

PROBABILITÉS ET STATISTIQUE POUR L'ENSEIGNEMENT SECONDAIRE

Ce document a été rédigé à l'occasion d'un stage de formation continue de professeurs de mathématiques de troisième et seconde en décembre 2009 à Toulouse, suite à l'introduction des probabilités en troisième.

Plan de la présentation

0. Prérequis de statistique - Liens entre statistique et probabilités.
1. Probabilités sur un ensemble fini
2. Probabilité conditionnelle, indépendance en probabilité
3. Variable aléatoire, distribution de probabilité d'une variable aléatoire ; transport de loi
4. Modélisation et simulation
5. Variable aléatoire réelle ; lois discrètes et lois continues, espérance mathématique et variance
6. Théorèmes limites des probabilités (loi des grands nombres et théorème limite central)
7. Estimation par intervalle de confiance

Annexes

1. Taux de reçus au BAC Toulouse. Liens entre fréquences et probabilités, entre fréquences conditionnelles et probabilités conditionnelles.
2. Distributions de fréquence et distributions de probabilité discrètes et continues
3. Arbre de probabilité - construction
4. Arbre de dénombrement et arbre de probabilité - exemple
5. Lancer d'une pièce de monnaie équilibrée 10 fois de suite
6. Lois de probabilité usuelles et applications
7. Tableau des approximations de lois de probabilité

Prérequis de statistique

Liens entre statistique et probabilités

Proposition de navigation sur le site pédagogique : <http://finestat.free.fr>

En page d'accueil, on trouvera deux documents :

- *La statistique : connaissance et maîtrise de l'aléatoire*

Il s'agit d'une présentation de la statistique en quatre parties :

- 1) Statistique descriptive élémentaire (une ou deux variables) et statistique descriptive multidimensionnelle
- 2) Calcul des probabilités
- 3) Statistique inférentielle (échantillonnage, estimation, tests)
- 4) Modélisation aléatoire

- *Importance du mode de pensée statistique dans le regard scientifique sur le monde*

Il s'agit du premier des six thèmes de convergence disciplinaire au collège.

Ce document de deux pages se termine par le problème de la variabilité des mesures en physique et chimie et par un lien sur un document du site intitulé *Modélisation des erreurs de mesure*.

Page de la Statistique Descriptive

Ouvrir les documents **Vocabulaire** puis **Pourcentages** pour des mises en garde préalables : des difficultés d'apprentissage de la statistique et des probabilités sont dues à des problèmes de vocabulaire, en particulier sur les pourcentages.

St@ternet : Introduction à la statistique

<http://www2.toulouse.iufm.fr/mathematiques/>

Statistique Descriptive Élémentaire (une ou deux variables), cours et exercices corrigés, en ligne, en libre accès et téléchargeable.

Comme on a deux types de variables, catégorielle (ou qualitative) et réelle (ou quantitative), la statistique uni et bidimensionnelle est composée de cinq parties :

- traitement statistique d'une variable catégorielle
- traitement statistique d'une variable réelle
- traitement statistique de deux variables catégorielles
- traitement statistique de deux variables réelles
- traitement statistique d'une variable catégorielle et d'une variable réelle

Lien entre fréquences en statistique et probabilités

Le traitement statistique de deux variables catégorielles est particulièrement important à introduire dès le collège.

Les notions de fréquences de sous-ensembles, de fréquences conditionnelles dans le cas de deux variables permettent d'introduire :

- la logique et les opérations "et", "ou", "non" (et le travail sur le vocabulaire de la logique et le vocabulaire courant)
- les ensembles et les opérations "intersection", "union", "passage au complémentaire" (et l'utilisation des outils de dénombrement, diagrammes, tableaux, arbres)

- les importantes distinctions à faire entre "fréquence de A et B", "fréquence de A par rapport à B", "fréquence de B par rapport à A", pour la formation des citoyens et pour aborder les probabilités conditionnelles et la redoutable formule de Bayes.

Un premier lien entre fréquences et probabilités, entre fréquences conditionnelles et probabilités conditionnelles peut être présenté. (cf. annexe 1. Taux de reçus au BAC Toulouse).

Parallèle entre variable réelle en statistique et variable aléatoire réelle en probabilités

L'étude des variables réelles discrètes et continues (distributions de fréquences et diagrammes associés, fréquences cumulées et fonction de répartition, résumés numériques) est une entrée privilégiée pour l'étude des variables aléatoires réelles discrètes et continues (cf. annexe 2, Distributions de fréquence et distributions de probabilité discrètes et continues).

1. Probabilités sur un ensemble fini

1.1. Expérience aléatoire

On lance un dé cubique équilibré dont les faces sont numérotées de 1 à 6 et on observe le numéro de la face supérieure.

On ne connaît pas à l'avance le résultat de cette expérience ; on sait seulement que c'est un nombre entier compris entre 1 et 6. Cette expérience est appelée "*expérience aléatoire*". Le mot "aléatoire" vient du latin "alea" qui signifie justement "coup de dé". Le mot "hasard" vient de "az zahr", mot arabe qui signifie "dé à jouer" ! Le mot "expérience" renvoie au fait qu'elle est reproductible.

1.2. Vocabulaire des événements et vocabulaire des ensembles

L'ensemble des résultats possibles d'une expérience aléatoire est appelé "*univers*" et souvent noté Ω . Ici on a : $\Omega = \{1, 2, \dots, 6\}$.

Les "*événements*" liés à l'expérience sont identifiés à des sous-ensembles (ou parties) de l'univers, par exemple "obtenir un nombre pair" est identifié à l'ensemble $A = \{2, 4, 6\}$. On dit que l'événement A est *réalisé* si le résultat de l'expérience appartient à A .

Les événements qui n'ont qu'un seul élément, c'est-à-dire les singletons de Ω , sont appelés "*événements élémentaires*". Ici, on a 6 événements élémentaires $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$.

L'événement Ω est appelé "*événement certain*" ; en effet, comme c'est l'ensemble des résultats possibles, on est certain qu'il sera réalisé à chaque expérience.

L'ensemble vide $\emptyset$ est appelé "*événement impossible*" car on suppose qu'un (et un seul) résultat de Ω est *observé* à chaque expérience.

Il n'est pas possible que se réalisent simultanément les événements $\{1, 2, 3\}$ et $\{5, 6\}$ car leur intersection est vide (ensembles disjoints). On dit que deux événements identifiés à des ensembles disjoints sont "*incompatibles*".

L'événement "*contraire*" à un événement A est son complémentaire dans Ω , noté $\bar{A}$. Par exemple, l'événement contraire à l'événement $A = \{2, 4, 6\}$ est $\bar{A} = \{1, 3, 5\}$

On appelle *système complet d'événements* de Ω une partition de Ω , c'est-à-dire un ensemble de parties non vides de Ω , deux à deux disjoints et dont la réunion est Ω . Par exemple $\{A, \bar{A}\}$ et $\{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\}$ sont deux systèmes complets d'événements de Ω .

1.3. Probabilité d'un événement, distribution (ou loi) de probabilité sur un ensemble, espace probabilisé

Le fait de dire que le dé est "équilibré" signifie que l'on a autant de "chance" d'obtenir le 1 que le 2 ... ou que le 6. On appelle "*probabilité*" d'un événement la part de "chance" de l'événement en décidant de donner la probabilité 1 à Ω (100% de chance).

Ainsi, on peut en déduire que l'on a 1 "chance" sur 6 d'obtenir le 1 ou le 2 ... ou le 6 ce que l'on exprimera par : la "probabilité" d'avoir le 1 est $1/6$, de même pour le 2, ..., de même pour le 6.

La probabilité de l'événement $A = \{2, 4, 6\}$ est alors $3/6$ car cet événement est la réunion de 3 événements élémentaires de probabilité $1/6$ chacun.

On peut ainsi définir la probabilité de tout événement (c'est-à-dire, de chacun des sous-ensembles de Ω).

Il faut distinguer la **probabilité d'un événement** particulier $P(A)$ (réel compris entre 0 et 1) et **l'application P définie sur l'ensemble des parties de Ω** qui à un événement A associe sa probabilité $P(A)$. Lorsque Ω est de cardinal n , l'ensemble des parties de Ω est de cardinal 2^n . Dans notre exemple, l'ensemble des parties de Ω est de cardinal 2^6 .

Cette application P est appelée indifféremment **probabilité** sur Ω ou **distribution de probabilité** sur Ω ou **loi de probabilité** sur Ω et elle est complètement déterminée par sa restriction aux événements élémentaires (ou singletons) de Ω , c'est-à-dire dans notre exemple, par l'ensemble :

$$\left\{ \left(1, \frac{1}{6}\right), \left(2, \frac{1}{6}\right), \left(3, \frac{1}{6}\right), \left(4, \frac{1}{6}\right), \left(5, \frac{1}{6}\right), \left(6, \frac{1}{6}\right) \right\}.$$

Un ensemble Ω muni d'une probabilité P sur Ω est appelé **espace probabilisé**. Il s'agit donc d'un ensemble pondéré qui peut être représenté par un diagramme en barres ou en bâtons.

Commentaires.

La notion de probabilité se construit sur la notion ambiguë de « chance », ambiguë car ce mot est utilisé indifféremment pour désigner des probabilités (nombres compris entre 0 et 1) ou des éventualités (qu'il s'agit de dénombrer).

L'idée d'attribuer un nombre compris entre 0 et 1 pour « mesurer » la chance d'obtenir telle ou telle éventualité n'est pas évidente et doit être précisée explicitement. Historiquement, la notion d'espérance mathématique a émergé avant la notion de probabilité.

1.4. Distribution de probabilité sur un ensemble fini ; exemples

Voici la définition adoptée au lycée.

Définitions. Soit $\{(\omega_i, p_i); i \in \{1, \dots, n\}\}$ où $\{\omega_i; i \in \{1, \dots, n\}\}$ est un ensemble fini Ω de cardinal n et $\{p_i; i \in \{1, \dots, n\}\}$ un ensemble de nombres réels positifs ou nuls de somme égale à 1. Pour tout sous-ensemble A de Ω , on pose $P(A) = \sum_{\{i; \omega_i \in A\}} p_i$.

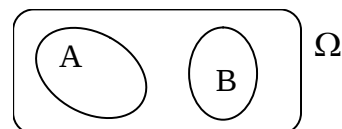
Ce nombre est appelé *probabilité de A* et l'application P ainsi définie sur l'ensemble des parties de Ω est appelée **probabilité** (ou **distribution de probabilité** ou **loi de probabilité**) sur Ω .

Convention. On a en particulier : pour tout $i = 1, \dots, n$, $p_i = P(\{\omega_i\})$ mais, afin d'alléger l'écriture, on pourra écrire $p_i = P(\omega_i)$.

Propriétés. On vérifie aisément les propriétés suivantes (visualisation des propriétés grâce aux diagrammes de Venn) :

- $P(\Omega) = 1$,

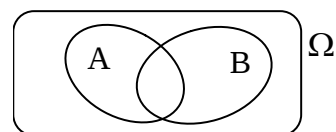
- si A et B sont deux événements incompatibles (deux parties disjointes), alors $P(A \cup B) = P(A) + P(B)$,



sinon $P(A \cup B) = P(A) + P(B) - P(A \cap B)$,

- $P(\bar{A}) = 1 - P(A)$,

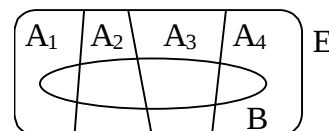
- $P(\emptyset) = 0$,



- si $A \subset B$ alors $P(A) \leq P(B)$.

Si $\{A_1, \dots, A_r\}$ est un système complet d'événements (une partition)

de Ω et B un événement, alors $P(B) = P(A_1 \cap B) + \dots + P(A_r \cap B)$



Cas particulier : l'équiprobabilité

Dans le cas particulier où tous les p_i sont égaux (équiprobabilité), leur somme étant égale à 1, ils sont tous égaux à $1/n$ et, pour toute partie A de Ω , la probabilité de A est alors égale à k/n où k est le cardinal de A ("nombre de cas favorables" sur "nombre de cas possibles"). Cette formule est connue sous le nom de **formule de Laplace** car Laplace est le premier à avoir explicitement formulé que la probabilité d'un événement est le "nombre de cas favorables" sur le "nombre de cas possibles" à condition que tous les cas possibles aient les mêmes chances d'advenir.

On dit que la probabilité P est *équirépartie* sur Ω ou *uniforme* sur Ω ou qu'il s'agit de l'**équiprobabilité** sur Ω .

Commentaires.

1) Dans le cas de l'équiprobabilité, le calcul des probabilités se ramène alors à un calcul de dénombrement (qui peut vite devenir compliqué).

Rappel : si E est un ensemble fini non vide de cardinal n et p un entier strictement positif,

- le nombre de p -listes (c'est-à-dire de p -uplets) d'éléments de E est n^p

- le nombre d'arrangements de p éléments de E (c'est-à-dire de p -listes d'éléments distincts), noté A_n^p , est $n(n-1)\dots(n-p+1)$ si $p \leq n$, 0 sinon ;

en utilisant la notation factorielle $n! = 1 \times \dots \times n$ et la convention $0! = 1$, on a : $A_n^p = \frac{n!}{(n-p)!}$

- le nombre de permutations de E (c'est-à-dire de n -arrangements de E) est $n!$

- le nombre de combinaisons ou parties ou sous-ensembles de p éléments de E , noté $\binom{n}{p}$

(ou C_n^p) est $\binom{n}{p} = \frac{A_n^p}{p!} = \frac{n!}{p!(n-p)!}$ si $p \leq n$, 0 sinon.

La convention permet de poser : $\binom{n}{0} = 1$ et $\binom{0}{0} = 1$.

2) Les échantillons aléatoires simples de taille n dans une population finie de taille N sont tirés avec équiprobabilité, on a :

- N^n échantillons dans le cas d'un tirage successif avec remise,
- A_N^n (avec $n \leq N$) dans le cas d'un tirage successif sans remise,
- $\binom{N}{n}$ (avec $n \leq N$) dans le cas d'un tirage simultané.

Premiers exemples d'équiprobabilité

1. Lancer d'un dé équilibré

Le couple (Ω, P) avec $\Omega = \{1, 2, \dots, 6\}$ et P équirépartie sur Ω est un espace probabilisé associé au lancer d'un dé équilibré.

2. Lancer d'une pièce de monnaie équilibrée

Un espace probabilisé associé au lancer d'une pièce de monnaie équilibrée est (Ω, P) avec $\Omega = \{p, f\}$ et P équirépartie sur Ω .

3. Tirage au hasard d'une boule dans un sac (ou une urne)

Un sac opaque contient n boules indiscernables au toucher numérotées de 1 à n . On tire "au hasard" une boule de l'urne et on observe le numéro marqué sur la boule.

Par convention, tirer "au hasard" signifie que l'on a autant de "chance" de tirer une boule qu'une autre.

Un espace probabilisé associé à cette expérience aléatoire est alors (Ω, P) avec $\Omega = \{1, 2, \dots, n\}$ et P équirépartie sur Ω .

Exemples de non équiprobabilité (obtenue à partir d'une équiprobabilité)

1. On lance un dé équilibré dont une face porte le chiffre 1, deux faces portent le chiffre 2 et trois faces le chiffre 3. Un espace probabilisé associé au lancer de ce dé est $\Omega = \{1, 2, 3\}$ avec

$P(\{1\}) = \frac{1}{6}$, $P(\{2\}) = \frac{2}{6}$ et $P(\{3\}) = \frac{3}{6}$. La loi de probabilité sur Ω n'est plus l'équiprobabilité mais on a utilisé l'équiprobabilité sur les sorties des 6 faces de ce dé équilibré pour construire l'espace probabilisé associé à notre expérience aléatoire.

2. Une urne contient 10 boules indiscernables au toucher, 3 vertes et 7 blanches. On tire au hasard une boule de l'urne et on observe sa couleur. Un espace probabilisé associé à cette

expérience est $\Omega = \{v, b\}$ avec $P(\{v\}) = \frac{3}{10}$ et $P(\{b\}) = \frac{7}{10}$.

On n'a plus équiprobabilité sur Ω mais on s'est servi de l'équiprobabilité sur les sorties des 10 boules de l'urne pour construire l'espace probabilisé associé à l'expérience aléatoire.

Exemples de non équiprobabilité : approche fréquentiste de la probabilité

1. On lance un dé dont les faces sont numérotées de 1 à 6 mais on a de sérieuses raisons de penser que le dé a été pipé. En effet, lors de 100 lancers de ce dé, on a observé les fréquences de sorties suivantes :

chiffres	1	2	3	4	5	6
fréquences	0.31	0.29	0.14	0.17	0.04	0.05

Un espace probabilisé associé au lancer de ce dé pipé peut être le couple (Ω, P) avec $\Omega = \{1, 2, \dots, 6\}$ et en choisissant comme distribution de probabilité P (ou loi de probabilité P) la distribution des fréquences observées, quitte à modifier ce choix si l'on réalise d'autres lancers de ce dé.

2. On lance une punaise qui peut retomber soit sur la tête soit sur la pointe. L'ensemble des résultats possibles est $\Omega = \{t, p\}$ mais il n'y a aucune raison que les résultats t et p aient les mêmes chances de sortir. Cela dépendra de la punaise.

Supposons qu'après 1 000 lancers d'une punaise, elle soit retombée 364 fois sur la tête et 636 fois sur la pointe. On pourra choisir comme espace probabilisé associé au lancer de cette punaise $\Omega = \{t, p\}$, avec $P(\{t\}) = 0.364$ et $P(\{p\}) = 0.636$, quitte à modifier ce choix si l'on réalise d'autres lancers de cette punaise.

Commentaire

Intuitivement, on estime la probabilité d'un événement lors d'une expérience par la fréquence d'apparition de l'événement sur une longue série d'expériences aléatoires réalisées indépendamment et dans les mêmes conditions. On verra que cette **approche fréquentiste de la probabilité** repose sur un théorème de probabilités, appelé la **loi des grands nombres**.

3. Chaque matin depuis cinq semaines un collégien relève sa durée d'attente du bus pour aller au collège. Les données sont présentées dans le tableau suivant :

durée en min	$0 \leq t < 2$	$2 \leq t < 4$	$4 \leq t < 6$	$6 \leq t < 8$	$8 \leq t < 10$	$10 \leq t < 12$
nombre d'observations	28	22	17	20	24	14
fréquence	0.224	0.176	0.136	0.160	0.192	0.112

Quelle est la probabilité qu'il attende demain moins de 2min ? entre 2 et 4 min ? ... plus de 8 min ? On peut les évaluer à 22.4%, 17.6%, ..., 11.2%. On suppose que les 125 durées sont des observations de variables aléatoires réelles (*v.a.r.*) indépendantes et de même distribution de probabilité (**i.i.d. indépendantes et identiquement distribuées**). Cette distribution de probabilité est alors approchée par la distribution de fréquence observée. La loi des grands nombres est appliquée à un système complet d'événements (c'est ainsi qu'est présentée la loi "vulgarisée" des grands nombres au lycée).

2. Probabilité conditionnelle, indépendance en probabilité

Cette partie est au programme de terminale. Elle n'est indiquée ici que pour discuter de la notion d'*indépendance en probabilité* qui est utilisée comme hypothèse de travail dans les situations proposées dès le collège (cf. approche fréquentiste de la probabilité et *v.a.r. i.i.d.* du § précédent). Lorsque l'on répète une expérience aléatoire dans les mêmes conditions, on utilise le fait que les résultats de la 2^{ème} expérience *ne dépendent pas* des résultats de la 1^{ère} expérience (absence de mémoire).

Cette partie peut être utile pour repérer, pour l'enseignement en troisième et en seconde, quelles sont les expériences à deux épreuves qui ne relèvent pas du programme.

2.1. Probabilité conditionnelle ; définition

Soit $\Omega = \{\omega_i ; i \in \{1, \dots, n\}\}$ un ensemble fini muni d'une distribution de probabilité $\{(\omega_i, p_i) ; i \in \{1, \dots, n\}\}$ et A un sous-ensemble de Ω de probabilité non nulle.

L'information apportée par la réalisation de A modifie la probabilité des événements ; on définit une nouvelle distribution de probabilité sur Ω , nulle sur le complémentaire de A et répartie proportionnellement aux p_i sur A , c'est-à-dire :

$$\{(\omega_i, q_i) ; i \in 1, n\} \text{ avec } q_i = \frac{p_i}{P(A)} \text{ si } \omega_i \in A, \quad q_i = 0 \text{ sinon.}$$

Cette nouvelle distribution de probabilité sur Ω est appelée **probabilité conditionnelle** à A (ou probabilité sachant A) et notée P_A .

$$\text{On a alors, pour toute partie } B \text{ de } \Omega, \quad P_A(B) = \frac{P(A \cap B)}{P(A)}.$$

Propriétés

$$P(A \cap B) = P(A)P_A(B) \text{ (si } A \text{ est de probabilité non nulle)}$$

Formule des probabilités totales :

- si A et $\bar{A}$ sont de probabilités non nulles,

$$P(B) = P(A \cap B) + P(\bar{A} \cap B) = P(A)P_A(B) + P(\bar{A})P_{\bar{A}}(B)$$

- si $\{A_1, \dots, A_r\}$ est un système complet d'événements de Ω , chacun de probabilité non nulle

$$P(B) = \sum_{i=1}^r P(A_i \cap B) = \sum_{i=1}^r P(A_i)P_{A_i}(B).$$

2.2 Indépendance en probabilité

Si A est de probabilité non nulle, on dira que B est indépendant en probabilité de A si la probabilité sachant A de B est égale à la probabilité de B (l'information selon laquelle A est réalisé ne modifie pas la probabilité de B).

On montre alors que l'on a $P(A \cap B) = P(A)P(B)$ et que, si B est indépendant en probabilité de A , la relation est symétrique, A et B sont indépendants en probabilité, d'où les définitions suivantes.

Définitions

1. Deux événements A et B de Ω sont indépendants en probabilité si $P(A \cap B) = P(A)P(B)$.

2. Deux partitions de Ω , $\{A_1, \dots, A_r\}$ et $\{B_1, \dots, B_c\}$, sont indépendantes en probabilité si pour tout $i = 1, \dots, r$ et tout $j = 1, \dots, c$, $P(A_i \cap B_j) = P(A_i)P(B_j)$.

Propriétés

1. Si A et B sont de probabilités non nulles alors :

$$A \text{ et } B \text{ sont indépendants en probabilité} \Leftrightarrow P_A(B) = P(B) \Leftrightarrow P_B(A) = P(A)$$

2. A et B sont indépendants en probabilité $\Leftrightarrow \bar{A}$ et B sont indépendants en probabilité

$$\Leftrightarrow A \text{ et } \bar{B} \text{ sont indépendants en probabilité} \Leftrightarrow \bar{A} \text{ et } \bar{B} \text{ sont indépendants en probabilité}$$

Les partitions $\{A, \bar{A}\}$ et $\{B, \bar{B}\}$ sont alors indépendantes en probabilité.

3. Si A et $\bar{A}$ sont de probabilités non nulles, alors :
 A et B sont indépendants en probabilité $\Leftrightarrow P_A(B) = P_{\bar{A}}(B)$.

Cf. annexe 3, Arbre de probabilité - construction

Cf. annexe 4, Arbre de dénombrement - arbre de probabilité - exemple.

Commentaire

La notion d'indépendance en *probabilité* dépend de la probabilité.

Pour illustrer ce commentaire, considérons l'exemple suivant.

On tire successivement et avec remise trois boules d'une urne contenant une proportion p ($0 < p < 1$) de boules blanches.

Soit A l'événement « la première boule du tirage est blanche » et B l'événement « le tirage comprend une et une seule boule blanche ». On montre que A et B sont indépendants en probabilité si et seulement si $p = 1/3$.

On a en effet : $P(A) = p$; $P(B) = 3 p (1 - p)^2$ et $P(A \cap B) = p (1 - p)^2$.

A et B sont indépendants en probabilité si et seulement si $P(A \cap B) = P(A) \times P(B)$, soit $p (1 - p)^2 = 3 p^2 (1 - p)^2$, soit $p = 1/3$.

3. Variable aléatoire, distribution (ou loi) de probabilité d'une variable aléatoire ; transport de loi

Définition

Soit (Ω, P) un espace probabilisé associé à une expérience aléatoire (on suppose ici Ω fini), on appelle *variable aléatoire* (v.a.) associée à l'expérience toute application X définie sur Ω .

Soit $X(\Omega) = \{x_1, \dots, x_r\}$ l'ensemble image (fini de cardinal inférieur ou égal à celui de Ω). On désigne par $[X = x_i]$, $i = 1, \dots, r$, l'image réciproque du singleton $\{x_i\}$ par l'application X , c'est-à-dire, l'ensemble des éléments de Ω qui ont pour image x_i .

L'ensemble $\{[X = x_i], i = 1, \dots, r\}$ est la partition de Ω engendrée par l'application X (c'est-à-dire, en utilisant le vocabulaire des événements, le système complet d'événements de Ω engendré par la variable aléatoire X).

Proposition et définition

On pose, pour tout x de $X(\Omega)$, $P_X(\{x\}) = P([X = x])$, alors P_X définit une probabilité sur $X(\Omega)$ appelée **distribution (ou loi) de probabilité de X** .

La probabilité P_X est appelée *probabilité image de P par X* . L'application X nous a permis de *transporter* la probabilité P sur Ω en une probabilité P_X sur $X(\Omega)$.

Commentaire

Comme on ne retient de la variable aléatoire que sa *loi de probabilité*, c'est-à-dire, l'ensemble des valeurs qu'elle prend avec les probabilités correspondantes, on introduit parfois les variables aléatoires sans faire référence à l'espace probabilisé de départ, comme une simple notation. On a déjà introduit implicitement des variables aléatoires dans le paragraphe précédent. Soit X la durée d'attente du bus, on suppose qu'il s'agit d'une variable aléatoire

réelle continue dont la loi de probabilité peut être approchée par la distribution de fréquences des durées observées.

On retiendra que les applications ou variables aléatoires permettent de **transporter des distributions de probabilités**.

Exemples

1. Une urne contient 10 boules indiscernables au toucher, 3 vertes et 7 blanches. On tire au hasard une boule de l'urne et on observe sa couleur. Avec des notations évidentes, on note $\Omega' = \{v_1, v_2, v_3, b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$ l'ensemble des résultats de l'expérience aléatoire et on choisit comme espace probabilisé associé (Ω', P') avec P' équirépartie sur Ω' . Soit X l'application définie sur Ω' qui à la sortie d'une boule verte associe v et à la sortie d'une boule blanche b . On a alors : $X(\Omega') = \{v, b\}$ et :

$$P_X(\{v\}) = P'(\{v_1, v_2, v_3\}) = \frac{3}{10} \text{ et } P_X(\{b\}) = P'(\{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}) = \frac{7}{10}.$$

On retrouve évidemment l'espace probabilisé proposé précédemment.

Ici, la variable aléatoire X est catégorielle ; on s'intéressera plus particulièrement aux variables aléatoires réelles, c'est-à-dire, à valeurs dans $\mathbf{R}$.

2. On lance un dé cubique équilibré dont les faces sont numérotées de 1 à 6 et on note X le nombre marqué par la face supérieure du dé. Dans cet exemple l'univers Ω est égal à $X(\Omega)$, la variable aléatoire X est l'application identique et la loi de probabilité P_X de X est identique à la probabilité P équirépartie sur Ω .

Cet exemple justifie d'appeler *distribution (ou loi) de probabilité* la *probabilité* définie sur l'univers pour décrire une expérience aléatoire, comme cela est proposé pour l'enseignement des probabilités et des variables aléatoires au lycée.

3. On lance un dé cubique équilibré deux fois de suite ; soit X la valeur absolue de la différence des points marqués. Quelle est la distribution de probabilité de X ? (utilisation d'un tableau de dénombrement, équiprobabilité sur les 36 cases du tableau).

4. On lance un dé cubique équilibré trois fois de suite ; soit X la somme des points marqués. Quelle est la distribution de probabilité de X ? (utilisation d'un arbre de dénombrement ébauché, 216 résultats équiprobables)

5. On lance une pièce de monnaie 10 fois de suite. On appelle *idemséquence* une suite ininterrompue de "pile" ou une suite ininterrompue de "face".

Soit X le nombre de "pile", soit Y le nombre d'idemséquences et Z la longueur de l'idemséquence la plus longue.

Pour préciser les définitions, au résultat suivant du lancer de la pièce de monnaie 10 fois de suite P P P F F P P F P P, on associe le nombre 7 pour la variable X , le nombre 5 pour la variable Y et le nombre 3 pour la variable Z .

Quelle est la distribution de probabilité de X ? de Y ? de Z ?

On peut utiliser un tableur pour obtenir les lois de probabilités *exactes* de ces variables aléatoires. Cf. annexe 5, Lancer d'une pièce de monnaie équilibrée 10 fois de suite

4. Modélisation et simulation

Nous n'avons encore parlé ni de modélisation, ni de simulation.

Exemples introductifs

1. a) On lance une pièce de monnaie équilibrée 100 fois de suite dans les mêmes conditions. On obtient 52 "pile" et 48 "face". Que parier sur le prochain lancer ? Pile ? Face ? Indifférent ?

1. b) On lance une punaise 100 fois de suite dans les mêmes conditions. Elle tombe sur la pointe 52 fois, elle tombe sur la tête 48 fois. Que parier sur le prochain lancer ? Pointe ? Tête ? Indifférent ?

2. a) On lance deux fois de suite une pièce de monnaie, quelle est la probabilité d'obtenir deux "pile" ?

2. b) On lance simultanément deux pièces de monnaie identiques, quelle est la probabilité d'obtenir deux "pile" ?

3. a) On lance deux fois de suite un dé équilibré dont les faces sont numérotées de 1 à 6 et on s'intéresse à la somme des points marqués. Quelle est la probabilité d'obtenir le 11 ? le 12 ?

3. b) On lance simultanément deux dés cubiques équilibrés indiscernables à l'œil ; les faces sont numérotées de 1 à 6 et on s'intéresse à la somme des points marqués. Quelle est la probabilité d'obtenir le 11 ? le 12 ?

4. Paradoxe du jeu de Passe-Dix

À la fin du XVI^{ème} siècle, le Duc de Toscane aurait posé à Galilée le problème suivant au sujet du jeu de Passe-Dix. Ce jeu consiste à lancer trois dés et à observer la somme des points marqués.

« Au jeu de Passe-Dix, on a 6 manières possibles de faire le 11 comme le 12. En effet :

Somme = 11 : "6,4,1" "6,3,2" "5,5,1" "5,4,2" "5,3,3" "4,4,3"

Somme = 12 : "6,5,1" "6,4,2" "6,3,3" "5,5,2" "5,4,3" "4,4,4"

Pourtant, lorsqu'on joue un grand nombre de fois, on s'aperçoit que le 11 sort plus souvent que le 12. Qu'en pensez-vous ? »

Commentaires succincts sur ces exemples

1) *Simulation* versus *modélisation et estimation*.

Dans le cas du lancer d'une "pièce équilibrée", hypothèse qui n'est pas l'objet de la discussion ici, on *simule* l'équiprobabilité sur {pile, face}. Les expériences étant faites "indépendamment dans les mêmes conditions", les 100 résultats de cette suite de lancers sont les observations de 100 v.a.r. i.i.d., chacune de même loi : l'équiprobabilité sur {pile, face}. Le résultat 52 "pile" et 48 "face" n'apporte aucune information, la différence avec 50 "pile" et 50 "face" vient de la fluctuation de l'échantillonnage. Le processus est sans mémoire ; pour le prochain lancer, il est indifférent de parier sur la sortie de "pile" ou sur la sortie de "face".

En revanche, dans le cas du lancer de la punaise, on *modélise* la situation : soit p la probabilité que la punaise tombe sur la "pointe", et donc $1 - p$ qu'elle tombe sur la "tête". Les expériences étant faites "indépendamment et dans les mêmes conditions", les 100 résultats de cette suite de lancers sont les observations de 100 v.a.r. i.i.d., chacune de même loi : p pour "pointe" et $1 - p$ pour "tête". On peut *estimer* p par la fréquence observée sur ces 100 lancers 0.52. Pour le prochain lancer, on a un léger avantage à parier pour "pointe".

2) *Qualité d'un modèle*

Pour le lancer d'une pièce équilibrée deux fois de suite, l'espace probabilisé associé est sans ambiguïté l'équiprobabilité sur l'ensemble des quatre résultats $\{(P,P), (P,F), (F,P), (F,F)\}$.

En revanche, pour le lancer simultané de deux pièces équilibrées indiscernables, deux distributions de probabilité sont en compétition sur l'ensemble des trois résultats $\{PP, FF, PF\}$ (où PF désigne 1 pile et 1 face sans ordre particulier puisque les deux pièces sont lancées simultanément et sont indiscernables) : $1/3, 1/3, 1/3$ ou $1/4, 1/4, 1/2$.

Se pose le problème du choix de modèle et de la qualité d'un modèle : un grand nombre d'expériences réelles réalisées dans les mêmes conditions permettra de constater que la distribution de fréquence des trois résultats est plus proche de la seconde distribution de probabilité que de la première.

La simulation ne résout pas le problème du choix de modèle car on ne sait simuler que le second modèle, on introduit nécessairement dans la programmation un ordre qui n'apparaît pas dans l'expérience réelle.

3) Variante de l'exemple 2

Pour le lancer d'un dé cubique équilibré deux fois de suite on a équiprobabilité sur les 36 résultats possibles.

Pour le lancer simultané de deux dés cubiques équilibrés indiscernables à l'œil, on n'a pas équiprobabilité sur les 21 résultats visibles mais une probabilité $1/36$ pour chacun des six "doubles" (deux chiffres identiques) et une probabilité $1/18$ pour chacun des 15 "simples" (deux chiffres distincts).

4) Variante des exemples 2 et 3

Les résultats amenant le 11 (ou le 12) ne sont pas équiprobables. La probabilité est de $1/216$ pour les "triples", de $3/216$ pour les "doubles" et de $6/216$ pour les "simples, ce qui donne une probabilité de $27/216$ pour une somme égale à 11 et $25/216$ pour une somme égale à 12.

L'intérêt de cet exemple historique est de voir que le paradoxe repose sur les deux approches intuitives des probabilités : équiprobabilité pour des raisons de symétrie et approche fréquentiste des probabilités. Pour arriver expérimentalement à repérer que "le 11 sort plus souvent que le 12", il a fallu jouer un très grand nombre de fois !!

5. Variable aléatoire réelle ; lois discrètes et lois continues, espérance mathématique et variance

Dans le cas où l'application X est à valeurs dans l'ensemble des nombres réels, on parle de *variable aléatoire réelle* (v.a.r.). On note P sa loi de probabilité sur R , c'est-à-dire l'ensemble des valeurs prises par la variable aléatoire et les probabilités correspondantes.

Fonction de répartition

On pose, pour tout réel x , $F(x) = P([-\infty; x])$.

L'application F de R dans $[0; 1]$ est appelée *fonction de répartition* (f.d.r.) de X et caractérise la distribution (ou loi) de probabilité de X .

Lois de probabilités discrètes

Lorsque $X(\Omega) = \{x_i; i \in I\}$ avec I partie finie ou dénombrable de R , la v.a.r. ainsi que sa loi de probabilité P sur R , $\{(x_i, p_i); i \in I\}$, est dite *discrète*.

Les lois de probabilités discrètes usuelles présentées en annexe 6, finies (loi de Bernoulli, loi binomiale, loi hypergéométrique) ou dénombrables (loi géométrique, loi de Poisson), sont à valeurs entières positives.

Lois de probabilités continues

Lorsque $X(\Omega)$ est un intervalle (ou une réunion d'intervalles) de R , la v.a.r. ainsi que sa loi de probabilité P sur R est dite *absolument continue* ou à *densité* s'il existe une application f positive et continue (sauf éventuellement en un nombre fini de points) d'intégrale sur R égale à 1 telle que l'on ait, pour tout x de R , $F(x) = \int_{-\infty}^x f(u)du$. L'application f (qui n'est pas unique) est appelée densité de probabilité de la loi de probabilité P .

Les lois de probabilités continues usuelles présentées en annexe 6 sont les lois uniformes sur des intervalles bornés de R , les lois normales définies sur R et les lois exponentielles définies sur R^+ .

Espérance et variance de lois de probabilités sur R

Cas où la loi de probabilité P sur R est discrète $\{(x_i, p_i); i \in I\}$ avec I dénombrable.

À condition que la somme $\sum_{i \in I} p_i x_i$ soit finie, on définit l'*espérance mathématique* de P ou d'une variable aléatoire réelle X ayant P pour loi de probabilité par $E(P) = E(X) = \sum_{i \in I} p_i x_i$

À condition que la somme $\sum_{i \in I} p_i x_i^2$ soit finie, on définit la *variance* de P ou d'une variable aléatoire réelle X ayant P pour loi de probabilité par $V(P) = V(X) = \sum_{i \in I} p_i (x_i - E(X))^2$

Cas où la loi de probabilité P sur R est continue de densité f .

À condition que la somme $\int_{-\infty}^{+\infty} x f(x)dx$ soit finie, on définit l'*espérance mathématique* de P ou d'une variable aléatoire réelle X ayant P pour loi de probabilité par $E(P) = E(X) = \int_{-\infty}^{+\infty} x f(x)dx$.

À condition que la somme $\int_{-\infty}^{+\infty} x^2 f(x)dx$ soit finie, on définit la *variance* de P ou d'une variable aléatoire réelle X ayant P pour loi de probabilité par $V(P) = V(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 f(x)dx$

Lorsque la variance existe, l'*écart-type* est défini comme la racine carrée de la variance.

Propriétés de l'espérance et de la variance

Si X et Y sont des variables aléatoires réelles définies sur un même espace probabilisé, a et b des nombres réels, on a :

- $E(aX + bY) = a E(X) + b E(Y)$ (l'espérance est linéaire) et $V(aX + b) = a^2 V(X)$

- si les variables aléatoires réelles X et Y sont indépendantes en probabilité, on a :

$$V(X + Y) = V(X) + V(Y)$$

6. Théorèmes limites des probabilités (loi des grands nombres et théorème limite central)

Soit X une variable aléatoire réelle (v.a.r.) admettant une moyenne (ou espérance mathématique) μ et un écart-type σ . Soit $(X_i)_{i \in \mathbb{N}_*}$ une suite de v.a.r., indépendantes et identiquement distribuées (i.i.d.) comme X ("identiquement distribuée comme X " signifie "de même distribution (ou loi) de probabilité que X "); $(X_i)_{i=1, \dots, n}$ est appelé *échantillon statistique* de taille n .

On pose $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$, cette v.a.r. est appelée moyenne d'échantillonnage.

Les propriétés de l'espérance mathématique et de la variance permettent d'établir les égalités suivantes : $E(\bar{X}_n) = \mu$ et $\text{var}(\bar{X}_n) = \frac{\sigma^2}{n}$; l'espérance de la suite $(\bar{X}_n)$ est une suite constante égale à μ et la variance de la suite $(\bar{X}_n)$ converge vers 0 lorsque n tend vers l'infini.

Le premier théorème limite (*loi faible des grands nombres*) consiste à dire que la suite de v.a.r. $(\bar{X}_n)$ converge en probabilité vers une v.a.r. d'espérance μ et de variance nulle, c'est-à-dire, vers la v.a.r. constante (dégénérée) μ .

(Définition : une suite de v.a.r. (Y_n) converge en probabilité vers une v.a.r. Y (noté $(Y_n) \xrightarrow{\text{Pr}} Y$) si $\forall \varepsilon > 0, \lim_{n \rightarrow \infty} P(|Y_n - Y| > \varepsilon) = 0$.)

Le deuxième théorème limite (*théorème central limite*) permet de préciser la vitesse de convergence de la suite $(\bar{X}_n)$ vers μ . Pour cela on définit la variable $\bar{X}_n$ standardisée en

posant : $\bar{X}_n^* = \frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}}$; la v.a.r. $\bar{X}_n^*$ est obtenue par centrage et réduction de la v.a.r. $\bar{X}_n$.

Il s'agit donc d'une v.a.r. centrée réduite : $E(\bar{X}_n^*) = 0$ et $\text{var}(\bar{X}_n^*) = 1$

Le deuxième théorème limite (*théorème central limite*) consiste à dire que la suite $(\bar{X}_n^*)$ converge en loi vers la loi normale centrée réduite, notée $N(0,1)$.

(Définition : une suite de v.a.r. (Y_n) converge en loi vers une v.a.r. Y (noté $(Y_n) \xrightarrow{\text{Loi}} Y$) si on a, en tout point de continuité y de la fonction de répartition F_Y de Y , $\lim_{n \rightarrow \infty} F_{Y_n}(y) = F_Y(y)$.)

En résumé, les deux théorèmes limites des probabilités sont :

Loi faible des grands nombres :	$(\bar{X}_n) \xrightarrow{\text{Pr}} \mu$
Théorème central limite :	$\left(\frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} \right) \xrightarrow{\text{Loi}} N(0,1)$

Définition : L'intervalle de fluctuation de la variable aléatoire réelle $\bar{X}_n$ au niveau de probabilité de 95% est (asymptotiquement) :

$$\left[\mu - 1.96 \frac{\sigma}{\sqrt{n}} ; \mu + 1.96 \frac{\sigma}{\sqrt{n}} \right]$$

La preuve vient de la lecture de la table d'une loi normale centrée réduite ; si U est une v.a.r. normale centrée réduite la probabilité qu'elle prenne ses valeurs entre -1.96 et $+1.96$ est 0.95.

Cas particulier: la fréquence d'échantillonnage

Si X est une v.a.r. de loi de Bernoulli de probabilité p (probabilité de succès lors d'une épreuve aléatoire). On a alors $\mu = p$ et $\sigma = \sqrt{p(1-p)}$, $\sum_{i=1}^n X_i$ (nombre de succès lors de n épreuves aléatoires indépendantes) suit une loi binomiale de paramètres n et p et $\bar{X}_n$ (notée F_n) est la fréquence de succès lors des n épreuves.

Les deux théorèmes limites deviennent dans ce cas particulier :

Loi faible des grands nombres : Théorème central limite :	$(F_n) \xrightarrow{\text{Pr}} p$ $\left(\frac{F_n - p}{\sqrt{p(1-p)/n}} \right) \xrightarrow{\text{Loi}} N(0,1)$
--	--

Définition : L'intervalle de fluctuation de la variable aléatoire F_n au niveau de probabilité de 95% est (asymptotiquement) :

$$\left[p - 1.96 \sqrt{\frac{p(1-p)}{n}} ; p + 1.96 \sqrt{\frac{p(1-p)}{n}} \right]$$

Cf. annexe 7, Tableau des approximations de lois de probabilité

Extension du théorème central limite

On pose : $V_n = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ (variance d'échantillonnage).

Les propriétés sur les espérances et variances permettent d'établir l'égalité : $E(V_n) = \frac{n-1}{n} \sigma^2$.

Pour $n > 1$, on pose : $S_n^2 = \frac{n}{n-1} V_n = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$, on a alors $E(S_n^2) = \sigma^2$.

La formule définissant V_n est la "formule de variance pour une population finie", alors que celle définissant S_n^2 est la "formule de variance pour un échantillon" (appelée parfois variance corrigée) car, lorsque l'on travaille sur un échantillon, c'est pour estimer au mieux les paramètres inconnus de la loi de probabilité sous-jacente ; l'espérance de la v.a.r. S_n^2 est égale à σ^2 ; on dit que la v.a.r. S_n^2 est un estimateur sans biais de σ^2 ce qui n'est pas le cas de V_n .

Le théorème central limite peut être généralisé au cas où l'on remplace le paramètre inconnu σ par la v.a.r. S_n (racine carrée de S_n^2) ou par la v.a.r. S'_n (racine carrée de V_n) :

$$\left(\frac{\bar{X}_n - \mu}{S_n / \sqrt{n}} \right) \xrightarrow{\text{Loi}} N(0,1) \quad \text{et} \quad \left(\frac{F_n - p}{\sqrt{F_n(1-F_n)/n}} \right) \xrightarrow{\text{Loi}} N(0,1)$$

7. Estimation par intervalle de confiance

La statistique inférentielle a pour objectif d'obtenir de l'information et de prendre des décisions à partir d'un échantillon. Elle est constituée de trois parties : échantillonnage, estimation et tests.

La partie échantillonnage consiste à chercher les distributions de probabilité de variables aléatoires définies sur des échantillons (moyenne d'échantillonnage, variance d'échantillonnage, fréquence d'échantillonnage). Cette partie repose sur le calcul des probabilités. Dans le paragraphe précédent, nous avons obtenu la loi de probabilité (approchée) de la moyenne d'échantillonnage et de la fréquence d'échantillonnage.

La partie estimation consiste, à partir de l'observation d'UN échantillon à inférer des informations sur la distribution de probabilité dont est issue l'échantillon.

On suppose que les valeurs $x_1, \dots, x_n$ sont les observations de n v.a.r. i.i.d. $X_1, \dots, X_n$ de loi de probabilité inconnue mais dont on suppose l'existence d'une moyenne μ et d'un écart-type σ .

Alors la moyenne $\bar{x}_n$ et l'écart-type s_n sont les observations de variables aléatoires $\bar{X}_n$ et S_n qui, pour n suffisamment grand (dans la pratique supérieur à 30), vérifient la propriété suivante (cf. dernier résultat du paragraphe précédent) :

$$P\left(\bar{X}_n - 1.96 \frac{S_n}{\sqrt{n}} \leq \mu \leq \bar{X}_n + 1.96 \frac{S_n}{\sqrt{n}}\right) \approx 0.95$$

On en déduit l'estimation de la moyenne μ par intervalle à 95% de confiance :

$$\left[\bar{x}_n - 1.96 \frac{s_n}{\sqrt{n}} ; \bar{x}_n + 1.96 \frac{s_n}{\sqrt{n}} \right]$$

De la même manière, on obtient l'estimation de la probabilité p par intervalle à 95% de confiance :

$$\left[f_n - 1.96 \sqrt{\frac{f_n(1-f_n)}{n}} ; f_n + 1.96 \sqrt{\frac{f_n(1-f_n)}{n}} \right]$$

Commentaires

1) On parle ici de "confiance" car on ne peut plus parler de "probabilité", n'ayant qu'UN échantillon, on construit les intervalles sur des observations de variables aléatoires.

2) Dans le cas où l'échantillonnage est réalisé dans une population finie (approche sondage), la moyenne μ n'est autre que la moyenne arithmétique sur la population du caractère quantitatif étudié et la probabilité p n'est autre que la proportion par rapport à la population de la sous-population d'intérêt.

3) L'approche sondage de la statistique inférentielle est beaucoup plus concrète que l'approche modèle. Cf., à la page "Statistique Inférentielle" du site, les documents :

De la statistique descriptive élémentaire à la statistique inférentielle en passant par le calcul des probabilités (7 pages)

La statistique inférentielle par une approche sondage : échantillonnage dans une population finie (5 pages)

4) Pour approfondir ses connaissances en statistique inférentielle cf., à la page "Statistique Inférentielle" du site, les documents :

Échantillonnage, estimation et tests : 2 fiches (2 pages)

- l'échantillonnage, l'estimation et le test d'égalité à une valeur fixée pour une proportion

- l'échantillonnage et le test d'adéquation à l'équiprobabilité

Vers les tests du χ^2 (3 pages)

Petit cours de statistique inférentielle (17 pages, prépa capes de maths).

Résultats au baccalauréat d'enseignement général dans l'académie de Toulouse, session 2000

Section	Nombre de candidats		Taux de reçus	
	Filles	Garçons	Filles	Garçons
Econom. et Social	2114	1301	82.2%	76.7%
Littéraire	2464	535	83.8%	79.4%
Scientifique	3171	3890	85.2%	81.3%
Ensemble	7749	5726	83.9%	80.1%

(Source : Rectorat de l'Académie de Toulouse)

Ensemble des reçus selon la section et le sexe (distributions d'effectifs) :

Section	Filles	Garçons	Ensemble
Econom. et Social	1738	998	2736
Littéraire	2065	425	2490
Scientifique	2702	3163	5864
Ensemble	6505	4585	11090

Ensemble des reçus selon la section et le sexe (distributions de fréquences) :

Section	Filles	Garçons	Ensemble
Econom. et Social	16%	9%	25%
Littéraire	19%	4%	22%
Scientifique	24%	29%	53%
Ensemble	59%	41%	100%

Ensemble des reçus selon le sexe (distributions conditionnelles à la section) :

Section	Filles	Garçons	Ensemble
Econom. et Social	64%	36%	100%
Littéraire	83%	17%	100%
Scientifique	46%	54%	100%
Ensemble	59%	41%	100%

Ensemble des reçus selon la section (distributions conditionnelles au sexe) :

Section	Filles	Garçons	Ensemble
Econom. et Social	27%	22%	25%
Littéraire	32%	9%	22%
Scientifique	42%	69%	53%
Ensemble	100%	100%	100%

Lien entre fréquence et probabilité, entre fréquence conditionnelle et probabilité conditionnelle

Résultats au baccalauréat de l'enseignement général selon le sexe
dans l'académie de Toulouse, session 2000

Enseignement général	Répartition	Taux de reçus
Filles	57.5%	84%
Garçons	42.5%	81%
Ensemble	100.0%	

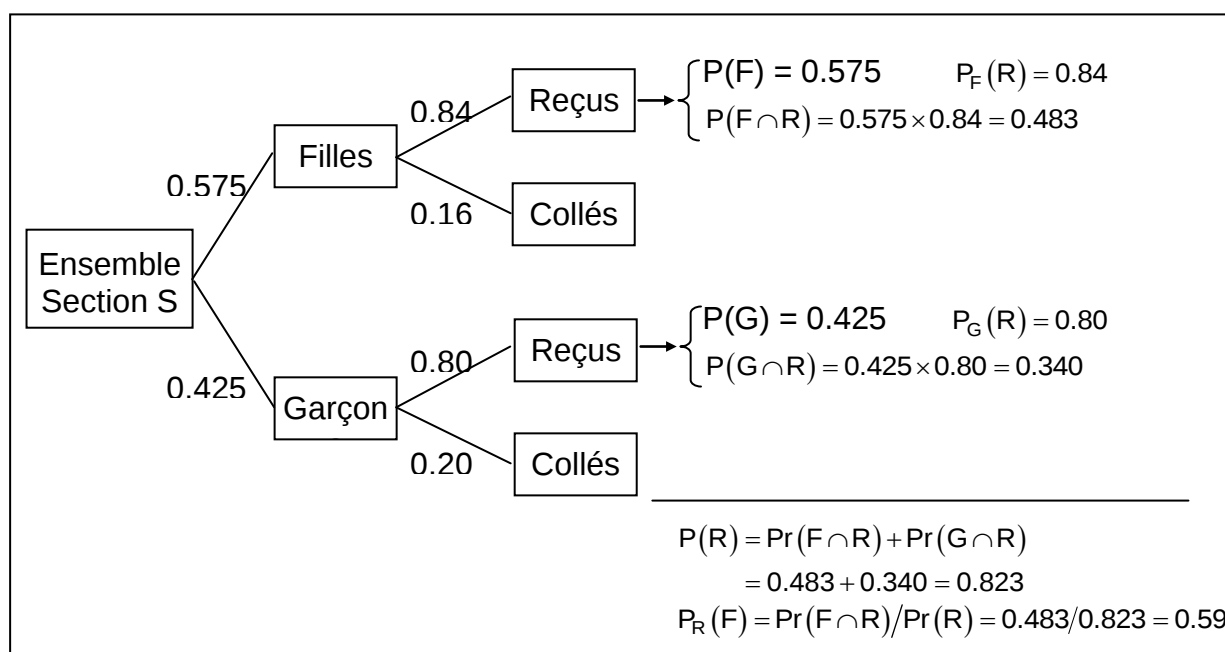
(Source : Rectorat de l'Académie de Toulouse)

On choisit *au hasard* (c'est-à-dire avec équiprobabilité) une personne candidate au baccalauréat de l'enseignement général de l'académie de Toulouse en 2000.

- Quelle est la probabilité que ce soit une fille ?
- Quelle est la probabilité sachant que c'est une fille qu'elle ait été reçue ?
- Quelle est la probabilité que la personne candidate ait été reçue ?
- Quelle est la probabilité sachant que la personne candidate a été reçue que ce soit une fille ?

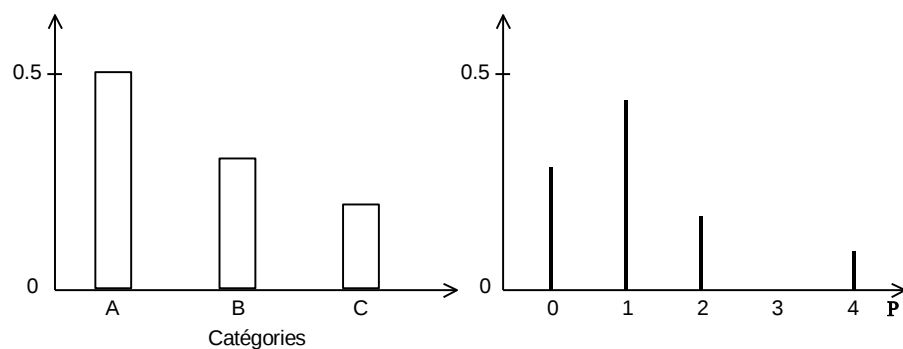
L'ensemble des résultats de cette expérience aléatoire est l'ensemble des candidats au baccalauréat de l'enseignement général et P est l'équiprobabilité. La probabilité d'un événement est le nombre de cas favorables sur le nombre de cas possibles, c'est-à-dire, la proportion (ou fréquence) de l'ensemble qui représente l'événement.

Arbre de probabilité conditionnelle (ou de fréquence conditionnelle) :

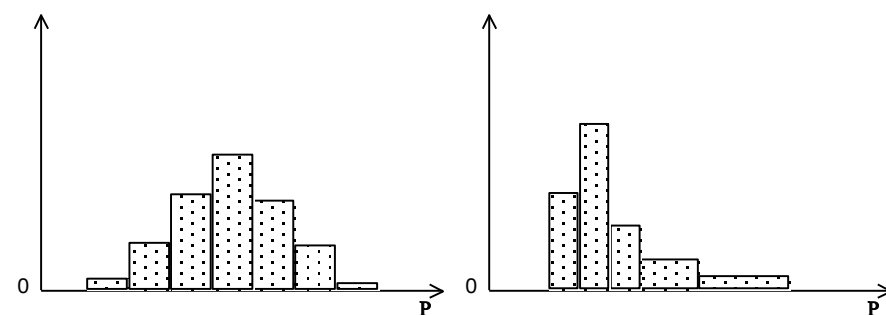


On utilise les mêmes arguments et les mêmes formules (règle du produit, règle de la somme) dans le cadre des fréquences conditionnelles ou dans le cadre des probabilités conditionnelles. On reconnaît ici la formule de Bayes.

Distributions d'effectifs et de fréquences et représentations graphiques associées ("histogrammes")

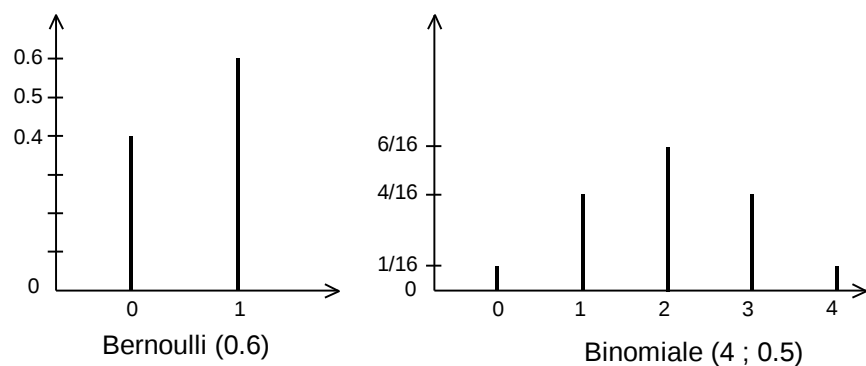


Hauteurs proportionnelles aux effectifs et aux fréquences



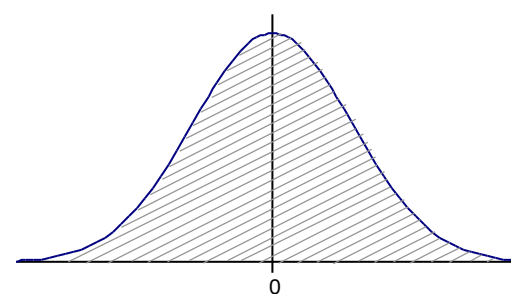
Aires proportionnelles aux effectifs et aux fréquences

Distributions de probabilité

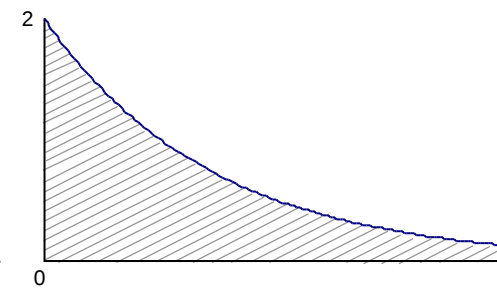


Bernoulli (0.6)

Binomiale (4 ; 0.5)



Normale (0 ; 1)



Exponentielle (2)

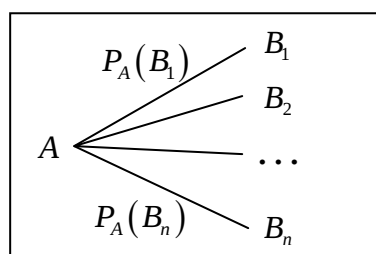
Règles de construction d'un arbre de probabilité

Soit $(\Omega, P(\Omega), P)$ un espace probabilisé. On note A un événement et $\bar{A}$ l'événement contraire, tous deux supposés de probabilité non nulle, $\{B_1, \dots, B_n\}$ un système complet d'événements tous de probabilité non nulle (noté $\{B, \bar{B}\}$ dans le cas $n = 2$).

Un *arbre de probabilité* (ou arbre "pondéré", les pondérations étant ici des probabilités) est en général présenté de haut en bas ou de gauche à droite avec pour racine Ω .

Les sommets (ou nœuds) correspondent à des événements et les branches issues d'un sommet, A par exemple, ont pour extrémité les événements d'un système complet, $\{B_1, \dots, B_n\}$ par exemple ; les branches sont pondérées par les probabilités conditionnelles à A de ces événements et la somme des poids des branches issues d'un même sommet est égale à 1 :

$$\sum_{i=1}^n P_A(B_i) = 1.$$



On note que la probabilité conditionnelle à Ω n'est autre que P .

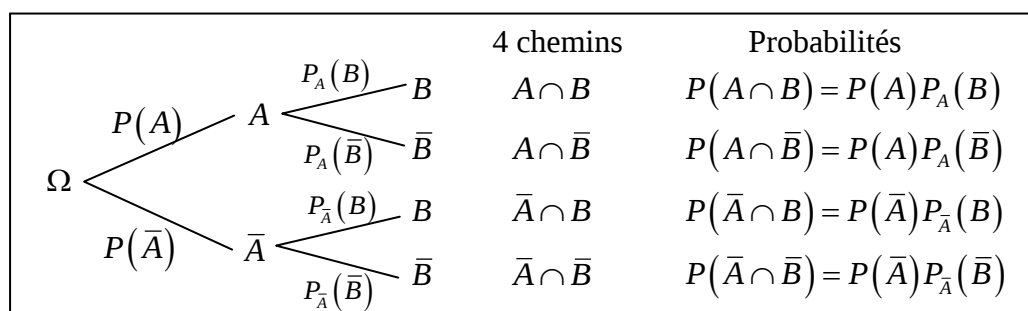
Un *chemin* parcouru de la racine Ω jusqu'à l'extrémité des branches correspond à la conjonction (l'intersection) des événements rencontrés ; la probabilité associée est égale au produit des probabilités rencontrées (règle du produit pour un arbre pondéré) :

$$P(A \cap B) = P(A)P_A(B).$$

La probabilité d'un événement égal à la réunion d'événements deux à deux disjoints situés à l'extrémité de plusieurs chemins est égale à la somme des probabilités de ces événements (règle de la somme pour un arbre pondéré) :

$$B = (A \cap B) \cup (\bar{A} \cap B) \text{ et } (A \cap B) \cap (\bar{A} \cap B) = \emptyset$$

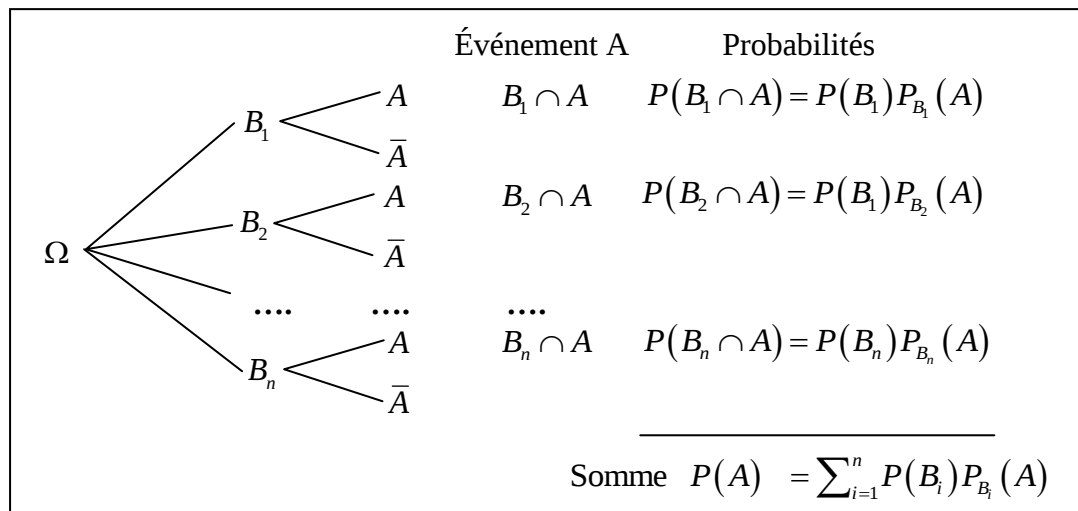
$$P(B) = P(A \cap B) + P(\bar{A} \cap B)$$



Remarque : A et B sont indépendants en probabilité si et seulement si $P_A(B) = P_{\bar{A}}(B)$.

On a en effet :
$$\frac{P(A \cap B)}{P(A)} = \frac{P(\bar{A} \cap B)}{P(\bar{A})} = \frac{P(A \cap B) + P(\bar{A} \cap B)}{P(A) + P(\bar{A})} = P(B).$$

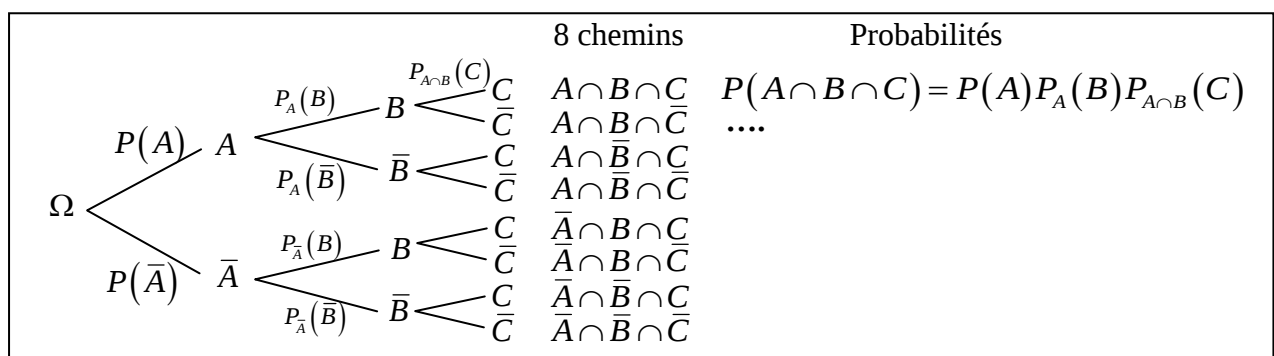
Formule des probabilités totales : $P(A) = \sum_{i=1}^n P(B_i \cap A) = \sum_{i=1}^n P(B_i)P_{B_i}(A)$.



Formules de Bayes : $\forall j \in 1, n, P_A(B_j) = \frac{P(B_j \cap A)}{P(A)} = \frac{P(B_j)P_{B_j}(A)}{\sum_{i=1}^n P(B_i)P_{B_i}(A)}$.

Si l'arbre est à trois niveaux, par exemple, $A, \bar{A}$, puis $B, \bar{B}$, enfin $C, \bar{C}$, alors les pondérations du dernier niveau sont les probabilités de C et $\bar{C}$ conditionnellement à respectivement $A \cap B, A \cap \bar{B}, \bar{A} \cap B, \bar{A} \cap \bar{B}$ (on suppose que ces quatre événements sont de probabilité non nulles) ; on dénombre alors 8 chemins et on a, par exemple pour le premier chemin, la propriété :

$$P(A \cap B \cap C) = P(A)P_A(B)P_{A \cap B}(C)$$



Cette propriété se généralise à n événements ($n > 3$), $A_1, \dots, A_n$, qui ne forment évidemment pas ici un système complet. Si $P(A_1 \cap \dots \cap A_{n-1}) \neq 0$ alors on a :

$$P(A_1 \cap \dots \cap A_n) = P(A_1)P_{A_1}(A_2)P_{A_1 \cap A_2}(A_3) \dots P_{A_1 \cap \dots \cap A_{n-1}}(A_n)$$

Arbre de dénombrement et arbre de probabilité - Exemple

On tire "au hasard" (c'est-à-dire avec équiprobabilité) successivement et avec remise (resp. sans remise) deux boules d'une urne contenant 3 boules numérotées de 1 à 3.

Pour tout i de $\{1, 2, 3\}$, on note A_i (resp. B_i) l'événement "obtenir la boule i au premier tirage" (resp. "obtenir la boule i au deuxième tirage").

Proposition

Que le tirage soit avec ou sans remise, on a, pour tout i de $\{1, 2, 3\}$, $P(A_i) = \frac{1}{3}$ et $P(B_i) = \frac{1}{3}$.

Dans le cas du tirage avec remise, on a, pour tout i de $\{1, 2, 3\}$ et tout j de $\{1, 2, 3\}$, $P(A_i \cap B_j) = \frac{1}{9}$; les événements A_i et B_j sont indépendants en probabilité.

Dans le cas du tirage sans remise, on a pour tout i de $\{1, 2, 3\}$ et tout j de $\{1, 2, 3\}$, $P(A_i \cap B_j) = \frac{1}{6}$ si $i \neq j$, 0 sinon; les événements A_i et B_j ne sont pas indépendants en probabilité.

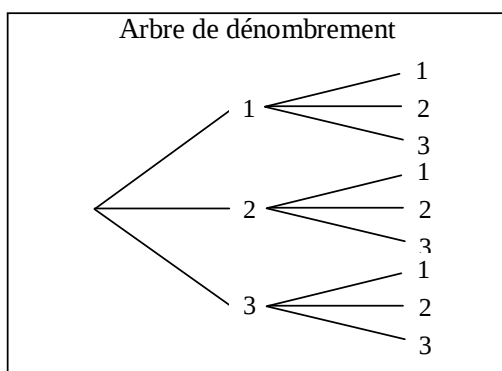
Preuves

Que ce soit pour le tirage avec remise ou le tirage sans remise,

- on peut utiliser un *arbre de dénombrement* pour lister l'ensemble des résultats possibles et utiliser la proposition selon laquelle l'équiprobabilité à chaque tirage entraîne l'équiprobabilité sur l'ensemble des résultats des tirages successifs (cette proposition est admise à tous les niveaux d'enseignement secondaire)
- on peut utiliser l'information fournie sous forme de probabilité conditionnelle (*arbre de probabilité*) et en déduire la probabilité des événements élémentaires.

Tirage avec remise

- Arbre de dénombrement



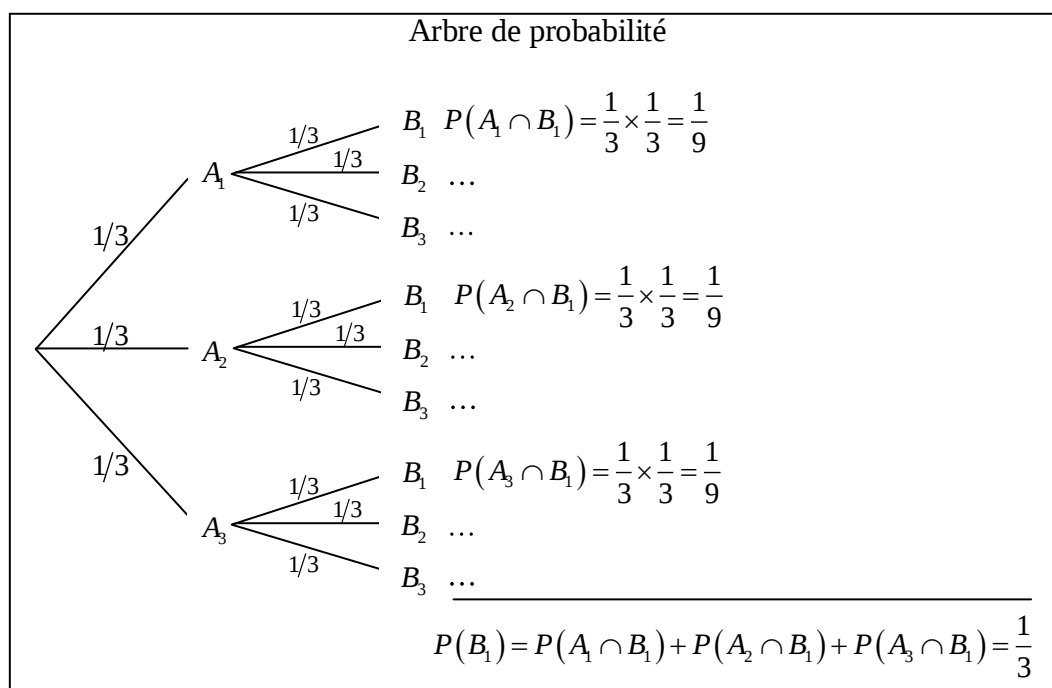
$\Omega = \{(1,1), (1,2), (1,3), (2,1), (2,2), (2,3), (3,1), (3,2), (3,3)\}$ et P est l'équiprobabilité sur Ω .

On a pour tout i de $\{1, 2, 3\}$ et tout j de $\{1, 2, 3\}$:

$$A_i = \{(i,1), (i,2), (i,3)\}, B_j = \{(1,j), (2,j), (3,j)\}, A_i \cap B_j = \{(i,j)\}$$

dont on déduit : $P(A_i) = \frac{1}{3}$, $P(B_j) = \frac{1}{3}$ et $P(A_i \cap B_j) = \frac{1}{9}$ et le résultat sur l'indépendance.

Arbre de probabilité



On a par hypothèse, pour tout i de $\{1, 2, 3\}$, $P(A_i) = \frac{1}{3}$ et pour tout j de $\{1, 2, 3\}$ $P_{A_i}(B_j) = \frac{1}{3}$.

On en déduit, pour tout i de $\{1, 2, 3\}$ et tout j de $\{1, 2, 3\}$

$P(\{(i, j)\}) = P(A_i \cap B_j) = P(A_i)P_{A_i}(B_j) = \left(\frac{1}{3}\right)^2 = \frac{1}{9}$ (règle du produit pour un arbre pondéré). On retrouve l'équiprobabilité des événements élémentaires.

Enfin on en déduit : pour tout j de $\{1, 2, 3\}$, $P(B_j) = \sum_{i \in \{1, 2, 3\}} P(A_i \cap B_j) = \sum_{i \in \{1, 2, 3\}} \frac{1}{9} = \frac{1}{3}$ (règle de la somme pour un arbre pondéré) et le résultat sur l'indépendance.

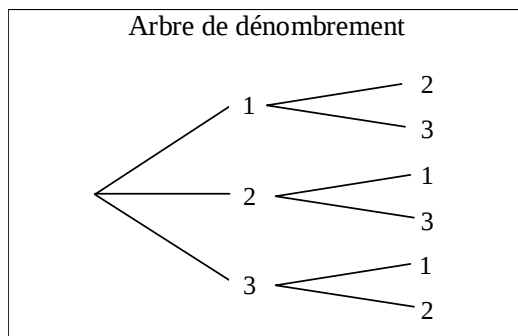
- **Tableau : distribution de probabilité conjointe et de distributions marginales des résultats des premier et deuxième tirages**

1 ^{er} tirage \ 2 ^o tirage	B_1	B_2	B_3	
A_1	1/9	1/9	1/9	1/3
A_2	1/9	1/9	1/9	1/3
A_3	1/9	1/9	1/9	1/3
	1/3	1/3	1/3	1

La distribution conjointe est égale au produit des marges : indépendance en probabilité des deux tirages.

Tirage sans remise

- Arbre de dénombrement

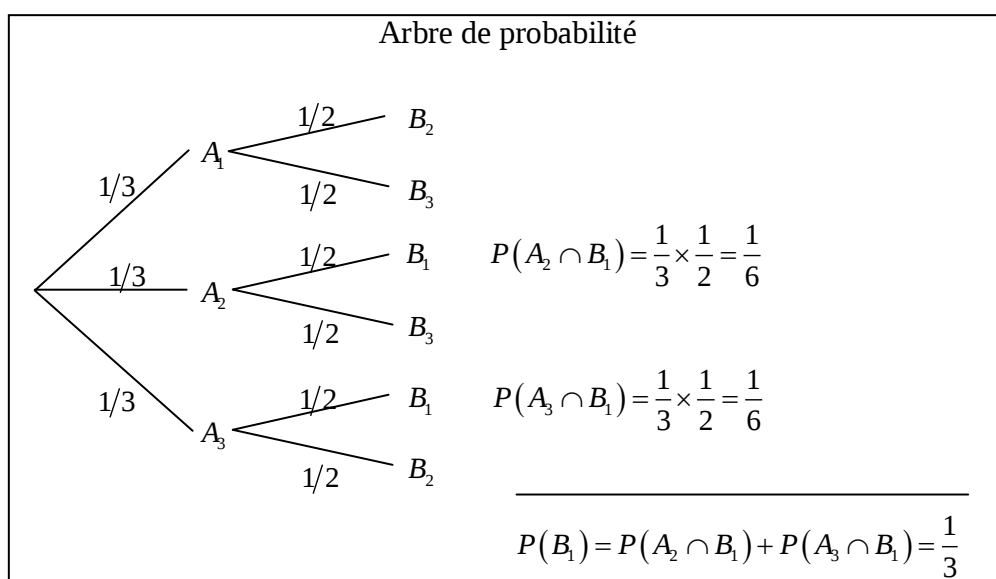


$\Omega = \{(1, 2), (1, 3), (2, 1), (2, 3), (3, 1), (3, 2)\}$ et P est l'équiprobabilité sur Ω .

On a : $A_1 = \{(1, 2), (1, 3)\}$, $A_2 = \{(2, 1), (2, 3)\}$, $A_3 = \{(3, 1), (3, 2)\}$,
 $B_1 = \{(2, 1), (3, 1)\}$, $B_2 = \{(1, 2), (3, 2)\}$, $B_3 = \{(1, 3), (2, 3)\}$, et, pour tout i de $\{1, 2, 3\}$ et tout j de $\{1, 2, 3\}$: $A_i \cap B_j = \{(i, j)\}$ si $i \neq j$, $\emptyset$ sinon.

On déduit : $P(A_i) = \frac{1}{3}$, $P(B_j) = \frac{1}{3}$ et $P(A_i \cap B_j) = \frac{1}{6}$ si $i \neq j$, 0 sinon.

- Arbre de probabilité



On a par hypothèse, pour tout i de $\{1, 2, 3\}$, $P(A_i) = \frac{1}{3}$ et pour tout j de $\{1, 2, 3\}$

$P_{A_i}(B_j) = \frac{1}{2}$ si $i \neq j$, 0 sinon. On en déduit, pour tout i de $\{1, 2, 3\}$ et tout j de $\{1, 2, 3\}$

avec $i \neq j$, $P(\{(i, j)\}) = P(A_i \cap B_j) = P(A_i)P_{A_i}(B_j) = \left(\frac{1}{3}\right) \times \left(\frac{1}{2}\right) = \frac{1}{6}$ (règle du produit pour un arbre pondéré). On retrouve l'équiprobabilité des événements élémentaires.

Enfin on en déduit : pour tout j de $\{1, 2, 3\}$, $P(B_j) = \sum_{i \in \{1, 2, 3\}} P(A_i \cap B_j) = \sum_{i \in \{1, 2, 3\} \setminus \{j\}} \frac{1}{6} = \frac{1}{3}$ (règle de la somme pour un arbre pondéré).

- **Tableau : distribution de probabilité conjointe et de distributions marginales des résultats des premier et deuxième tirages.**

$\begin{matrix} 2^{\circ} \text{ tirage} \\ 1^{\text{er}} \text{ tirage} \end{matrix}$	B_1	B_2	B_3	
A_1	0	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$
A_2	$\frac{1}{6}$	0	$\frac{1}{6}$	$\frac{1}{3}$
A_3	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{3}$
	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

La distribution conjointe n'est pas égale au produit des marges, non indépendance en probabilité des deux tirages.

On lance une pièce de monnaie équilibrée 10 fois de suite (1024 séries équiprobables).

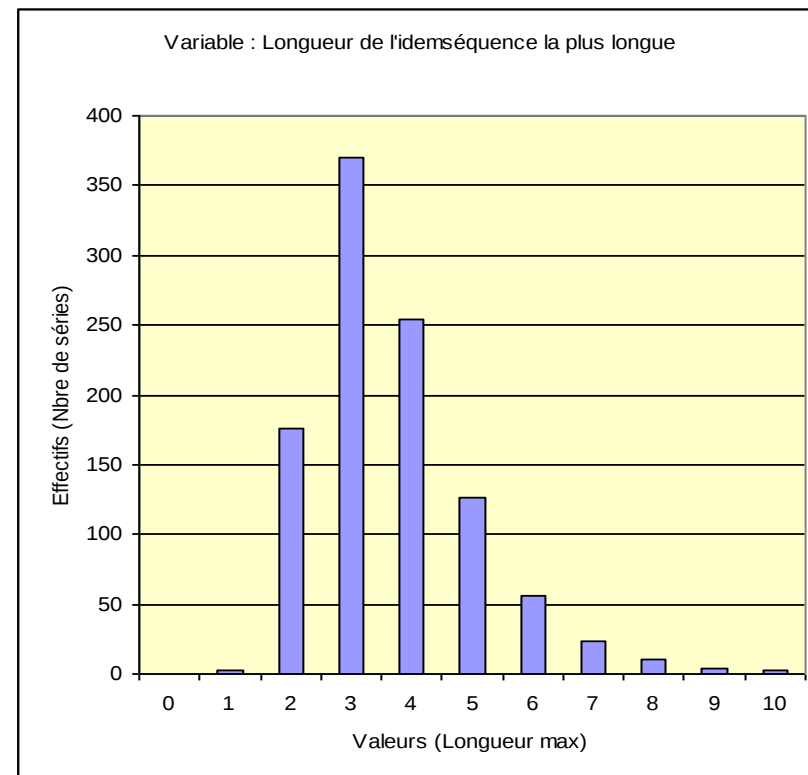
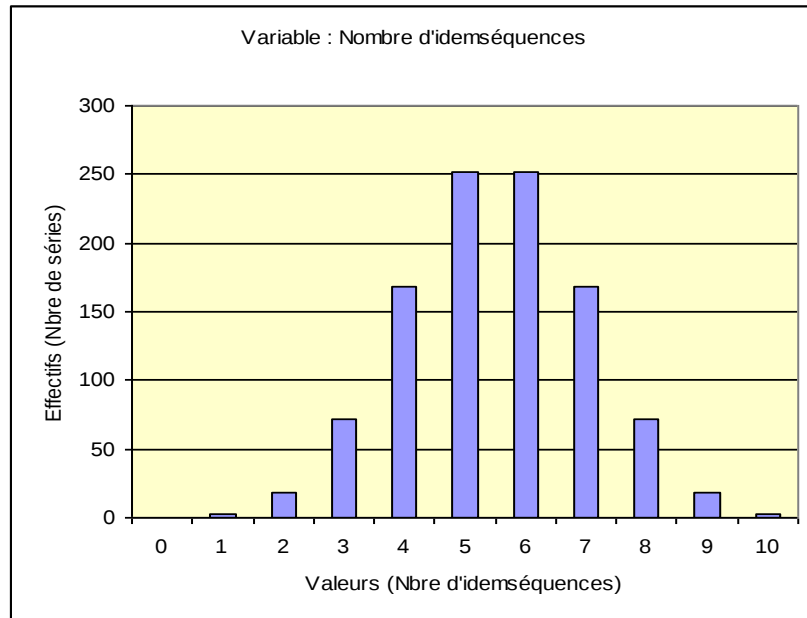
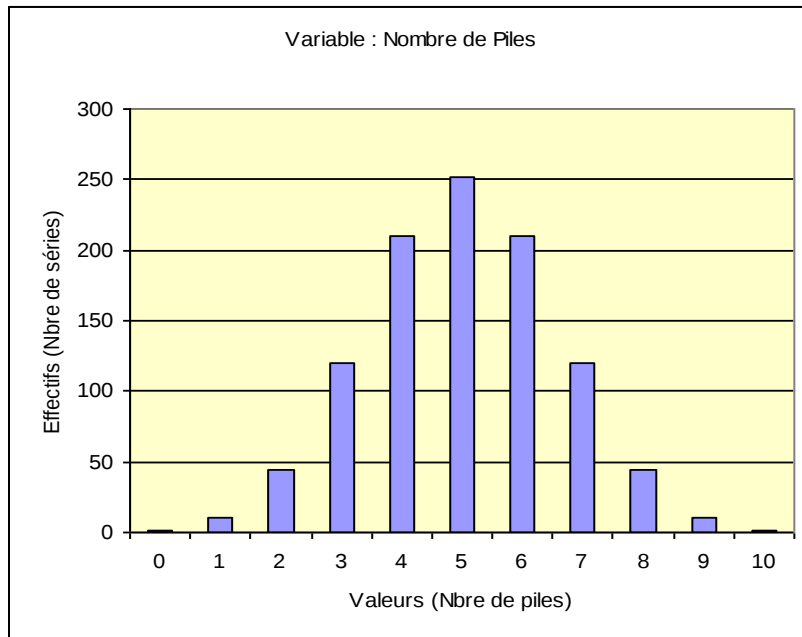
Distribution d'effectifs (et de probabilité en divisant par 1024) des variables :

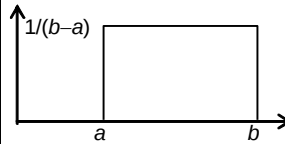
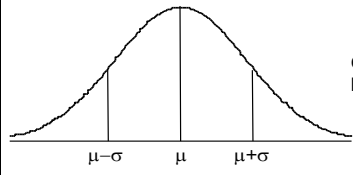
X : nombre de piles

Y : nombre d'identséquences

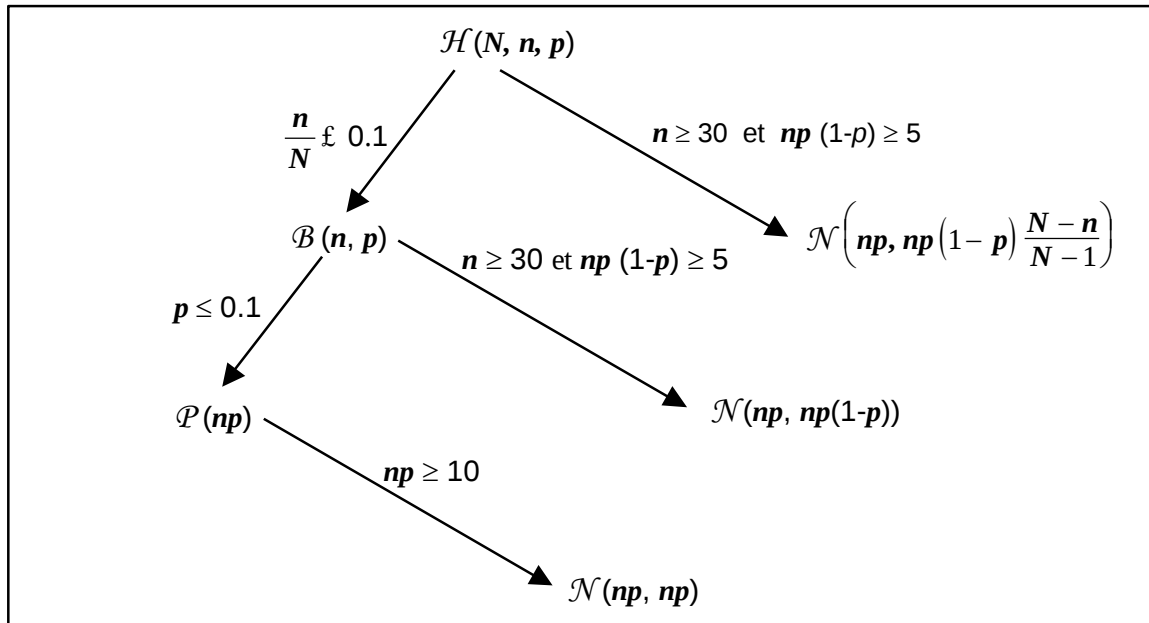
Z : longueur de l'identséquence la plus longue

Valeurs	Eff. X	Eff. Y	Eff. Z
0	1	0	0
1	10	2	2
2	45	18	176
3	120	72	370
4	210	168	254
5	252	252	126
6	210	252	56
7	120	168	24
8	45	72	10
9	10	18	4
10	1	2	2
Total	1024	1024	1024



NOM	NOTATION	VALEURS	LOI DE PROBABILITÉ	MOYENNE	VARIANCE	APPLICATIONS
Loi uniforme sur $\{1, \dots, n\}$	$\mathcal{U}_n$ $n \in \mathbf{N}^*$	$\{1, 2, \dots, n\}$	$p_k = 1/n$	$(n+1)/2$	$(n^2-1)/12$	On tire "au hasard" une boule dans une urne contenant n boules numérotées de 1 à n ; X : numéro de la boule tirée $X \propto \mathcal{U}_n$
Loi de Bernoulli	$\mathcal{B}(1, p)$ $p \in]0, 1[$	$\{0, 1\}$	$p_0 = 1-p \quad p_1 = p$	p	$p(1-p)$	a) On tire "au hasard" une boule dans une urne dont la proportion de boules rouges est p ; X : nombre de boules rouges tirées $X \propto \mathcal{B}(1, p)$; b) Indicatrice du succès dans une épreuve de Bernoulli
Loi binomiale	$\mathcal{B}(n, p)$ $n \in \mathbf{N}^* \quad p \in]0, 1[$	$\{0, 1, 2, \dots, n\}$	$p_k = \binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$	a) On tire "au hasard" et avec remise n boules dans une urne dont la proportion de boules rouges est p ; X : nombre boules rouges tirées $X \propto \mathcal{B}(n, p)$ b) Nombre de succès dans un schéma de Bernoulli de taille n
Loi hypergéométrique	$\mathcal{H}(N, n, p)$ où $p=K/N$ avec $n \in \mathbf{N}^* \quad n \leq N$ $K \in \mathbf{N}^* \quad N \in \mathbf{N}^*, K < N$	$\{0, 1, 2, \dots, n\}$	$p_k = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$	np	$np(1-p) \frac{N-n}{N-1}$	On tire "au hasard" et sans remise (ou simultanément) n boules dans une urne contenant N boules dont K rouges ; X : nombre de boules rouges tirées $X \propto \mathcal{H}(N, n, p)$ avec $p = K/N$
Loi de Poisson	$\mathcal{P}(\alpha) \quad \alpha \in \mathbf{R}_*^+$	$\mathbf{N}$	$p_k = e^{-\alpha} \frac{\alpha^k}{k!}$	α	α	Loi des événements rares (approximation de $\mathcal{B}(n, p)$ pour p petit, $np = \alpha$)
Loi géométrique	$\mathcal{G}(p) \quad p \in]0, 1[$	$\mathbf{N}^*$	$p_k = p(1-p)^{k-1}$	$1/p$	$(1-p)/p^2$	Modèle de durée discret (sans mémoire) X : nombre de tirages avec remise nécessaires pour obtenir la 1ère boule rouge dans une urne dont la proportion de boules rouges est p ; $X \propto \mathcal{G}(p)$
Loi uniforme sur $[a, b]$	$\mathcal{U}[a, b]$ $a \in \mathbf{R} \quad b \in \mathbf{R}, b > a$	$[a, b]$	$f(x) = 1/(b-a)$	$(a+b)/2$	$(b-a)^2/12$	 On choisit "au hasard" un point x de $[a, b]$ $X \propto \mathcal{U}_{[a, b]}$
Loi normale ou de Laplace-Gauss	$\mathcal{N}(\mu, \sigma^2)$ $\mu \in \mathbf{R} \quad \sigma \in \mathbf{R}_*^+$	$\mathbf{R}$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	μ	σ^2	 Cf. Théorème central limite
Loi exponentielle	$\mathcal{E}(\lambda)$ $\lambda \in \mathbf{R}_*^+$	$\mathbf{R}^+$	$f(x) = \lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	 Modèle de durée continue (sans mémoire)

Approximations des lois de probabilités



Formule de Stirling :

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$$