,33].

Cas où la proportion est connue :

Propriété :

Connaissant la proportion p ($0,2 \leq p \leq 0,8$) d'individus ayant un certain caractère A dans une population, la fréquence de ce caractère dans un échantillon de taille n, ($n \geq 25$) appartient à l'intervalle [p-1/rac(n) ; p+1/rac(n)] avec une probabilité d'au moins 95%

Démonstration …

on fait comment ??

Exemple : « La parité, c’est quoi ? »

Deux entreprises A et B recrutent dans une région où il y a autant d’hommes que de femmes. Dans l’entreprise A, il y a 100 employées dont 43 femmes. Dans l’entreprise B, il y a 2500 employés dont 1150 femmes.

Quelle entreprise respecte le mieux la parité ?

Dans l’entreprise A, l’intervalle de fluctuation est [0,4 ; 0,6] et 43% appartient à cet intervalle.

Dans l’entreprise B, l’intervalle de fluctuation est [0,48 ; 0,52] et 46% n’appartient pas à cet intervalle.

Dans l’entreprise B la proportion 46% s’observe dans moins de 5% des échantillons, donc on peut rejeter l’hypothèse qu’elle respecte mieux la parité que l’entreprise A, contrairement aux apparences

Cas où la proportion n'est pas connue :

Propriété :

Dans une population, on désire estimer la proportion inconnue p d'un caractère donné. On étudie, pour cela, un échantillon de taille n, $n \geq 25$. Le caractère étudié apparait avec la fréquence f ($0,2 \leq f \leq 0,8$). On peut estimer, que la proportion p est dans l'intervalle

[f-1/rac(n) ; f+1/rac(n)] avec une probabilité d'au moins 95%

Nouveau programme en 1ère S
Statistiques et probabilité :

Les points majeurs :

La loi géométrique tronquée

La loi binomiale

Échantillonnage et prise de decision

Modèle de la répétition d'expériences identiques et indépendantes à deux ou trois issues.	• Représenter la répétition d'expériences identiques et indépendantes par un arbre pondéré. • Utiliser cette représentation pour déterminer la loi d'une variable aléatoire associée à une telle situation.	Pour la répétition d'expériences identiques et indépendantes, la probabilité d'une liste de résultats est le produit des probabilités de chaque résultat. La notion de probabilité conditionnelle est hors programme. On peut aussi traiter quelques situations autour de la loi géométrique tronquée. ◇ On peut simuler la loi géométrique tronquée avec un algorithme.

1. Loi géométrique tronquée

Les situations de répétitions d'une expérience aléatoire, dans des conditions d'indépendance constituent un élément fort du programme de première.

L'introduction de la loi géométrique tronquée présente de nombreux avantages :

– travailler sur des répétitions d'une expérience de Bernoulli ;

– envisager ces répétitions sous l'angle algorithmique ;

– présenter une situation d'arbre pour lequel tous les chemins n'ont pas la même longueur ;

– exploiter dans un autre cadre les propriétés des suites géométriques ;

– exploiter dans un autre cadre des résultats sur la dérivation.

Définition

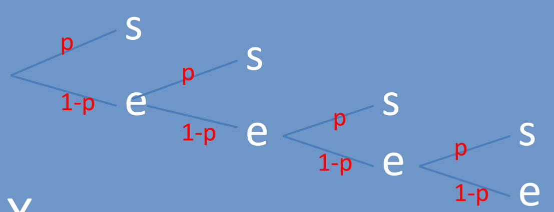
Soit p un réel de l'intervalle]0, 1[et n un entier naturel non nul. On considère l'expérience aléatoire qui consiste à répéter de manière indépendante une expérience de Bernoulli de paramètre p avec au maximum n répétitions et arrêt du processus au premier succès.

On appelle loi géométrique tronquée de paramètres n et p la loi de la variable aléatoire X définie par :

$X = 0$ si aucun succès n'a été obtenu ;

pour $1 \leq k \leq n$, $X = k$ si le premier succès est obtenu à l'étape k.

Exemple pour n=4



Déterminons la loi de X.

$X = 0$ si aucun succès n'a été obtenu donc avec l'outil arbre: $P(X = 0) = (1-p)^n$

Pour $1 \leq k \leq n$, avec l'arbre, le premier succès est obtenu à l'étape k pour le chemin qui présente dans l'ordre $(k – 1)$ échecs et un succès d'où :

$$P(X = k) = (1 – p)^{k-1} p$$