Chapitre 10

Loi binomiale - Échantillonnage

I. Loi binomiale

1) *Schéma de Bernoulli*

Épreuve de Bernoulli

Définition :
Une **épreuve de Bernoulli** est une expérience aléatoire à deux issues appelées « Succès » et « Échec ». On dit qu'une épreuve de Bernoulli est de paramètre *p* si la probabilité de l'issue « Succès » est *p*.

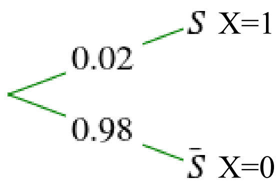
Exemple :
On lance un dé équilibré à six faces et on considère comme un succès d'obtenir un 1.
Cette expérience aléatoire est une épreuve de Bernoulli de paramètre $\frac{1}{6}$.

Remarque :
On utilisera communément la lettre *q* pour désigner la probabilité d'un échec.
« Succès » et « Échec » étant des événements contraires, on a donc $q=1-p$.
Dans l'exemple précédent, un échec a la probabilité $q=1-\frac{1}{6}=\frac{5}{6}$.

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

Exemple :
Dans une usine, la probabilité qu'un article fabriqué présente un défaut est 0,02.
Soit *X* la variable aléatoire qui vaut 1 si l'article présente un défaut et 0 sinon.
X suit une loi de Bernoulli de paramètre $p=0{,}02$.

x_i	0	1
$p(X=x_i)$	0,98	0,02



Propriété :
Soit *X* une variable aléatoire de Bernoulli de paramètre *p*.
- L'**espérance** de X est $E(X)=p$.
- La **variance** de X est $V(X)=p(1-p)$.
- L'**écart-type** est $\sigma(X)=\sqrt{p(1-p)}$.

Démonstrations :
$E(X)=0\times(1-p)+1\times p=p$ et $V(X)=0^2\times(1-p)+1^2\times p-p^2=p-p^2=p(1-p)$.

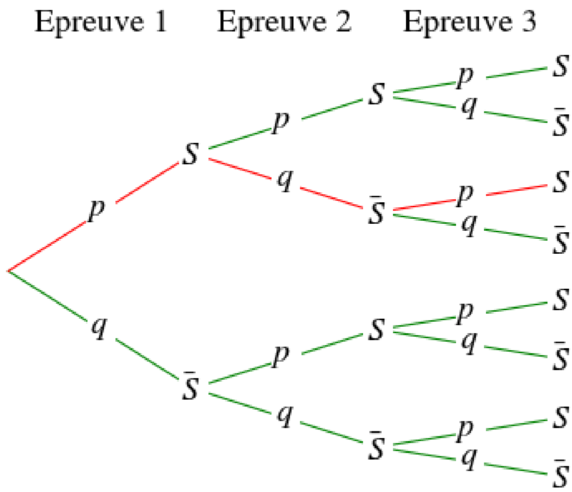
Schéma de Bernoulli

Définition :
Soit n un entier naturel non nul et p un nombre réel appartenant à l'intervalle $[0;1]$.
Un **schéma de Bernoulli** est une expérience consistant à répéter n fois, de manière **indépendante**, la même épreuve de Bernoulli.
Un schéma de Bernoulli a deux paramètres : n le nombre de répétitions de l'épreuve et p le paramètre de l'épreuve répétée.

Propriété :
On peut représenter un schéma de Bernoulli de paramètre n et p par un **arbre de probabilité** à 2^n branches.
Les issues sont des n-uplets dont les n termes sont S pour « Succès » ou $\overline{S}$ pour « Échec ».

Exemple :
Ci-contre un schéma de Bernoulli pour $n=3$.

L'issue correspondant au chemin rouge peut être notée $(S, \overline{S}, S)$.



Propriété :
La probabilité d'une issue d'un schéma de Bernoulli s'obtient en faisant le produit des probabilités des issues obtenues à chaque épreuve de Bernoulli.

Exemple :
Si un schéma de Bernoulli a pour paramètre $n=4$ et $p=0{,}3$, alors l'issue $(S, \overline{S}, \overline{S}, S)$ a pour probabilité $0{,}3\times0{,}7\times0{,}7\times0{,}3=0{,}0441$.

2) Coefficients binomiaux

Définition :
Une expérience suit un schéma de Bernoulli de paramètre n et p.
k est un entier naturel tel que $0 \leqslant k \leqslant n$.
On appelle **coefficient binomial**, ou combinaison de k parmi n, le nombre de chemins conduisant à k succès sur l'arbre représentant l'expérience.
Ce nombre se note $\binom{n}{k}$ et se lit « k parmi n ».

Remarque :
Une issue de l'expérience ayant k succès étant un n-uplet de la forme $(S, S, \overline{S}, \ldots, \overline{S}, S)$, elle contient k termes S et $n-k$ termes $\overline{S}$. Compter les issues ayant k succès revient à dénombrer les façons de choisir les places des succès S dans la liste des n termes.
D'où le terme « **combinaison** » car on cherche à dénombrer toutes les combinaisons possibles de S et $\overline{S}$.

Propriétés :
Une expérience suit un schéma de Bernoulli de paramètre n et p.
k est un entier naturel tel que $0 \leqslant k \leqslant n$.
On a les résultats suivants :

$$\binom{n}{0} = 1 \quad ; \quad \binom{n}{n} = 1 \quad ; \quad \binom{n}{1} = n \quad ; \quad \binom{n}{n-k} = \binom{n}{k}$$

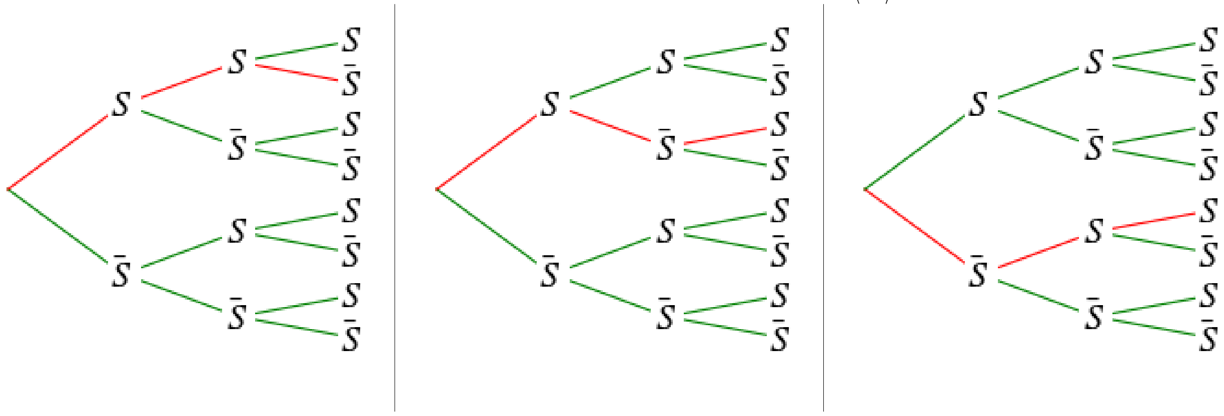
Démonstrations :

- $\binom{n}{0} = 1$ car il n'y a qu'un seul chemin réalisant 0 succès : celui ne comportant que des échecs.
- $\binom{n}{n} = 1$ car il n'y a qu'un seul chemin réalisant n succès : celui ne comportant que des succès.
- $\binom{n}{1} = n$ car il y a n chemins réalisant 1 succès. En effet, les n-uplets réalisant un seul succès ne diffèrent que par la place qu'occupe l'unique succès dans la liste des issues. Il y a n choix possibles pour placer S.
- $\binom{n}{n-k} = \binom{n}{k}$ car lorsqu'il y a $n-k$ succès, il y a k échecs. Compter les chemins menant à $n-k$ succès revient à compter ceux menant à k échecs. Dénombrer les façons de placer k échecs parmi n termes revient à calculer la combinaison $\binom{n}{k}$.

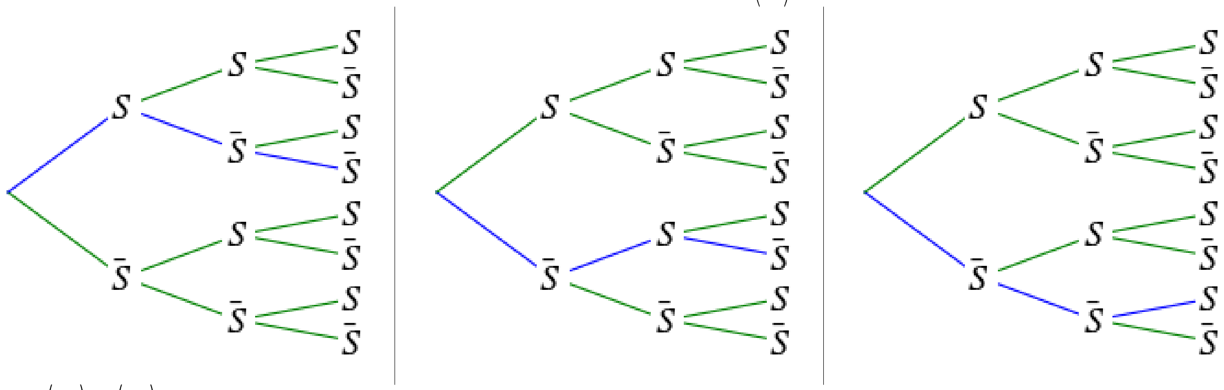
Exemple :

On considère un schéma de Bernoulli pour $n=3$.

On peut noter en rouge les chemins correspondant à la combinaison $\binom{3}{2}$.



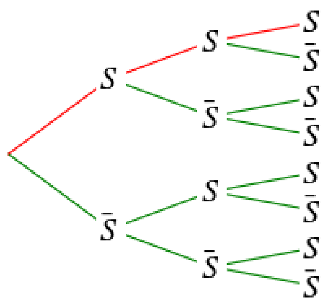
On peut noter en bleu, ceux correspondant à la combinaison $\binom{3}{1}$.



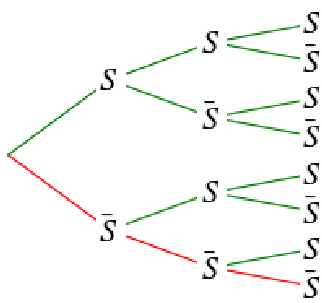
On a $\binom{3}{2}=\binom{3}{1}=3$.

On retrouve aussi les résultats suivants :

- $\binom{3}{3}=1$ en suivant le chemin (S,S,S) .



- $\binom{3}{0}=1$ en suivant le chemin $(\bar{S},\bar{S},\bar{S})$.



Théorème :
Une expérience suit un schéma de Bernoulli de paramètre n et p.
k est un entier naturel tel que $0 \leqslant k \leqslant n$.
On a la relation suivante :
$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$$

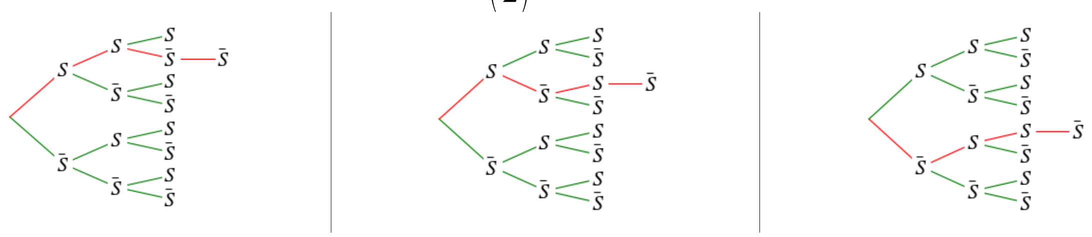
Démonstration :
Les chemins comportant $k+1$ succès après $n+1$ répétitions de l'épreuve de Bernoulli sont de deux types : ceux pour lesquels la dernière épreuve $(n+1)^{ième}$ donne un succès et ceux pour lesquels la dernière épreuve donne un échec.

- Si la $(n+1)^{ième}$ épreuve donne un succès, alors pour avoir un total de $k+1$ succès, il faut que les n épreuves précédentes aient données k succès. Il y a donc $\binom{n}{k}$ combinaisons possibles.
- Si la $(n+1)^{ième}$ épreuve donne un échec, alors pour avoir un total de $k+1$ succès, il faut que les n épreuves précédentes aient déjà donné $k+1$ succès. Il y a donc $\binom{n}{k+1}$ combinaisons possibles.

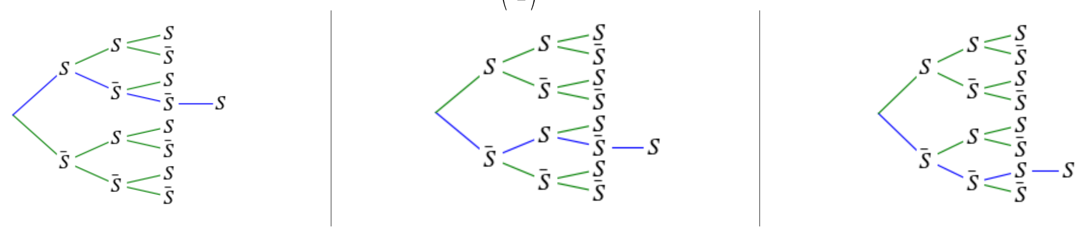
Les ensembles de ces deux types de chemins sont disjoints, on en déduit donc que :
$$\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}$$

Exemple :
Pour calculer $\binom{4}{2}$, on utilise le schéma de Bernoulli avec $n=3$.
Les chemins comportant 2 succès parmi 4 proviennent :

- des chemins en rouge comportant 2 succès parmi les 3 premières épreuves ; la quatrième épreuve sera alors un échec. Il y en a $\binom{3}{2}$.



- des chemins en bleu comportant 1succès parmi les 3 premières épreuves ; la quatrième épreuve sera alors un succès. Il y en a $\binom{3}{1}$.



On obtient :
$$\binom{4}{2} = \binom{3}{1} + \binom{3}{2} = 3+3=6$$

Théorème :
Une expérience suit un schéma de Bernoulli de paramètre n et p.
k est un entier naturel tel que $0 \leqslant k \leqslant n$.
On peut déterminer de proche en proche toutes les combinaisons à l'aide de la relation $\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$, en construisant le **triangle de Pascal** partiellement représenté ci-contre.

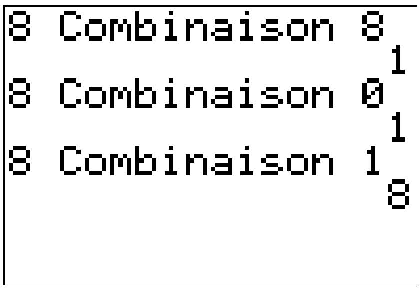
n \ k	0	1	2	3	4	5
1	1	1				
2	1	2	1			
3	1	3	3	1		
4	1	4	6	4	1	
5	1	5	10	10	5	1

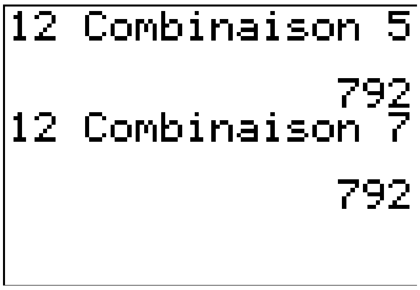
Démonstration :
Si l'on connaît tous les coefficients binomiaux pour une répétition de n épreuves de Bernoulli, on peut calculer tous ceux correspondant à $n+1$ répétitions de l'épreuve grâce à la relation $\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$. On peut donc calculer n'importe quel coefficient binomial en itérant le processus de 1 à n. On peut initier le processus car on sait que $\binom{1}{0} = 1$ et $\binom{1}{1} = 1$.

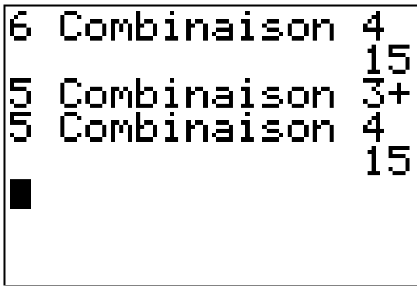
Exemple :
En prolongeant le tableau précédent, on peut calculer toutes les combinaisons correspondant à $n = 6$. On a par exemple :

$$\binom{6}{4} = \binom{5}{3} + \binom{5}{4} = 10 + 5 = 15$$

Calculatrice :







3) Loi binomiale

Définition :
Une expérience suit un schéma de Bernoulli de paramètre n et p.
k est un entier naturel tel que $0 \leqslant k \leqslant n$.
On associe à l'expérience la variable aléatoire X qui donne le nombre total de succès.
La loi de probabilité de X est appelée **loi binomiale** de paramètre n et p.
On la note $\mathscr{B}(n,p)$.

Propriété :
Si une variable aléatoire X suit une loi binomiale $\mathscr{B}(n,p)$, alors pour tout entier k compris entre 0 et n, la probabilité que X soit égal à k est :

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

Démonstration :
On sait que tous les chemins comportant k succès sont équiprobables car, en faisant le produit des probabilités des n issues de chaque épreuve de Bernoulli, on obtient k facteurs p (pour k succès) et $n-k$ facteurs $(1-p)$ (pour $n-k$ échecs). Leur probabilité est donc $p^k(1-p)^{n-k}$.
Il suffit ensuite de compter les chemins menant à k succès : il y en a $\binom{n}{k}$. On obtient donc :

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

Exemple :
On lance 3 fois un dé équilibré à six faces et on considère comme un succès d'obtenir un 6.
En nommant X la variable aléatoire donnant le nombre de succès, X suit une loi binomiale $\mathscr{B}(n,p)$.

On peut calculer $p(X=1)=\binom{3}{1}\times\left(\frac{1}{6}\right)^1\times\left(\frac{5}{6}\right)^2=3\times\frac{1}{6}\times\frac{25}{36}=\frac{75}{216}=\frac{25}{72}$.

Nombre de chemins avec un seul succès

Probabilité d'un succès

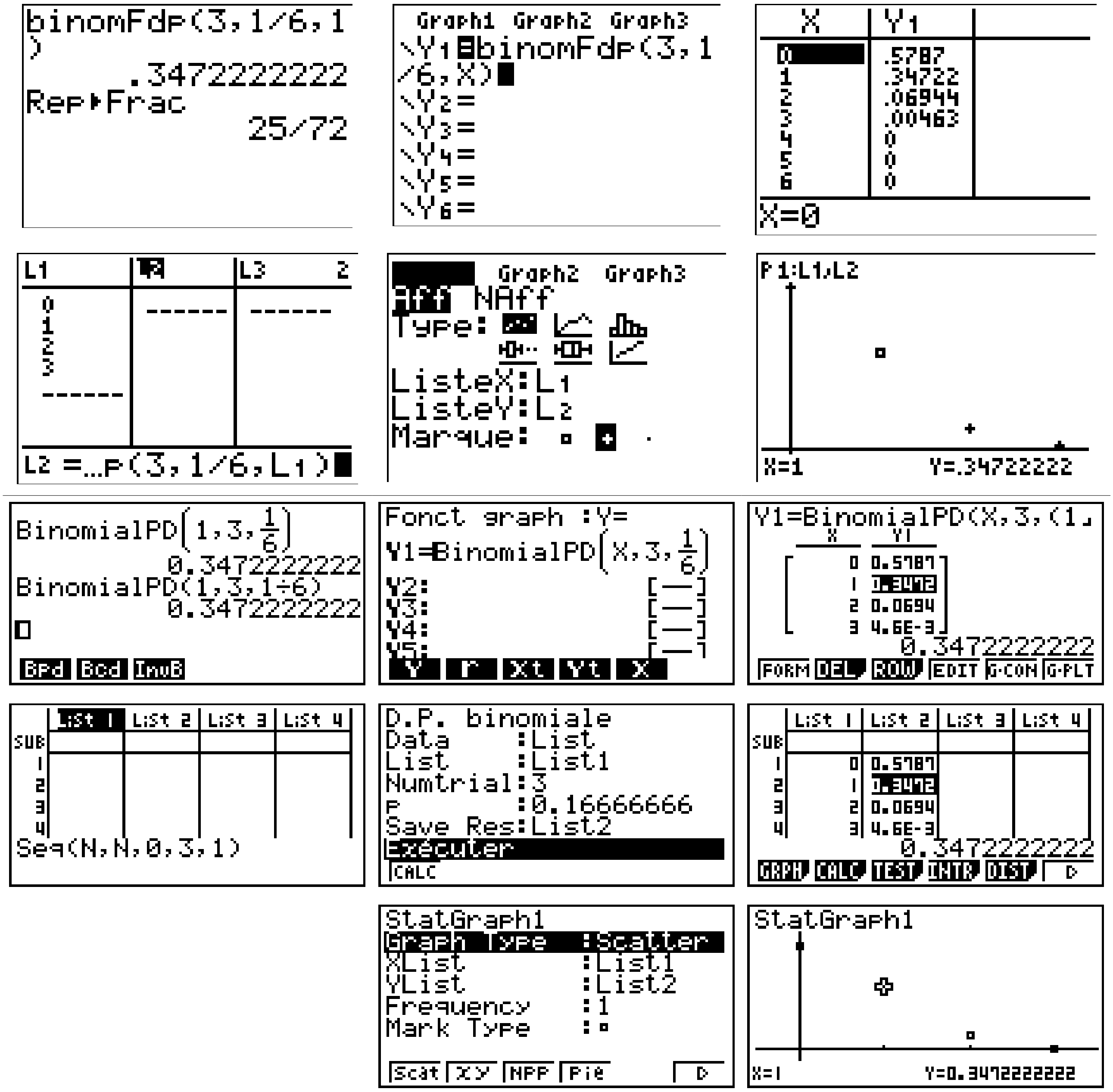
Probabilité de deux échecs

La probabilité d'obtenir exactement un 6 en trois lancers est $\frac{25}{72}$.

On peut de même déterminer la loi de probabilité de X :

x_i	0	1	2	3
$p(X=x_i)$	$\frac{125}{216}$	$\frac{75}{216}$	$\frac{15}{216}$	$\frac{1}{216}$

Calculatrice :



Propriétés :
Soit une variable aléatoire X qui suit une loi binomiale $\mathscr{B}(n,p)$.
$E(X)=n\times p$; $V(X)=n\times p\times q$ où $q=1-p$

Exemple :
En reprenant l'exemple précédent, on peut calculer l'espérance grâce à la loi de probabilité de X.
$E(X)=0\times\frac{125}{216}+1\times\frac{75}{216}+2\times\frac{15}{216}+3\times\frac{1}{216}=\frac{108}{216}=\frac{1}{2}$. On vérifie que $n\times p=3\times\frac{1}{6}=\frac{1}{2}$.

Remarque :
On peut remarquer que la formule de l'espérance peut s'expliquer sans calcul. En effet, chaque épreuve de Bernoulli a pour espérance de succès p donc, en la répétant n fois, on peut espérer obtenir en moyenne $n\times p$ succès.
Dans l'exemple précédent, le 6 sort avec la probabilité $\frac{1}{6}$, donc on peut espérer tripler ce résultat en triplant l'expérience car ce sont des expériences successives et indépendantes et $E(X)=3\times\frac{1}{6}=\frac{1}{2}$.

II. Échantillonnage

1) Échantillon

Définition :
Un **échantillon** de taille n est obtenu en prélevant au hasard, **successivement** et **avec remise**, n éléments d'une population.

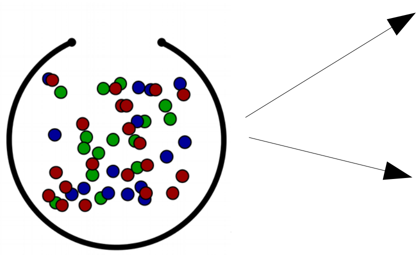
Remarques :

- Après prélèvement d'un échantillon, on s'intéresse à la valeur d'un caractère des éléments de la population. Nous nous restreignons ici à des populations dont le caractère étudié n'a que **deux valeurs possibles**.
- **Exemples de situations correspondants à un prélèvement d'échantillon** :
  - Prélever des pièces dans une production de manière identique et indépendante, noter à chaque fois si la pièce présente un défaut ou non et la remettre dans la production.
  - Lancer plusieurs fois un dé et noter à chaque fois si la face supérieure est un 6 ou non.
  - Lancer plusieurs fois de manière indépendante une pièce de monnaie et noter si elle affiche « pile » ou « face ».
  - Sortir au hasard de manière indépendante une boule dans une urne qui ne contient que des boules rouges et des boules d'autres couleurs et noter à chaque fois si elle est rouge ou non.
- Souvent, il n'y a pas de remise lors du prélèvement. Mais lorsque l'effectif total est très grand par rapport au nombre d'objets prélevés, on considère néanmoins que l'échantillon est constitué, au sens de la définition donnée, avec remise.
  Exemples de situations assimilées à un prélèvement d'échantillon :
  - Lors d'un prélèvement de pièces dans une production, après avoir constaté qu'une pièce a un défaut, il n'est pas envisageable de la remettre parmi l'ensemble des pièces produites.
  - Lors d'un sondage « sortie des urnes », on ne peut pas attendre que tout le monde ait voté et réunir toutes les personnes avant de choisir au hasard des individus pour obtenir un échantillon.

2) Intervalle de fluctuation

On souhaite donner les caractéristiques d'un échantillon à partir de celles connues de la population dans laquelle il a été prélevé. C'est ce qu'on nomme l'**échantillonnage**.

Exemple :
Urne avec 40 % de boules rouges.
Le caractère « rouge » a, dans la population, une proportion de $p=0,4$.



Premier échantillon de taille $n=15$.
Nombres de boules ayant le caractère « rouge » : $k=6$.
Fréquence observée : $f=\frac{6}{15}=0,4$.

Deuxième échantillon de taille $n=15$.
Nombres de boules ayant le caractère « rouge » : $k=4$.
Fréquence observée : $f=\frac{4}{15}\simeq 0,27$.

Propriété :
Soit une population dont une proportion p des éléments admet un caractère donné.
Dans un **échantillon de** taille n prélevé dans cette population, l'**effectif** des éléments qui présentent ce **caractère** est une variable aléatoire qui suit la **loi binomiale** de paramètres n et p.

Démonstration :

- Le prélèvement d'un échantillon de taille n peut être assimilé à un tirage successif avec remise dès lors que l'effectif de la population est assez grand par rapport à n.
  On peut donc considérer que l'expérience est une répétition de n tirages identiques à deux issues : avoir ou ne pas avoir le caractère choisi.
- La variable aléatoire X donnant le nombre d'éléments qui ont le caractère choisi suit alors une loi binomiale de paramètre n (nombre de répétition d'un tirage) et p (probabilité de prélever un élément ayant le caractère choisi).

Remarque :
Pour cette loi binomiale de paramètre n et p, la somme des probabilités

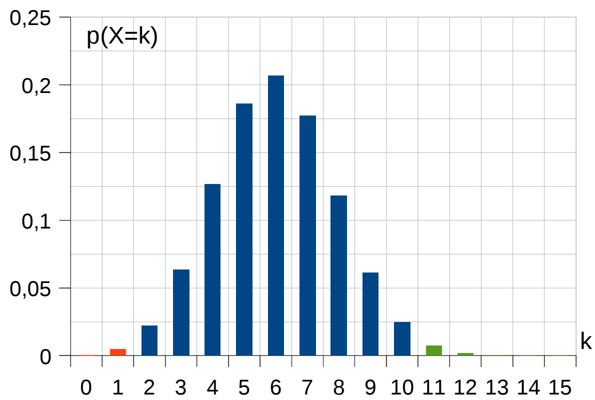
$\sum_{k=0}^{n} p(X=k)=p(X=0)+p(X=1)+\ldots+p(X=n)$ vaut 1.

Sur la représentation graphique ci-contre :

- La somme des probabilités des parties vertes est inférieure à 0,025.
- La somme des probabilités des parties rouges est inférieure à 0,025.

Pour que **la somme des probabilités des parties bleues soit supérieure où égale à 0,95 tout en étant la plus petite possible**, on détermine :

- a le plus petit entier tel que $p(X\leqslant a)>0{,}025$.
- b le plus petit entier tel que $p(X\leqslant b)\geqslant 0{,}975$.



Loi binomiale de paramètre $n=15$ et $p=0{,}4$

Définition :
Soit X la loi binomiale qui correspond à la réalisation d'un échantillon de taille n dans une population ayant une proportion p d'éléments avec un caractère donné.
On s'intéresse aux fréquences de tels échantillons, c'est-à-dire aux proportions des éléments des échantillons ayant le caractère donné.
L'**intervalle de fluctuation au seuil de 95 % de la fréquence observée** est l'intervalle $\left[\frac{a}{n};\frac{b}{n}\right]$ où a est le plus petit entier tel que $p(X\leqslant a)>0{,}025$ et b le plus petit entier tel que $p(X\leqslant b)\geqslant 0{,}975$.

Remarques :

- L'intervalle $[a;b]$ est appelé **intervalle de fluctuation au seuil de 95 % de l'effectif**.
- On peut dire « au seuil de 95 % » ou « à 95 % » ou « au seuil 0,95 ».
- La probabilité pour que la fréquence d'un échantillon de taille n, pris parmi tous les échantillons de taille n, soit dans l'intervalle $\left[\frac{a}{n};\frac{b}{n}\right]$ est environ de 0,95.
- On peut définir, de même, des intervalles de fluctuation à un seuil autre que 95 %.

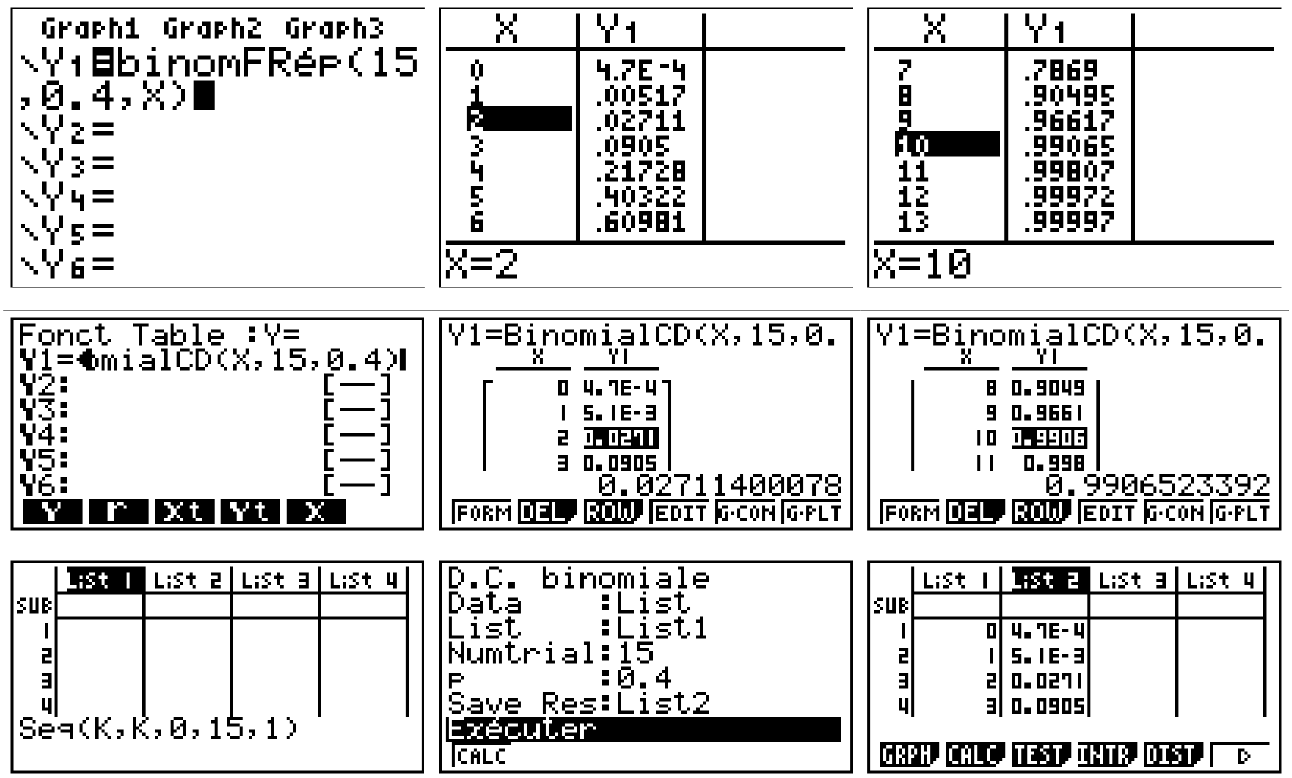
Exemple :
Dans une urne contenant 40 % de boules rouges, on prélève un échantillon de taille $n=15$.
Soit X la variable aléatoire qui compte le nombre de boules rouges dans l'échantillon prélevé.
Les valeurs possibles de X sont les entiers compris entre 0 et 15.
La loi de probabilité de X est donnée par :

			Intervalle de fluctuation													
k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$p(X=k)$	0,000	0,005	0,022	0,063	0,127	0,186	0,207	0,177	0,118	0,061	0,025	0,007	0,002	0,000	0,000	0,000
	0,027															
	0,991															

- $a=2$ car $p(X\leqslant 1)\simeq 0{,}005$ et $p(X\leqslant 2)\simeq 0{,}027$. D'où $\frac{a}{n}\simeq 0{,}13$.
- $b=10$ car $p(X\leqslant 9)\simeq 0{,}966$ et $p(X\leqslant 10)\simeq 0{,}991$. D'où $\frac{b}{n}\simeq 0{,}67$.

L'intervalle $[0{,}13;0{,}67]$ est un intervalle de fluctuation de la fréquence observée au seuil 0,95.

Calculatrice :



Propriété :
Pour un échantillon de grande taille $(n\geqslant 30)$ ayant une proportion du caractère p comprise entre 0,2 et 0,8, l'intervalle $\left[p-\frac{1}{\sqrt{n}};p+\frac{1}{\sqrt{n}}\right]$ est une bonne approximation de l'intervalle de fluctuation au seuil de 95 % de la fréquence observée f du caractère.

Exemple :
Dans l'exemple précédent, l'intervalle de fluctuation au seuil de 95 % de la fréquence observée du caractère est donc $[0{,}13;0{,}67]$.

L'intervalle donné par la propriété ci-dessus est $\left[0{,}4-\frac{1}{\sqrt{15}};0{,}4+\frac{1}{\sqrt{15}}\right]$, ce qui donne ici $[0{,}14;0{,}66]$.

C'est une bonne approximation de l'intervalle de fluctuation au seuil de 95 % de la fréquence observée. Mais cette approximation n'est pas très bonne car n n'est pas supérieur à 30 comme demandé dans les hypothèses de la propriété.

3) Estimation

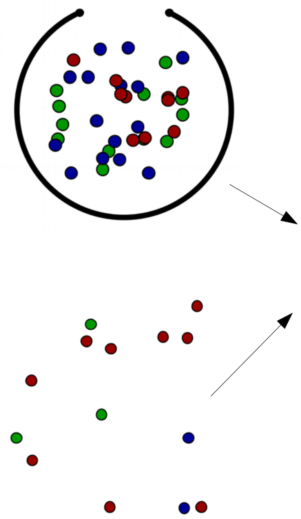
Contrairement à l'échantillonnage, l'estimation consiste à induire des caractéristiques de la population à partir de celles d'un échantillon.

Urne avec une proportion p, supposé connue, de boules rouges.
On émet l'hypothèse que la proportion est p.

Un échantillon de taille $n=15$.

Nombre de boules ayant le « caractère rouge » : $k=9$.

Fréquence observée : $f=\frac{9}{15}=0{,}6$.



Accepter ou rejeter, au seuil de 0,95, l'hypothèse (la valeur de p).

Exemples :

- Dans une production, on pense qu'il y a une proportion p de pièces défectueuses. Après un prélèvement de pièces, la proportion de pièces défectueuses trouvées dans l'échantillon peut permettre d'accepter ou de rejeter l'hypothèse, c'est-à-dire la valeur de p.
- Après avoir lancé n fois un dé supposé équilibré, la proportion de « six » trouvée va permettre d'accepter ou de rejeter l'hypothèse que le dé soit équilibré.
- Avant une élection, on annonce qu'un candidat aura une proportion p de voix. Un résultat d'un sondage « sortie des urnes » permet d'accepter ou de rejeter l'hypothèse, c'est-à-dire que p % des voix se porteront sur lui.

Prise de décision :

Situation :

Soit une population et un caractère présent ou non pour chaque individu.
On émet l'hypothèse que la proportion d'individus de la population ayant ce caractère est p.
On souhaite accepter ou refuser cette hypothèse.

Démarche :

La loi binomiale de paramètre n et p permet de trouver l'intervalle de fluctuation de 95 % de la fréquence observée.

(Éventuellement, on peut prendre l'intervalle $\left[p-\frac{1}{\sqrt{n}};p+\frac{1}{\sqrt{n}}\right]$ comme intervalle de fluctuation).

Énoncé de la règle de décision :

Dans un échantillon, notons f la fréquence du caractère étudié.

- Lorsque f appartient à l'intervalle I, on dit qu'au seuil de 95 %, on accepte l'hypothèse de la proportion dans la population est p.
- Lorsque f n'appartient pas à l'intervalle I, on dit qu'au seuil de 95 %, on réfute l'hypothèse de la proportion dans la population est p.

Prise de décision :

On prélève un échantillon de taille n dont on détermine la fréquence f du caractère étudié.
On applique la règle énoncée ci-dessus.

Exemple :
Une urne contient une proportion p de boules rouges.
On fait l'hypothèse que cette proportion vaut $p=0{,}4$.
Pour accepter ou rejeter cette hypothèse, on prélève un échantillon de taille $n=15$ et on constate que, pour cet échantillon, la fréquence f de boules ayant le caractère « rouge » vaut $\frac{9}{15}=0{,}6$.
L'intervalle de fluctuation de f au niveau 0,95 est $[0{,}13\,;0{,}67]$ (intervalle obtenu à l'aide de la loi binomiale de paramètres $n=15$ et $p=0{,}4$).
Comme 0,6 se trouve dans cet intervalle, on peut, au seuil de 95 %, accepter l'hypothèse que p est égale à 0,4.

Remarques :

- La méthode utilisée repose sur la propriété suivante : la probabilité pour que la fréquence d'un échantillon de taille n, pris parmi tous les échantillons de taille n, soit dans l'intervalle de fluctuation I au seuil de 95 % est de 0,95. Donc f appartient à I avec une probabilité de 0,95 et f n'appartient pas à I avec une probabilité de 0,05.
- Il ne faut jamais affirmer que l'on est certain que l'hypothèse faite sur p est vraie car la probabilité de se tromper dans l'affirmation est de 0,05.