Index

L'exemple est celui de l'activité 1 page 226 du livre

I- Activité du livre: La loi du nombre de succès

A/ Une expérience aléatoire élémentaire.

Vers la définition d'une épreuve de Bernoulli

(Jacob Bernoulli : (1654-1705))

$\mathrm{P}(S) = \frac{1}{3}$, d'où, $\mathrm{P}(E) = 1 - \frac{1}{3} = \frac{2}{3}$

Bilan : On considère une épreuve à deux issues (S et $\overline{S}$) ou (S et E) (*Comme Succès et Échec*)

On pose $\mathrm{P}(S) = p$ et, par conséquent $\mathrm{P}(\overline{S}) = 1 - p = q$

Un telle épreuve à deux issues S et $\overline{S}$ est **une épreuve de Bernoulli** de paramètre p.

Remarques :

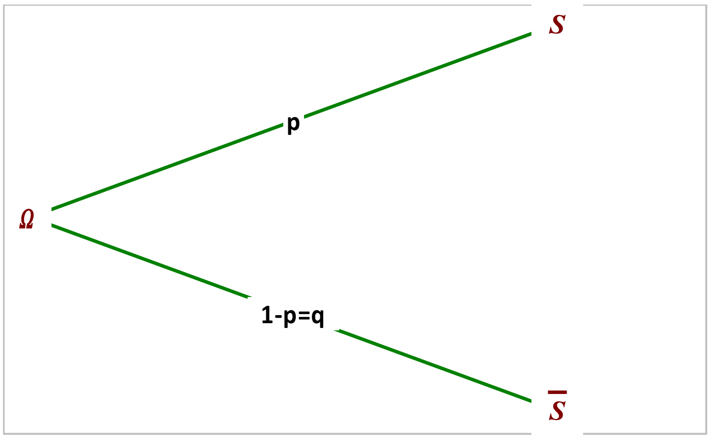
Ne pas confondre pour les calculs à venir : les événements (ensembles), les probabilités (nombres), et, lors de répétition d'une épreuve l'étape à laquelle on réalise l'événement.

Par exemple, on jette le dé 1000 fois de suite et on s'intéresse à l'obtention du n°6.

S : " Obtenir le n°6 " est un événement lors d'un lancer.

S_2 : " obtenir le n°6 au 2ème lancer " et S_{500} : " obtenir le n°6 au 500ème lancer " sont des événements distincts.

Si le dé est équilibré, on a : $P(S)= P(S_2) = P(S_{500}) = \frac{1}{6}$



On définit la variable aléatoire X prenant la valeur 1 si S est réalisé et prenant la valeur 0 sinon.

X est une variable de Bernoulli qui suit la loi de probabilité de Bernoulli de paramètre p.

k	**0**	**1**
$P(X = k)$	$1 - p$	p

L'espérance mathématique de X est $E(X) = p$.

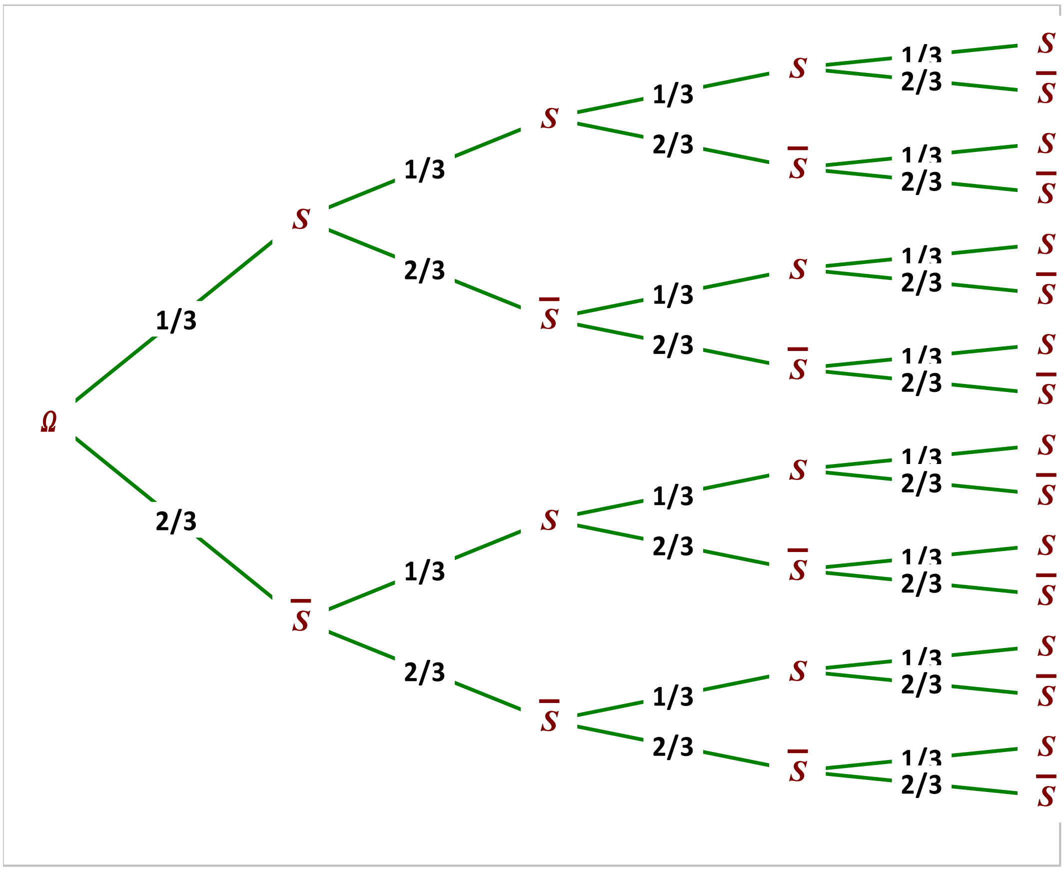
La variance de X est : $V(X) = p(1 - p) = pq$

L'écart-type de X est $\sigma(X) = \sqrt{pq}$

B/ Quatre répétitions

Un schéma de Bernoulli.

1a)



b) L'issue $S\overline{S}S\overline{S}$ a pour probabilité $\left(\frac{1}{3}\right)^2\times\left(\frac{2}{3}\right)^2$

L'issue $\overline{S}SS\overline{S}$ a pour probabilité $\left(\frac{1}{3}\right)^2\times\left(\frac{2}{3}\right)^2$

c) P(" Obtenir deux succès ") = P($SS\overline{S}\overline{S}$) + P($S\overline{S}S\overline{S}$) + P($S\overline{S}\overline{S}S$) + P($\overline{S}SS\overline{S}$) + P($\overline{S}S\overline{S}S$) + P($\overline{S}\overline{S}SS$)

$$= 6\times\left(\frac{1}{3}\right)^2\times\left(\frac{2}{3}\right)^2$$

2)

k	0	1	2	3	4	***Total***
P($X = k$)	$\left(\frac{2}{3}\right)^4$	$4\times\left(\frac{1}{3}\right)\times\left(\frac{2}{3}\right)^3$	$6\times\left(\frac{1}{3}\right)^2\times\left(\frac{2}{3}\right)^2$	$4\times\left(\frac{1}{3}\right)^3\times\left(\frac{2}{3}\right)$	$\left(\frac{1}{3}\right)^4$	1

Bilan : On appelle **schéma de Bernoulli** une expérience aléatoire où on répète n épreuves de Bernoulli de paramètre p **identiques et indépendantes.**

(C'est le fait d'avoir des épreuves ***identiques et indépendantes*** *qui permet de ne pas avoir besoin d'indicer les étapes …, ce serait nécessaire sinon.*

par exemple, on suppose une personne qui tire 10 fois de suite aux fléchettes :

" tirer une fléchette dans la cible " est l'événement S.

On suppose que l'expérience permet un apprentissage et que la réussite est meilleure après les premiers tirs …

Il y répétition d'épreuves identiques mais elles ne sont pas indépendantes... Le modèle décrit ne s'applique pas)

On peut associer un arbre où un chemin est une liste de n éléments S et $\overline{S}$.

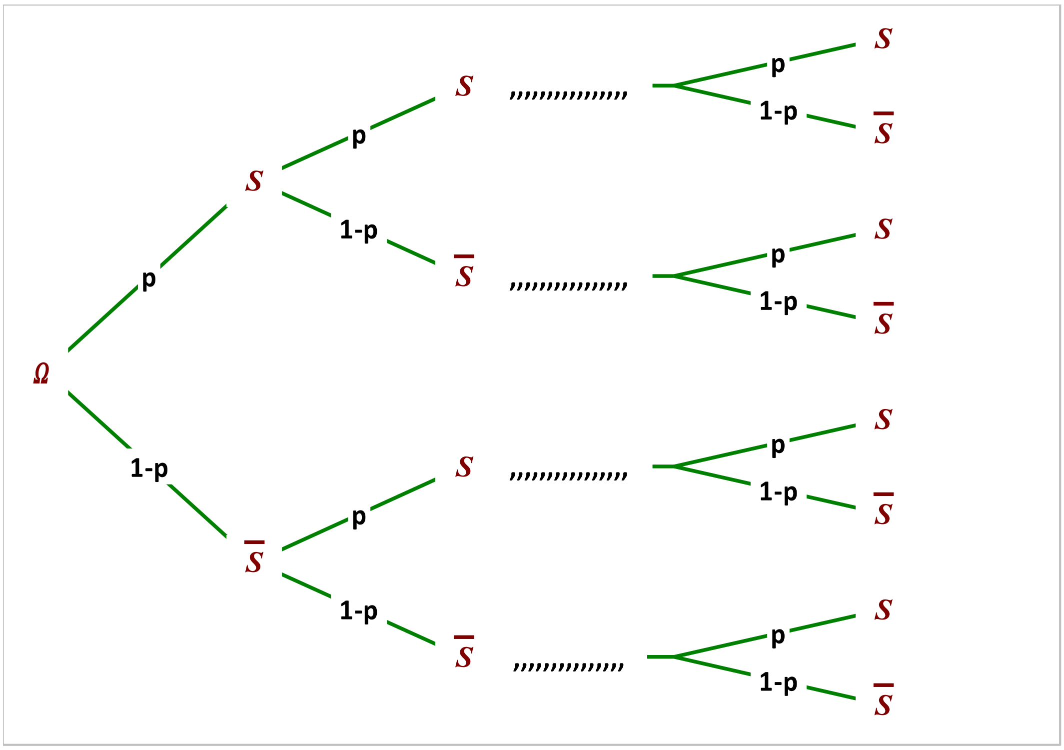
Un chemin ayant k succès a donc $n – k$ échecs. $\left(\overbrace{S\ldots S}^{k}\ \overbrace{\overline{S}\ldots\overline{S}}^{n-k}\right)$

On considère la **variable aléatoire *X*** prenant pour valeurs **le nombre de succès** lors de ces n répétitions.

X prend par conséquent $n + 1$ valeurs entières de 0 à n .

La probabilité d'un chemin ayant k succès et $n – k$ échecs est $p^k\times q^{n-k}$ où $q = 1 – p$.

Pour déterminer la loi de probabilité de X, on est amené à compter **le nombre de chemins ayant *k* succès** et $n – k$ échecs.



Notation du nombre de chemins, les coefficients binomiaux. (Combinaisons)

Le **nombre de chemins** donnant k succès et $n-k$ échecs est noté $\binom{n}{k}$ et se lit : k parmi n

Ces nombres $\binom{n}{k}$ sont appelés **coefficients binomiaux**

D'après l'exemple : $\binom{4}{0} = 1$, $\binom{4}{1} = 4$, $\binom{4}{2} = 6$, $\binom{4}{3} = 4$, $\binom{4}{4} = 1$

Pour ceux qui veulent en savoir plus : $\binom{n}{k} = \dfrac{n(n-1)\times\ldots\times(n-k+1)}{k\times\ldots\times 1}$

C/ Cinq répétitions

Une propriété des coefficients binomiaux

1) On a un et un seul chemin pour l'événement $(Y=0)$. $\quad P(Y=0) = \left(\frac{2}{3}\right)^5$

On a un et un seul chemin pour l'événement $(Y=5)$. $\quad p(Y=5) = \left(\frac{1}{3}\right)^5$

2) a) La probabilité d'un chemin ayant deux succès (et donc trois échecs) est $\left(\frac{1}{3}\right)^2\times\left(\frac{2}{3}\right)^3$

b) Si un tel chemin se termine par S, on doit avoir avant un seul Succès lors des quatre premières répétitions.

On a en ce cas : 4 chemins. (Voir § B-2, $\binom{4}{1} = 4$)

Si un tel chemin se termine par $\overline{S}$, on doit avoir avant deux Succès lors des quatre premières répétitions.

On a en ce cas : 6 chemins. (Voir § B-2, $\binom{4}{2} = 6$)

Finalement, on a 4 + 6 = 10 chemins donnant deux succès en cinq répétitions. (c-à-d : $\binom{5}{2} = \binom{4}{2} + \binom{4}{1}$).

c) $P(Y = 2) = 10 \times \left(\frac{1}{3}\right)^2 \times \left(\frac{2}{3}\right)^3$

Bilan : Pour calculer le nombre de chemins $\binom{n+1}{k}$ représentant k succès lors de $n + 1$ répétitions à partir du nombre de chemins lors des n répétitions, on peut distinguer deux cas

- soit on a eu k succès, $n - k$ échecs pour n répétitions et 1 échec supplémentaire,
- soit on a eu $k - 1$ succès, $n - k + 1$ échecs pour n répétitions et 1 succès supplémentaire.

On a donc : $\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}$

Dans l'exemple : $\underbrace{\binom{5}{2}}_{\text{2 succès parmi 5}} = \underbrace{\binom{4}{2}}_{\substack{\text{2 succès parmi 4} \\ \text{1 échec supplémentaire}}} + \underbrace{\binom{4}{1}}_{\substack{\text{1 succès parmi 4} \\ \text{1 succès supplémentaire}}}$

D/ Dix répétitions

1) Z peut prendre toutes les valeurs entières de 0 à 10.

$P(Z = 0) = \left(\frac{2}{3}\right)^{10}$ et $P(Z = 10) = \left(\frac{1}{3}\right)^{10}$

2 a) Une issue contenant 4 succès et 6 échecs a pour probabilité : $\left(\frac{1}{3}\right)^4 \times \left(\frac{2}{3}\right)^6$ quelque soit l'ordre des succès et des échecs.

b) c) il faut connaître le nombre de chemins donnant 4 succès et 6 échecs pour déterminer $P(Z = 4)$.

$P(Z = 4) = \binom{10}{4} \left(\frac{1}{3}\right)^4 \times \left(\frac{2}{3}\right)^6$ où $\binom{10}{4}$ est le nombre de chemins.

d) Comme $\binom{10}{4} = 210$, $P(Z = 4) \approx 0{,}2276$ à 10^{-4} près

Propriété des coefficients binomiaux:

On a n répétitions.

Pour tout $n \in \mathbb{N}$, et, pour tout $k \in \mathbb{N}$ tel que $k \leqslant n$, on a :

$\binom{n}{0} = 1$ (un seul chemin ayant 0 succès) et $\binom{n}{n} = 1$ (un seul chemin ayant n succès)

$\binom{n}{1} = n$ Il y a exactement n chemins avec un seul succès (donc $n - 1$ échecs)

$\binom{n}{n-1} = n$ et exactement n chemins ayant un seul échec (donc $n - 1$ succès)

$\binom{n}{k} = \binom{n}{n-k}$ Il y a autant de chemins ayant k succès que de chemins ayant $n - k$ succès.

$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}$ voir § précédent et § suivant (triangle de Pascal).

Triangle de Pascal.

À l'intersection d'une colonne k et d'une ligne n , on lit le nombre $\binom{n}{k}$.

À la ligne suivante $n + 1$, on a donc la somme des deux nombres de la ligne n au-dessus et à gauche.
Tous les nombres de la colonne 0 sont égaux à 1.
Tous les nombres de la diagonale sont égaux à 1

n \ k	0	1	2	3	4	5	6	7	8
0	1								
1	1	1							
2	1	2	1						
3	1	3	3	1					
4	1	4	6	4	1				
5	1	5	10	10	5	1			
6	1	6	15	20	15	6	1		
7	1	7	21	35	35	21	7	1	
8	1	8	28	56	70	56	28	8	1

$\binom{6}{2} + \binom{6}{3} = \binom{7}{3}$

II- Loi binomiale

Soit un schéma de n épreuves de Bernoulli de paramètre p.

(épreuves à deux issues ***identiques et <u>indépendantes</u>****)*

X est la variable aléatoire qui donne le nombre de succès.

La probabilité d'avoir k succès (k prend les valeurs entières de 0 à n) est :

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

Son espérance est : $E(X) = np$.

Sa variance est : $V(X) = npq$ et son écart-type est : $\sigma(X) = \sqrt{npq}$

On dit que **X suit la loi binomiale de paramètres n et p**, notée $\mathcal{B}(n ; p)$.

k	0	1	...	k	...	$n-1$	n
$P(X=k)$	q^n	npq^{n-1}		$\binom{n}{k} p^k q^{n-k}$		$np^{n-1}q$	p^n

La représentation graphique d'une loi binomiale donne une courbe qui tend vers une courbe en cloche (courbe de Gauss).

Voir en annexe quelques exemples de représentations graphiques

III- Échantillon, simulation, intervalle de fluctuation (Rappel de seconde)

Échantillon

Un **échantillon** de taille n est une liste de n résultats obtenus par n répétitions indépendantes de la même expérience aléatoire permettant d'étudier un caractère.

On cherche à déterminer la fréquence f d'apparition de ce caractère pour déterminer la proportion p de ce caractère dans la population.

Remarques:

On mesure la fréquence f. (La proportion p est en général inconnue).

f et p sont des nombres compris entre 0 et 1.

Simulation

Lorsqu'on ne peut pas étudier effectivement des échantillons, on réalise une simulation.

Il existe des tables de nombres aléatoires ou bien, on utilise une machine.

Les fonctions ALEA ou RAND des tableurs permettent de faire ces simulations.

Exemple:

Pour simuler une proportion de 35 % dans une population, on peut entrer dans une cellule d'un tableur:

=SI(ALEA.ENTRE.BORNES(1;100)<=35;1;0)

La fréquence d'apparition du 1 est de $\frac{35}{100}$ et celle du 0 de $\frac{65}{100}$.

ou encore

=ENT(ALEA()+0,35) retourne 0 dans 65 % des cas et 1 dans 35 % des cas

En effet, ENT désigne la partie entière d'un réel. Si $n \leqslant x < n+1$ avec $n \in \mathbb{Z}$ alors ENT(x) = n.

ALEA() retourne au hasard un nombre x tel que $0 \leqslant x < 1$

ALEA()+0,35 est donc un nombre y compris entre 0,35 et 1,35

Lorsque $0{,}35 \leqslant y < 1$, sa partie entière est 0

Lorsque $1 \leqslant y < 1{,}35$ sa partie entière est 1

La longueur de l'intervalle [0,35 ; 1[est 0,65, celle de [1 ; 1,35[est 0,35

Intervalle de fluctuation

Lorsque $0{,}2 \leqslant p \leqslant 0{,}8$ et $n \geqslant 25$, on montre que dans 95 % des cas au moins, $f \in \left[p-\frac{1}{\sqrt{n}} ; p+\frac{1}{\sqrt{n}}\right]$.

Cet intervalle $I = \left[p-\frac{1}{\sqrt{n}};p+\frac{1}{\sqrt{n}}\right]$ est appelé intervalle de fluctuation au seuil de 95 %.

Remarque:

$f \in \left[p-\frac{1}{\sqrt{n}};p+\frac{1}{\sqrt{n}}\right] \Leftrightarrow p \in \left[f-\frac{1}{\sqrt{n}};f+\frac{1}{\sqrt{n}}\right]$

IV- Loi binomiale et prise de décision.

Le problème :

Dans une population, on veut étudier un caractère présent dans une proportion p (a priori inconnue).

On prélève au hasard un échantillon de taille n dans cette population. (Si la population est très grande, le fait de procéder avec ou sans remise n'a pas d'influence).

On observe alors la fréquence f du caractère (f est un nombre que l'on calcule).

On cherche à savoir pour quelles valeurs de *f*, on va rejeter l'hypothèse : " la proportion du caractère est *p* " au risque d'erreur de 5%.

Comprendre : Les risques de se tromper

Il y a deux façons de se tromper ....

1) On rejette à tort l'hypothèse " la proportion est p " au seuil α.

α est la probabilité de rejeter une hypothèse alors qu'elle est vraie.

2) on accepte à tort l'hypothèse " la proportion est p " au seuil β.

β est la probabilité d'accepter une hypothèse alors qu'elle est fausse.

Seul le premier cas nous concerne en 1S.

C'est-à-dire : La probabilité de rejeter cette hypothèse alors qu'elle est vraie est inférieure à 5%.

On ne prend pas le risque de retenir une hypothèse alors qu'elle est vraie

Intervalle de fluctuation à 95 %

Dans une population, on veut étudier un caractère présent dans une proportion p

Soit un échantillon de taille n.

Soit la variable aléatoire X définie par :

X prend pour valeurs le nombre d'individus de l'échantillon ayant le caractère étudié.
X suit la loi binomiale $\mathscr{B}(n ; p)$ de paramètres n et p.
On cherche alors les entiers a et b dans $[0 ; n]$ tels que :
2,5 % des valeurs de X sont dans $[0 ; a - 1]$ (On ne dépasse pas 2,5 %)
95 % des valeurs de X sont dans $[a ; b]$
2,5 % des valeurs de X sont dans $[b + 1 ; n]$ (On ne dépasse pas 2,5 %)

Définition : Intervalle de fluctuation au seuil de 95%

L'intervalle de fluctuation au seuil de 95% de la fréquence f est l'intervalle $\left[\frac{a}{n};\frac{b}{n}\right]$ tel que :

* a est le plus petit entier tel que $P(X \leqslant a) > 2{,}5$ %

* b est le plus petit entier tel que $P(X \leqslant b) \geqslant 97{,}5$ %

Conséquences directes : (Revoir si besoin l'exercice 46 page 219)

$P(X \leqslant a - 1) \leqslant 2{,}5\ \%$, $P(x \geqslant b + 1) \leqslant 2{,}5\ \%$

$P(a \leqslant X \leqslant b) \geqslant 95\ \%$

En pratique : établir la distribution des fréquences cumulées. (Voir exemple ci-dessous)

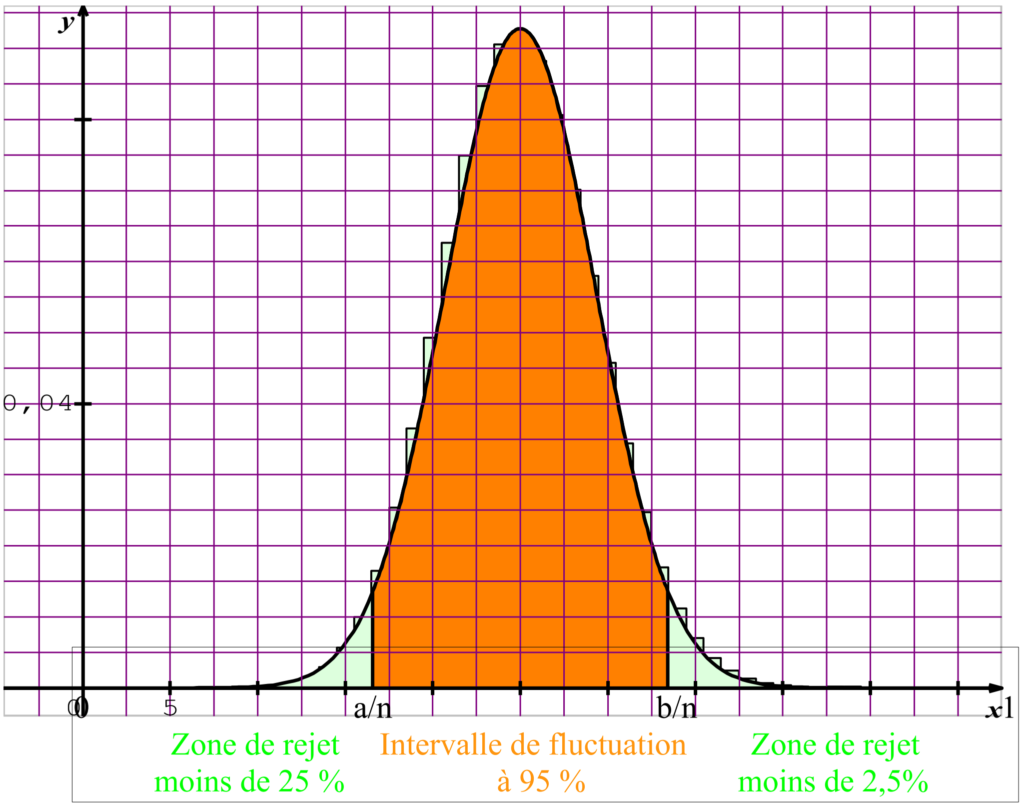
Prise de décision

On fait l'hypothèse : la proportion du caractère dans la population est p

* Si $f \in \left[\frac{a}{n};\frac{b}{n}\right]$ alors on ne peut pas rejeter l'hypothèse " la proportion du caractère dans la population est p "

* Si $f \notin \left[\frac{a}{n};\frac{b}{n}\right]$ alors on rejette cette hypothèse au risque d'erreur de 5%.

Schématiquement :



Un exemple :

X est une variable aléatoire qui suit la loi binomiale de paramètres $n = 20$ et $p = 0{,}4$
À la calculatrice ou au tableur, on fait le tableau donnant les probabilités $P(X \leqslant k)$ de k allant de 0 à 20.
On cherche le plus petit entier a tel que $P(X \leqslant k) > 0{,}025$
On cherche le plus petit entier b tel que $P(X \leqslant k) \geqslant 0{,}975$
L'intervalle de fluctuation au seuil de 95 % est $\left[\frac{a}{n};\frac{b}{n}\right]$

k	P(X<=k)
0	3,65616E-005
1	0,0005240494
2	0,0036114721
3	0,0159611628
4	0,0509519532
5	0,1255989727
6	0,2500106719
7	0,4158929376
8	0,5955987253
9	0,7553372033
10	0,8724787539
11	0,943473633
12	0,9789710725
13	0,9935341246
14	0,9983884754
15	0,9996829689
16	0,999952655
17	0,9999949587
18	0,9999996592
19	0,999999989
20	1

$a = 4$

car $P(X \leqslant 3) = 0{,}016$ et $P(X \leqslant 4) = 0{,}05$

$b = 12$, car $P(X \leqslant 11) = 0{,}943$ et $P(X \leqslant 12) = 0{,}978$ ...

L'intervalle de fluctuation au seuil de 95 % est $\left[\frac{4}{20};\frac{12}{20}\right] = [0{,}2 ; 0{,}6]$

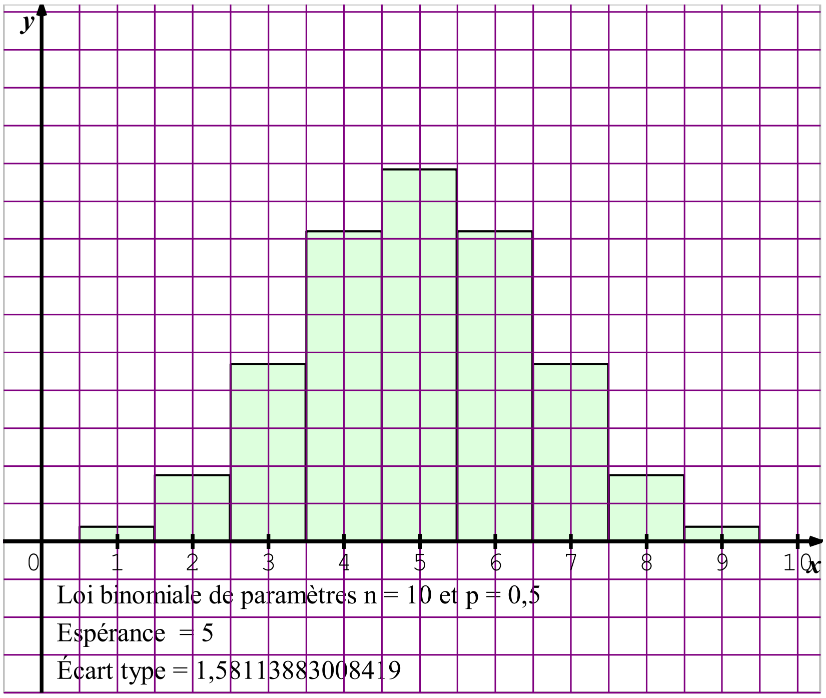
Quelques exemples de représentations graphiques réalisées avec sinequanon

$\mathscr{B}(10\ ;\ 0{,}5)$

$P(X=0) = 0{,}5^{10} \approx 10^{-3}$

$P(X = 10) = 0{,}5^{10} \approx 10^{-3}$

$P(X = 5) = \binom{10}{5}\ 0{,}5^5 \times 0{,}5^5 \approx$ 0,25



$\mathscr{B}(100\ ;\ 0{,}2)$

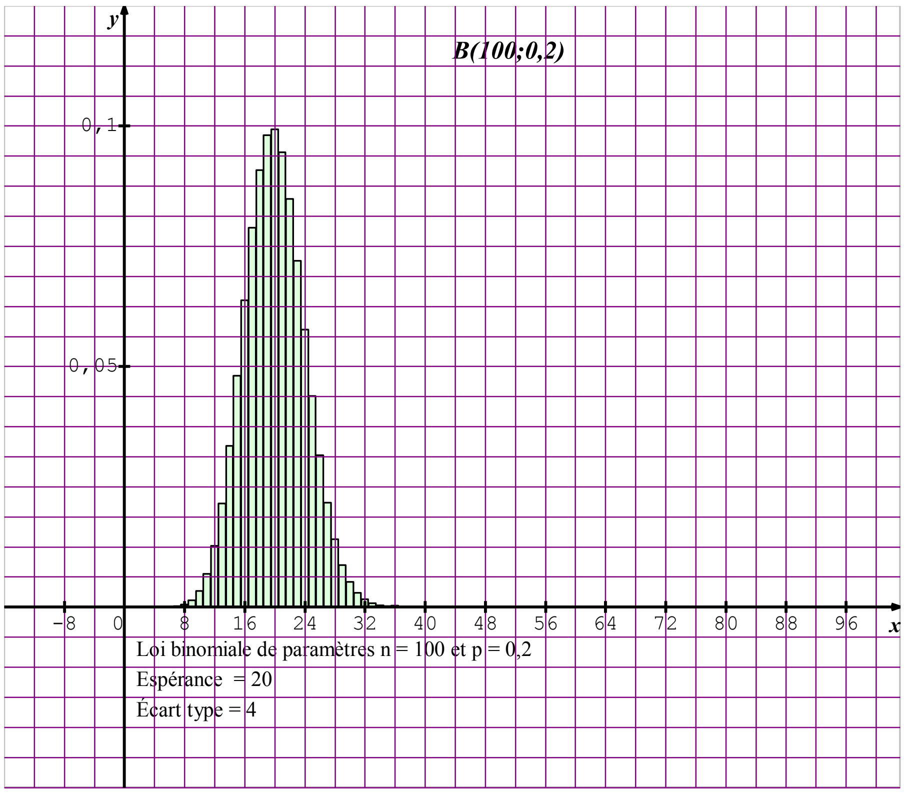
$P(X=0) = 0{,}8^{100} \approx 2\times 10^{-10}$

$P(X = 10) = \binom{100}{10}\ 0{,}2^{10} \times 0{,}8^{90} \approx$ 0,003

$P(X = 20) = \binom{100}{20}\ 0{,}2^{20} \times 0{,}8^{80} \approx$ 0,1

$P(X = 30) = \binom{100}{30}\ 0{,}2^{30} \times 0{,}8^{70} \approx$ 0,003

$P(X = 40) = \binom{100}{40}\ 0{,}2^{40} \times 0{,}8^{60} \approx$ 2×10^{-6}



Quelques exemples de représentations graphiques réalisées avec sinequanon

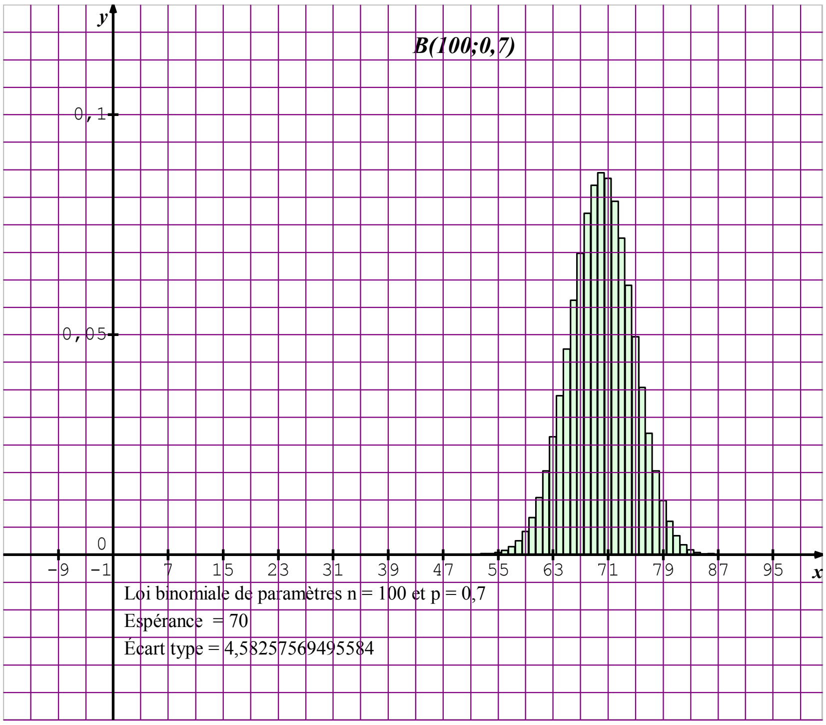
$\mathscr{B}(100 ; 0{,}7)$

$P(X=0) = 0{,}3^{100} \approx 5^{-53}$

$P(X = 10) = \binom{100}{10} 0{,}7^{10} \times 0{,}3^{90} \approx 4 \times 10^{-36}$

$P(X = 60) = \binom{100}{60} 0{,}7^{60} \times 0{,}3^{40} \approx 0{,}008$

$P(X = 70) = \binom{100}{70} 0{,}7^{70} \times 0{,}3^{30} \approx 0{,}087$



Retour au paragraphe : loi binomiale