Chapitre 7

Loi binomiale. Échantillonnage

I Schéma de Bernoulli

I - 1) épreuve de Bernoulli

* lorsque, dans une expérience aléatoire, on s'intéresse uniquement à la réalisation d'un certain événement S (appelé « succès ») ou à sa non réalisation $\bar{S}$ (appelé « échec »), on dit que cette expérience est une **épreuve de Bernoulli**.

* notons X la variable aléatoire prenant la valeur 1 en cas de succès et la valeur 0 en cas d'échec.
On dit que X suit une **loi de Bernoulli**.

exemple : un jeu de dé est tel que le joueur gagne lorsque le 6 sort et perd dans le cas contraire.
Appelons « succès »l'événement S « Sortie du 6 » ; l'échec $\bar{S}$ est donc l'événement « Le 6 ne sort pas ».
Si le dé n'est pas pipé : $p(S) = \dfrac{1}{6}$ et $p(\bar{S}) = 1 - p(S) = \dfrac{5}{6}$.

La variable aléatoire X qui prend la valeur 1 si le 6 sort et la valeur 0 dans les cinq autres cas suit une loi de Bernoulli.

x_i	1	0
$P(X = x_i)$	$\frac{1}{6}$	$\frac{5}{6}$

I - 2) schéma de Bernoulli

Lorsqu'on effectue plusieurs épreuves de Bernoulli successives, indépendantes les unes des autres, on dit qu'il s'agit d'un **schéma de Bernoulli**.

exemple : l'expérience qui consiste à effectuer trois fois de suite l'épreuve de Bernoulli de l'exemple précédent (paragraphe I.1) est un schéma de Bernoulli.

On note Y la variable aléatoire qui compte le nombre de succès après les trois lancers.

Les valeurs possibles prises par Y sont :

* $Y = 0$: les trois lancers n'ont jamais donné le 6. Il y a eu trois échecs.

* $Y = 1$: un des trois lancers a donné le 6, les deux autres ont donné un nombre autre que le 6. Il y a eu un succès et deux échecs.

* $Y = 2$: un des trois lancers n'a pas donné le 6, les deux autres ont donné le 6. Il y a eu deux succès et un échec.

* $Y = 3$: les trois lancers ont donné le 6. Il y a eu trois succès.

Le but du paragraphe qui suit est de déterminer rapidement la probabilité de chacun de ces évènements, autrement dit de mettre en place la loi de probabilité de la variable aléatoire Y.

II Loi binomiale

II - 1) loi binomiale de paramètres n et p

définition 1 :

On considère un schéma de Bernoulli constitué par la répétition de n épreuves de Bernoulli identiques. Pour chacune d'elles, on note p la probabilité d'obtenir un succès S.

La loi de probabilité de la variable aléatoire X comptant le nombre de succès est appelée **loi binomiale de paramètres n et p**.

On note cette loi : $\mathcal{B}(n, p)$

exemple : la variable aléatoire Y du paragraphe I.2 comptant le nombre de sorties du 6 après trois lancers d'un dé bien équilibré suit une loi binomiale de paramètres :

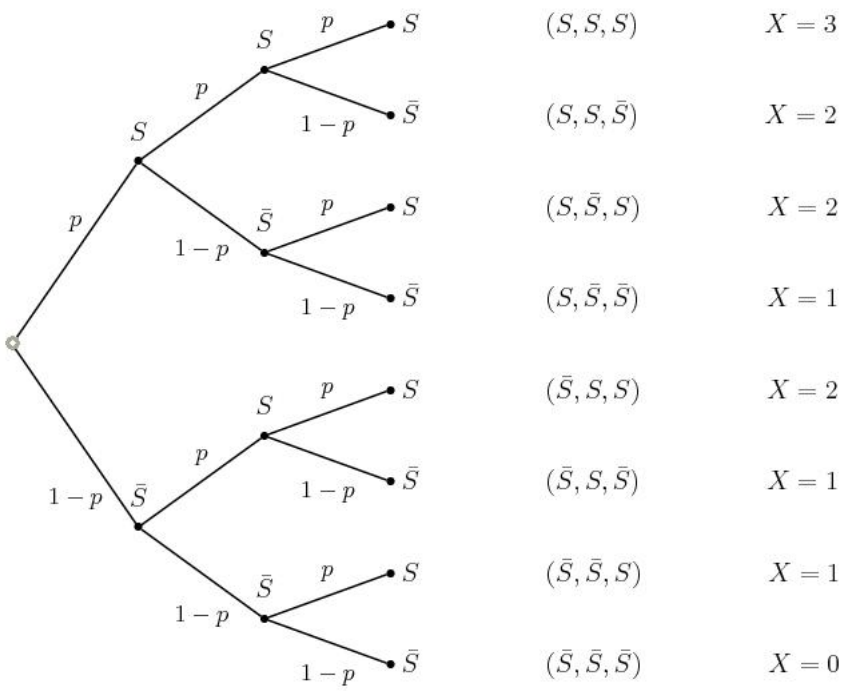
* $n = 3$: il s'agit en effet de la répétition de **trois** épreuves de Bernoulli.

* $p = \frac{1}{6}$: en effet, la probabilité du succès de chaque épreuve *(ici la probabilité que le 6 sorte)* est égale à $\frac{1}{6}$.

On dit que cette variable aléatoire suit la loi $\mathcal{B}(3, \frac{1}{6})$

II - 2) coefficients binomiaux

exemple : on considère un schéma de Bernoulli constitué de trois épreuves de Bernoulli identiques. On pose $p = Prob(S)$. X est la variable aléatoire comptant le nombre de succès. Cette situation peut être représentée par l'arbre ci dessous :



* probabilité que trois succès aient lieu :

Il n'y a qu'une seule manière d'obtenir trois succès (chemin du haut) ; les trois succès sont obtenus à la probabilité p^3 ; on note : $p(X = 3) = p^3$

* probabilité que deux succès aient lieu :

Il y a trois manières d'obtenir cette probabilité (trois chemins sur l'arbre) : on note $\binom{3}{1}$ pour signifier qu'on comptabilise le nombre de succès (1 ici) parmi le nombre d'essais (3 ici).

Ce nombre de chemins $\binom{3}{1}$ est appelé **coefficient binomial**.

Chaque issue comptant deux succès et un échec a une probabilité égale à $p^2(1-p)$.

Finalement, $p(X = 1) = \binom{3}{1}p^2(1-p) = 3p^2(1-p)$

définition 2 :

On considère l'arbre associé à un schéma de Bernoulli constitué par la répétition de n expériences.

On note $\binom{n}{k}$ le nombre de chemins de l'arbre réalisant k succès.

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

$\binom{n}{k}$ se lit : « k *parmi* n ».

En pratique : les calculatrices usuelles permettent d'obtenir les nombres $\binom{n}{k}$.

Pour les petites valeurs de n, ces nombres peuvent être calculés directement à l'aide d'un arbre.

Par exemple, en utilisant l'arbre précédent, on obtient :

$\binom{3}{3} = 1 \quad \binom{3}{2} = 3 \quad \binom{3}{1} = 3 \quad \binom{3}{0} = 1$

II - 3) formule de la loi binomiale

théorème 1 :

Si X est une variable aléatoire suivant la loi binomiale $\mathcal{B}(n,p)$, alors :

pour tout entier $k = 0, 1, 2, ..., n$, $\boxed{Prob(X = k) = \binom{n}{k} p^k (1-p)^{n-k}}$

démonstration :

L'événement $(X = k)$ est associé à l'ensemble des chemins pour lesquels il y a k succès et $n-k$ échecs. Chacun de ces chemins a une probabilité égale au produit des probabilités inscrite sur les branches qui le constituent, c'est-à-dire à $p^k(1-p)^{n-k}$.
Or, il a $\binom{n}{k}$ chemins de ce type. D'où : $Prob(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$

exemple : on reprend l'exemple du paragraphe I.2 concernant la variable aléatoire Y.

Y suit une loi $\mathcal{B}(3, \frac{1}{6})$.

On peut directement déterminer les probabilités suivantes :

* réalisation de trois échecs : $P(Y=0) = \binom{3}{0} \left(\frac{1}{6}\right)^0 \left(1-\frac{1}{6}\right)^{3-0} = 1 \times 1 \times \left(\frac{5}{6}\right)^3 \approx 57,9\%$

* réalisation d'un succès : $P(Y=1) = \binom{3}{1} \left(\frac{1}{6}\right)^1 \left(1-\frac{1}{6}\right)^{3-1} = 3 \times \frac{1}{6} \times \left(\frac{5}{6}\right)^2 \approx 34,7\%$

* réalisation de deux succès : $P(Y=2) = \binom{3}{2} \left(\frac{1}{6}\right)^2 \left(1-\frac{1}{6}\right)^{3-2} = 3 \times \left(\frac{1}{6}\right)^2 \times \frac{5}{6} \approx 6,9\%$

* réalisation de trois succès : $P(Y=3) = \binom{3}{3} \left(\frac{1}{6}\right)^3 \left(1-\frac{1}{6}\right)^{3-3} = 1 \times \left(\frac{1}{6}\right)^3 \times 1 \approx 0,5\%$

On peut vérifier que la somme de ces probabilités est égale à 100% (autrement dit à 1).

II - 4) espérance d'une loi binomiale

théorème 2 :

Si une variable aléatoire X suit la loi binomiale $\mathcal{B}(n,p)$, alors :

$$\boxed{E(X) = np}$$

démonstration :

Ce théorème est admis.

exemple : l'espérance de la loi Y qui suit une loi binomiale $\mathcal{B}(3, \frac{1}{6})$ est égale à :

$E(Y) = 3 \times \frac{1}{6} = \frac{1}{2}$.

On peut interpréter ce résultat en disant que la « moyenne »de la variable aléatoire Y est égale à $\frac{1}{2}$, autrement dit, que si on répète un grand nombre de fois l'expérience aléatoire consistant à lancer trois fois de suite un dé, la valeur moyenne du nombre de 6 sortis sera proche de $\frac{1}{2}$.

III Échantillonnage

III - 1) un exemple pour comprendre

On reprend le jeu présenté au début de ce cours, à savoir un lancer de dé. On appellera succès le fait d'obtenir un 6.
A présent, on va répéter cette épreuve de Bernoulli 100 fois.
On note X la variable aléatoire qui compte le nombre de succès ; X suit une loi $\mathcal{B}(100, \frac{1}{6})$.

Jean dit : « *sur 100 lancers de dé, je suis sûr d'obtenir entre 7 fois et 27 fois le 6.* »
Paul affirme lui : « *sur 100 lancers, je suis certain d'obtenir entre 15 fois et 20 fois le 6.* »
Que veut dire « être sûr » ? Dans ce cas, la seule chose dont on est sûr à 100% est d'obtenir entre 0 et 100 fois le 6 !

On va essayer d'affiner ce raisonnement, en se donnant un **seuil** de 95 % par exemple.
Cela signifie que l'on va chercher l'intervalle donnant à une probabilité de 95% le nombre de sorties de 6.

Méthode :

1. on calcule $p(X = k)$ pour k allant de 0 à 100.

2. on calcule les probabilités $p(X \leq k)$ pour k allant de 0 à 100 : il suffit d'additionner les probabilités calculées précédemment.

3. on cherche la plus petite valeur k telle que $p(X \leq k) > 2,5\%$: on note k_1 cette valeur. *(2,5% correspond à la moitié de 5%)*

4. on cherche la plus petite valeur de k telle que $p(X \leq k) \geq 97,5\%$: on note k_2 cette valeur. *(97,5% correspond à 100%-2,5%)*

5. l'intervalle $[k_1; k_2]$ est l'intervalle de fluctuation au seuil de 95%.

Table de la loi $\mathcal{B}(100, \frac{1}{6})$:

On représente ici un extrait de calculs des probabilités d'une variable aléatoire suivant la loi $\mathcal{B}(100, \frac{1}{6})$.

La ligne correspondant à $X = 16$ signifie :

* la probabilité que le 6 sorte exactement 16 fois sur 100 lancers est égale à 10,65%.

* la probabilité que le 6 sorte moins de 16 fois *(y compris 16 fois)* sur 100 lancers est égale à 49,42%.

X	p(X=k)	p(X≤k)
3	0,00%	0,00%
4	0,01%	0,01%
5	0,03%	0,04%
6	0,09%	0,13%
7	0,25%	0,38%
8	0,58%	0,95%
9	1,18%	2,13%
10	2,14%	4,27%
11	3,50%	7,77%
12	5,20%	12,97%
13	7,03%	20,00%
14	8,74%	28,74%
15	10,02%	38,77%
16	10,65%	49,42%
17	10,52%	59,94%
18	9,71%	69,65%
19	8,38%	78,03%
20	6,79%	84,81%
21	5,17%	89,98%
22	3,71%	93,69%
23	2,52%	96,21%
24	1,62%	97,83%
25	0,98%	98,81%
26	0,57%	99,38%
27	0,31%	99,69%
28	0,16%	99,85%
29	0,08%	99,93%
30	0,04%	99,97%
31	0,02%	99,99%
32	0,01%	100,00%

Recherche de l'intervalle de fluctuation au seuil de 95% pour la loi $\mathcal{B}(100, \frac{1}{6})$:

* le plus petit entier k tel que $p(X \leq k) > 2,5\%$ est égal ici à 10.

* le plus petit entier k tel que $p(X \leq k) \geq 97,5\%$ est égal à ici à 24.

* pour la loi $\mathcal{B}(100, \frac{1}{6})$, l'intervalle de fluctuation au seuil de 95% est $[10\,;\,24]$: cela signifie que la probabilité pour que le 6 sorte entre 10 et 24 fois sur 100 lancers est environ égale à 95%.

Autrement dit, « *on est sûr à 95%* »que sur 100 lancers, le 6 sortira entre 10 et 24 fois.

remarque : Jean avait raison (au seuil de 95% mais son intervalle pouvait être affiné).
Paul a tort au seuil de 95% ... il n'a raison qu'au seuil de 56%... [1]

Interprétation graphique :

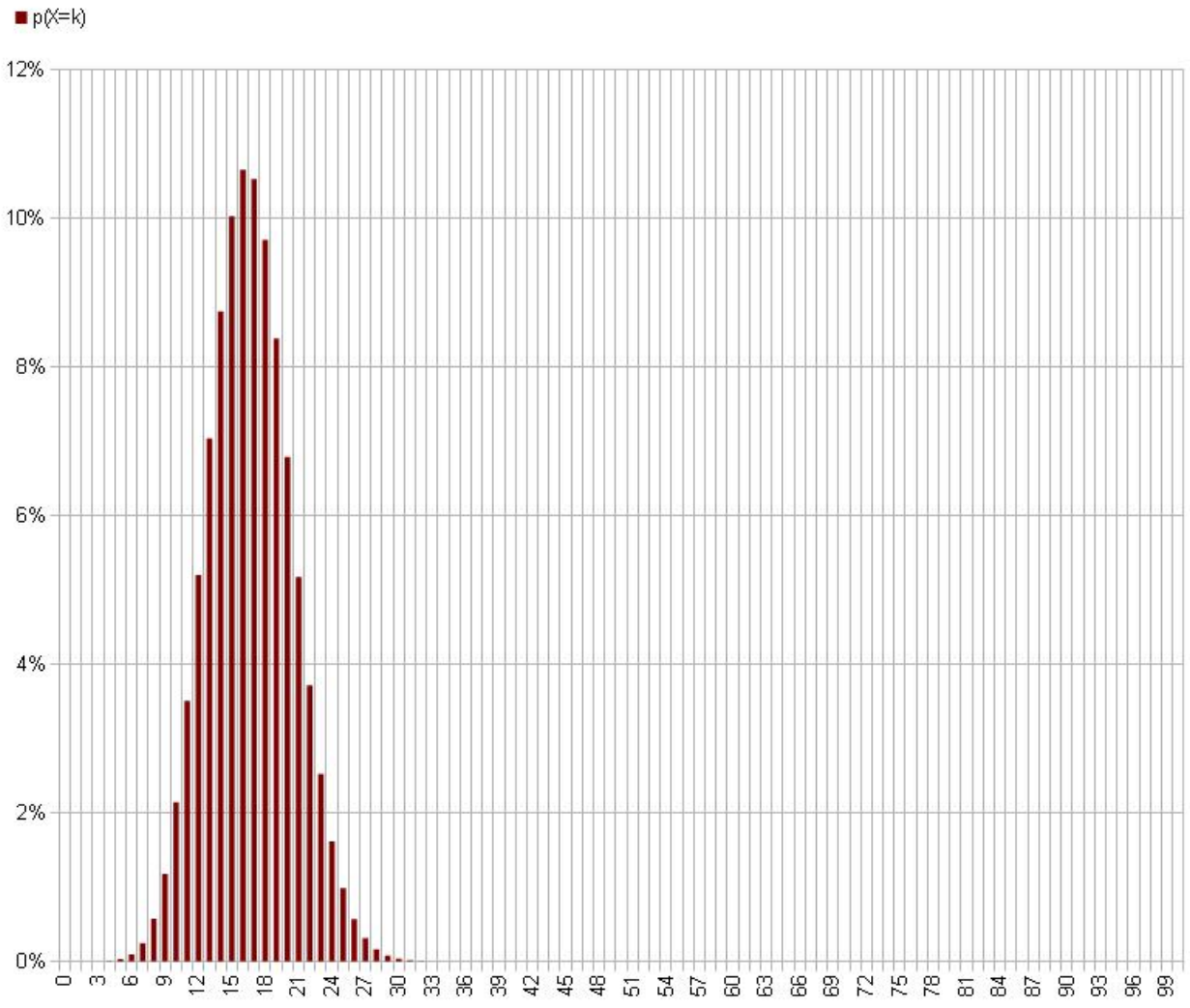


diagramme représentant la loi $\mathcal{B}(100, \frac{1}{6})$

remarque : certains bâtons ne sont pas visibles car leur hauteur est trop petite.

1. la somme des probabilités $p(X = k)$ pour k allant de 15 à 20 est environ égale à 56%

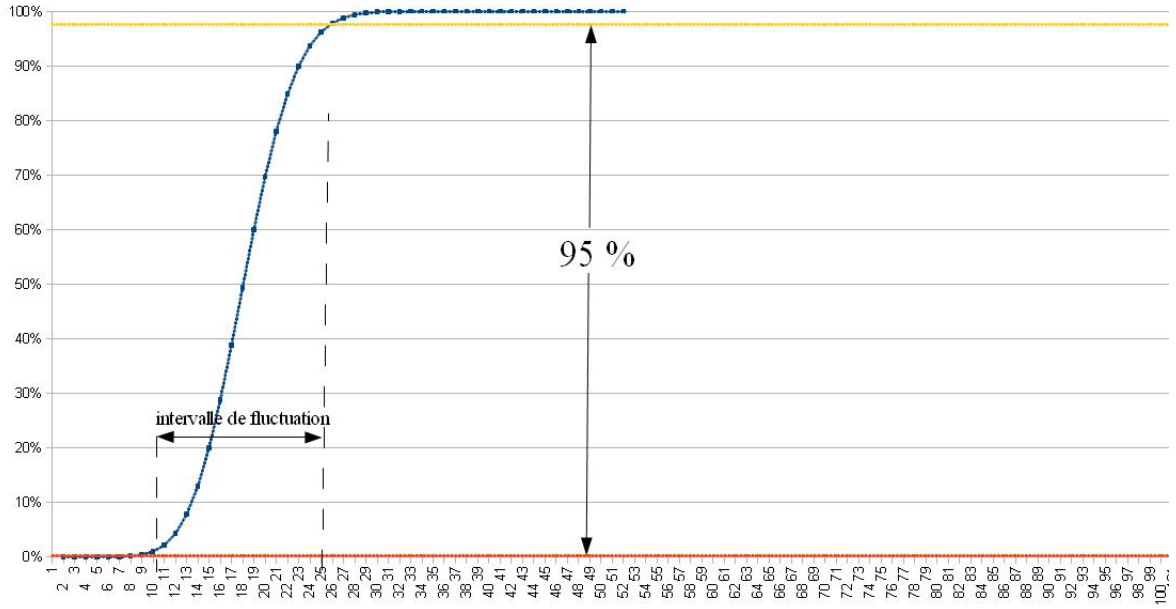


diagramme représentant les probabilités cumulées de la loi $\mathcal{B}(100, \frac{1}{6})$

Pour déterminer l'intervalle de fluctuation au seuil de 95%, on peut chercher l'intersection avec la probabilité 2,5% et l'intersection avec la probabilité 97,5%. L'intervalle restant correspond à une probabilité totale proche de 95%.

III - 2) prise de décision

Pierre fait l'expérience : il lance 100 fois un dé et obtient 28 fois le 6.

Or, 28 ne fait pas partie de l'intervalle de fluctuation déterminé précédemment (qui était $[10\,;\,24]$).

On peut donc affirmer que son dé n'est pas équilibré, **au risque de 5% de se tromper**.

III - 3) cas des grands échantillons

On considère une population où la proportion d'un caractère est p comprise entre 0,2 et 0,8.

On considère un échantillon issu de cette population, échantillon de taille n, avec $n \geq 25$.

On note f la fréquence observée du caractère dans cet échantillon.

L'intervalle de fluctuation au seuil de 95% de la fréquence observée peut être approximé par l'intervalle $\left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}\right]$.

Application :

On ne peut pas appliquer ce résultat à l'exemple décrit dans les paragraphes précédents de ce cours, car dans cet exemple, $p = \frac{1}{6} < 0,2$ *: on n'est pas dans les conditions d'application de la propriété.*

On prend un autre exemple : on lance une pièce et on note si elle tombe sur PILE ou sur FACE.

La probabilité « théorique »pour que la pièce tombe sur PILE est $p = \frac{1}{2} = 0,5$.

On fait 100 essais : *cela signifie que l'on prend un échantillon de taille* $n = 100$.

On est bien dans les conditions d'application de la propriété précédente ($0,2 \leq p \leq 0,8$ et $n \geq 25$).

L'intervalle de fluctuation au seuil de 95% peut être approximé par : $\left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}\right]$, ce qui donne ici :

$$\left[0,5 - \frac{1}{\sqrt{100}}; 0,5 + \frac{1}{\sqrt{100}}\right] = [0,5 - 0,1; 0,5 + 0,1] = [0,4; 0,6]$$

C'est-à-dire que sur 100 lancers, on a 95% de chances d'obtenir entre 40 et 60 fois PILE.

Si on fait l'expérience et que l'on obtient moins de 40 PILE ou plus de 60 PILE, on pourra conclure (avec un risque d'erreur de 5%) que la pièce n'est pas bien équilibrée.