Probabilité et statistique

Table des matières

Probabilité : rappels 2
Probabilité: variable aléatoire. 3
Probabilité: Répétition d'expériences identiques 4
Probabilité: Le schéma de Bernoulli 5
Probabilité: les coefficients binomiaux 6
Statistique: tendance centrale et dispersion 7
Statistique: Intervalle de fluctuation et loi binomiale 8

Probabilité : rappels

En classe de seconde on a défini:

Une expérience aléatoire Une expérience (souvent un tirage) dont seul le hasard détermine le résultat.
Les issues: tous les résultats possibles de l'expérience aléatoire
Issues équiprobables**:** dont aucune n'est favorisée ou défavorisée par le protocole de l'expérience aléatoire.
L'Univers Ω Ensemble des issues
Une tribu famille de sous-ensembles de Ω comprenant leur réunion, leur complémentaire, Ω, ∅
Les évènements Sous-ensembles appartenant à la tribu
Evènements incompatibles évènements dont l'intersection et nulle
Evènements composites: A ou B (A U B), A et B (A ∩ B), non A (Ω – A ou $\bar{A}$)
Evènement mesurable: ensemble dont on peut évaluer la longueur, l'aire, le volume, le nombre d'éléments…
Probabilité d'un évènement rapport de la mesure de ce sous-ensemble à la mesure de Ω

On a vu quelques propriétés élémentaires de la probabilité

P(Ω)=1 P(∅) = 0
P(A ou B) = P(A U B) = P(A) + P(B) – P(A ∩ B) P(A ou B) = (P(A) + P(B) pour des évènements incompatibles
P(non A) = 1 – P(A)

On a étudié quelques techniques de dénombrement en vue d'étudier les tirages multiples.

■ On a défini **l'ensemble produit** de n ensembles E1, E2, …., En, noté E1xE2x..xEn

E1xE2x..xEn = ensemble des n-uplets (X_1,X_2,..,X_n) tels que X_1∈E1 , X_2∈E2, …..,X_n∈En

En probabilité, quand on tire, par exemple, un élément dans 3 ensembles différents (E, F, G), qui peuvent être des sacs, des urnes, des jeux de cartes, des dés …. on peut considérer qu'on a tiré un élément de ExFxG.
Quand on tire 1 éléments de E puis un autre après remise on peut considérer qu'on a tiré un élément de ExE encore noté E^2.
Donc pour savoir calculer les probabilités dans ces tirages multiples, il faut savoir dénombrer les issues possibles dont le nombre est **le nombre d'éléments (ou cardinal) N de l'ensemble produit.**

Si E compte e éléments, F compte f éléments et G compte g éléments,
ExFxG compte efg éléments. N= Card(ExFxG) = efg
E^2 compte e^2 éléments. N= Card(E^2) = e^2
E^n compte e^n éléments. N= Card(E^n) = e^n

■ Quand on prélève p éléments dans un ensemble qui en contient n (simultanément ou consécutivement sans remise mais sans considérer que l'ordre a de l'importance issue {a,b,c} identique à {b,a,c}) on dit qu'on a prélevé **une combinaison de p éléments parmi n.** (Un sous-ensemble de p éléments parmi n)

On note **comb(n,p)** ou $\binom{n}{p}$ le nombre de combinaisons (de sous-ensembles) possibles de n objets p à p.
Les calculatrices scientifiques donnent ce nombre (souvent touche nCr).
Combien de combinaisons de 3 boules je peux faire dans un sac qui en contient 7 ? comb(7,3) = 35.

■ Quand on prélève p éléments consécutivement dans un ensemble qui en contient n en considérant que l'ordre a de l'importance issue (a,b,c) différente de (b,a,c) on dit qu'on a prélevé
un arrangement de p objets parmi n
On note **Arr(n,p)** le nombre d'arrangements possibles de n objets p à p.
Les calculatrices scientifiques donnent ce nombre (souvent touche nPr).
Combien d'arrangements (de mots) de 3 lettres puis-je faire avec 7 lettres différentes? Arr(7,3)= 210.

■ Si on se demande combien d'arrangements différents on peut faire avec une combinaison de n objets on dit qu'on cherche le **nombre de permutations possibles de n objets** (le nombre de façons dont on peut ordonner ces n objets)
On note **n!** ce nombre qu'on appelle "factorielle n" et qui est égal à n! = n(n–1) (n–2)….(3)(2)1
Les calculatrices scientifiques donnent ce nombre (souvent touche X! ou N!).
Combien de permutations possibles puis-je faire avec 3 lettres ? 3! = (3)(2)1 = 6.
On remarque que Arr(n,p) = (p!) [comb(n,p)]

Lors de tirages multiples une issue est notée (a_1,a_2,…a_n) où a_i , $i^{ème}$ composante d'une issue est le résultat de la $i^{ème}$ épreuve. La probabilité d'une issue (a_1,a_2,…a_n) est égale au produit des probabilités de chacune de ses composantes <u>au moment de l'épreuve dont elle résulte.</u>

$$P(a_1,a_2,\dots,a_n) = P(a_1).P(a_2)\dots.P(a_n)$$

Probabilité: variable aléatoire.

Une variable aléatoire désigne un nombre réel associé à chaque issue d'une expérience aléatoire.
On la note par une lettre majuscule (souvent X).

Par exemple les faces d'un dé, l'abscisse d'un nombre tiré au hasard sur un segment, l'âge d'un individu tiré au hasard, son poids ou une note qu'il a obtenue à un examen sont des variables aléatoires.
Dans ce qui suit, on ne s'intéresse qu'aux variables aléatoires à valeurs discrètes, c'est-à-dire dont on peut compter les valeurs.

La loi de probabilité d'une variable aléatoire associe à chacune de ses valeurs possibles {x1, x2, ..,xn} la probabilité de l'évènement correspondant P(X=x1) = P1 , P(X=x2)=P2 , …., P(X=xn)=Pn.

On doit donc avoir P(X=x1) + P(X=x2) + ….+ P(X=xn)= 1
On va désormais noter Pi la probabilité P(X=Xi)

L'espérance mathématique d'une variable aléatoire X, notée E(X), est la somme des valeurs que peut prendre la variable X, chacune d'elle étant pondérée (multipliée) par sa probabilité.

E(X) = x1.P1 +x2.P2+……+Xn.P2 ce que l'on note E(X)=$\sum_{i=1}^{i=n} Xi.Pi$

Si toutes les valeurs de X avaient une probabilité égale (1/n), E(X) serait la moyenne des valeurs de X.
Une probabilité a les mêmes caractéristiques que la fréquence f i de la classe Xi dans une série statistique ($fi \in [0,1]$, $Pi \in [0,1]$, $\sum fi = 1$, $\sum Pi = 1$). D'ailleurs la probabilité d'une modalité dans une population est souvent la fréquence de cette modalité dans cette population. La définition de la moyenne est $\bar{X} = \sum Xifi$, celle de l'espérance mathématique est E(X) =$\sum XiPi$. On peut donc assimiler l'espérance mathématique de X à une moyenne de X.

Quand X est un nombre relatif représentant un gain ou une perte, E(X) évalue le bilan moyen du joueur par partie sur un très grand nombre de parties. Par exemple on jette un dé équilibré et il faut deviner le chiffre sortant.
On mise 1€ par partie. Si on gagne, on gagne 3€ et on garde la mise, et si on perd, on perd la mise.
On a 5 chances sur 6 de perdre 1€ et une chance sur 6 d'en gagner 3. Donc la loi de probabilité est

X	–1	+3
P(X)	5/6	1/6

E(X) = $-1(\frac{5}{6}) + 3(\frac{1}{6}) = -\frac{5}{6} + \frac{3}{6} = \frac{-2}{6} = -\frac{1}{3}$ = – 0,33. En moyenne sur 6 coups je perds 5€ et j'en gagne 3 donc je perds 2€ en 6 parties ce qui fait en moyenne 2/6 d'euro= 0,33 euro par partie.
Si le jeu est équitable on a E(X)=0. Une espérance de gain négative signifie que le jeu défavorise le parieur.

La variance V(X) d'une variable aléatoire X d'espérance mathématique E est

V(X) = P1 (X1 –E) ^{2} + P2 (X2 –E) ^{2} + ….+Pn(Xn–E) ^{2} ce que l'on note V(X) = $\sum_{i=1}^{i=n} Pi(Xi-E)^2$

L'écart type σ (X) de X est la racine carrée de la variance.

σ(X) = $\sqrt{V(X)}$

σ(X) et V(X) sont des nombres positifs.
Si on assimile E à une moyenne de X, V(X) est la moyenne des carrés des distances de X à la valeur moyenne E.
Et l'écart type peut être assimilé à la distance moyenne des valeurs de X à E.
Autrement dit, plus les valeurs de X seront resserrées autour de E (valeur moyenne) plus **σ(X)** sera petit.
σ(X) et V(X) sont des paramètres de dispersion.

En remplaçant dans V(X) → $(Xi – E)^2$ par $(Xi^2 – 2XiE +E^2)$ on trouve que V(X) = $\sum Pi.Xi^2 - 2E\sum XiPi + E^2\sum Pi$
Et comme $\sum XiPi$ = E et $\sum Pi$ = 1 on peut dire que

$\boxed{V(X) = (\sum PiXi^2) - E^2}$

Soient a et b 2 nombres réels quelconques. Si on remplace X par aX +b on a:

E(aX + b) = a .E(X) + b
V(aX+b) = a^2.V(X)
(aX + b) = |a|. σ(X)

Probabilité: Répétition d'expériences identiques

On procède à une expérience aléatoire formée de la répétition d'une même épreuve aléatoire. Comment évaluer la probabilité d'une issue particulière de ce type d'expérience?

Soit une épreuve aléatoire E.
Son univers Ω est formé de 2 issues {a ; b}
Sa loi de probabilité est

issue	a	b
probabilité	p	q

Avec p + q = 1

Si on répète n fois cette épreuve aléatoire, cela est considéré comme une expérience aléatoire composite dont le résultat, l'issue, sera (x1,x2, ….,xn) avec xi ∈ Ω.
Exemple si n =5 →(a,b,b,a,b) constitue une issue possible.
xi, résultat de la n^{ième} épreuve est la i^{eme} composante de l'issue.
L'univers de cette expérience est l'ensemble produit Ω^n qui contient 2^n issues différentes.

La probabilité d'une issue particulière (x1,x2,…,xn) est égale au produit des probabilités de chacune de ses composantes dans l'épreuve qui l'a produite.
P(x1,x2,….xn)=P(x1).P(x2)…..P(xn) avec P(xi)=p ou q.

Si l'issue (x1,x2,….,xn) est formée de x composantes = a et de y composantes = b (avec x + y = n) sa probabilité est donc
P(x1,x2,….,xn) = $p^xq^y = p^x(1-p)^y$.

Exemple on jette 5 fois un dé, les issues sont {multiple de 3 ; non multiple de 3}
Loi de probabilité

$3m$	$\overline{3m}$
1/3	2/3

La probabilité d'obtenir un multiple de 3 aux seules épreuves de rang 2 et 4 est $(\frac{1}{3})^2(\frac{2}{3})^3 = \frac{8}{243}$.

Soit une épreuve aléatoire E.
Son univers Ω est formé de 3 issues {a ; b; c}
Sa loi de probabilité est

issue	a	b	c
probabilité	p	q	r

Avec p + q + r = 1

Si on répète n fois cette épreuve aléatoire, cela est considéré comme une expérience aléatoire composite dont le résultat, l'issue, sera (x1,x2, ….,xn) avec xi ∈ Ω.
Exemple si n =5 →(a,b,c,c,b) constitue une issue possible.
xi, résultat de la n^{ième} épreuve est la i^{eme} composante de l'issue.
L'univers de cette expérience est l'ensemble produit Ω^n qui contient 3^n issues différentes.

La probabilité d'une issue particulière (x1,x2,…,xn) est égale au produit des probabilités de chacune de ses composantes dans l'épreuve qui l'a produite.
P(x1,x2,.xn) = P(x1).P(x2)…P(xn) avec P(xi)=p ou q ou r

Si l'issue (x1,x2,….,xn) est formée de x composantes = a, y composantes = b et z composantes = c (avec x+y+z =n) sa probabilité est donc:
P(x1,x2,….,xn) = $p^xq^yr^z$.

Exemple une urne contient 2 boules rouges, 5 vertes, 3 blanches. On tire 5 fois dans cette urne avec remise.
Loi de probabilité:

R	V	B
0,2	0,5	0,3

La probabilité de (B,V,V,R,V) est $(0,3)(0,5)^3(0,2)=0,0075$

Si dans une épreuve les seules issues possibles **A**, **B** ont pour probabilité respective **P** et **1–P**, dans N épreuves semblables, la probabilité d'une issue comportant **x fois A** et **n–x fois B** à des rangs bien spécifiques est $\mathbf{P^x(1-P)^{n-x}}$

Dans le cas où l'on répète n fois une épreuve à deux issues, A et B, comment compter dans l'univers les issues qui comportent x composantes égales à A et n–x composantes égales à B?

Supposons que n = 5 et qu'on veuille compter les issues dont seulement 2 des 5 composantes sont égales à A.
Vérifient ce critère les issues (B,A,B,B,A) → A aux rangs 2 et 5, (A,A,B,B,B) → A aux rangs 1 et 2, etc .
On va imaginer toutes les façons dont le A peut occuper 2 des 5 rangs de composantes.

Rangs des 2 composantes = A			
1 et 2	1 et 3	1 et 4	1 et 5
	2 et 3	2 et 4	2 et 5
		3 et 4	3 et 5
			4 et 5

Le nombre cherché est donc le nombre de combinaisons de 5 rangs 2 à 2. Ou encore le nombre de combinaisons de 2 rangs parmi 5.
C'est-à-dire Comb(5,2), qu'on peut désigner par un raccourci: "2 parmi 5".
La calculatrice et notre décompte trouvent ce nombre égal à 10.

On peut donc construire le tableau suivant:

Nombre de composantes égales à A	0	1	2	3	4	5
Nombre d'issues	Comb(5,0)	Comb(5,1)	Comb(5,2)	Comb(5,3)	Comb(5,4)	Comb(5,5)
	1	5	10	10	5	1

On a bien un univers Ω^5 dont le nombre d'issues est 2^5 soit 32. Et donc $2^5 = \sum_{i=0}^{i=5} comb(5,i)$
Les issues comportant x fois A étant les mêmes que celles qui comportent n – x fois B.
Si le nombre des premières est comb (5,x) le nombre des secondes doit être comb(5,n–x) et comb(5,x) =comb(5,n–x)

Quand on répète **n** fois une épreuve à deux issues A et B, le nombre d'issues composées de **x** fois A et de n–x fois B est **Comb(n, x)**, encore noté $\binom{n}{x}$

On a $\binom{n}{x} = \binom{n}{n-x}$ et $2^n = \sum_{i=0}^{i=n} \binom{n}{i}$

Probabilité: Le schéma de Bernoulli

Une **épreuve de Bernoulli** est une épreuve aléatoire comportant seulement 2 issues appelées "succès" (probabilité p) et "échec" (probabilité 1 – p)

Une **variable de Bernoulli** x prend la valeur **x = 1** si le résultat de l'épreuve aléatoire est "**succès**" et la valeur **x=0** si le résultat est "**échec**"

Si x est une variable de Bernoulli
E(x) = 1.p + 0.(1– p) **= p**
V(x) = $p(1-p)^2 + (1-p)(0-p)^2$ **= p(1 – p)**

La répétition de **n** épreuves de Bernoulli identiques avec Prob(succès) =**p** est un **schéma de Bernoulli**.
A un schéma de Bernoulli on associe la variable **X** qui est **le nombre de succès sur n épreuves**.
La loi de probabilité de X s'appelle la **loi binomiale de paramètres n et p** et on la note **B(n,p)**.

Donc l'issue d'un schéma de Bernoulli est une suite de n chiffres égaux à 0 ou à 1 et X est le nombre de 1 de la suite. Or, on a vu que
Si S =(01011…001) est une issue particulière comportant k composantes à 1, sa probabilité est

$P(S)=p^k(1-p)^{n-k}$

Et le nombre d'issues comportant k composantes à 1 et de probabilité égale à P(S) est le nombre de combinaisons possibles de k rangs parmi n soit

N(k)=Comb (n,k) ou $\binom{n}{k}$ qu'on peut désigner par un raccourci: "**k parmi n**"

On en déduit que

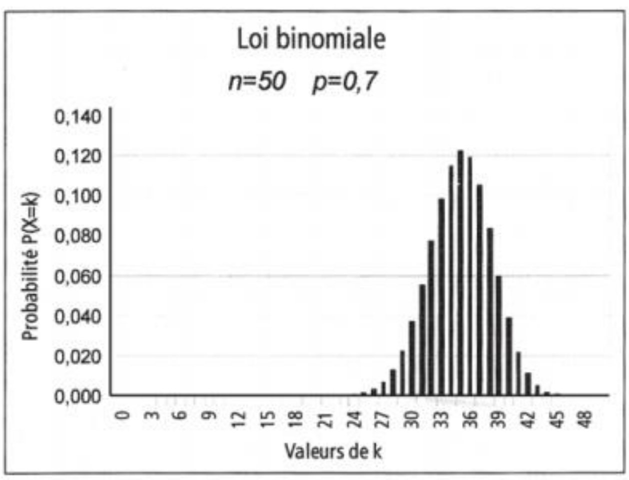
La probabilité de K succès dans un schéma de Bernoulli est $P(X=K) = \binom{n}{k} p^k(1-p)^{n-k}$

Ce qui nous donne la loi de probabilité de X.
$\binom{n}{k}$ joue ici le rôle d'un **coefficient binomial**.

X étant la somme de n variables de Bernoulli d'espérance mathématique p et de variance p(1–p) on a

E(X) = np
V(X) =np(1–p)

Représentation graphique l'une loi binomiale



On voit ici le diagramme de la loi B(50, 0,7)
La probabilité du succès étant p =0,7 on procède à 50 épreuves et le diagramme donne la probabilité de k succès, P(X=k) en fonction de k qui peut varier de 0 à 50.
Si on fait la somme des probabilités représentées par la hauteur de tous les bâtons, on doit trouver 1.

On a E(X) =35, or k= 35 est la valeur pour laquelle P(X=k) est maximale, et il semble que la probabilité se répartisse équitablement de part et d'autre de la droite X =35, ce qui confirme le rôle de moyenne dévolu à E(X).
Par ailleurs la variance de X est V(X) = 10,5 et on voit que l'essentiel du graphe est concentré dans l'intervalle [E(X)–10,5 ; E(X) +10,5] autrement dit [24,5 ; 45,5], les probabilités P(X=k) pour k extérieur à cet intervalle étant pratiquement nulles.

Probabilité: les coefficients binomiaux

On rappelle qu'une issue (1,0,0,…1,1,0) du schéma de Bernoulli est une suite de n variables de Bernoulli, donc elle est formée de n composantes toutes égales à 0 ou à 1.
$\binom{n}{k}$ (Avec k ≤ n) **est le nombre de combinaisons (de sous-ensembles) de k éléments parmi n.**

$\binom{n}{k}$ **est le nombre d'issues ayant k composantes égales à 1 (ou à 0)**.
En effet, chaque composante dont la valeur est 1 occupe un rang dans l'issue de 1 à n.
Et $\binom{n}{k}$ est le nombre de combinaisons possibles de k rangs parmi n (chacune correspondant à une issue)

$\binom{n}{0}$ **= 1** car il y a une seule issue ne comprenant aucun 1. Une issue formée essentiellement de 0.

$\binom{n}{n}$ **= 1** car il y a une seule issue ne comprenant aucun 0. Une issue formée essentiellement de 1.

$\binom{n}{k} = \binom{n}{n-k}$ car toute issue ayant k composantes égales à 1 a aussi n–k composantes égales à 0.
Et réciproquement. Or $\binom{n}{n-k}$ est le nombre d'issues ayant n –k composantes égales à 0.

$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \ldots + \binom{n}{n-1} + \binom{n}{n} = 2^n$ parce que 2^n est le nombre total d'issues de Ω^n.

En outre 2^n est aussi le nombre de sous-ensembles qu'on peut faire dans un ensemble E de n éléments (en comptant ∅ et E) et $\binom{n}{k}$ est le nombre de sous-ensembles de k éléments qu'on peut faire avec n éléments.

$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$

En effet, soit l'ensemble des n+1 premiers entiers naturels, zéro excepté {1, 2, …., n+1} .
$\binom{n+1}{k+1}$ est le nombre de sous-ensembles de k+1 éléments qu'on peut faire dans cet ensemble.
Parmi ces sous-ensembles il y a ceux qui ne contiennent pas n+1, c'est les sous-ensembles de k+1 éléments qu'on peut faire avec {1,2,….n} et donc leur nombre est $\binom{n}{k+1}$
Et ceux qui ceux qui contiennent n+1. Mais pour faire tous les sous-ensembles de k+1 éléments qui contiennent n+1, il faut prendre tous les sous-ensembles de k éléments qu'on peut faire avec {1,2,…,n} et leur adjoindre n+1.
Leur nombre est donc $\binom{n}{k}$ et en ajoutant $\binom{n}{k}$ et $\binom{n}{k+1}$ on trouve bien $\binom{n+1}{k+1}$.

Binôme de Newton et triangle de Pascal. (Complément au programme)

Comment développer facilement $(a+b)^n = (a+b)(a+b)\ldots(a+b)$? n facteurs égaux à (a+b)

On sait qu'on va trouver une somme de termes en a^n, $a^{n-1}b$, $a^{n-2}b^2$, …..,ab^{n-1}, b^n .
Donc des termes de la forme $a^{n-k}\,b^k$, k variant de 0 à n.
Mais le problème est combien de termes de chaque sorte va-t-on trouver?
Il suffit de considérer que chacun de ces termes provient de la multiplication du a de n-k facteurs par le b des k facteurs restants et de compter les façons dont on peut associer k facteurs parmi n. Or on sait que ce nombre est $\binom{n}{k}$.
Par exemple pour $(a+b)^3=(a+b)(a+b)(a+b)$ pour compter le nombre de a^2b que je dois trouver, il me faut compter toutes les façons dont je peux associer 2 facteurs contenant a parmi 3. Je trouve 3 et le coefficient $\binom{3}{2}$ est bien égal à 3.
(a du facteur 1 x a du facteur 2 x b du facteur 3) (a du facteur 1 x a du facteur 3 x b du facteur 2) (a du facteur 2 x a du facteur 3 x b du facteur 1)
Le terme général de notre somme est donc $\binom{n}{k} a^{n-k} b^k$ et

$(a+b)^n = \sum_{k=0}^{k=n}\binom{n}{k}a^{n-k}b^k$ (En déduire que $\sum_{k=0}^{k=n} P(X=k) = \sum_{k=0}^{k=n}\binom{n}{k}p^{n-k}(1-p)^k = 1$)

Pour retrouver facilement la valeur des premiers coefficients binomiaux:

n=0	1					
n=1	1	1				
n=2	1	2	1			
n=3	1	3	3	1		
n=4	1	4	6	4	1	
n=5	1	?				1

On sait que $(a+b)^2=a^2+2ab+b^2$. Donc les coefficients 1, 2 ,1 forment la ligne de n = 2.
Pour remplir la ligne suivante, on sait qu'elle commence et se termine par 1
Pour remplir chacune des autres cases, il faut faire la somme de la case qui est juste au-dessus de la case à remplir avec la case qui est immédiatement à sa gauche.
On va donc remplacer le ? par 4+1 =5 et les cases suivantes recevront 10 , 10, et 5.
On utilise la propriété $\binom{n}{k}=\binom{n-1}{k}+\binom{n-1}{k-1}$ qu'on peut aussi écrire $\binom{n+1}{k+1} = \binom{n}{k+1} + \binom{n}{k}$

On voit que dans le tableau les coefficients binomiaux dessinent un triangle: **le triangle de Pascal**.

Statistique: tendance centrale et dispersion

En classe de seconde on a vu…

La définition d'une série numérique statistique qui découpe une population d'effectif global N selon la valeur d'un nombre X qui caractérise chacun de ses éléments. A chaque valeur de X ou intervalle de X (rangés dans le tableau par valeur croissante) correspond une classe dont la série indique l'effectif.

Classes	X_1	X_2	…	X_{n-1}	X_n
Effectifs	N_1	N_2	…	N_{n-1}	N_n

On peut compléter ce tableau par des lignes indiquant la fréquence de chaque classe ($f_i = N_i / N$), l'effectif cumulé croissant de la classe Xi qui est la somme des effectifs pour lesquels X ≤ Xi ou la fréquence cumulée croissante de Xi qui est la somme des fréquences des classes pour lesquelles X ≤ Xi

La définition de la médiane

- Si l'effectif N de la série est un nombre impair, $N=2n+1$, la médiane de la série est la valeur centrale du caractère, celle qui est numérotée $n+1$.
- Si l'effectif N de la série est un nombre pair, $N=2n$, la médiane est le nombre égal à la demi-sommes des deux valeurs centrales, celles qui sont numérotées n et $n+1$.

Ou

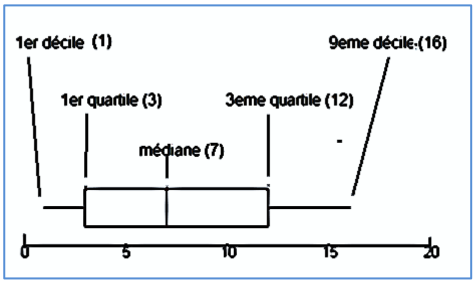
La première classe pour laquelle la fréquence cumulée croissante dépasse 50% s'appelle la **classe médiane.**

La définition des quartiles et des déciles

- **Premier quartile Q_1** : c'est la plus petite valeur de la série telle qu'au moins 25% des données soient inférieures à Q_1.
- **Troisième quartile Q_3** : c'est la plus petite des valeurs de la série telle qu'au moins 75% des données soient inférieures à Q_3.
- **Premier décile D_1** : c'est le plus petit élément des valeurs de la série tel qu'au moins 10% des données soient inférieures à D_1.
- **Neuvième décile D_9**: c'est le plus petit élément des valeurs de la série tel qu'au moins 90% des données soient inférieures à D_9.

La définition de l'intervalle inter-déciles
[D1; D9]
Et des écarts interquartiles et inter-déciles de la série statistique
Q3 –Q1 et D9 – D1

Le diagramme en boîte résume la plupart des paramètres de la série



Précisons que si on fait subir à toutes les valeurs de la série une transformation du type X → aX+b (avec a et b réels)
La médiane m devient am + b
Un quartile Q devient aQ + b
Un décile D devient aD + b
Un écart E devient aE.

A ces paramètres déjà vus on ajoute

La moyenne de la série

$\overline{X}=\frac{\sum XiNi}{N}$ **ou** $\overline{X} = \sum Xifi$ **(**fi **étant la fréquence de la classe)**

La variance et l'écart type de la série

V(X) = $\frac{\sum Ni(Xi-\overline{X})^2}{N}$ $\quad ou \quad V(X) = \sum fi(Xi-\overline{X})^2 \quad ou\, V(X) = \left[\sum fi(Xi)^2\right] - \overline{X}^2 \quad et \quad \sigma(X) = \sqrt{V(X)}$

Statistique: Intervalle de fluctuation et loi binomiale

En seconde, on a vu comment on pouvait estimer la représentativité d'un échantillon de taille n en mesurant la fréquence f de la modalité d'un caractère parmi ses éléments et en la comparant à la fréquence p de la même modalité dans la population, la valeur p étant réputée fiable.

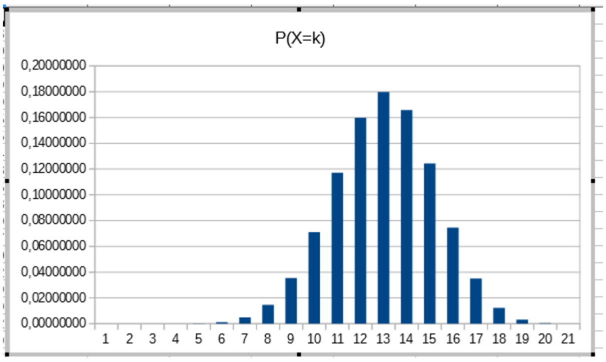
Si f ∈ [p – $\frac{1}{\sqrt{n}}$; p + $\frac{1}{\sqrt{n}}$] on dit que l'échantillon est acceptable au risque de 5%.

Si ce n'est pas le cas, l'échantillon est suspect de non représentativité et nous le rejetons.

On va voir maintenant comment l'échantillonnage peut, au contraire, être utilisé pour tester la fiabilité d'une fréquence mesurée dans la population et comment on peut utiliser la loi binomiale pour jouer ce rôle.

On pense que la fréquence de la modalité d'un caractère donné dans la population est f.
Par exemple parmi les élèves de première on pense que 60% ont la moyenne ou plus en mathématiques.
Si c'est le cas, quand on choisit un élève au hasard parmi les élèves de première, la probabilité pour qu'il ait au moins la moyenne en math est 0,6. Si on attribue à cet élève une variable de Bernoulli égale à 1 quand il a au moins la moyenne en math et égale à 0 quand il ne l'a pas, on procède à une épreuve de Bernoulli.
Quand on remet cet élève dans la population et qu'on procède à 20 épreuves du même type, on obtient un schéma de Bernoulli et si X est le nombre des élèves sur 20 ayant au moins la moyenne en math, la loi de probabilité de X devrait être B(20; 0,6).

K	P(X=k)	P(X≤ k)
0	0,00000001	0,00000001
1	0,00000033	0,00000034
2	0,00000470	0,00000504
3	0,00004230	0,00004734
4	0,00026969	0,00031703
5	0,00129449	0,00161152
6	0,00485435	0,00646588
7	0,01456305	0,02102893
8	0,03549744	0,05652637
9	0,07099488	0,12752125
10	0,11714155	0,24466280
11	0,15973848	0,40440127
12	0,17970579	0,58410706
13	0,16588227	0,74998933
14	0,12441170	0,87440103
15	0,07464702	0,94904805
16	0,03499079	0,98403884
17	0,01234969	0,99638853
18	0,00308742	0,99947595
19	0,00048749	0,99996344
20	0,00003656	1



Ce graphique illustre la série statistique présente dans la seconde colonne du tableau à savoir la probabilité de compter k notes supérieures ou égales à la moyenne dans un échantillon de 20 individus. B(20: 0.6)

La 3eme colonne nous donne la probabilité pour que X soit inférieur à k.
Le contenu d'une cellule de cette colonne est donc la somme des contenus des cellules de rangs inférieurs ou égal de la colonne précédente. C'est une probabilité cumulée.

On cherche le plus petit entier **a** tel que P(X ≤ a) soit > 0,025 c'est **8**.
On cherche le plus petit entier **b** tel que P(X ≤ b) soit > 0,975 c'est **16**.

On en déduit que la probabilité pour que le X de notre échantillon soit dans l'intervalle **[8 ; 16]** est 95% .
C'est **l'intervalle de fluctuation au risque de 5%.**

Notre but est juste d'illustrer à travers tableaux et graphiques, la méthode que nous devons mettre en œuvre pour résoudre notre problème. Mais ici, l'effectif n= 20 de notre échantillon est nettement insuffisant. Pour un échantillonnage fiable, il faudrait que n soit au moins égal à 25 et que p ne soit pas trop éloigné de la moyenne. Plus n sera grand, plus notre intervalle de fluctuation au risque de 5% sera large et plus notre échantillons sera fiable. Mais nous pensons que le faible nombre de valeurs contribue à la clarté de l'exposé.

Si le X de notre échantillon n'est pas dans l'intervalle de fluctuation, et si cela se reproduit dans plus de 5% des cas quand nous renouvelons l'expérience, c'est probablement parce que la fréquence de 0,6 que nous avons prise pour base de notre raisonnement n'est pas fiable. Il va falloir procéder à sa réévaluation.
Dans certains cas, quand n est très grand, on peut même adopter la fréquence mesurée dans notre échantillon qui est $\frac{X}{n}$ comme fréquence de la modalité testée dans la population.