Les Probabilités

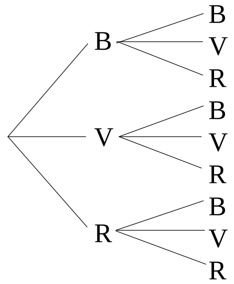
I) Le vocabulaire des probabilités

1) Exemple

Une urne contient 3 boules : une bleue, une rouge, une verte. On tire au hasard une boule de cette urne, on note sa couleur puis on la replace dans l'urne que l'on agite. On tire alors une deuxième boule de l'urne en notant sa couleur.

On dit que l'ensemble des opérations ainsi réalisées constitue une **épreuve** ou **expérience aléatoire**. On peut représenter l'ensemble des résultats possibles ou **éventualités** par un tableau ou un arbre :

1 / **2**	**B**	**V**	**R**
B	(B,B)	(V,B)	(R,B)
V	(B,V)	(V,V)	(R,V)
R	(B,R)	(V,R)	(R,R)



2) Evénement

Considérons Ω = {(B,B);(B,V);(B,R);(V,B);(V,V);(V,R);(R,B);(R,V);(R,R)} l'**univers** ou **ensemble des éventualités** de cette expérience.

L'événement A_1 = {(B,B);(V,V);(R,R)} est l'événement "Obtenir 2 boules de même couleur"

A_2 = {(B,R);(B,V);(B,B)} est l'événement "Obtenir une boule bleue au premier tirage"

Remarque

Si une épreuve donne comme résultat (B,R), nous pouvons dire que l'événement A_2 est réalisé mais par contre A_1 ne l'est pas.

3) Evénement élémentaire

A_3 = {(B,B)} est l'événement "obtenir deux boules bleues".

C'est un événement qui ne contient qu'une seule éventualité. On dit que c'est un **événement élémentaire**.

4) Evénement "A et B"

Considérons les événements A_1, A_2 et A_3 précédemment définis : A_1 : "Obtenir 2 boules de même couleur"

A_2 : "Obtenir une boule bleue au premier tirage"

A_3 : "obtenir deux boules bleues".

Une épreuve étant accomplie, nous disons que l'événement A_3 est réalisé lorsque les événements A_1 et A_2 sont simultanément réalisés.

L'événement A_3 est l'événement "A_1 et A_2", noté A_1 et A_2 mais aussi $A_1 \cap A_2$.

$A_3 = A_1$ et A_2 = {(B,B)}

5) Evénement incompatible

Reprenons l'événement A_2 : "Obtenir une boule bleue au premier tirage" et soit A_4 = {(R,R)} : "obtenir deux boules rouges".

Aucune éventualité ne réalise simultanément A_2 et A_4; on dit que A_2 et A_4 sont incompatibles et on note $A_2 \cap A_4 = \varnothing$

Remarque

Les événements élémentaires sont deux à deux incompatibles

7) Evénements contraires

Soit A_1 : "Obtenir 2 boules de même couleur".

On note $\overline{A_1}$ l'événement contraire, c'est-à-dire "obtenir un tirage bicolore".

On a A_1 = {(B,B);(V,V);(R,R)} et donc $\overline{A_1}$ = {(B,V);(B,R);(V,B);(V,R);(R,B);(R,V)}

Ainsi, si une épreuve est accomplie, l'un et l'un seulement des deux événements A_1 et $\overline{A_1}$ est réalisé.

II) Probabilité

1) La notion de probabilité

Considérons l'épreuve suivante : elle consiste à lancer un dé ; l'univers associé est Ω = {1;2;3;4;5;6}. Notons $\{\omega_i\}$ l'événement élémentaire "obtenir le chiffre *i*", $1 \leq i \leq 6$

4 séries d'épreuves comportant 1 000, 5 000, 10 000 et 20 000 lancers ont été résumées pour l'apparition du chiffre 1 par le tableau ci-contre :

Nombre de lancers	**Nombre d'apparitions du chiffre 1**	**Fréquence**
1 000	173	0,1730
5 000	844	0,1688
10 000	1 650	0,1650
20 000	3 320	0,1660

On remarque que les résultats sont très proches. On supposera, comme en physique, que l'on admet "une valeur exacte" (ou théorique) de la mesure.

Nous admettrons que la fréquence de $\{\omega_i\}$ est une valeur approchée d'un nombre réel compris entre 0 et 1 que l'on appelle la probabilité de l'événement $\{\omega_i\}$ que l'on note $p(\{\omega_i\})$. Dans le cas présent, on supposera $p(\{\omega_i\}) = \frac{1}{6}$.

Définition

Soit E une épreuve aléatoire. L'ensemble des éventualités est $\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$.

- A chaque événement élémentaire $\{\omega_i\}$ est associé un nombre réel, **élément de [0;1]** appelé probabilité de l'événement élémentaire (représentation idéale de la fréquence) tel que : $p(\{\omega_1\}) + p(\{\omega_2\}) + ... + p(\{\omega_n\}) = 1$
- La probabilité de tout événement A est la somme des probabilités des événements élémentaires qui le composent.

En particulier, $p(\Omega) = 1$.

- Si A = $\emptyset$ alors $p(A) = 0$

2) L'hypothèse d'équiprobabilité

Supposons que tous les événements élémentaires soient tels que $p(\{\omega_1\}) = p(\{\omega_2\}) = ... = p(\{\omega_n\})$. On dit alors qu'ils sont équiprobables. L'égalité $p(\{\omega_1\}) + p(\{\omega_2\}) + ... + p(\{\omega_n\}) = 1$ entraîne $p(\{\omega_n\}) = \frac{1}{n}$

Les mots "au hasard", "indiscernable au toucher", "bien équilibré", ... sous-entendent des expériences équiprobables.

Propriété

Dans l'hypothèse d'équiprobabilité, le nombre total d'éventualités étant n, si un événement A est constitué de m éventualités alors sa probabilité est $p(A) = \frac{m}{n} = \frac{\text{effectif de A}}{\text{effectif de } \Omega} = \frac{\text{Card(A)}}{\text{Card}(\Omega)} = \frac{\text{nbre de cas favorables à A}}{\text{nbre de cas possibles}}$

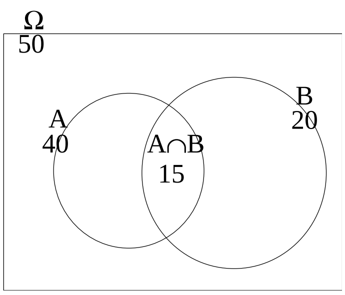
3) Propriétés des probabilités

Probabilité de la réunion de deux événements

Exemple

Un sac contient un ensemble de 50 jetons de formes et de couleurs différentes. 20 sont ronds, 40 sont rouges et 15 sont à la fois ronds et rouges.

On tire au hasard un jeton du sac. Calculons la probabilité pour qu'il soit rond ou rouge.



Désignons par A et B les événements :

A : "obtenir un jeton rouge" et B : "obtenir un jeton rond"

On peut représenter la situation par le schéma ci-contre :

Par lecture du schéma, le nombre d'éléments de $A \cup B$ est 40 + 20 - 15 (on note $\text{Card}(A \cup B) = 40 + 20 - 15$)

et $p(\text{A ou B}) = \frac{(40 - 15) + 15 + (20 - 15)}{50} = \frac{40}{50} + \frac{20}{50} - \frac{15}{50} = p(A) + p(B) - p(\text{A et B})$

Propriété

Si A et B sont deux événements quelconques alors $p(\text{A ou B}) = p(A) + p(B) - p(\text{A et B})$

Probabilité de la réunion de deux événements incompatibles

Si A et B sont deux événements incompatibles, on a $A \cap B = \emptyset$ et $p(\text{A et B}) = 0$

Dans ce cas, $p(\text{A ou B}) = p(A) + p(A)$

Conséquence

Soit A un événement quelconque. A et $\overline{A}$ sont deux événements contraires donc $A \cap \overline{A} = \emptyset$ et $A \cup \overline{A} = \Omega$.

L'application directe de la dernière propriété donne $1 = p(\Omega) = p(\text{A ou } \overline{A}) = p(A) + p(\overline{A})$.

Soit $p(\overline{A}) = 1 - p(A)$

III) Variable aléatoire

1) Définition. Exemple

$\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$ est l'univers d'une expérience aléatoire sur lequel est définie une probabilité.
Une **variable aléatoire** X est une fonction de Ω dans $\mathbb{R}$. Notons x_1, x_2, x_q les valeurs de X (en général, $q < n$).
L'événement "X prend la valeur x_i" est noté $X = x_i$ et sa probabilité p_i. L'événement $X = x_i$ est l'événement qui contient toutes les issues dont l'image par X est x_i.

Exemple : Chacun des mots de l'expression "par le plus grand des hasards" est inscrit sur un carton. On tire au hasard l'un des cartons et on considère la variable aléatoire X qui à chaque mot tiré associe le nombre de lettres de ce mot.
Ici, Ω est l'ensemble des six cartons ou des six mots. Un événement élémentaire est un mot, les événements élémentaires sont équiprobables (tirage au hasard). L'ensemble des valeurs de X est $\{2,3,4,5,7\}$ et est noté $X(\Omega)$.
La loi de probabilité de X est donnée par le tableau suivant :

x_i	2	3	4	5	7
p_i	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

ex : L'événement $X = 3$ contient deux éventualités provenant des deux mots "par" et "des".

2) Paramètres d'une variable aléatoire

Espérance : $E(X) = x_1p_1 + x_2p_2 + \ldots + x_qp_q$
Variance : $V(X) = [x_1 - E(X)]^2p_1 + [x_2 - E(X)]^2p_2 + \ldots + [x_q - E(X)]^2p_q = E(X^2) - [E(X)]^2$
Ecart-type : $\sigma(X) = \sqrt{V(X)}$

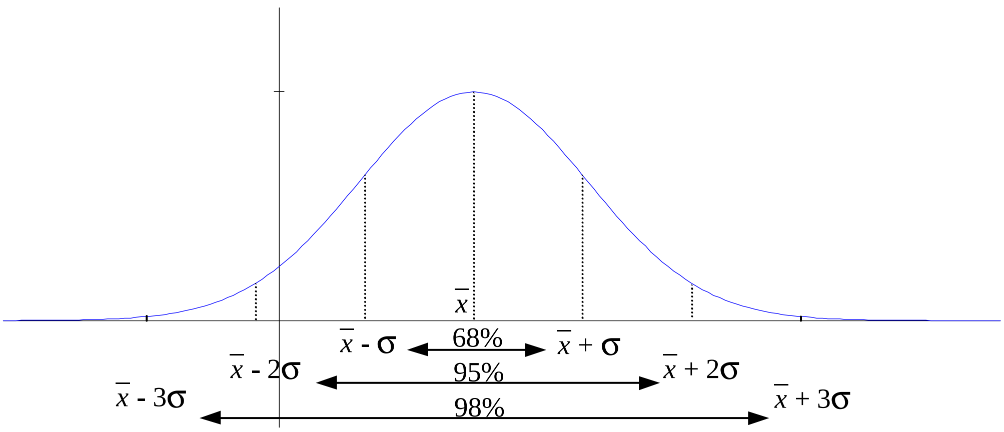
Liens entre loi de probabilité et distribution de fréquences en statistique :

Distribution de fréquences sur $E = \{x_1, ..., x_r\}$	Loi de probabilité sur $E = \{x_1, ..., x_r\}$
$(f_1, ..., f_r)$ $f_i \geq 0$; $\sum f_i = 1$ $A \subset E$, fréquence de A : $f(A) = \sum_{x_i \in A} f_i$ Événement complémentaire : $f(\bar{A}) = 1 - f(A)$ Événements A et B disjoints : $f(A \cup B) = f(A \text{ ou } B) = f(A) + f(B)$ Cas numérique : Moyenne empirique : $\bar{x} = \sum f_i x_i$ Variance empirique : $s^2 = \sum f_i (x_i - \bar{x})^2$ Écart type empirique : $s = \sqrt{\sum f_i (x_i - \bar{x})^2}$	$(p_1, ..., p_r)$ $p_i \geq 0$; $\sum p_i = 1$ $A \subset E$, probabilité de A : $P(A) = \sum_{x_i \in A} p_i$ Événement complémentaire : $P(\bar{A}) = 1 - P(A)$ Événements A et B disjoints : $P(A \cup B) = P(A \text{ ou } B) = P(A) + P(B)$ Cas numérique : Espérance d'une loi P : $\mu = \sum p_i x_i$ Variance d'une loi P : $\sigma^2 = \sum p_i (x_i - \mu)^2$ Écart type d'une loi P : $\sigma = \sqrt{\sum p_i (x_i - \mu)^2}$

3) Interprétation de la moyenne et de l'écart-type d'une série statistique

La répartition normale

D'une façon générale, en statistique, une répartition est dite **normale**, si son polygone des effectifs (courbe obtenue en reliant les sommets des "bâtons" d'un diagramme en bâtons" a l'aspect d'une courbe en cloche (courbe de Gauss), centrée en $\bar{x}$, la moyenne de la série.



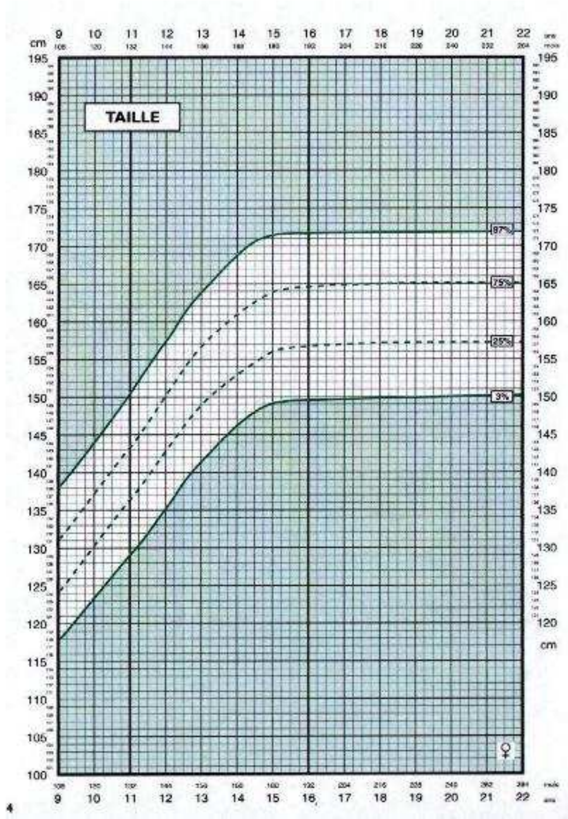
Dans ce cas, on trouve dans chacun des intervalles :

$[\bar{x} - \sigma ; \bar{x} + \sigma]$ 68 % de l'effectif total

$[\bar{x} - 2\sigma ; \bar{x} + 2\sigma]$ 95 % de l'effectif total

$[\bar{x} - 3\sigma ; \bar{x} + 3\sigma]$ 98 % de l'effectif total

Les diagrammes de croissance des carnets de santé

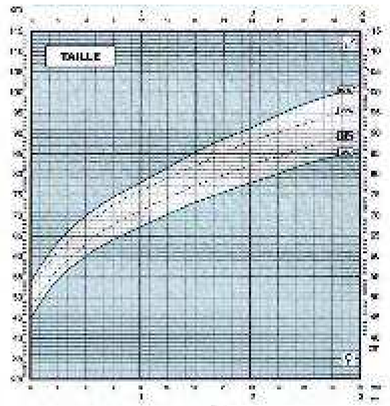


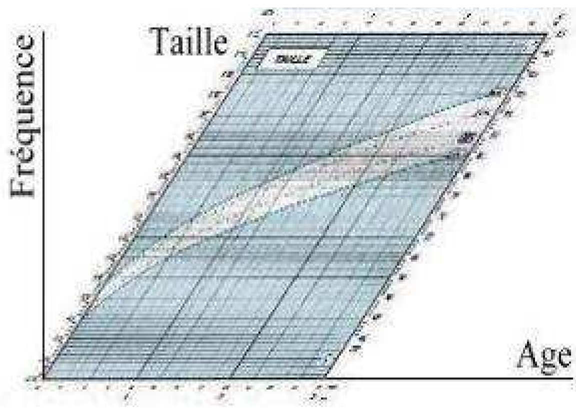
Dans les carnets de santé qui sont distribués par les maternités, on trouve des diagrammes sur lesquels figurent des courbes qui représentent la croissance somatique des enfants par tranches d'âge. L'image ci-contre est celle des filles de 9 à 22 ans.

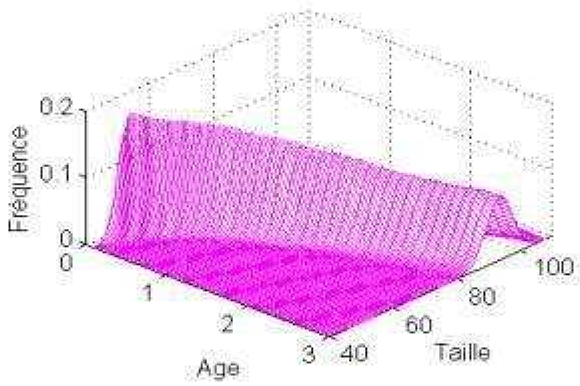
Ces courbes représentent en fonction de l'âge, les quantiles 3, 25, 75, 97 pour cent (répartition en pourcentages de la population), de la distribution empirique des tailles.

Elles sont basées sur un modèle probabiliste de distribution gaussienne des tailles.

Enfants de 0 à 3 ans :







d'après un article "Courbes de croissance" de M. Perreau Guimaraes et B. Ycart

IV) Quelques rappels de statistique

La médiane

Lorsque le caractère étudié est ordonné, la **médiane** est une valeur Me du caractère qui partage la population en deux sous-ensembles de même effectif.
La médiane est donc une valeur centrale de l'échantillon : il y a autant de valeurs qui lui sont inférieures que supérieures
Contrairement à la moyenne, la médiane est insensible aux valeurs aberrantes

Exemple
Une série de notes est la suivante : 12, 9, 10, 16, 8, 11, 12.
Après avoir ordonné la série 8,9,10,11,12,12,16, la médiane est 11

Si nous ajoutons la note 5, la série ordonnée devient 5,8,9,10,11,12,12,16,
Une médiane est alors toute note strictement comprise entre 10 et 11. On appellera l'intervalle]10;11[l'*intervalle médian*, et on définit parfois la médiane comme le milieu de l'intervalle médian.

Construction d'un diagramme en boîte (ou boîte à moustache)

Même sur de très gros échantillons, les quantiles sont peu coûteux à calculer puisqu'il suffit de trier l'échantillon par ordre croissant pour calculer les statistiques d'ordre, et donc tous les quantiles simultanément. Ils fournissent une visualisation facile de la distribution empirique. Nous savons que la médiane est une valeur centrale. Pour mesurer la dispersion, on peut calculer l'*étendue*, qui est la différence entre la plus petite et la plus grande valeur. Mais cette étendue reflète plus les valeurs extrêmes que la localisation de la majorité des données. On appréhende mieux la dispersion d'un échantillon par les intervalles *inter-quartiles* et *inter-déciles*.

Définition

On appelle intervalle inter-quartiles l'intervalle $[Q_{0,25}; Q_{0,75}]$ qui contient la moitié centrale des valeurs de l'échantillon. On appelle intervalle inter-déciles l'intervalle $[Q_{0,1}; Q_{0,9}]$, qui contient 80% des valeurs centrales de l'échantillon.

Diagramme en boîte

Ces intervalles sont à la base d'une représentation très compacte de la distribution empirique : le *diagramme en boîte* ou boîte à moustaches. Il n'y a pas de définition standardisée de cette représentation. Elle consiste en une boîte rectangulaire, dont les deux extrémités sont les quartiles. Ces extrémités se prolongent par des traits terminés par des segments orthogonaux (les moustaches). La longueur de ces segments varie selon les auteurs. Nous proposons de la fixer aux déciles extrêmes. On représente aussi la médiane par un trait dans la boîte, et parfois les valeurs extrêmes par des points (voir figure ci-dessous).