

L'enseignement des statistiques et probabilités en premières S, ES et STI2D-STL¹

philippe.dutarte@ac-creteil.fr

I – Modèle de la répétition d'expériences identiques et indépendantes à deux ou trois issues

1 – Arbre pondéré

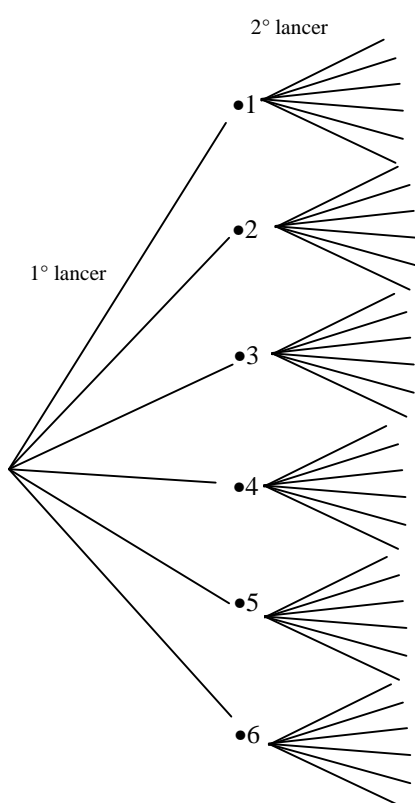
La formulation du programme « répétition d'expériences identiques et indépendantes à deux ou trois issues » exprime que les situations envisagées répondent aux critères suivants :

- définition précise d'une expérience aléatoire, l'univers associé Ω étant composé de deux ou trois éléments,
- répétition de cette expérience dans des conditions qui garantissent qu'à chaque fois Ω est toujours l'univers associé à l'expérience, quelle que soit la répétition.

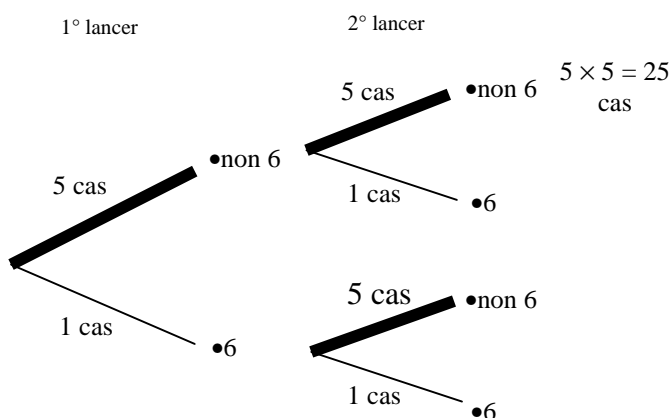
Dans ces conditions l'univers associé à n répétitions de l'expérience est le produit cartésien $\Omega \times \Omega \times \dots \times \Omega = \Omega^n$.

Une introduction par le recours aux situations d'équiprobabilité permet de justifier les règles d'utilisation des arbres pondérés lorsque les probabilités considérées sont rationnelles (on admet qu'il en est de même pour des probabilités réelles).

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 ».

Univers associé à un lancer correspondant à l'arbre équiprobable : $\Omega_1 = \{1, 2, 3, 4, 5, 6\}$.

Univers associé à un lancer correspondant à l'arbre simplifié : $\Omega_2 = \{\text{non}6, 6\}$.

On peut ensuite construire l'arbre simplifié associé à trois lancers du même dé et calculer toujours dans le cadre de l'équiprobabilité la probabilité de quelques issues ou événements.

L'étape suivante est la construction d'un arbre pondéré. Pour plus de commodité, on peut noter A l'issue « obtenir un numéro 6 » et B l'issue « ne pas obtenir un numéro 6 ». On indique alors sur l'arbre les issues A et B avec sur chaque branche les probabilités $1/6$ et $5/6$. Dans ce cas, la règle du produit pour trouver la probabilité d'un chemin se trouve validée par l'arbre précédent qui lui est bâti sur le modèle d'équiprobabilité.

2 – 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.

Une situation d'étude

La probabilité qu'un atome se désintègre par unité de temps est 0,07. On décide d'observer cette désintégration en limitant le temps d'attente à 100 unités de temps.

On peut concevoir un algorithme qui affiche une série de 200 valeurs du temps d'attente avant la désintégration ainsi que le temps moyen d'attente calculé à partir de ces 200 valeurs.

On conviendra de noter 0 lorsqu'après 100 unités de temps l'atome n'est pas encore désintégré. On distingue ainsi cette situation de la désintégration lors de la 100^e unité de temps.

On constate que les temps d'attentes avant désintégration sont, individuellement, extrêmement imprévisibles. En revanche, la moyenne sur 200 expériences est relativement stable avec des valeurs autour de 13, 14 ou 15. Il est donc sans doute intéressant d'étudier de plus près la loi de la variable aléatoire « temps d'attente ». On montrera que l'espérance de cette variable aléatoire est environ 14,2.

L1	L2	L3	1
7			
1			
4			
33			
1			
15			
13			
L1(200) = 15			

```

PrgmTPSATTEN
13.815
Fait
15.98
Fait
15.275
Fait

```

```

PROGRAM:TPSATTEN
: suite(0,1,1,200
:)→L1
: For(N,1,200)
: 0→K
: 0→A
: While A=0 et K<
100
:  ent(NbrAléat+0.
07)→A
: K+1→K
: End
: If A=0
: Then
: A→L1(N)
: Else
: K→L1(N)
: End
: End
: Disp moyenne(L1
)

```

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 dans des conditions identiques 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 .

Loi de la variable X

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 : $P(X = k) = (1 - p)^{k-1} p$.

On vérifie facilement que $\sum_{k=0}^n P(X = k) = 1$ (exploitation des sommes géométriques)

Au niveau première, il est possible de déterminer l'espérance de la loi géométrique tronquée de paramètres n et p . Cette activité permet de mobiliser à la fois le cours sur les suites géométriques et sur la dérivation.

Sans être exigible, elle peut faire l'objet d'un travail de recherche.

Activité : exploiter les résultats d'analyse

1. Montrer que $E(X) = p \left[1 + 2(1 - p) + 3(1 - p)^2 \cdots + n(1 - p)^{n-1} \right]$. On pose alors $q = 1 - p$ et on écrit plus simplement : $E(X) = p \left[1 + 2q + 3q^2 \cdots + nq^{n-1} \right] = p \sum_{k=1}^n kq^{k-1}$.
2. Soit f la fonction définie sur $]0, 1[$ par : $f(x) = 1 + x + x^2 + \cdots + x^n$.
 - a. Pour tout réel x de $]0, 1[$, écrire $f(x)$ sous la forme d'un quotient.
 - b. Vérifier que f est dérivable sur $]0, 1[$ et calculer deux expressions différentes de $f'(x)$ pour tout réel x élément de $]0, 1[$.
 - c. En déduire le calcul de la somme $1 + 2x + 3x^2 \cdots + nx^{n-1} = \sum_{k=1}^n k x^{k-1}$ pour tout x de $]0, 1[$.
3. Calculer $E(X)$ et vérifier que : $E(X) = \frac{1}{p} [1 - (1 + np)(1 - p)^n]$.
4. Utiliser une calculatrice ou un tableur pour émettre une conjecture sur la limite de $E(X)$ lorsque n tend vers l'infini.

Il semble que la suite $[1 - (1 + np)(1 - p)^n]$ admette pour limite 0 quelle que soit la valeur de p choisie.

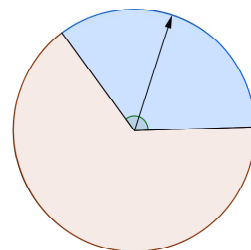
La limite de $E(X)$ semble donc être $\frac{1}{p}$.

II – Loi binomiale**1 – Découverte de la loi binomiale et introduction des coefficients binomiaux****Exemple 1 : répétition d'une épreuve de Bernoulli de paramètre p quelconque**

On fait tourner la roue de loterie présentée ci-contre : on obtient la couleur « Bleu » avec une probabilité qui dépend de l'angle indiqué sur la figure et qui est notée p . On obtient donc la couleur « Rouge » avec une probabilité de $1 - p$.

On décide de noter S, comme succès, l'éventualité « la flèche tombe sur la zone bleue » et de noter E, comme échec, l'éventualité contraire c'est-à-dire « la flèche tombe sur la zone rouge ».

1. On répète quatre fois cette épreuve de Bernoulli de paramètre p .



Représenter cette répétition par un arbre pondéré à quatre niveaux.

2. On définit la variable aléatoire X correspondant au nombre de succès obtenus à l'issue des quatre répétitions.

En utilisant l'arbre, déterminer la probabilité des événements suivants :

« $X = 0$ » ; « $X = 4$ » ; « $X = 1$ » et « $X = 2$ ».

Il faut noter que $P(X = 1)$ et $P(X = 2)$ s'obtiennent en comptant les chemins qui conduisent respectivement à 1 et 2 succès.

On note $\binom{4}{1}$ et on lit « 1 parmi 4 »

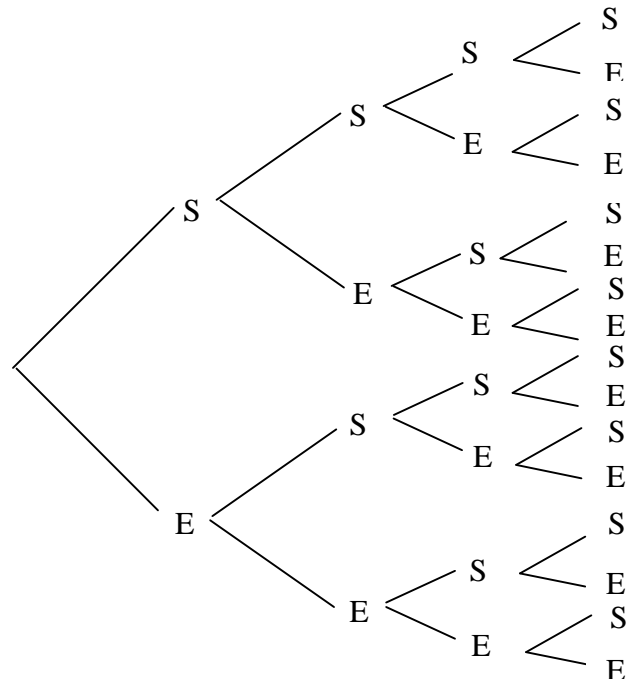
le nombre de chemins qui conduisent à 1 succès exactement.

Ici, $\binom{4}{1} = 4$.

On note $\binom{4}{2}$ et on lit « 2 parmi 4 »

le nombre de chemins qui conduisent à 2 succès exactement.

Ici, $\binom{4}{2} = 6$.



Enfin la réponse attendue est : $P(X = 2) = 10 p^2 (1 - p)^3$ ou $P(X = 2) = \binom{5}{2} p^2 (1 - p)^3$.

Un schéma de Bernoulli associé à n répétitions peut être représenté par un arbre pondéré qui comporte n niveaux.

Par définition la loi binomiale de paramètres n et p , notée $\mathcal{B}(n, p)$, est la loi de la variable aléatoire X correspondant au nombre de succès dans la répétition de n épreuves de Bernoulli de paramètre p .

Coefficients binomiaux

Pour déterminer par exemple $P(X = 2)$, dans le cas d'un schéma de Bernoulli associé à n répétitions, on procéderait de la même façon : la probabilité de chaque chemin qui réalise exactement deux succès est : $p^2 (1 - p)^{n-2}$. Il faut ensuite multiplier cette probabilité par le nombre de chemins qui présentent exactement deux succès. Ce nombre est noté $\binom{n}{2}$ et on

lit « 2 parmi n ». Il peut être obtenu avec une calculatrice ou avec un tableur.

Plus généralement :

Si n est un entier naturel et si k est un entier compris entre 0 et n , on note $\binom{n}{k}$ et on lit « k parmi n » le nombre de chemins qui réalisent exactement k succès dans l'arbre à n niveaux, associé à un schéma de Bernoulli.

Formule générale de la loi binomiale

La probabilité de chacun des chemins qui réalisent exactement k succès est $p^k (1 - p)^{n-k}$. On obtient donc :

Soient un entier naturel n et un réel p de l'intervalle $[0, 1]$. La variable aléatoire X correspondant au nombre de succès dans la répétition de n épreuves de Bernoulli de paramètre p suit la loi binomiale $\mathcal{B}(n, p)$ avec pour tout entier k compris entre 0 et n :

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Remarque : les nombres $\binom{n}{k}$ sont souvent appelés les coefficients binomiaux. Ils interviennent en effet comme coefficients dans la formule générale ci-dessus mais aussi dans la formule du binôme de Newton qui donne le développement de $(a + b)^n$ pour tous réels a et b .

Propriétés des coefficients binomiaux

– symétrie

Le nombre de chemins réalisant $(n - k)$ succès est aussi le nombre de chemins réalisant k échecs. Par symétrie, on obtient le nombre de chemins réalisant k succès.

Si n est un entier naturel et si k est un entier compris entre 0 et n , on note $\binom{n}{k} = \binom{n}{n-k}$.

– formule de Pascal

Il s'agit ici de calculer un coefficient binomial associé à $(n + 1)$ répétitions à partir des coefficients calculés sur l'arbre réalisé au niveau n .

Le coefficient binomial $\binom{n+1}{k+1}$ donne le nombre de chemins qui réalisent exactement $(k + 1)$ succès.

Il y a deux façons d'obtenir $(k + 1)$ succès suivant qu'à la dernière étape on obtient un succès ou un échec.

– Si la dernière étape donne un échec, il faut compter les chemins qui au niveau précédent conduisaient déjà à $(k + 1)$ succès. On sait que $\binom{n}{k+1}$ chemins sont dans ce cas.

– Si la dernière étape donne un succès, il faut compter les chemins qui au niveau précédent conduisaient à exactement k succès. On sait que $\binom{n}{k}$ chemins sont dans ce cas.

En conclusion, $\binom{n}{k} + \binom{n}{k+1}$ chemins de l'arbre des $(n + 1)$ répétitions conduisent à $(k + 1)$ succès, d'où le résultat :

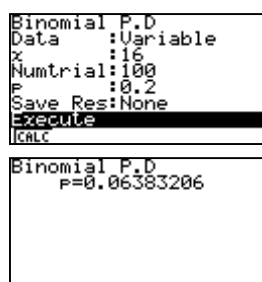
Si n est un entier naturel et si k est un entier compris entre 0 et $n - 1$, on note $\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}$.

2 – Calcul et représentations graphiques

Les moyens de calcul actuels permettent d'obtenir directement les probabilités $P(X = k)$ ou $P(X \leq k)$, où la variable aléatoire X suit une loi binomiale, de dresser des tables de ces valeurs et d'obtenir des représentations graphiques correspondantes. D'avantage que les calculs en eux-mêmes, l'accent peut être mis sur l'exploitation des tables ou des graphiques et sur l'interprétation des résultats.

Exemple du calcul de $P(X = 16)$ et de $P(X \leq 16)$ où X suit la loi binomiale de paramètres $n = 100$ et $p = 0,2$.

Sur un modèle CASIO, on fait Menu STAT / DIST / BINM / Bpd ou Bcd ; sur un modèle TI (à partir de la TI82 Stats), la procédure est 2nde / Distrib (ou DISTR) / binomFdp ou binomPRép (ou binompdf ou binomcdf).



Images d'écrans CASIO

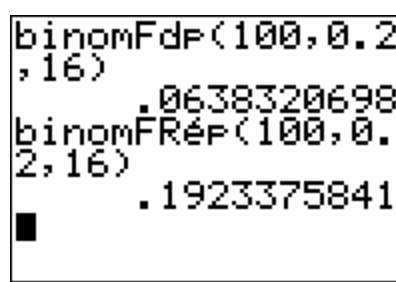
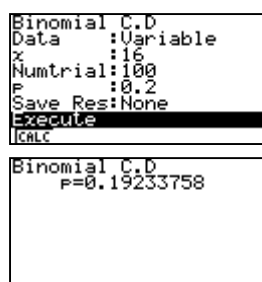
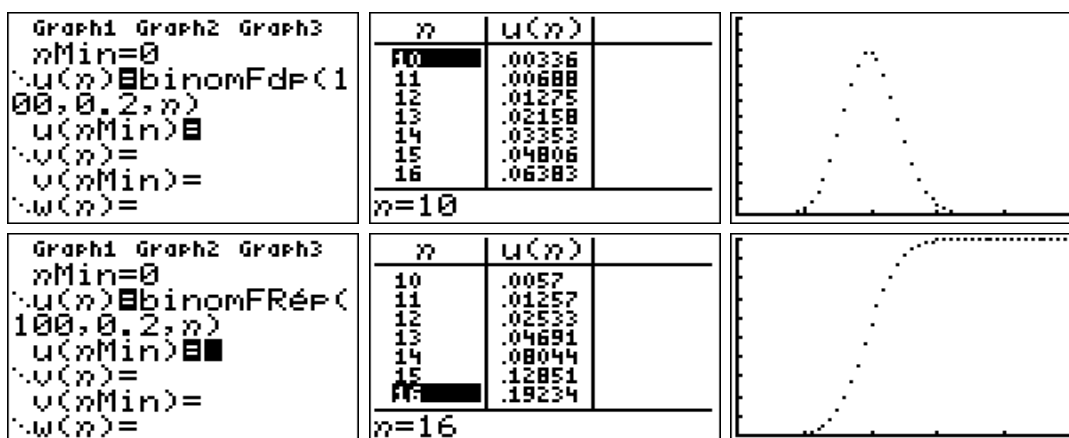


Image d'écran TI

Tabulation et illustration sur calculatrice

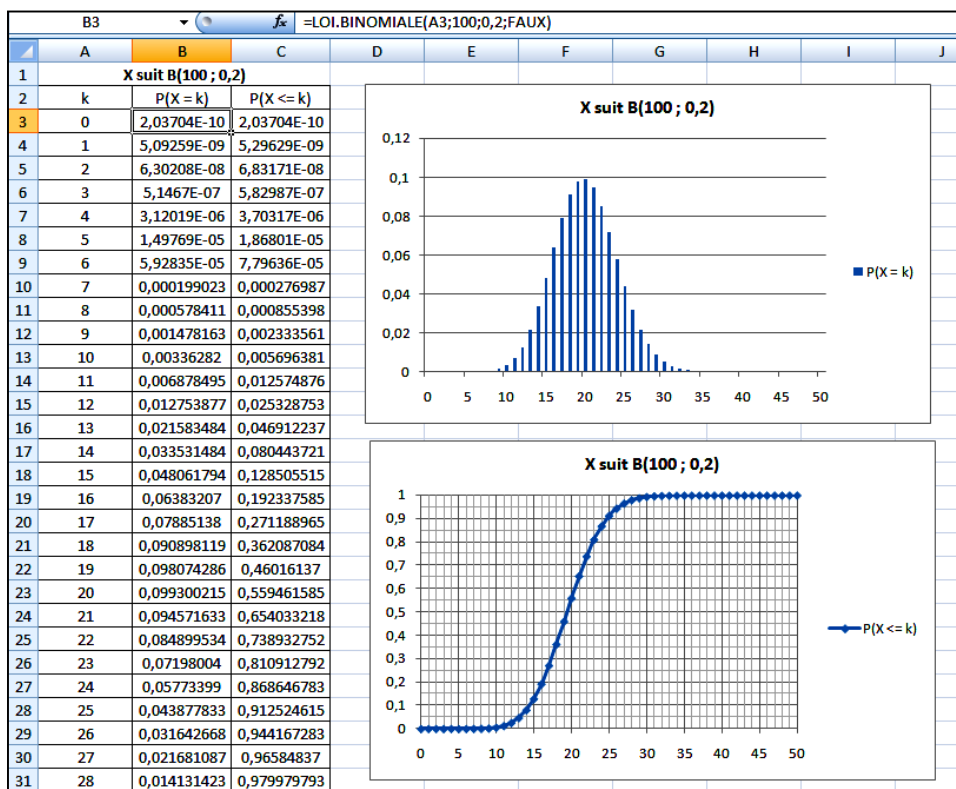
Sur calculatrice, en utilisant le mode « suite », on peut aisément dresser des tables de probabilités ou de probabilités cumulées (croissantes), et les illustrer graphiquement.



Captures d'écrans de calculatrice (modèle T.I.) : tabulation et illustration de la loi binomiale $B(100; 0,2)$, probabilités et probabilités cumulées

Tabulation et illustration sur tableur

Le tableur est l'outil idéal pour tabuler une loi binomiale et obtenir des représentations graphiques de qualité. Pour réaliser les tables ci-dessous, il suffit d'entrer en B3 la formule =LOI.BINOMIALE(A3;100;0,2;FAUX) , en C3 la formule =LOI.BINOMIALE(A3;100;0,2;VRAI) , puis de recopier vers le bas.



Remarque :

Le diagramme des probabilités cumulées est à rapprocher de celui des fréquences cumulées, utilisé en seconde, mais ne doit pas être confondu avec la représentation de la fonction de répartition de la loi binomiale. Cette dernière est une fonction en escalier, qui n'est pas au programme.

III – Fluctuations d'échantillonnage et prise de décision

1 – Position du problème et définition de l'intervalle de fluctuation à 95 %

Position du problème

On considère une population dans laquelle on suppose que la proportion d'un certain caractère est p . Dans les situations de prise de décision étudiées en classe de première, on a une idée de la valeur p_0 que pourrait avoir p , de sorte que l'on formule l'hypothèse : $p = p_0$.

Pour juger de cette hypothèse, on prélève, au hasard et avec remise, un échantillon de taille n sur lequel on observe une fréquence f du caractère.

Hypothèse :
proportion $p = p_0$?

Taille n
Observation :
fréquence f

On rejette l'hypothèse $p = p_0$ lorsque la fréquence f observée est trop éloignée de p_0 , dans un sens ou dans l'autre (pour être dans la continuité du programme de seconde, on se situe, au moins dans un premier temps, dans des cas bilatéraux). On choisit de fixer le seuil de décision de sorte que la probabilité de rejeter l'hypothèse, alors qu'elle est vraie, soit inférieure à 5 %.

Un premier exemple

Un médecin de santé publique veut savoir si, dans sa région, le pourcentage d'habitants atteints d'hypertension artérielle est égal à la valeur de 16 % récemment publiée pour des populations semblables. En notant p la proportion d'hypertendus dans la population de sa région, le médecin formule l'hypothèse $p = 0,16$. Pour vérifier cette hypothèse, le médecin constitue un échantillon de $n = 100$ habitants de la région, dont il détermine la fréquence f d'hypertendus (l'échantillon est prélevé au hasard et la population est suffisamment importante pour considérer qu'il s'agit de tirages avec remise).

Lorsque la proportion dans la population vaut $p = 0,16$, la variable aléatoire X correspondant au nombre d'hypertendus observé dans un échantillon aléatoire de taille $n = 100$, suit la loi binomiale de paramètres $n = 100$ et $p = 0,16$. On cherche à partager l'intervalle $[0, 100]$, où X prend ses valeurs, en trois intervalles $[0, a - 1]$, $[a, b]$ et $[b + 1, 100]$ de sorte que X prenne ses valeurs dans chacun des intervalles extrêmes avec une probabilité proche de 0,025, sans dépasser cette valeur.

En tabulant les probabilités cumulées $P(X \leq k)$, pour k allant de 0 à 100, il suffit de déterminer le plus petit entier a tel que $P(X \leq a) > 0,025$ et le plus petit entier b tel que $P(X \leq b) \geq 0,975$.

On lit $a = 9$ et $b = 23$.

La règle de décision est la suivante : si la fréquence observée f appartient à l'intervalle de

fluctuation à (au moins) 95 % $\left[\frac{a}{n}, \frac{b}{n} \right] = [0,09 ; 0,23]$, on considère que l'hypothèse selon

laquelle la proportion d'hypertendus dans la population est $p = 0,16$ n'est pas remise en question et on l'accepte ; sinon, on rejette l'hypothèse selon laquelle cette proportion vaut $p = 0,16$.

k	$P(X \leq k) \approx$
8	0,0147
9	0,0316
10	0,0607
...	...
22	0,9572
23	0,9754
24	0,9865

Définition de l'intervalle de fluctuation à 95 % déterminé dans une situation bilatérale à l'aide de la loi binomiale

Définition : l'intervalle de fluctuation à 95 % d'une fréquence correspondant à la réalisation, sur un échantillon aléatoire de taille n , d'une variable aléatoire X de loi binomiale, est l'intervalle $\left[\frac{a}{n}, \frac{b}{n}\right]$ défini par :

- a est le plus petit entier tel que $P(X \leq a) > 0,025$;
- b est le plus petit entier tel que $P(X \leq b) \geq 0,975$.

2 – Détermination de l'intervalle de fluctuation à 95 % à l'aide d'un algorithme

Le calcul, à l'aide de la loi binomiale, de l'intervalle de fluctuation à 95 %, $\left[\frac{a}{n}, \frac{b}{n}\right]$, de la fréquence des échantillons aléatoires de taille n , correspondant à la zone d'acceptation d'une hypothèse sur une proportion dans le cadre bilatéral, peut faire l'objet d'une recherche d'algorithme.

On peut, par exemple, procéder sur un tableur comme le montre l'image d'écran suivante.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Intervalle de fluctuation à 95 % d'une fréquence donné par la loi binomiale												
2	n = taille de l'échantillon		p = proportion supposée dans la population										
3	n= 50		p= 0,35		Intervalle classe de seconde :				p-1/rac(n)= 0,209		p+1/rac(n)= 0,491		
4													
5	k	P(X<=k)	recherche a	recherche b		Intervalle de fluctuation à environ 95 % (selon la loi binomiale):							
6	0	4,4225E-10										0,22	0,48
7	1	1,2349E-08											
8	2	1,69426E-07											
9	3	1,52271E-06											
10	4	1,00848E-05											
11	5	5,25002E-05											
12	6	0,000223793											
13	7	0,000803553											
14	8	0,002481512											
15	9	0,006697923											
16	10	0,016006462											
17	11	0,03423297	0,22										
18	12	0,066129359	0,24										
19	13	0,116333143	0,26										
20	14	0,18777699	0,28										
21	15	0,280104423	0,3										
22	16	0,388855486	0,32										
23	17	0,505972015	0,34										
24	18	0,621587051	0,36										
25	19	0,726436313	0,38										
26	20	0,813945121	0,4										
27	21	0,881259589	0,42										
28	22	0,929038739	0,44										
29	23	0,96035885	0,46										
30	24	0,97933161	0,48	0,48									
31	25	0,989956356	0,5	0,5									
32	26	0,995457334	0,52	0,52									
33	27	0,99809028	0,54	0,54									
34	28	0,999254853	0,56	0,56									

On a entré en B6 la formule =SI(A6<=B\$3;LOI.BINOMIALE(A6;B\$3;D\$3;VRAI);"") pour tabuler les probabilités $P(X \leq k)$ lorsque X suit la loi binomiale de paramètres n et p entrés en cellules B3 et D3.

On a entré en C6 la formule =SI(B6>0,025;A6/B\$3;"") pour afficher les valeurs de k telles que $P(X \leq k)$ dépasse 0,025.

On a entré en D6 la formule =SI(B6>=0,975;A6/B\$3;"") ;"") pour afficher les valeurs de k telles que $P(X \leq k)$ dépasse 0,975.

Ces trois formules ont été recopiées vers le bas jusqu'à la ligne 1 006 (l'algorithme fonctionne pour une valeur maximale de n égale à 1 000).

L'intervalle de fluctuation à 95 % est affiché en cellules M5 et N5 contenant les formules =MIN(C6:C1006) et =MIN(D6:D1006).

3 – Quelques exemples

Politique dans un pays lointain

Monsieur Z, chef du gouvernement d'un pays lointain, affirme que 52 % des électeurs lui font confiance. On interroge 100 électeurs au hasard (la population est suffisamment grande pour considérer qu'il s'agit de tirages avec remise) et on souhaite savoir à partir de quelles fréquences, au seuil de 5 %, on peut mettre en doute le pourcentage annoncé par Monsieur Z, dans un sens, ou dans l'autre.

1. On fait l'hypothèse que Monsieur Z dit vrai et que la proportion des électeurs qui lui font confiance dans la population est $p = 0,52$. Montrer que la variable aléatoire X , correspondant au nombre d'électeurs lui faisant confiance dans un échantillon de 100 électeurs, suit la loi binomiale de paramètres $n = 100$ et $p = 0,52$.

2. On donne ci-contre un extrait de la table des probabilités cumulées $P(X \leq k)$ où X suit la loi binomiale de paramètres $n = 100$ et $p = 0,52$.

k	$P(X \leq k) \approx$
40	0,0106
41	0,0177
42	0,0286
43	0,0444
...	...
61	0,9719
62	0,9827
63	0,9897
64	0,9941

a. Déterminer a et b tels que :

- a est le plus petit entier tel que $P(X \leq a) > 0,025$;
- b est le plus petit entier tel que $P(X \leq b) \geq 0,975$.

b. Comparer l'intervalle de fluctuation à 95 %, $\left[\frac{a}{n}, \frac{b}{n}\right]$, ainsi obtenu grâce à la loi

binomiale, avec l'intervalle $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]$.

3. Énoncer la règle décision permettant de rejeter ou non l'hypothèse $p = 0,52$, selon la valeur de la fréquence f des électeurs favorables à Monsieur Z obtenue sur l'échantillon.

4. Sur les 100 électeurs interrogés au hasard, 43 déclarent avoir confiance en Monsieur Z. Peut-on considérer, au seuil de 5 %, l'affirmation de Monsieur Z comme exacte ?

Éléments de réponse

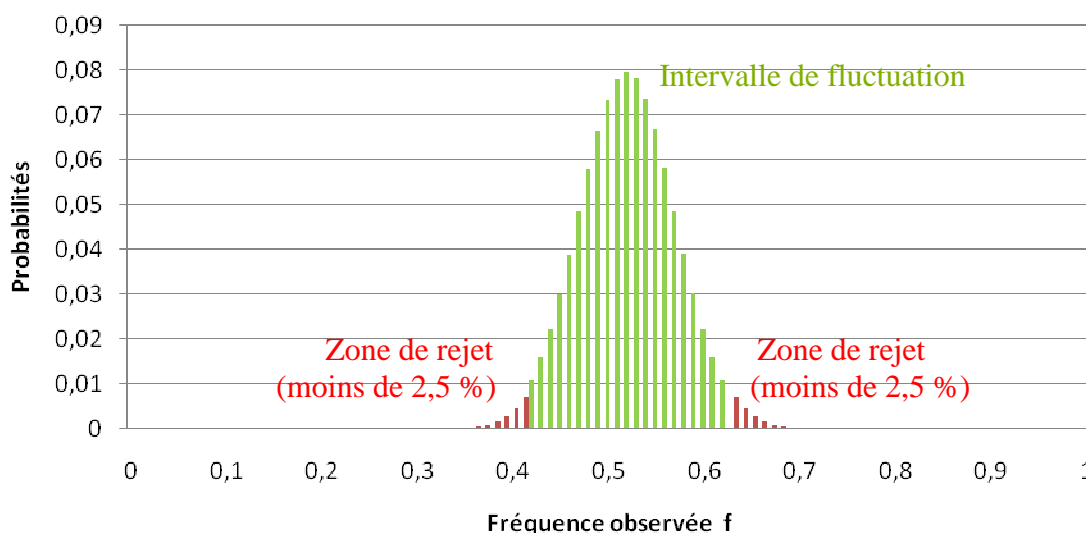
2. a. On lit $a = 42$ et $b = 62$.

b. Les intervalles sont identiques.

3. Si f appartient à l'intervalle $[0,42 ; 0,62]$, l'hypothèse $p = 0,52$ est acceptable, sinon, l'hypothèse $p = 0,52$ est rejetée, au seuil de 5 %.

4. On considère que l'affirmation de Monsieur Z est exacte.

Remarque : la recherche de l'intervalle de fluctuation peut-être illustrée par le diagramme en bâtons de la loi binomiale de paramètres $n = 100$ et $p = 0,52$.



Discrimination

En Novembre 1976 dans un comté du sud du Texas, Rodrigo Partida est condamné à huit ans de prison. Il attaque ce jugement au motif que la désignation des jurés de ce comté est, selon lui, discriminante à l'égard des Américains d'origine mexicaine. Alors que 80 % de la population du comté est d'origine mexicaine, sur les 870 personnes convoquées pour être jurés lors des années précédentes, il n'y a eu que 339 personnes d'origine mexicaine. Devant la Cour Suprême, un expert statisticien produit des arguments pour convaincre du bien fondé de la requête de l'accusé. En vous situant dans le rôle de cet expert, utilisez la loi binomiale pour montrer que la sous-représentation des américains d'origine mexicaine dans les jurys de ce comté est « significative ».

Éléments de réponse

On suppose que les 870 jurés sont tirés au sort dans la population du comté (la population étant très importante, on peut considérer qu'il s'agit de tirages avec remise). Sous cette hypothèse, la variable aléatoire X correspondant au nombre de jurés d'origine mexicaine suit la loi binomiale de paramètres $n = 870$ et $p = 0,8$.

On peut alors rechercher l'intervalle de fluctuation à 95 % correspondant.

Une tabulation de la loi binomiale de paramètres $n = 870$ et $p = 0,8$ fournit les résultats suivants :

k	$P(X \leq k)$	fréquence k / n
672	0,0245	0,772
673	0,0296	0,774
...
718	0,9733	0,825
719	0,9783	0,826

L'intervalle de fluctuation à 95 % de la fréquence des jurés d'origine mexicaine est : $[0,774 ; 0,826]$.

La fréquence observée est $f = \frac{339}{870} \approx 0,39$. Cette valeur ne se situe pas dans l'intervalle de fluctuation. La différence est significative et l'hypothèse selon laquelle les jurés sont tirés au hasard est rejetée.

De fait, l'accusé a obtenu gain de cause et a été rejugé par un autre jury.

Sécurité au carrefour

Un groupe de citoyens demande à la municipalité d'une ville la modification d'un carrefour en affirmant que 40 % des automobilistes tournent en utilisant une mauvaise file. Un officier de police constate que sur 500 voitures prises au hasard, 190 prennent une mauvaise file.

1. Déterminer, en utilisant la loi binomiale sous l'hypothèse $p = 0,4$, l'intervalle de fluctuation à 95 %.

2. D'après l'échantillon, peut-on considérer, au seuil de 5 %, comme exacte l'affirmation du groupe de citoyens ?

Éléments de réponse

1. $[0,358 ; 0,444]$.

2. $f = 0,38$. L'affirmation est considérée comme exacte.

Carte de contrôle

Dans l'industrie automobile, certains véhicules, après leur passage en peinture, présentent un défaut de type « grains ponctuels ». Ce défaut est pratiquement imperceptible, mais constitue un témoin de la qualité du processus de peinture.

Lorsque le processus est « sous contrôle », 20 % des capots produits ont ce type de défaut. Des modifications apportées au processus de fabrication sont susceptibles de modifier ce pourcentage, dans un sens ou dans l'autre.

On contrôle la production en prélevant des échantillons de $n = 50$ capots. La production est suffisamment importante pour considérer qu'il s'agit de tirages au hasard avec remise.

On choisit de fixer le seuil de décision de sorte que la probabilité de rejeter l'hypothèse à tort soit inférieure à 5 %.

On accepte l'hypothèse selon laquelle la proportion dans la production est $p = 0,2$, lorsque la fréquence f observée sur l'échantillon se situe dans l'intervalle de fluctuation à plus de 95 % correspondant au complémentaire de la zone de rejet.

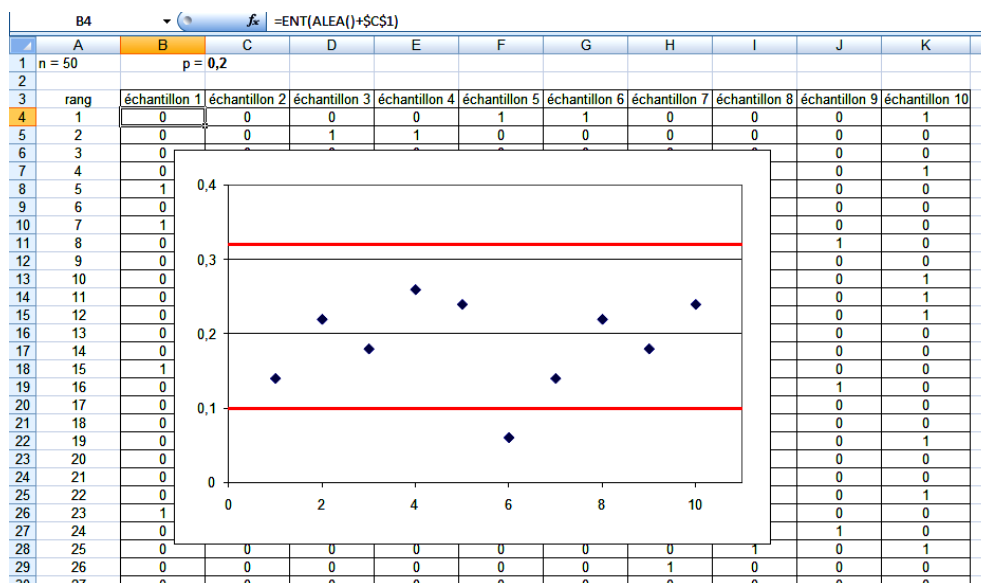
Les « limites de contrôle » sont les bornes de l'intervalle de fluctuation à 95 %, calculées en considérant la variable aléatoire X correspondant au nombre de capots présentant le défaut sur un échantillon de taille 50. Sous l'hypothèse $p = 0,2$, cette variable aléatoire suit la loi binomiale de paramètres $n = 50$ et $p = 0,2$.

1. Calculer les limites de contrôle.
2. Simuler le fonctionnement de cette carte de contrôle lorsque le processus est sous contrôle.
3. Le processus est sous contrôle, quelle est la probabilité de commettre une erreur de décision à partir d'un échantillon ?
- 4*². La proportion de capots présentant le défaut dans la production est $p = 0,3$. Quelle est la probabilité de commettre une erreur de décision à partir d'un échantillon ?

Éléments de réponse

1. Les limites de contrôle sont 0,1 et 0,32 correspondant à des effectifs de 5 et 16 capots présentant le défaut.

2.



² Cette question, de réflexion, n'est pas un attendu du programme.

3. Si X suit la loi binomiale de paramètres $n = 50$ et $p = 0,2$, on a $P(X \leq 4) \approx 0,0185$ et $P(X \geq 17) \approx 0,0144$. La probabilité de commettre une erreur de décision lorsque $p = 0,2$ correspond à la probabilité de la zone de rejet soit environ 3,3 %.

4. La probabilité d'erreur, avec un seul échantillon, est $P(5 \leq X \leq 16)$ où X suit la loi binomiale de paramètres $n = 50$ et $p = 0,3$. On trouve environ 68,4 %.

Remarque :

Il est difficile de distinguer $p = 0,2$ et $p = 0,3$ sur un seul échantillon de taille 50. La procédure de décision a tendance à être « conservatrice » et privilégie l'hypothèse $p = 0,2$ qui n'est rejetée que si la différence observée est réellement significative.

Gauchers

Dans le monde, la proportion de gauchers est 12 %.

Soit n le nombre d'élèves dans votre classe.

1. Déterminer, à l'aide de la loi binomiale, l'intervalle de fluctuation à 95 % de la fréquence des gauchers sur un échantillon aléatoire de taille n .

2. Votre classe est-elle « représentative » de la proportion de gauchers dans le monde ?

Éléments de réponse

1. En prenant $n = 30$, l'intervalle de fluctuation à 95 % est $[0,033 ; 0,233]$ (entre 1 et 7 gauchers).

En prenant $n = 25$, l'intervalle de fluctuation à 95 % est $[0 ; 0,24]$ (entre 0 et 6 gauchers).

4 – Comparaison de l'intervalle de fluctuation de première avec l'intervalle de fluctuation exploité en classe de seconde

On considère une population où la fréquence d'un caractère est p , dans laquelle on prélève, au hasard et avec remise, un échantillon de taille n .

La variable aléatoire X correspondant au nombre d'observations du caractère sur un échantillon suit la loi binomiale de paramètres n et p .

On peut calculer l'intervalle $\left[\frac{a}{n}, \frac{b}{n} \right]$ de fluctuation à 95 % de la fréquence observée sur un

échantillon de taille n , où a est le plus petit entier tel que $P(X \leq a) > 0,025$ et b est le plus petit entier tel que $P(X \leq b) \geq 0,975$.

Le programme des classes de premières S, ES et STI2D-STL, demande de comparer, pour une taille de l'échantillon importante, cet intervalle avec l'intervalle de fluctuation

$\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}} \right]$ exploité en classe de seconde. Cet intervalle, dont l'expression est plus simple, résulte d'une double approximation.

D'une part, le théorème de de Moivre - Laplace permet d'approcher la loi de la variable aléatoire $F = \frac{1}{n}X$, correspondant à la fréquence observée sur un échantillon de taille n , par

la loi normale de moyenne p et d'écart type $\sqrt{\frac{p(1-p)}{n}}$. En pratique³, cette approximation

est possible dès que $n \geq 30$, $n \times p \geq 5$ et $n \times (1-p) \geq 5$. Les propriétés de la loi normale permettent alors d'affirmer que la variable aléatoire F prend ses valeurs dans l'intervalle

centré sur p et de rayon $1,96 \times \sqrt{\frac{p(1-p)}{n}}$ avec une probabilité égale à 0,95.

³ Il s'agit de critères empiriques dont la pertinence dépend du domaine d'application.

D'autre part, la majoration du polynôme $p(1-p)$ par $\frac{1}{4}$ et de la valeur 1,96 par 2, permet de majorer le rayon de l'intervalle par $\frac{1}{\sqrt{n}}$. Cette majoration augmente la longueur de

l'intervalle et donc la probabilité pour la variable aléatoire F d'y prendre ses valeurs. Dans la pratique, la majoration du polynôme $p(1-p)$ est raisonnable lorsque p est compris entre 0,2 et 0,8.

Il s'agit, en classe de première, d'utiliser l'algorithme de calcul, à l'aide de la loi binomiale, de l'intervalle de fluctuation à 95 % $\left[\frac{a}{n}, \frac{b}{n}\right]$, pour comparer ses bornes avec

celles de l'intervalle $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]$.

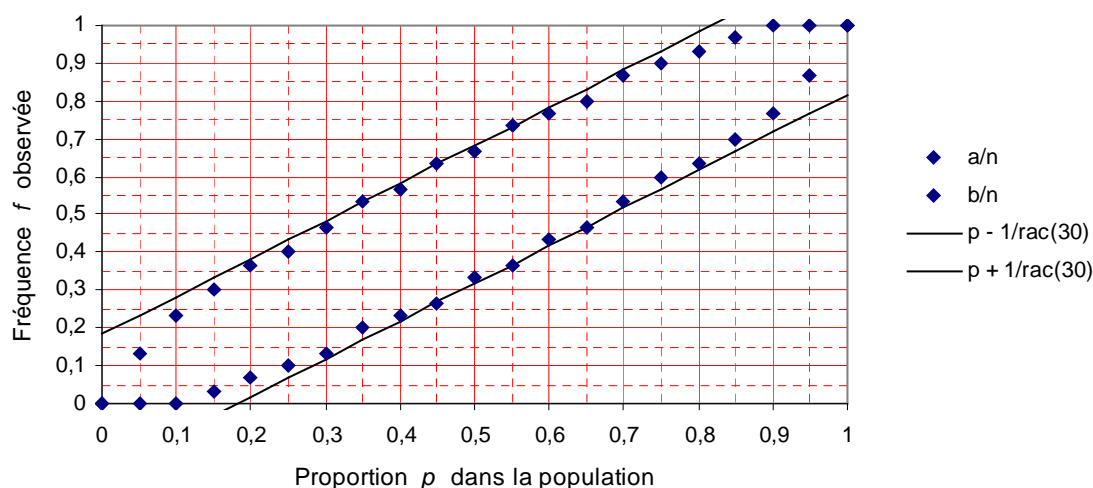
Étude pour des échantillons de taille $n = 30$

Pour $n = 30$, on observe que l'intervalle de fluctuation $\left[\frac{a}{n}, \frac{b}{n}\right]$ est sensiblement le même

que l'intervalle $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]$ lorsque p est compris entre 0,25 et 0,75.

$n =$	30																				
p	0	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95	1
a/n	0	0	0	0,03	0,07	0,1	0,13	0,2	0,23	0,27	0,33	0,37	0,43	0,47	0,53	0,6	0,63	0,7	0,77	0,87	1
b/n	0	0,13	0,23	0,3	0,37	0,4	0,47	0,53	0,57	0,63	0,67	0,73	0,77	0,8	0,87	0,9	0,93	0,97	1	1	1
$p - 1/\text{rac}(30)$	-0,18	-0,13	-0,08	-0,03	0,02	0,07	0,12	0,17	0,22	0,27	0,32	0,37	0,42	0,47	0,52	0,57	0,62	0,67	0,72	0,77	0,82
$p + 1/\text{rac}(30)$	0,18	0,23	0,28	0,33	0,38	0,43	0,48	0,53	0,58	0,63	0,68	0,73	0,78	0,83	0,88	0,93	0,98	1,03	1,08	1,13	1,18

Intervalle de fluctuation à 95 % pour des échantillons de taille 30



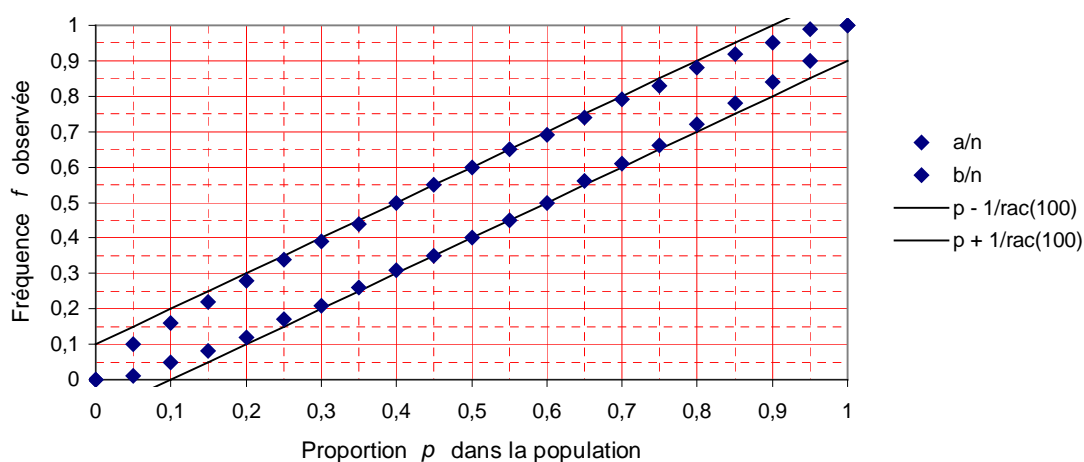
Étude pour des échantillons de taille $n = 100$

Pour $n = 100$, on observe que l'intervalle de fluctuation $\left[\frac{a}{n}, \frac{b}{n}\right]$ est sensiblement le même

que l'intervalle $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]$ lorsque p est compris entre 0,15 et 0,85.

$n = 100$

p	0	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95	1
a/n	0	0,01	0,05	0,08	0,12	0,17	0,21	0,26	0,31	0,35	0,4	0,45	0,5	0,56	0,61	0,66	0,72	0,78	0,84	0,9	1
b/n	0	0,1	0,16	0,22	0,28	0,34	0,39	0,44	0,5	0,55	0,6	0,65	0,69	0,74	0,79	0,83	0,88	0,92	0,95	0,99	1
$p - 1/\text{rac}(100)$	-0,1	-0,05	0	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9
$p + 1/\text{rac}(100)$	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95	1	1,05	1,1

Intervalle de fluctuation à 95 % pour des échantillons de taille 100**Étude pour des échantillons de taille $n = 1\,000$**

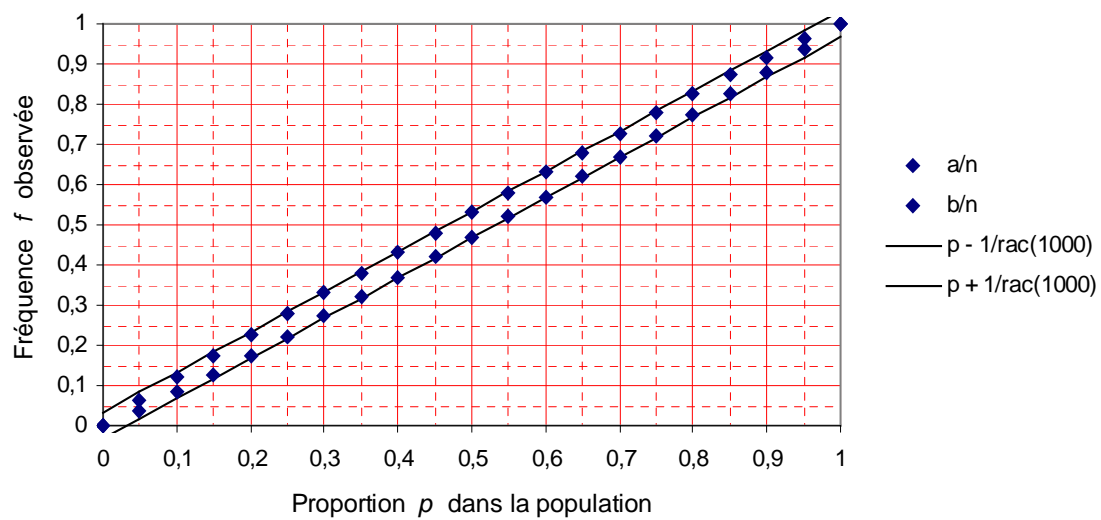
Pour $n = 1\,000$, on observe que l'intervalle de fluctuation $\left[\frac{a}{n}, \frac{b}{n}\right]$ est sensiblement le même

même que l'intervalle $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]$ lorsque p est compris entre 0,05 et 0,95.

$n = 1000$

p	0	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95	1
a/n	0	0,04	0,08	0,13	0,18	0,22	0,27	0,32	0,37	0,42	0,47	0,52	0,57	0,62	0,67	0,72	0,78	0,83	0,88	0,94	1
b/n	0	0,06	0,12	0,17	0,23	0,28	0,33	0,38	0,43	0,48	0,53	0,58	0,63	0,68	0,73	0,78	0,82	0,87	0,92	0,96	1
$p - 1/\text{rac}(1000)$	-0,03	0,02	0,07	0,12	0,17	0,22	0,27	0,32	0,37	0,42	0,47	0,52	0,57	0,62	0,67	0,72	0,77	0,82	0,87	0,92	0,97
$p + 1/\text{rac}(1000)$	0,03	0,08	0,13	0,18	0,23	0,28	0,33	0,38	0,43	0,48	0,53	0,58	0,63	0,68	0,73	0,78	0,83	0,88	0,93	0,98	1,03

Intervalle de fluctuation à 95 % pour des échantillons de taille 1 000



5 – Approfondissement : étude d'une situation unilatérale

L'activité suivante est inspirée⁴ de présentations pédagogiques effectuées, pour des physiciens, par Alain VIVIER, responsable pédagogique à l'INSTN⁵, institut dépendant du Commissariat à l'énergie atomique à Saclay. Convaincu de l'intérêt pédagogique des expérimentations sur tableur, Alain VIVIER déclare : « pour ma part je n'aurais jamais approfondi ces aspects [de statistique et probabilités], indispensables en physique, sans cet outil [le tableur]. Cela m'a été utile non seulement pour des aspects enseignements, mais aussi de recherche (notamment dans le domaine de la problématique du seuil de décision, qui peut s'avérer parfois difficile) ».

Une activité sur le thème de la radioactivité, consistant à rechercher un seuil de différence significative à l'aide de la loi binomiale⁶, peut se présenter comme suit.

On mesure en laboratoire, avec un compteur Geiger, un objet pouvant être « radioactif ». Le compteur est réglé selon une certaine sensibilité et on effectue une mesure à un mètre de l'objet, pendant dix secondes. L'instrument compte 37 désintégrations ou « coups ». Cependant, avec ce réglage et dans ces conditions, une mesure de « bruit de fond » (correspondant à l'environnement du laboratoire) donne en moyenne un comptage de 30 coups. La question qui se pose est de savoir si la différence observée est assez importante pour considérer l'objet comme « radioactif ».

On suppose que dans le laboratoire, durant chaque centième de seconde, le compteur est aléatoirement susceptible de compter un coup de bruit de fond avec une probabilité 0,03, ce que l'on simule à l'aide d'un tableur avec l'instruction =ENT(ALEA()+0,03).

1. Simuler en colonne A un comptage de bruit de fond pendant 10 secondes, puis recopier vers la droite pour obtenir la simulation de 100 comptages.
2. Calculer la moyenne des 100 comptages simulés. Est-elle proche de 30 coups ? (Faire F9 pour obtenir d'autres simulations.)
3. Un comptage supérieur ou égal à 37 coups vous semble-t-il exceptionnel ?
4. Déterminer les paramètres n et p de la loi binomiale suivie par la variable aléatoire X modélisant un comptage de bruit de fond.
5. Sur une nouvelle feuille, calculer une table fournissant $P(X \leq k)$ pour k allant de 0 à 1 000.
6. On considérera qu'il y a radioactivité (comptage « significatif ») à partir d'un comptage de $N + 1$ coups, où N est le plus petit entier tel que : $P(X \leq N) > 0,95$.
 - a. Déterminer la valeur de N .
 - b. Contrôler la valeur de N obtenue en vérifiant qu'environ 95 % des comptages simulés (en moyenne) sont inférieurs ou égaux à N (utiliser la fonction NB.SI et faire plusieurs fois F9).
 - c. Un comptage de 37 coups est-il « significatif » ? Quel est le comptage minimal à partir duquel on considère qu'il y a une radioactivité significative ?
 - d. Quelle est la probabilité de commettre l'erreur suivante : on considère que le comptage est « significatif » alors que c'est un bruit de fond (il n'y a pas de radioactivité) ?

⁴ Avec l'aimable autorisation de Monsieur Alain Vivier.

⁵ Institut national des sciences et techniques nucléaires.

⁶ Le comptage du nombre de désintégrations durant un intervalle de temps donné suit, a priori, une loi de Poisson de paramètre λ . Cependant, la loi binomiale de paramètres n et p (beaucoup plus facile à simuler) est, sous certaines conditions, proche de la loi de Poisson de paramètre $\lambda = n \times p$. C'est le cas ici.