La loi binomiale

Capacités attendues en fin de chapitre :

- Modéliser une situation par une succession d'épreuves indépendantes ou une succession de deux ou trois épreuves quelconques.
- Calculer une probabilité en utilisant l'indépendance, une probabilité conditionnelle ou la formule des probabilités totales.
- Modéliser une situation par un schéma de Bernoulli, par une loi binomiale.
- Dans le cadre d'une résolution de problème modélisé par une variable binomiale X, calculer numériquement $P(X = k)$, $P(X \leq k)$, $P(k \leq X \leq k')$ en s'aidant au besoin d'un algorithme.
- Déterminer un intervalle I pour lequel $P(X \in I)$ est inférieur à une valeur donnée α ou 1-α (intervalle de fluctuation)

Le mathématicien du chapitre :

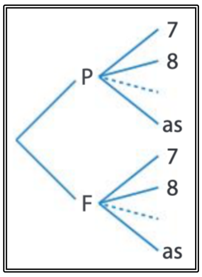
Pas moins de 8 Bernoulli ont donné leur nom à des découvertes en mathématiques. Jacob (francisé en Jacques) Bernoulli (1654-1705) est celui qui a travaillé sur les probabilités en démontrant notamment la loi faible des grands nombres pour le jeu de pile ou face. Il travailla également sur les séries (somme des termes d'une suite) en démontrant rigoureusement la convergence de la série $\sum_{n=1}^{\infty} \frac{1}{n^2}$.

1) Succession d'épreuves indépendantes

a) Univers d'une succession d'épreuves

Exemples introductifs :

- Le Yam's est un jeu de dé. On lance 5 dés et on tente, comme au poker, d'effectuer des combinaisons (3 dés identiques, 4 dés identiques, etc…). Le résultat de chaque dé est indépendant des autres. Il s'agit donc bien d'une succession d'épreuves identiques et indépendantes.
- Une expérience aléatoire consiste à lancer une pièce équilibrée et à noter sa face puis à piocher une carte dans un jeu de huit cartes et à noter sa hauteur. Les deux épreuves sont indépendantes puisque la face sortie n'influence pas le choix de la carte tirée. L'univers Ω de la succession des deux épreuves est le produit cartésien des deux épreuves $\Omega_1 = \{P; F\}$ et $\Omega_2 = \{As\ ;\ R\ ;\ D\ ;\ V\ ;\ 10\ ;9\ ;\ 8\ ;7\}$. On obtient 16 issues correspondants aux 16 chemins sur l'arbre de probabilité. Sur l'arbre ci-contre, toutes les issues sont équiprobables.



Définition :

Lorsqu'une expérience aléatoire se compose d'une succession de n épreuves indépendantes notées $E_1, E_2, \dots, E_n$, l'univers des issues possibles est le produit cartésien $\Omega_1 \times \Omega_2 \times \dots \times \Omega_n$ où Ω_i est l'univers de l'épreuve E_i avec i allant de 1 à n.
Une issue de la succession d'épreuves est donc le n-uplet $(e_1, e_2, \dots, e_n)$ ou e_i est une issue de l'épreuve E_i avec i allant de 1 à n.

Remarque :

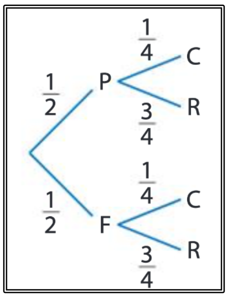
Lorsque les épreuves sont indépendantes, les deuxièmes branches sur l'arbre de probabilité ne dépendent pas des premières. Contrairement au chapitre de première sur les probabilités conditionnelles, les probabilités qui pondèrent l'arbre ne sont pas conditionnelles. La réalisation de la première épreuve n'exerce aucune influence sur la deuxième.

b) Calcul de probabilités

Définition :

Lors d'une succession de n épreuves indépendantes, la probabilité d'une issue $(e_1, e_2, \ldots, e_n)$ est égale au produit des probabilités de chacune des issues du n-uplet.

Exemple :



Un jeu consiste à lancer une pièce de monnaie (jeu de pile ou face) puis à piocher un jeton dans une urne qui contient un jeton carré et trois jetons ronds. On a les univers : $\Omega_1 = \{P; F\}$ et $\Omega_2 = \{C; R\}$. Le produit cartésien des deux univers donne donc quatre 2-uplets différents.

$P\big((F;C)\big) = \frac{1}{2} \times \frac{1}{4} = \frac{1}{8}$. L'arbre étant pondéré, les issues ne sont pas équiprobables. Pour déterminer la probabilité d'obtenir un jeton rond, il faut ajouter les probabilités des chemins conduisant à R.

$P(R) = P\big((P;R)\big) + P\big((F;R)\big) = \frac{1}{2} \times \frac{3}{4} + \frac{1}{2} \times \frac{3}{4} = \frac{3}{4}$.

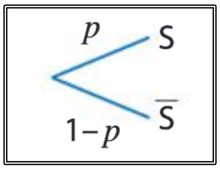
2) Le schéma de Bernoulli

a) Une épreuve de Bernoulli

Définition :

Une épreuve de Bernoulli est une expérience aléatoire à deux issues contraires appelées succès et échec. On note p la probabilité du succès et $1 - p$ la probabilité de l'échec

Exemples :



- Lancer un dé cubique et s'intéresser à l'apparition du 1.
- Un match de Tennis.
- Un match de foot n'est pas une épreuve de Bernoulli.

b) Loi de probabilité

On considère une épreuve de Bernoulli de paramètre p. La variable aléatoire X, prenant la valeur 1 en cas de succès et 0 sinon, suit la loi de probabilité donnée ci-dessous.

x_i	0	1	Total
p_i	$1-p$	p	1

On peut donc calculer l'espérance mathématique et la variance de X.

- $E(X) = 0 \times (1-p) + 1 \times p$ soit donc $E(X) = p$
- $V(X) = 0^2 \times (1-p) + 1^2 \times p - p^2$ soit $V(X) = p(1-p)$ Formule de Koenig
- L'écart-type s'écrit donc $\sigma(X) = \sqrt{p(1-p)}$

On dit donc que X suit une loi de Bernoulli de paramètre p, qui s'écrit alors $X \rightsquigarrow B(1;p)$

c) Exemple de loi de Bernoulli

Lors d'une évaluation, un élève ne connaît pas la réponse à une question d'un QCM. Cette question comporte 4 réponses possibles et on sait qu'il n'y a qu'une seule bonne réponse parmi les quatre. Il décide donc de répondre au hasard, le malheureux…
On construit donc la loi de probabilité.

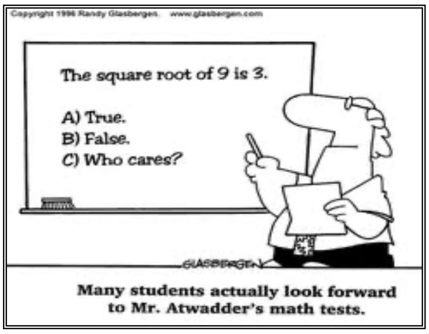
x_i	0	1	Total
p_i	$\frac{3}{4}$	$\frac{1}{4}$	1

Comme vu précédemment, on a $E(X) = \frac{1}{4}$

L'élève possède donc une chance sur quatre d'avoir la bonne réponse à la question s'il répond au hasard.

d) Le schéma de Bernoulli

Un QCM comporte maintenant 3 questions. Pour chaque question, 4 réponses sont proposées aux élèves et une seule est juste.
Un élève, qui a manifestement oublié de réviser décide de répondre au hasard aux 3 questions de manière indépendante.
On note X la variable aléatoire qui compte le nombre de bonne réponse.
Donner la loi de probabilité et l'espérance mathématique de X.

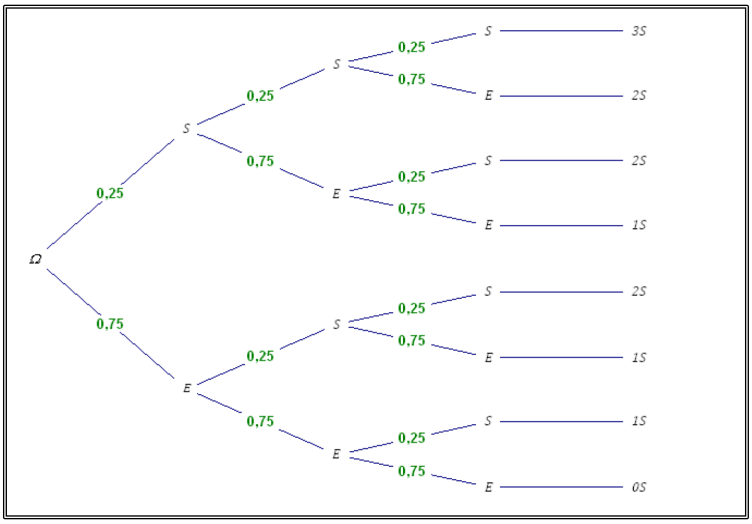


Remarque :

On suppose que l'élève répond vraiment au hasard aux questions. Il paraît souvent peu recevable aux élèves que la bonne réponse soit toujours à la même position dans un QCM.

Solution :

Les valeurs possibles pour X sont : 0, 1, 2 et 3.
On peut représenter la situation à l'aide d'un arbre de probabilité. Cet arbre va comporter au total 14 branches pour 8 chemins différents. On compte alors au bout de chaque chemin le nombre de succès présents. Il faudra ensuite pour chaque valeur de X, compter les chemins qui conduisent à cette valeur.



On peut donc répertorier les résultats afin d'obtenir la loi de probabilité de X.

x_i	0	1	2	3	Total
p_i	$1\times\left(\frac{3}{4}\right)^3=\frac{27}{64}$	$3\times\frac{1}{4}\times\left(\frac{3}{4}\right)^2=\frac{27}{64}$	$3\times\left(\frac{1}{4}\right)^2\times\frac{3}{4}=\frac{9}{64}$	$1\times\left(\frac{1}{4}\right)^3=\frac{1}{64}$	1
x_ip_i	0	$\frac{27}{64}$	$\frac{18}{64}$	$\frac{3}{64}$	$\frac{3}{4}$

Un élève qui n'aura pas révisé peut espérer obtenir en moyenne 0,75 bonne réponse…

Définition :

On définit un schéma de Bernoulli de paramètres n et p lorsqu'on répète n fois de manière indépendante la même épreuve de Bernoulli.

Remarques :

- L'objectif maintenant est de déterminer une formule générale de la loi de probabilité qui représente un schéma de Bernoulli.
- Il est clair que si $n \geq 4$, la construction d'un arbre de probabilité peut s'avérer très pénible.

Exercice :

Dans un jeu de 32 cartes, on tire successivement et avec remise trois cartes du jeu. On appelle *N* la variable aléatoire égale au nombre de figures (Roi, Dame, Valet) obtenues à l'issue de ces trois tirages.
Déterminer l'espérance mathématique de *N*, c'est à dire le nombre moyen de figures que l'on peut espérer si on joue un grand nombre de fois à ce jeu. On pourra s'aider d'un arbre.

Réponse :

n_i	0	1	2	3	Total
p_i	$1\times\left(\frac{5}{8}\right)^3=\frac{125}{512}$	$3\times\frac{3}{8}\times\left(\frac{5}{8}\right)^2=\frac{225}{512}$	$3\times\left(\frac{3}{8}\right)^2\times\frac{5}{8}=\frac{135}{512}$	$1\times\left(\frac{3}{8}\right)^3=\frac{27}{512}$	1
n_ip_i	0	$\frac{225}{512}$	$\frac{270}{512}$	$\frac{81}{512}$	$\frac{9}{8}$

3) *La loi binomiale*

a) *Définitions et propriétés*

Définition :

On considère un schéma de Bernoulli de paramètres *n* et *p* et *X* la variable aléatoire comptant le nombre de succès obtenus dans ce schéma.
On dit alors que *X* suit une loi binomiale de paramètres *n* et *p* et on note $X \rightsquigarrow B(n;p)$
La loi de la variable aléatoire *X* est donnée, pour tout entier *k* compris entre 0 et *n* par :
La probabilité d'obtenir *k* succès est donnée par : $P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$

Démonstration :

Pour calculer la probabilité de l'événement $(X=k)$, on fait la somme des probabilités des chemins menant à *k* succès. Chacun de ces chemins a la même probabilité $p^k(1-p)^{n-k}$ et il y a au total $\binom{n}{k}$ chemins ce cette sorte. D'où la formule ci-dessus.

Propriétés :

L'espérance mathématique de la variable est : $E(X)=np$
La variance de la variable est donnée par : $V(X)=np(1-p)$
L'écart-type de la variable est donné par : $\sigma(X)=\sqrt{np(1-p)}$

Exemple :

On lance un dé cubique non truqué 5 fois de suite. *X* est la variable aléatoire égale au nombre de 6 obtenus au cours de ces cinq lancers. On a alors un schéma de Bernoulli de paramètres $n=5$ et $p=\frac{1}{6}$. On obtient alors la loi de probabilité à l'aide de la TI-83 ou avec python

```python
from math import *
def binomfdp(n,p,k):
    c = factorial(n)/(factorial(k)*factorial(n-k))
    x = c*p**k*(1-p)**(n-k)
    return x

def binomfrep(n,p,k):
    S = binomfdp(n,p,0)
    for i in range(k):
        S = S+binomfdp(n,p,i+1)
    return S
```

```
> binomfdp(5,1/6,0)
0.401877572016461
> binomfdp(5,1/6,1)
0.40187757201646096
```

```
> binomfrep(5,1/6,0)
0.401877572016461
> binomfrep(5,1/6,1)
0.803755144032922
```

On lit alors sur la table : $P(X=2)=0{,}1608$ ou par exemple $P(X\leq 3)=0{,}9967$.
Si on souhaite déterminer $P(X\geq 3)$, on passe par le contraire : $P(X\geq 3)=1-P(X\leq 2)$
Soit $P(X\geq 3)=0{,}0355$

NORMAL FLOTT AUTO RÉEL RAD MP
APP SUR + POUR ΔTbl

X	Y1	Y2
0	0.4019	0.4019
1	0.4019	0.8038
2	0.1608	0.9645
3	0.0322	0.9967
4	0.0032	0.9999
5	1.3E-4	1

<u>Propriété :</u>

Si la variable aléatoire X suit une loi Binomiale de paramètres n et p, notée $B(n;p)$, où n est le nombre d'épreuves répétées et p la probabilité de succès à l'une de ces épreuves, alors :
$P(X=0)=(1-p)^n$, probabilité de n'avoir ***<u>aucun succès</u>***.
$P(X \geq 1)=1-P(X=0)=1-(1-p)^n$, probabilité d'avoir ***<u>au moins un succès.</u>***

<u>Exercice :</u>

Une entreprise de transport possède 10 camions. Une étude sur une période donnée de l'état de fonctionnement des camions a montré que, un jour donné, la probabilité qu'un camion tombe en panne est égale à 0,05. On admet que la panne d'un camion est indépendante des pannes survenues antérieurement et de celles des autres camions.

1. Montrer que cette situation peut être modélisée par un schéma de Bernoulli.
2. Soit X la variable aléatoire égale au nombre de camion en panne.
Justifier que cette loi suit une loi binomiale dont on précisera les paramètres n et p.

3. Déterminer la probabilité qu'un jour donné :
   a. Aucun camion ne soit en panne.
   b. au moins un camion soit en panne.
   c. 6 camions soient en état de fonctionnement.

4. La réparation d'un camion en panne coûte 500€. Quel budget le responsable de l'entreprise doit-il prévoir sur une année pour couvrir les frais éventuels dus aux pannes des camions ?

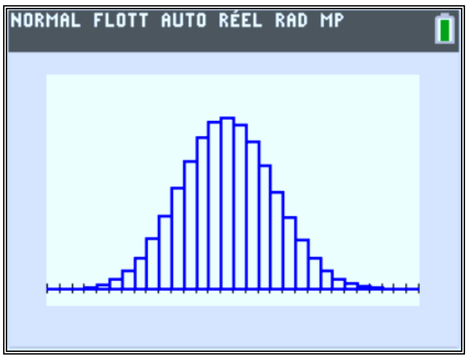
<u>Solution :</u>

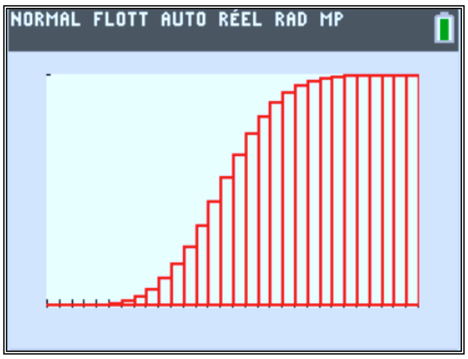
1. On note G l'évènement « le camion est en panne et doit rester au garage ». Il s'agit d'une répétition d'épreuves indépendantes à deux issues. C'est donc un schéma de Bernoulli.
2. On note X la variable aléatoire qui compte le nombre de succès, c'est à dire le nombre de camion au Garage. Alors $X \leadsto B(10; 0{,}05)$
3. $P(X=0)=(1-0{,}05)^{10}$ soit alors $P(X=0) \approx 0{,}599$
   $P(X \geq 1)=1-P(X=0)=1-0{,}599 \approx 0{,}401$
   Si 6 camions fonctionnent, c'est qu'il y en a 4 en panne : $P(X=4) \approx 0{,}00096$
4. L'espérance mathématique vaut $E(X)=10 \times 0{,}05=0{,}5$ En moyenne, 0,5 camion sont en panne ; sur une année, le budget à prévoir est donc : $0{,}5 \times 500 \times 365=91250€$

b) <u>Représentation graphique</u>

On peut représenter la loi binomiale à la calculatrice. On observe ci-dessous la forme caractéristique de la courbe en cloche.

Soit $X \leadsto B(60; 0{,}4)$. On peut alors représenter les probabilités ou les probabilités cumulées. Pour plus de précision, référez-vous à la fiche TI83 sur Moodle.

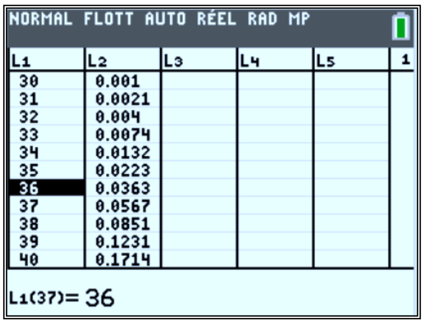




4) *Intervalle de Fluctuation*

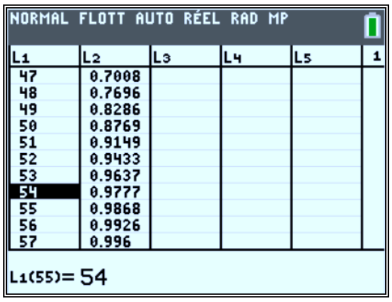
a) *Exemple introductif*

On cherche à déterminer si un dé est truqué. Pour cela on se demande si la fréquence d'un chiffre pair est normale. On lance ce dé 90 fois. Soit X le nombre de fois ou un chiffre pair apparaît. Si le dé n'est pas truqué, alors $X \sim B(90; 0{,}5)$. Alors, on a :$E(X) = 45$
Ceci signifie, qu'en moyenne, un chiffre pair doit apparaître 45 fois sur 90. Il est évident qu'on n'obtiendra jamais 45 chiffres pairs sur les 90. Mais jusqu'à quel nombre accepte-t-on d'aller pour justifier que le dé n'est pas truqué ?
Pour être dans la normalité, on cherche un intervalle centré sur $E(X)$ tel que 95 % des valeurs soient dans cet intervalle. On a donc deux zones de « rejets » de 2,5 % chacune à l'extérieure.
On utilise pour cela la formule des probabilités cumulatives de la loi binomiale.


NORMAL FLOTT AUTO RÉEL RAD MP

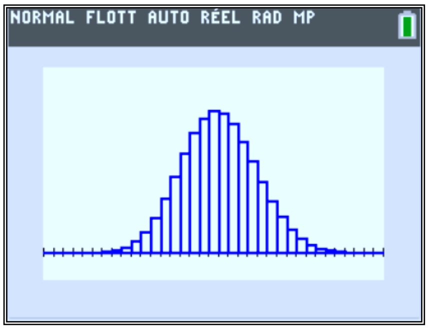
L1	L2	L3	L4	L5	1
30	0.001				
31	0.0021				
32	0.004				
33	0.0074				
34	0.0132				
35	0.0223				
36	0.0363				
37	0.0567				
38	0.0851				
39	0.1231				
40	0.1714				

L1(37)= 36


NORMAL FLOTT AUTO RÉEL RAD MP

L1	L2	L3	L4	L5	1
47	0.7008				
48	0.7696				
49	0.8286				
50	0.8769				
51	0.9149				
52	0.9433				
53	0.9637				
54	0.9777				
55	0.9868				
56	0.9926				
57	0.996				

L1(55)= 54



On obtient donc un intervalle dans lequel sont situées 95 % des valeurs. Cet intervalle est appelé intervalle de Fluctuation au seuil de 95 %.
Ici $I_F = \left[\frac{36}{90}; \frac{54}{90}\right]$ soit $I_F = [0{,}4; 0{,}6]$

b) *Intervalle de Fluctuation centré*

On a rappelé en préambule que $X \sim B(n; p)$. On partage alors $[0; n]$ en trois intervalles : $[0; a-1]$, $[a; b]$ et $[b+1; n]$.
a et b sont choisis de telle sorte que X prenne ses valeurs dans les deux intervalles extrêmes avec une probabilité proche de 2,5% sans jamais la dépasser.

Définition :

L'intervalle de fluctuation à 95% d'une fréquence correspondant à la réalisation, sur un échantillon de taille n, d'une variable aléatoire $X \sim B(n; p)$ est l'intervalle $I_F = \left[\frac{a}{n}; \frac{b}{n}\right]$.
a est le plus petit entier tel que $P(X \leq a) > 0{,}025$
b est le plus petit entier tel que $P(X \leq b) \geq 0{,}975$

Remarque :

- On peut alors montrer que $P(a \leq X \leq b) \geq 0{,}95$
- Dans au moins 95% des cas, la fréquence observée f_{obs} appartient à l'intervalle $\left[\frac{a}{n}; \frac{b}{n}\right]$
- Pour $n \geq 30$, $np \geq 5$, et $n(1-p) \geq 5$ l'intervalle de fluctuation à 95% est sensiblement le même que celui donné en seconde, défini par : $I_F = \left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}\right]$

Exemple :

Une urne contient 256 boules bleues et 144 boules rouges, toutes indiscernables au toucher. On tire au hasard et avec remise 200 boules de cette urne et on s'intéresse à la fréquence d'apparition des boules bleues obtenues.

1) Déterminer l'intervalle de fluctuation asymptotique au seuil de 95 % de la fréquence des boules bleues.
2) Donner alors une interprétation de cet intervalle.

Solution :

1) On a ici $n = 200$ et $p = 0{,}64$. A l'aide de la calculatrice, on obtient $I_F = \left[\frac{115}{200}; \frac{141}{200}\right]$ soit alors $I_F = [0{,}575; 0{,}705]$
2) Cela signifie que si on réitère cette expérience un grand nombre de fois, la fréquence d'apparition de boules bleues, dans ces échantillons de taille 200 sera comprise entre 0,575 et 0,705 dans environ 95 % des cas.

c) Cas général

Définition :

Soit $X \rightsquigarrow B(n;p)$, $\alpha \in]0;1[$, a et b des réels.
Un intervalle $\left[\frac{a}{n}; \frac{b}{n}\right]$ tel que $P(a \leq X \leq b) \geq 1 - \alpha$ est appelé intervalle de fluctuation au seuil de $1 - \alpha$ (ou au risque α) associé à X.

Remarque :

Le cas le plus fréquent au lycée est pour $\alpha = 0{,}05$ comme vu dans l'exemple introductif.

Propriété :

Soit $X \rightsquigarrow B(n;p)$, $\alpha \in]0;1[$, a et b des réels.
Si $P(X < a) \leq \frac{\alpha}{2}$ et $P(X > b) \leq \frac{\alpha}{2}$, alors l'intervalle $\left[\frac{a}{n}; \frac{b}{n}\right]$ est un intervalle de fluctuation au seuil de $1 - \alpha$ associé à X. On dit que l'intervalle est centré ou bilatéral.

d) Construire un intervalle en Python

On utilise les connaissances acquises en Python pour construire un intervalle centré. Pour cela, on complète avec les modules déjà codés. On peut demander la valeur de α afin de pouvoir utiliser le code dans toutes les situations.
Si on reprend les différents exemples du cours, on obtient les mêmes intervalles.

```python
from math import *
def binomfdp(n,p,k):
  c = factorial(n)/(factorial(k)*factorial(n-k))
  x = c*p**k*(1-p)**(n-k)
  return x

def binomfrep(n,p,k):
  S = binomfdp(n,p,0)
  for i in range(k):
    S = S+binomfdp(n,p,i+1)
  return S

def fluctuation(n,p,alpha):
  for k in range(1,n+1):
    if binomfrep(n,p,k)<=alpha/2:
      a = (1+k)/n
    if binomfrep(n,p,k)<1-alpha/2:
      b = (1+k)/n
  return [a,b]
```

```
> fluctuation(90,0.5,0.05)
[0.4, 0.6]
> fluctuation(200,0.64,0.05)
[0.575, 0.705]
> fluctuation(200,0.64,0.01)
[0.55, 0.725]
```

On peut également déterminer un intervalle de fluctuation bilatéral au seuil de 99% comme étudié dans le supérieur. C'est l'intervalle qui contient 99% des valeurs au moins. Il est donc nécessairement plus large que celui à 95 %.