I Loi de Bernoulli et loi binomiale

Loi de Bernoulli

Soit une expérience aléatoire présentant deux issues : l'une S que l'on appelle « succès » et l'autre $\overline{S}$ appelée « échec ». On note p la probabilité de succès et q celle d'échec, avec q = 1 – p.

Cette expérience aléatoire s'appelle une **épreuve de Bernoulli de paramètre p**.

<u>Définition</u>

- La variable aléatoire qui prend la valeur 1 en cas de succès et 0 en cas d'échec est appelée **variable aléatoire de Bernoulli**.
- La loi de probabilité de cette variable aléatoire est appelée **loi de Bernoulli de paramètre p.**

x_i	0	1
$P(X = x_i)$	1 - p	p

<u>Propriété</u>

Soit X suit une loi de Bernoulli de paramètre p, alors E(X) = p.

En effet, $E(X) = 0\times(1 - p) + 1\times p = p$.

Schéma de Bernoulli et loi binomiale

<u>Définitions</u>

- L'expérience aléatoire consistant à répéter n fois de manière indépendante une épreuve de Bernoulli de paramètre p s'appelle un **schéma de Bernoulli de paramètres n et p.**
- La loi de probabilité de la variable aléatoire X égale au nombre de succès au cours de ces n épreuves se nomme **la loi binomiale de paramètres n et p**. On la note $\mathscr{B}(n,p)$.

<u>Exemple</u> :

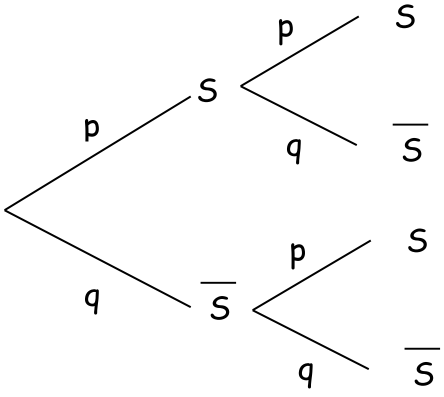
On lance un dé équilibré 100 fois de suite et on s'intéresse au nombre X de fois où l'on obtient la face 1.

La variable aléatoire X qui compte le nombre de succès (c'est-à-dire le nombre de fois que l'on obtient la face 1) suit la loi binomiale de paramètres n = 100 et p = $\frac{1}{6}$ notée $\mathscr{B}\left(100, \frac{1}{6}\right)$.

Loi binomiale pour n = 2 et n = 3

On peut modéliser la répétition de ces expériences par un arbre pondéré.

Cas n = 2



$P(X = 0) = P(\overline{S}\,\overline{S}) = q\times q = q^2$

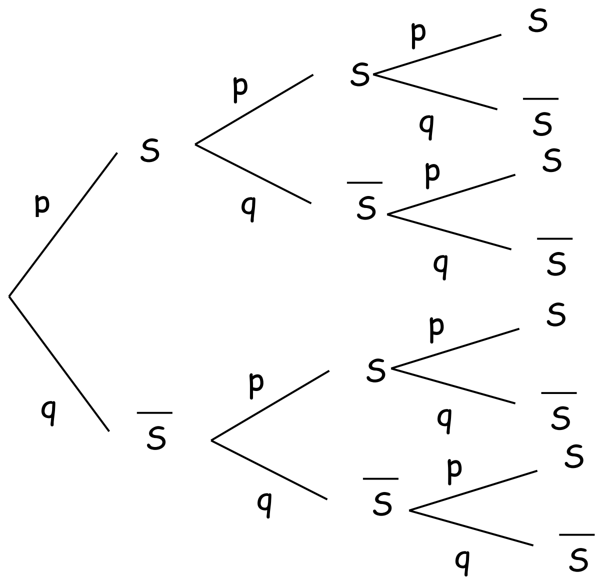
$P(X = 1) = P(S\overline{S}) + P(\overline{S}S) = pq + pq = 2pq$

$P(X = 2) = P(SS) = p^2$

<u>Loi de probabilité de B(2,p) avec q = 1 – p</u> :

x_i	0	1	2
$P(X = x_i)$	q^2	2pq	p^2

Cas n = 3



$P(X = 0) = P(\overline{S}\ \overline{S}\ \overline{S}) = q \times q \times q = q^3$

$P(X = 1) = P(S\ \overline{S}\ \overline{S}) + P(\overline{S}\ S\ \overline{S}) + P(\overline{S}\ \overline{S}\ S) = p \times q \times q + q \times p \times q + q \times q \times p = 3pq^2$

$P(X = 2) = P(SS\overline{S}) + P(S\overline{S}S) + P(\overline{S}SS) = p \times p \times q + p \times q \times p + q \times p \times p = 3p^2q$

$P(X = 3) = P(SSS) = p \times p \times p = p^3$

<u>Loi de probabilité de $\mathscr{B}(3,p)$ avec $q = 1 - p$</u> :

x_i	0	1	2	3
$P(X = x_i)$	q^3	$3pq^2$	$3p^2q$	p^3

II Coefficients binomiaux et loi binomiale

Coefficients binomiaux

Lorsque X suit une loi binomiale, pour calculer la probabilité d'avoir k succès, on note toutes les issues formées de k succès et de n – k échecs. Ces issues ont toutes la même probabilité $p^k \times q^{n-k}$. Il est ensuite nécessaire de connaître le nombre de chemins de l'arbre formé de k succès exactement (et donc de n – k échecs).

<u>Définition</u>

Soit n un entier naturel non nul et k un entier compris entre 0 et n.

Le **coefficient binomial** $\binom{n}{k}$ est le nombre de chemins réalisant k succès pour n répétitions dans l'arbre d'un schéma de Bernoulli.

<u>Exemple</u> :

$\binom{2}{1} = 2$, car pour deux répétitions, il y a deux chemins avec un succès, ceux associés à $S\,\overline{S}$ et à $\overline{S}\,S$.

<u>Vocabulaire</u> :

On appelle aussi $\binom{n}{k}$ le nombre de combinaisons de k éléments parmi n.

Calcul pratique des coefficients binomiaux

Les calculatrices et les tableurs permettent de calculer les coefficients binomiaux :

	Casio	Texas	Tableur
	Touche OPTN puis choisir ➢ puis PROB puis nCr	Touche MATH, puis choisir PRB, puis Combinaison	Fonction COMBIN()
Syntaxe	n **nCr** k	n **Combinaison** k	=COMBIN(n ;k)

Propriétés sur les coefficients binomiaux

Pour tous nombres entiers n et k :

Si $0 \leq k \leq n$ alors $\binom{n}{k} = \binom{n}{n-k}$

Si $0 \leq k \leq n - 1$ alors $\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$ (c'est la formule de Pascal

Triangle de Pascal

Le triangle de Pascal permet de calculer de proche en proche les coefficients binomiaux en utilisant la formule de Pascal.

Le coefficient binomial $\binom{n}{k}$ est à l'intersection de la ligne n et de la colonne k.

n\k	0	1	2	3	4	5
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

On place les valeurs évidentes $\binom{n}{0} = 1$ et $\binom{n}{n} = 1$

On complète le triangle avec la formule de Pascal.

Exemple : $\binom{4}{1} + \binom{4}{2} = \binom{5}{2}$

Formule générale de la loi binomiale

Propriété

Si la variable aléatoire X suit la loi binomiale de paramètres n et p, alors pour tout entier k compris entre 0 et n : $\mathbf{P(X = k) = \binom{n}{k} \times p^k \times q^{n-k}}$, où $q = 1 - p$.

Calcul pratique de P(X = k) et P(X ≤ k)

	Casio	Texas	Tableur
	Touche OPTN puis choisir STAT puis DIST puis BINM puis Bpd ou Bcd	Menu DISTR, puis choisir binomFdp ou binomFrép	Fonction LOI.BINOMIALE()
P(X = k)	BinomialPD(k,n,p)	BinomFdp(n,p,k)	=LOI.BINOMIALE(k ;n ;p ;FAUX)
P(X ≤ k)	BinomialCD(k,n,p)	BinomFRép(n,p,k)	=LOI.BINOMIALE(k ;n ;p ;VRAI)

Exemple :

Soit X une variable aléatoire qui suit la loi binomiale de paramètres n = 8 et p = 0,4.

Pour trouver P(X = 3), on utilise la calculatrice en remplaçant n par 8, k par 3 et p par 0,4.

On trouve environ 0,279 (Avec une calculatrice TI : BinomFdp(8,0.4,3))

Pour trouver P(X ≤ 3), avec une calculatrice TI : BinomFRép(8,0.4,3) :

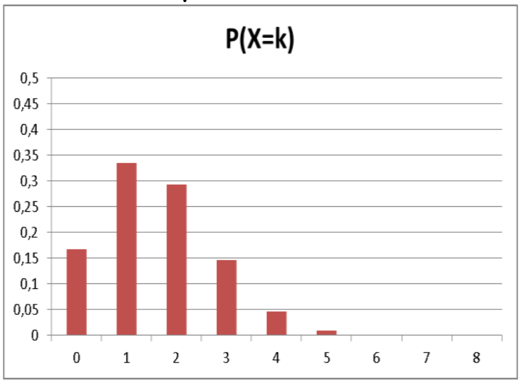
On trouve environ 0,594.

III Représentation graphique de la loi binomiale et espérance

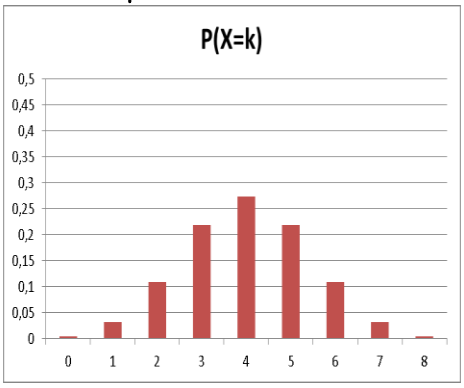
Représentation graphique

Exemple de représentation pour n = 8 et trois valeurs différentes de p.

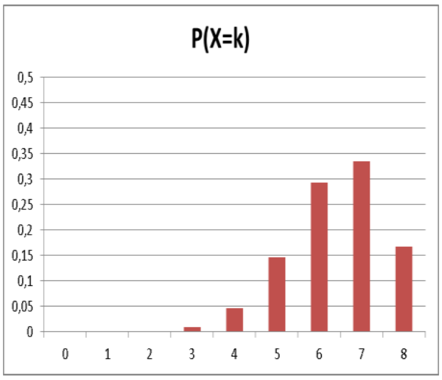
p = 0,2



p = 0,5



p = 0,8



Espérance mathématique, variance et écart-type

Propriété

X est une variable aléatoire qui suit la loi binomiale de paramètres n et p.
$E(X) = n\times p$
$V(X) = n\times p\times(1 - p) = npq$
$\sigma(X) = \sqrt{npq}$

IV Intervalle de fluctuation d'une fréquence

Propriété (vue en seconde)

Si p est la proportion d'un caractère dans une population (avec $0,2 \leq p \leq 0,8$), alors pour un échantillon de taille n ($n \geq 25$), la fréquence f du caractère dans l'échantillon appartient à l'intervalle $\left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}} \right]$ avec une probabilité d'au moins 95%.

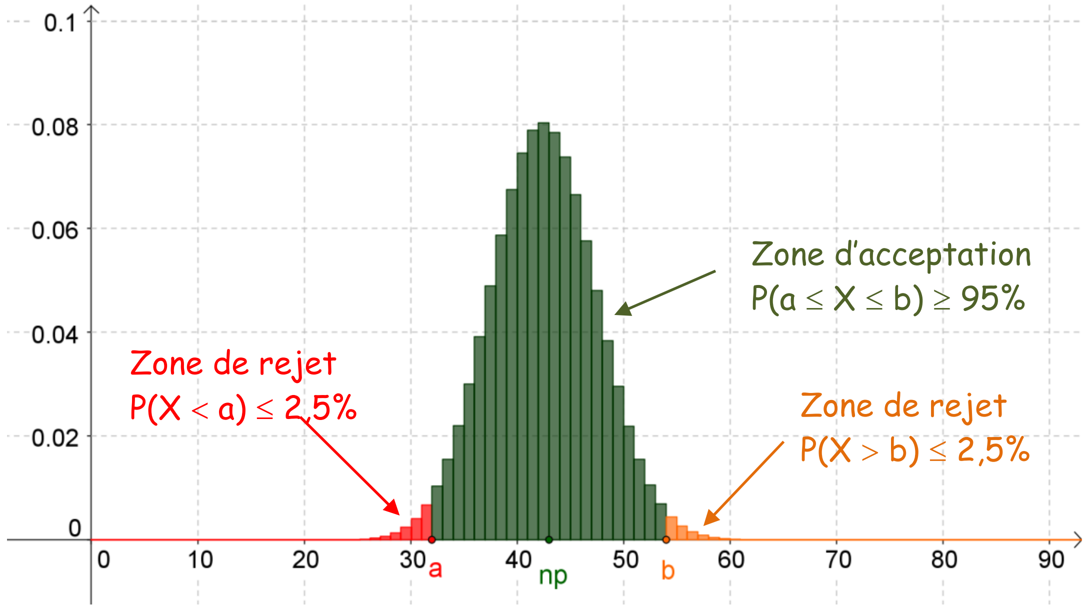
On peut améliorer ce résultat avec la loi binomiale.

Etude d'un exemple

Des études ont montré que la proportion des français qui font du sport au moins une fois par semaine est de 43%.
On veut déterminer un intervalle de fluctuation de la fréquence f des personnes faisant du sport au moins une fois par semaine dans les échantillons de taille 100.
On définit la variable aléatoire X égale au nombre de personnes faisant du sport au moins une fois par semaine.
On peut supposer que X suit une loi binomiale de paramètres 100 et 0,43.
X prenant des valeurs comprises entre 0 et 100, on va partager l'intervalle [0 ;100] en trois parties :

- A = [0 ;a – 1] avec a entier
- B = [a ;b] avec b entier
- C = [b + 1 ; 100]

On détermine ensuite a et b de façon que $P(X \in A) \leq 0{,}025$ et $P(X \in C) \leq 0{,}025$.
On aura alors $P(X \in B) = P(a \leq X \leq b) = 1 - P(X \in A) - P(X \in C) \geq 1 - 0{,}025 - 0{,}025 \geq 0{,}95$.



- Le plus petit entier a tel que $P(X < a) \leq 0{,}025$ est 33.
- $P(X > b) \leq 0{,}025 \Leftrightarrow 1 - P(X \geq b) \leq 0{,}025$
  $\Leftrightarrow P(X \leq b) \geq 1 - 0{,}025$
  $\Leftrightarrow P(X \leq b) \geq 0{,}975$
- Le plus petit entier b tel que $P(X \leq b) \geq 0{,}975$ est 53.

L'intervalle [a ;b] cherché est donc [33 ;53].
La fréquence f appartient donc à l'intervalle [0,33 ;0,53] avec une probabilité au moins égale à 0,95.

Intervalle de fluctuation

Propriété

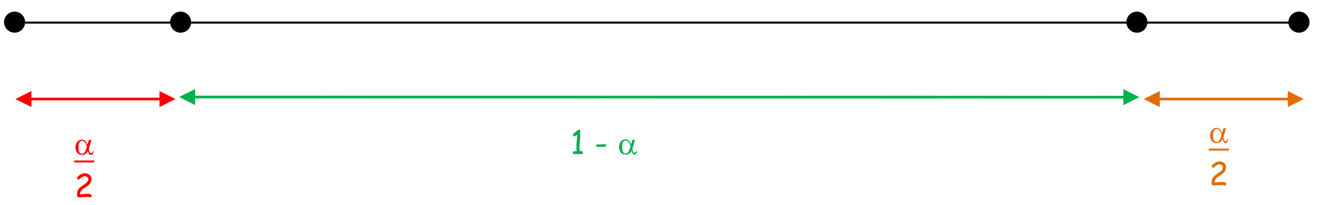
> L'intervalle de fluctuation au coefficient 95% de la fréquence correspondant à la réalisation, sur un échantillon aléatoire de taille n, d'une variable aléatoire X suivant une loi binomiale, est $\left[\frac{a}{n} ; \frac{b}{n}\right]$, où a est le plus petit entier tel que $P(X \leq a) \geq 0{,}025$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

V Prise de décision sur un échantillon

L'intervalle de fluctuation à 95% est un intervalle qui contient au moins 95% des fréquences observées sans les échantillons de taille n. Ce qui signifie qu'il y a un risque de 5% pour cette fréquence de ne pas se trouver dans l'intervalle.

Autres seuils possibles

On peut utiliser un autre coefficient que 95%.
Le plus fréquemment utilisé après 95% est 99%.
Si on choisit un seuil de risque α, c'est-à-dire un coefficient de confiance $1 - \alpha$.



Ainsi a et b sont les plus petits entiers tels que $P(X \leq a) \geq \frac{\alpha}{2}$ et $P(X \leq b) \geq 1 - \frac{\alpha}{2}$.

Exploitation d'un intervalle de fluctuation

La détermination d'un intervalle de fluctuation permet de prendre une décision lorsque l'on fait une hypothèse sur une proportion p dans une population.
Si la fréquence observée dans l'échantillon n'appartient à l'intervalle de fluctuation à 95%, on rejettera l'hypothèse faite sur p avec un risque de se tromper de 5%.

f appartient à I : on accepte l'hypothèse faite sur p

f_1 n'appartient pas à I :
on rejette l'hypothèse faite sur p.

f_2 n'appartient pas à I :
on rejette l'hypothèse faite sur p.

Propriété (vu en seconde)

On considère une population dans laquelle on **suppose** que la proportion d'un caractère est p.
Après expérience, on **observe** f comme fréquence de ce caractère dans un échantillon de taille n.
Soit l'hypothèse : « La proportion de ce caractère dans la population est p ».
Si I est l'intervalle de fluctuation de la fréquence à 95% dans les échantillons de taille n, alors :

- Si $f \notin I$: on rejette cette hypothèse au seuil de risque 5%.
- Si $f \in I$: on ne rejette pas cette hypothèse au seuil de risque 5%.