PROBA IS COMING

Cet article a pour objectif de vous donner les premières cordes à votre arc pour affronter le monde de la probabilité sans appréhension et avec confiance, tels de fiers Dothrakis.

Les bases de la probabilité

1. Les outils fondamentaux : Ensembles et Dénombrement

Dans un premier temps, nous souhaitons vous exposer les notions élémentaires de la probabilité pour que nous puissions parler le même langage tout au long de cet article.

1.a. Ensembles

Définition : “ Par ensemble, nous entendons toute collection M d'objets m [...] définis et distincts, ces objets étant appelés les éléments de M” (M Georg Cantor).

Exemple [Ensemble] : L’ensemble L des lecteurs de cet article est composé de tous les lecteurs de cet article.

Notation : L’élément m appartenant à l’ensemble M se note $m \in M$.

L’élément n n’appartenant pas à l’ensemble M se note $n \notin M$.

Un ensemble peut être fini ou infini.

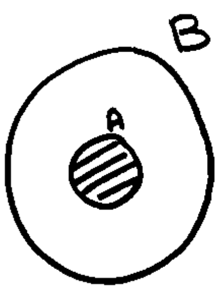
Exemple [Fini] : L’ensemble “Alphabet” noté A est composé de 26 lettres: A = {a,b,c … x,y,z}

Exemple [Infini] : L’ensemble des nombres réels compris entre 0 et 1.

L’ensemble $[0,1] \in \mathbb{R}$ est dénombrable mais infini car il existe une infinité de nombres réels qui peuvent être listés sans répétition ni omission entre les bornes de l’intervalle.

Et si nous rassemblons les éléments d’un ensemble ?

Nous pouvons grouper les éléments d’un ensemble en sous-ensemble. Nous disons alors que le sous-ensemble est inclus dans l’ensemble.



Notation :

$A \subset B$ signifie que A est inclus dans B.

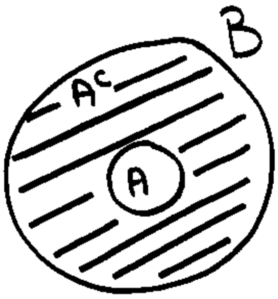
$A \not\subset B$ signifie que A n’est pas inclus dans B

Exemple **[Inclusion]** : L’ensemble “Voyelle” est un sous-ensemble de l’ensemble “Alphabet” : Voyelle$\in$Alphabet

Quand est-il pour les éléments exclus d’un ensemble ?

Le complément d’un ensemble est formé par tous les éléments n’appartenant pas à cet ensemble.

Notation : A^c est le complement de A



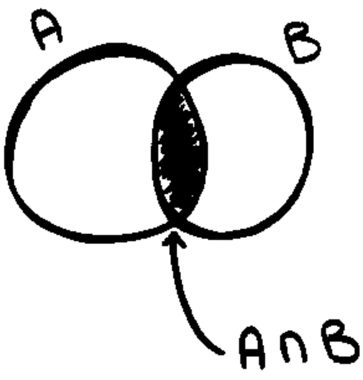
Exemples **[Complémentarité]**:

- Le complément de l’ensemble «Voyelle» comprend toutes les lettres sauf les voyelles.
- Soit l’ensemble N composé d’éléments suivants : $N=\{1,2,3,4,5,6,7,8,9\}$. Nous notons $P=\{2,4,6,8\}$ le sous-ensemble de N.
- Le complément de P est $P^c=\{1,3,5,7,9\}$

Et si nous nous intéressons aux éléments de deux ensembles?

L’intersection de deux ensembles A et B forme l’ensemble composé des éléments appartenant **à A et à B.**

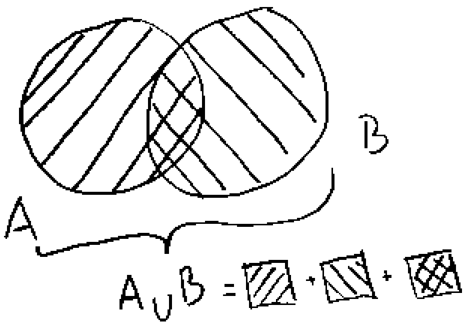
Notation : $A\cap B$ designe l'intersection de A et de B



Exemple **[Intersection]** : Soient l’ensemble A= {1,2,3,4,5,6} et l’ensemble B = {4,5,6,7,8,9}. L’ensemble C, intersection de A et de B, est tel que C = {4,5,6}

L’union de deux ensembles A et B est l’ensemble composé des éléments de **A ou B.**

Notation : $A \cup B$ désigne l'union de A et de B



Ainsi nous comprenons que $A \cup B = A + B$

> **Exemple** **[Union]** : Soient l’ensemble A= {1,2,3,4,5,6} et l’ensemble B = {4,5,6,7,8,9}. L’ensemble D, union de A et de B est tel que D = {1,2,3,4,5,6,7,8,9}

1.b. Dénombrement

Le dénombrement permet d’énumérer divers types de groupements que l’on peut faire à partir d’ensemble finis.

Le dénombrement peut aussi être appelé ‘Analyse combinatoire’. Il est né de l'étude des jeux de hasard. Il est par ailleurs lié à la théorie des nombres et à la théorie des graphes.

Permutation

Définition : Tout classement ordonné de n éléments distincts est une permutation de ces n éléments. Par exemple aebcd est une permutation des éléments a, b, c, d, e et dabce en est une autre.

Soit E un ensemble fini composé de n éléments. Une permutation des n éléments est tout classement dans un ordre souhaité que l’on peut faire avec ces n éléments.

Le nombre de permutations d’un ensemble E comprenant n éléments est égal à n!

> **Exemple** **[Permutation]** : Il y a 10 ! façons de ranger 10 livres sur un étagère.

Arrangement

Définition : Un arrangement est une collection de k objets pris ***successivement parmi n en tenant compte de l'ordre d'apparition****. Il est dit simple si on ne peut prendre chaque objet qu'une fois au plus.*

Dans un ensemble E de n éléments, un k-arrangement est un sous-ensemble ordonné de k éléments de E pris sans répétition.

Le nombre d'arrangements de k éléments distincts choisis dans un ensemble E de n éléments est donné par la formule $A_n^k = \frac{n!}{(n-k)!}$.

> **Exemple** **[Arrangement]** :Après les prolongations d'un match de football, l'entraîneur doit choisir les 5 tireurs de penalty parmi les onze joueurs et l'ordre de passage. Le nombre de choix qu'il a est : $A_{11}^5 = \frac{11!}{(11-5)!}$ = 55440

Combinaisons

Définition : Une combinaison est une collection de k objets pris ***simultanément parmi n, donc sans tenir compte de l'ordre d'apparition.*** *Elle est dite simple si on ne peut prendre chaque objet qu'une fois au plus.*

Dans un ensemble E comprenant n éléments, une combinaison est tout sous-ensemble de E comprenant k éléments.

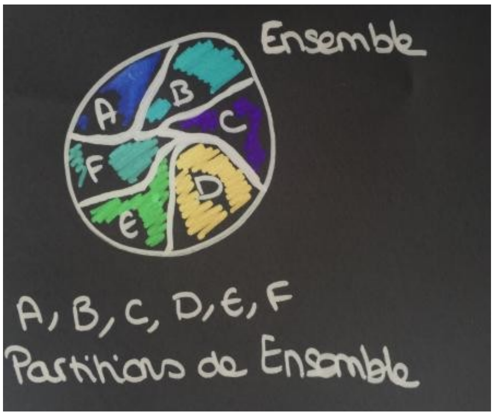
Le nombre de combinaisons de k éléments choisi parmi n est donné par la formule $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

> **Exemple** **[Combinaison]** : Nous tirons simultanément 9 cartes d'un jeu de 36 cartes. Nous pouvons avoir $\binom{36}{9} = \frac{36!}{9!27!}$ mains différentes.

Très utilisée pour le calcul des probabilités !

Partition

Dans un ensemble E , nous appelons P une partition de E l'ensemble de sous-ensembles P_i disjoints non vides tel que $\sum_i P_i = E$.



2. L'Univers des possibles et l'événement

En probabilité nous nous intéressons à des **événements** qui découlent d'une **expérience aléatoire**.

Définition : Une expérience aléatoire est une expérience dont le résultat ne dépend que du hasard.

Il s'agit d'une expérience renouvelable dont le résultat ne peut pas être prévu à l'avance. Cependant lorsque cette expérience est renouvelée, elle ne donnera pas forcément le même résultat.
Toutefois, nous supposons que nous sommes capable de connaître tout les résultats possibles de cette expérience, *a priori*.

Exemples **[Expérience aléatoire] :**

- On lance une pièce de monnaie. Le résultat « PILE » ou « FACE » ne dépend (en théorie) que du hasard.
- On lance un dé, l'apparition de l'une des six faces ne dépend que du hasard.
- On tire quatre cartes d'un jeu de 32 cartes ...

Notre expérience aléatoire donne lieu à un ensemble, dont les éléments de cet ensemble sont tous les résultats possibles de cette expérience.

Nous appelons **Ω** l'*ensemble des possibles*, ou l'*ensemble universel* constitué de tout les résultats possibles d'une expérience aléatoire. Tous ces résultats possibles sont notés ω_i et sont appelés « événements élémentaires ». Ce sont les événements qui composent un à un notre univers.

Nous appelons ω_i *le i-ème résultat possible, ou outcome, élément de Ω.*

Exemple **[Univers] :**

- Soit l'expérience aléatoire *E :* « On lance un dé équilibré».
  Ω={1;2;3;4 ;5 ;6}

Nous appelons ***événement*** un sous-ensemble de Ω. Un événement est une partie de l'univers, formée d'une ou de plusieurs issues possibles.

Un événement élémentaire est un événement composé d'un seul élément ω_i .

Le tableau suivant synthétise bien toutes les notions que nous venons d'aborder ensemble :

Notations	Vocabulaire ensembliste	Vocabulaire probabiliste	Exemple : on lance un dé
$\emptyset$	Ensemble vide	événement impossible	« obtenir un nombre négatif »
Ω	ensemble total	événement certain	$\Omega = \{1, 2, 3, 4, 5, 6\}$
ω	élément de Ω	événement élémentaire	« obtenir la face 2 »
A	sous-ensemble de Ω	événement	« obtenir un nombre pair »
$B \subset A$	B est inclus dans A	B implique A	$A =$ « obtenir un nombre pair » $B =$ « obtenir un multiple de 4 »
$A \cup B$	Union de A et B	A **ou** B	$A =$ « obtenir un nombre pair » $B =$ « obtenir un multiple de 3 » $A \cup B = \{2, 3, 4, 6\}$
$A \cap B$	Intersection de A et B	A **et** B	$A \cap B = \{6\}$
$\overline{A}$	Complémentaire de A	événement contraire de A	$\overline{A} =$ « obtenir un nombre impair »
$A \cap B = \emptyset$	A et B disjoints	A et B incompatibles	$A =$ « Obtenir un nombre pair » $B =$ « Obtenir un multiple de 5 »
$A \cap B = \emptyset$ $A \cup B = \Omega$	A et B complémentaires	A et B contraires	$A =$ « obtenir un nombre pair » $B =$ « obtenir un nombre impair »

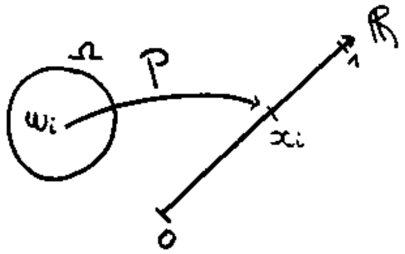
3. Outil de mesure

3.a. La probabilité, qu'est-ce donc cela ?

Nous allons à présent construire un outil mesurant la chance qu'un événement de Ω se réalise. Nous appellerons cette mesure : la probabilité de l' événement.

Une probabilité mesure la chance que l'on alloue à la réalisation d'un événement particulier de l'univers considéré. Cette mesure est comprise entre 0 et 1 . Une probabilité de zéro signifiant que l'événement considéré ne se réalisera jamais tandis qu'une probabilité de 1 signifie que l'événement considéré est sûre de se réaliser.

Cet outil de mesure nous permet de passer d'un ensemble Ω, formé de toutes les issues possibles de l'expérience aléatoire à un ensemble de réels compris en [0,1] :



A chaque élément ω_i est associé une probabilité $P(\omega_i)=x_i$.

Et si nous souhaitons mesurer une expérience aléatoire mettant en scène deux événements ?

Soient deux événements disjoints : A et B sont deux sous-ensembles de Ω. La probabilité de l'union de deux événements est égale à la somme de la probabilité de chacun des événements.

Les probabilités reposent sur une théorie axiomatique qui permet de cadrer les multiples calculs que l'on peut faire pour répondre à des problèmes de probabilités plus ou moins complexes.
Ainsi, nous posons les trois axiomes des probabilités :

- $Axiome\ de\ Non-Negativité : P(A) \geq 0$

Une probabilité est une mesure : c'est un nombre positif.

- $Axiome\ de\ Normativité : P(\Omega)=P(\sum \omega_i)=1$

La probabilité d'un événement absolument certain est 1.

Si A,B,C,D,E sont des événements qui partitionnent Ω alors P(A)+ P(B)+P(C)+P(D)+P(E)=1

- $Axiome\ d'additivité : si\ A \cap B = \emptyset\ alors\ P(A \cup B) = P(A) + P(B)$

Si deux événements sont bien distincts, la probabilité que ces deux événements se réalisent tous les deux est égale à la somme de leur probabilité respective.

Quelques conséquences :

- $P(A) \leq 1\,; P(\emptyset)=0\,; P(A \cup A^c)=P(A)+P(A^c)=1\,; P(A \cap A^c)=0$
- si $A \subset B$ alors $P(A) \leq P(B)\,; P(A \cup B)=P(A)+P(B)-P(A \cap B)\,; P(A \cup B) \leq P(A)+P(B)$
- $P(A \cup B \cup C)=P(A)+P(A^c \cap B)+P(A^c \cap B^c \cap C)$

3.b. Elle se calcule comment ?

Algorithme général de calcul d'une probabilité

1. Définir l'espace des possibles Ω.
2. Définir une loi de probabilité.
3. Identifier un événement intéressant
4. Utiliser la loi pour calculer la probabilité

Univers fini , événements discrets

Soit Ω , espace des possible finis, constitué de n outcomes discrets :

$$\Omega=\{\omega_1, \omega_2 \ldots, \omega_n\} \forall n \in \mathbb{N}$$

Nous supposons que chacun des outcomes a autant de probabilité de se réaliser :

C'est le cas de l'équiprobabilité !

$$P(\omega_i)=P(\omega_j) \forall i, j \in [1, n | \mathbb{N}]$$

Nous pouvons en déduire que : $P(A_i)=\frac{1}{n}$ où $A_i=\omega_i, \forall i \in [1, n | \mathbb{N}]$

Nous observons un événement intéressant, B , composé de k outcomes :

$$B=\{\omega_1, \omega_2, \ldots, \omega_k\}\ ;\ k \leq n\ ;\ k \in [1, n | \mathbb{N}]$$

Nous pouvons calculer la probabilité de B en rapportant la quantité d'outcomes de B à la quantité total d'outcomes de Ω (les ω_i):

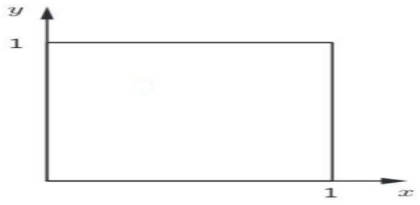
$$P(B)=\frac{k}{n}$$

Plus généralement nous écrivons : $P(B)=\frac{Card(B)}{Card(\Omega)}$

Card(B) et Card(Ω) sont deux nombres qui peuvent être déterminés par un dénombrement complet ou par analyse combinatoire (dans le cas où on est dans une expérience de tirage simultané).

__Univers fini, événements continus__

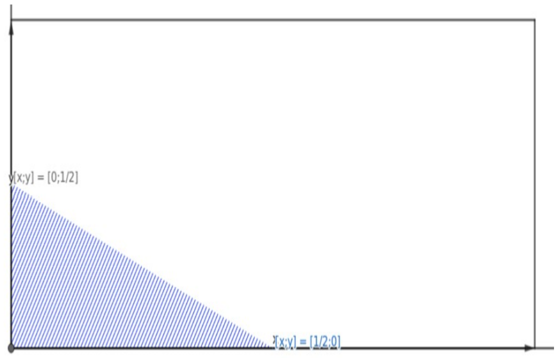
Soit Ω, espace des possibles finis, constitué de x et y compris entre 0 et 1 :
$\Omega=\{x,y\}\forall x,y\in[0,1|\mathbb{R}]$



Nous supposons que la probabilité d’un événement est égale à l’aire de cet événement dans Ω.

Cette loi est appelée la loi Uniforme continue.

Nous observons un événement intéressant, B, constitué de x, y tel que la somme de x et de y est inférieure à $\frac{1}{2}$:



Nous calculons l’aire de B : $P(B)=\left(\frac{1}{2}\right)^2\times(\frac{1}{2})=\frac{1}{8}$

__Univers infini, événements discrets__

Soit l’ensemble des possibles Ω tel que $\Omega=\mathbb{N}$.

Nous supposons que $P(n)=\frac{1}{2^n}\forall n\in\mathbb{N}$.

Nous observons l’événement B : « n est pair » : $B=\{2,4,6,\ldots\}$

Nous calculons la probabilité de B :

$$P(B)=P(\{2,4,6...\})=P(\{2\}\cup\{4\}\cup\{6\}....)$$

$$P(B)=\frac{1}{2^2}+\frac{1}{2^4}+\frac{1}{2^6}+....=\frac{1}{2^2}\left(1+\frac{1}{2^2}+\frac{1}{2^4}...\right)=\frac{1}{4}\left(\frac{1}{4^0}+\frac{1}{4^1}+\frac{1}{4^2}...\right)$$

Nous reconnaissons ici une suite géométrique :

$$P(B)=\frac{1}{4}\times\frac{1}{1-\frac{1}{4}}$$

$$P(B)=\frac{1}{3}$$

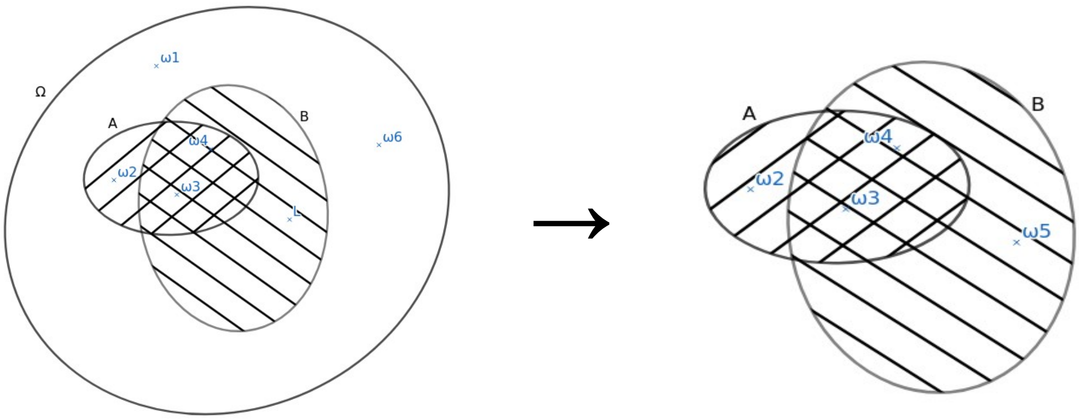
4.Condition et indépendance

4.a Formule des probabilités composées

Nous considérons les événements A,B de l’univers Ω .
Nous nous interrogeons sur un sous-ensemble de l’événement A: celui où l’on sait que B s’est réalisé.
$P(A|B)\Leftrightarrow$ Probabilité de A sachant B

P(A|B) est une **probabilité conditionnelle**: la probabilité qu’un événement se réalise à condition qu’un autre événement se soit déjà réalisé.

Dire que B s’est réalisé, c’est poser P(B) = 1. Nous révisons nos croyance, nous changeons d’univers.



Dans l’univers Ω, l’événement A a une probabilité de $\frac{3}{6}=\frac{1}{2}$ *de se réaliser contre* $\frac{2}{3}$ *dans l’univers B.*

Chercher la probabilité de A sachant B, c’est chercher $A\cap B$ dans l’univers B.

D’où $P(A|B)=\frac{P(A\cap B)}{P(B)}$ si $P(B)\neq 0$.

Diviser par zéro n’est pas tolérable pour les Dothrakis !

Nous en déduisons que : $P(A\cap B)=P(A|B)P(B)\,ou\,P(A\cap B)=P(B|A)P(A)$

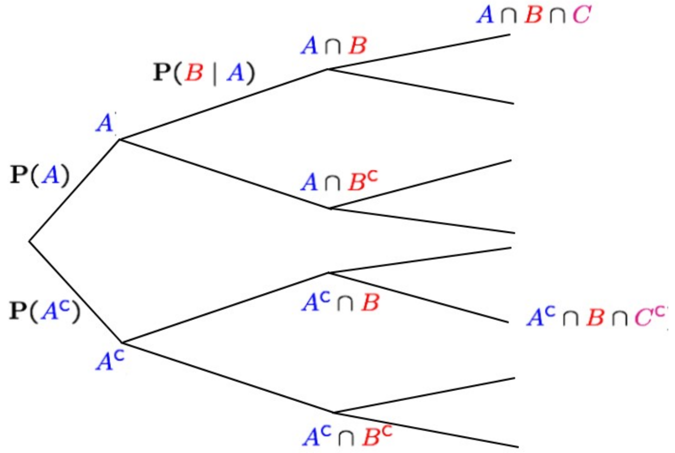
En effet, nous pouvons considérer $P(A\cap B)$ comme la probabilité de A rapportée à l’univers B ou comme la probabilité de B rapportée à l’univers A.

Les conditions peuvent être représentées sous forme d’arbre pondéré:

(Non, vous n’y trouverez pas la corneille à trois yeux)

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A \cap B) = P(B)\,P(A \mid B)$$
$$= P(A)\,P(B \mid A)$$

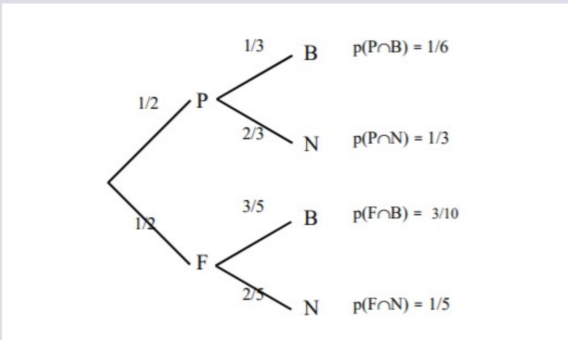


<u>*Méthode de calcul de l’arbre pondéré*</u>

- Règle 1 : La somme des probabilités des branches partant d'une même racine est toujours égale à 1
- Règle 2 : La probabilité d'un chemin est égale au produit des probabilités des branches de ce chemin.
- Règle 3 : La probabilité d'un événement est la somme des probabilités des chemins correspondant à cet événement.

Exemple **[Proba. conditionnelle]** : Soi une expérience aléatoire : « On jette une pièce ».
Si on obtient pile, on tire une boule dans l’urne P contenant 1 boule blanche et 2 boules noires.
Si on obtient face, on tire une boule dans l’urne F contenant 3 boules blanches et 2 boules noires.

On peut représenter cette expérience par l'arbre pondéré ci-dessous :



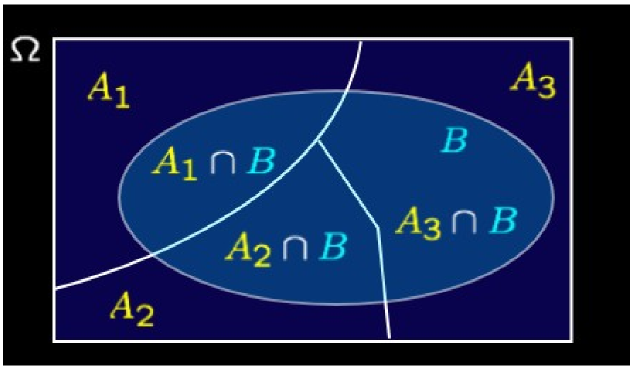
Et on a : P(B)=1/6 + 3/10 = 0.47

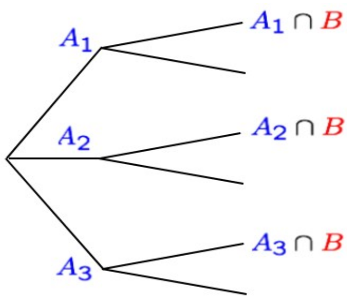
De manière plus générale, nous définissons ainsi la formule des probabilités composées :

Soient $A_{1,} A_2, \ldots, A_n$ évènements d'intersection non nulle
Alors :
$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)\,P(A_2 | A_1) \times \ldots \times P(A_n | A_1 \cap A_2 \cap \ldots \cap A_{n-1})$$

4.b Formules des probabilités totales

Supposons un univers Ω partitionné en A_i événements et incluant un événement B. $i.e: A_i \neq 0 \, \forall i ; \sum_i A_i = 1 ; A_i \cap A_j = 0 \, \forall i \neq j$, exemple en posant $i=3$:





Nous supposons connaître $P(A_i)$ et $P(B|A_i) \, \forall i$

La formule des probabilités totales nous permet de calculer P(B) à partir de ces informations :

$$P(B) = \sum_i P(A_i) P(B|A_i)$$

4.c Loi de Naïve Bayes

Supposons un univers Ω partitionné en A_i scénarios. Un événement B peut se réaliser dans chacun de ces scénarios :

$$i.e: A_i \neq 0 \, \forall i ; \sum_i A_i = 1 ; A_i \cap A_j = 0 \, \forall i \neq j$$

Nous supposons connaître $P(A_i)$ et $P(B|A_i) \, \forall i$,

$P(A_i)$ sont les probabilités initiales que nous accordons à chaque scénario A_i .

Nous apprenons que l'événement B se réalise. Cet apport d'informations nous pousse à réviser la probabilité qu'on accordait à chaque scenario.

Nous calculons ainsi la probabilité conditionnelle de A_i sachant B :

Nous savons que $P(A_i|B) = \frac{P(A_i \cap B)}{P(B)}$

Nous utilisons la formule des probabilités composées pour calculer le numérateur et la formule de probabilité totale pour calculer le dénominateur.

Nous obtenons la relation suivante, appelée Loi de Bayes :

$$P(A_i|B) = \frac{P(A_i) P(B|A_i)}{\sum_j P(A_j) P(B|A_j)}$$

4.c Exemple : un Test Médical

Énoncé:
Un médecin effectue le dépistage d'une maladie à l'aide d'un test fourni par un laboratoire.

Le test donne un résultat booléen : soit positif, soit négatif.

Lorsque le patient est porteur de la maladie, le test est positif dans 90 % des cas.

Lorsque le patient n'est pas atteint de la maladie, le test est positif dans 1% des cas.

Le médecin reçoit un résultat positif pour le test d'un patient.

Quelle est la probabilité que le patient soit réellement atteint de la maladie ?

Nous notons:

- M l'événement "le patient est atteint de la maladie"
  - $M^c = \Omega - M$ son complémentaire "le patient n'est pas atteint de la maladie"
- T l'événement "le test est positif"
- $P(T|M)$ = 0.9
- $P(T|M^c)$ = 0.01
- La grandeur recherchée est $P(M|T)$: la probabilité que le patient soit malade sachant que le test est positif.

Application du théorème de Bayes

Nous posons la probabilité conditionnelle recherchée : $P(M|T) = \frac{P(T \cap M)}{P(T)}$

Grâce à la formule des probabilités composées nous savons que : $P(T \cap M) = \left\{ \begin{array}{l} P(T|M)P(M) \\ P(M|T)P(T) \end{array} \right.$

Nous pouvons donc remplacer le numérateur de notre probabilité conditionnelle par l'expression $P(T|M)P(M)$.

Nous pouvons réécrire la probabilité conditionnelle : $P(M|T) = \frac{P(T|M)P(M)}{P(T)}$

Grâce à la formule de la probabilité totale on sait que :

$$P(T) = P(T|M)P(M) + P(T|M^c)P(M^c)$$

Nous réécrivons la probabilité conditionnelle : $P(M|T) = \frac{P(T|M)P(M)}{P(T|M)P(M) + P(T|M^c)P(M^c)}$

Application numérique

$$P(M|T) = \frac{0.9\,P(M)}{0.9\,P(M) + 0.01(P(\Omega) - P(M))} = \frac{0.9\,P(M)}{0.89\,P(M) + 0.01}$$

$P(M)$ est la probabilité globale que le patient soit malade i.e la proportion de malade dans la population à laquelle appartient le patient.

On pose $P(M) = 10^{-5}$: une personne parmi 100 000 est malade.

d'où:

$$P(M|T)=\frac{0.9\times10^{-5}}{0.89\times10^{-5}+0.01}=0.000899$$

La probabilité que le patient soit malade sachant que le test est positif est donc de 0.0899%.

Un taux très faible de Vrai Positif, c'est pour cela que nous faisons au moins deux test pour le test du sida.

<u>*Bayes au-delà des formules*</u>

Ce qu'il faut retenir de Bayes, c'est qu'il s'agit d'une technique mathématique permettant de réviser les probabilités que l'on accorde à un phénomène lors de nouvelles observations.

Exemple : La probabilité que l'on accorde à reconnaître l'image d'un chat .

Cette propriété est très utile en Machine Learning (apprentissage automatisé) et en Deep Learning (apprentissage profond).

5.Indépendance

Définition :On dit que deux événements A et B *sont indépendants si la réalisation de l'un ne nous apporte aucune information sur la réalisation de l'autre.*

Intuitivement nous pouvons l'écrire : $P(A|B)=P(A)$

Nous préférerons la notation suivante : $P(A\cap B)=P(A)P(B)$

Cette dernière possède trois avantages : elle est symétrique que l'on considère A ou B , elle implique la définition intuitive $P(A|B)=P(A)$ et elle reste valable même si $P(A)=0$

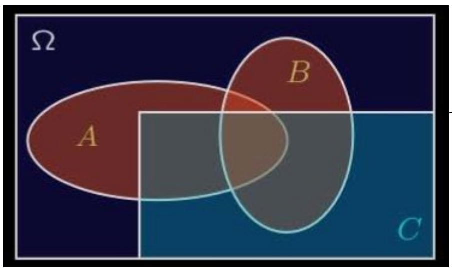
Rappelez-vous que les Dothrakis ne divisent PAS par zéro.

5.1 Indépendance et condition

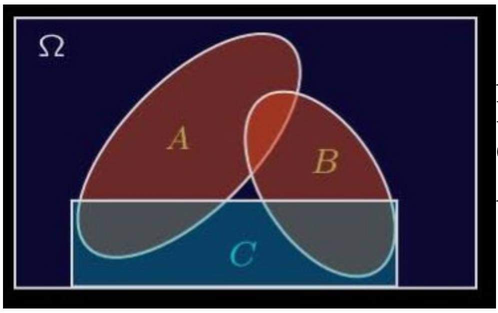
Le conditionnement de deux événements, indépendants entre eux, à un troisième événement peut modifier l'indépendance des deux premiers.

A et B independants sachant C si, pour P(C)>0 on a : $P(A\cap B|C)=P(A|C)P(B|C)$

De plus, si $P(B\cap C)>0$ alors on peut egalement écrire : $P(A|B\cap C)=P(A|C)$



A et B indépendants qu'on ce situe en Ω ou C



A et B indépendants en Ω
En C : A et B disjoint ($A \cap B = \emptyset$)
Ce qui implique que, sachant C, $P(A \cap B) = 0$
A et B non indépendant en C

Indépendance d'une collection d'événements

Nous généralisons la précédente définition d'indépendance.

Les événements $A_i, A_j, \ldots, A_n$ sont dit indépendants si

$$P(A_i \cap A_j \cap \ldots \cap A_n) = P(A_i) P(A_j) \ldots P(A_n) \forall i, j, n \, distincts$$

Indépendance paire à paire

Nous posons $n = 3$

$$\left\{ \begin{array}{l} P(A_1 \cap A_2) = P(A_1) P(A_2) \\ P(A_1 \cap A_3) = P(A_1) P(A_3) \\ P(A_3 \cap A_2) = P(A_3) P(A_2) \end{array} \right\} \text{Indépendance paire à paire de } A_1, A_2, A_3$$

$$\{ P(A_1 \cap A_2 \cap A_3) = P(A_1) P(A_2) P(A_3) \} \text{ indépendance de } A_{1,} A_{2,} A_3$$

L'indépendance paire à paire n'implique pas l' indépendance globale :

***Exemple** [Indépendance]* **:** Soit Ω, univers des possibles décrivant l'expérience consistant en deux pile-ou-face indépendants, avec une pièce équilibrée.

$$\Omega = \{ \{ F_1 F_2 \}, \{ F_1 P_2 \}, \{ P_1 P_2 \}, \{ P_1 F_2 \} \}$$

Soit $P(F_1) = P(\{FF\}, \{FP\}) = \frac{2}{4} = \frac{1}{2}$ probabilite que le premier pile-ou-face soit face

et $P(F_2) = P(\{FF\}, \{PF\}) = \frac{2}{4} = \frac{1}{2}$ probabilite que le second pile-ou-face soit face

et

$P(C) = P(\{HH\}, \{FF\}) = \frac{2}{4} = \frac{1}{2}$ probabilite que les deux pile-ou-face aient le meme resultat

Nous vérifions les indépendances paire à paire :

$$\left\{ \begin{array}{ll} P(F_1 \cap F_2) = \frac{1}{4} = P(F_1) P(F_2) & \text{Independance paire a paire de } F_1 \text{ et } F_2 \\ P(F_1 \cap C) = \frac{1}{4} = P(F_1) P(C) & \text{Independance paire a paire de } F_1 \text{ et } C \\ P(C \cap F_2) = \frac{1}{4} = P(C) P(F_2) & \text{Independance paire a paire de } F_2 \text{ et } C \end{array} \right\}$$

$$P(F_1 \cap F_2 \cap C) = P(FF) = \frac{1}{4} \neq P(F_1)P(F_2)P(C) = \frac{1}{8}$$

F_1, F_2, C indépendants paire à paire mais pas *globalement* indépendant

Nous savons à présent définir et mesurer une expérience aléatoire.

Si nous répétons n fois cette expérience aléatoire, comment modéliser la fréquence d'apparition d'un évènement ?

Les variables aléatoires

1. Kesako ?

La variable aléatoire nous permet de caractériser un événement aléatoire. C'est un outil qui modélise théoriquement des phénomènes aléatoires grâce à différentes mesures (telles que la probabilité (P), l'espérance (E), la variance (Var)).

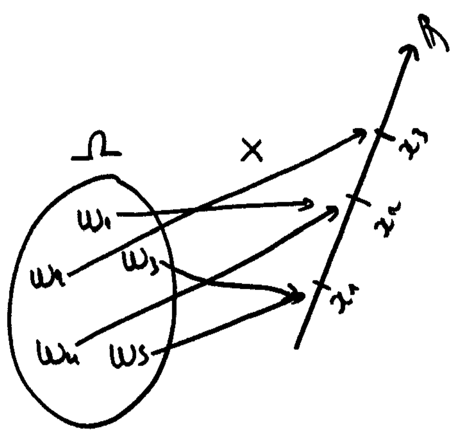
Nous savons qu'une expérience aléatoire donne lieu à un ensemble d'issues possibles ($\Omega = \{\omega_i\}$)
La variable aléatoire représente le résultat d'une expérience aléatoire. Elle peut prendre tous les résultats possibles considérés, associés à l'expérience aléatoire.

Définition : On appelle variable aléatoire une <u>*application*</u> *de l'ensemble des événements élémentaires (* Ω *) sur* $\mathbb{R}$ *:* $X : \Omega \to \mathbb{R}$.

Soit X cette application, l'ensemble des images des événements élémentaires par X constitue l'ensemble des valeurs prises par X : x_i .

A chaque événement élémentaire ω_i de Ω correspond un nombre réel x_i associé à la variable X. Le nombre de valeurs prises par X , n'est pas nécessairement égale au nombre d'événements élémentaires de Ω. L'application X est donc surjective.

Nous pouvons représenter l'application X grâce au schéma suivant :



Le support, noté $X(\Omega)$, est l'ensemble de toutes les valeurs prises par la variable aléatoire .
$X(\Omega)$ = L'image des événements élémentaires.
$X(\Omega)=\{X(\omega); \omega\in\Omega\}$.

Il ne faut pas confondre Ω qui est l'univers et $X(\Omega)$ qui est le support, un sous-ensemble de $\mathbb{R}$ et qui est composé de l'ensemble des valeurs prises par X dans $\mathbb{R}$.

On désignera la variable aléatoire par une majuscule et les valeurs prises par les applications dans $\mathbb{R}$ par des minuscules

Une variable aléatoire peut prendre un nombre élevé de valeurs possibles. Chaque valeur est prise avec une certaine probabilité.

A la différence d'une variable algébrique qui est une inconnue que l'on peut faire varier pour définir d'autres inconnues au sein d'un système d'équations.

Il existe deux types de variables aléatoires : les variables aléatoires qualitatives et quantitatives.
Une **variable aléatoire qualitative** qualifie ses issues possibles.
Soit E_1 , l'expérience aléatoire suivante : « Prendre au hasard un élève dans une classe de 30 personnes ». La variable aléatoire X= « Couleur des yeux » peut prendre les valeurs suivantes :
$X(\Omega)=\{\text{Bleu, Vert, Marron}\}$.

Pour une **variable aléatoire quantitative**, on associe à chaque résultat de l'expérience une valeur numérique.
Soit E_2 , l'expérience aléatoire suivante : « Lancer 7 fois un dés non pipé » :
La variable aléatoire Y = « Somme des 7 faces obtenues » est une variables aléatoire qui associe un nombre réel aux événements possibles de cette expérience.

Nous pouvons aussi encoder une variable qualitative afin de forcer son typage en variable quantitative.
Soit E_3 , l'expérience aléatoire suivante : « Lancer une pièce de monnaie non truquée »

Soit $Z=\begin{Bmatrix}\text{Pile}\\ \text{Face}\end{Bmatrix}$ revient à définir la variable aléatoire comme ceci $Z=\begin{Bmatrix}0 \text{ si Pile}\\ 1 \text{ si Face}\end{Bmatrix}$

Il existe deux types de variables aléatoires quantitatives :

- Les **variables aléatoires quantitatives discrètes :** X prend des valeurs distinctes ou séparées ($X\in\mathbb{N}$).
  Ces variables peuvent être **finies** :
  X = « Année de naissance d'un élève pris au hasard dans une classe de 30 élèves ».
  Elles peuvent également prendre une **infinité de valeurs** :
  Y = « Nombre de naissances avant d'avoir une fille ».

- Les **variables aléatoires quantitatives continues** : Y prend n'importe quelle valeur dans une intervalle ($Y\in\mathbb{R}$).
  Y = « Masse *exacte* d'un élève pris au hasard dans une classe »
  Dans ce cas il est impossible de lister toutes les masses possibles que peut avoir un élève.

Le type des variables aléatoires détermine la manière dont nous allons traiter et étudier ces variables. Les variables aléatoires quantitatives discrètes finies, infinies ou les variables aléatoires quantitatives continues ont chacune leur propres formules de calcul, lois et distributions. La distribution associée donne les probabilités que la variable réalise chacune des valeurs possibles.

2. Comment peut-on décrire une variable aléatoire discrète ?

Soit l'expérience aléatoire suivante : On lance trois fois une pièce équilibrée, on note X « le nombre de Piles obtenus ».
Le fait de répéter p fois l'expérience « Lancer une pièce de monnaie équilibrée » nous fait passer d'un élément élémentaire ω_i à un ensemble d'éléments appelé des p -listes de l'ensemble P .

Soit l'univers Ω, l'ensemble des issues possibles : $\{P,F\}^3=\{\omega_i\},1\leq\omega_i\leq 8$
C'est l'ensemble de tous les événements élémentaires ω_1, ω_2, …. ω_8 suivants :

ω_1= {PPP}	ω_4= {PFF}	ω_7= {FFP}
ω_2={PPF}	ω_5= {FPP}	ω_8= {FFF}
ω_3= {PFP}	ω_6= {FPF}	

<u>**Rappel sur la cardinalité de l'ensemble Ω**</u> : Ces solutions peuvent être trouvées avec un raisonnement de dénombrement :

- 2 choix pour le premier tirage ('Pile' ou 'Face'), 2 choix pour le suivant, 2 choix pour le dernier : Card(Ω)=2x2x2=8 solutions possibles.
- Nous pouvons voir ces solutions comme une 3-liste de l'ensemble $\{P,F\}$. En effet, pour un ensemble qui comporte n éléments, le nombre de p -listes est égale n^p , ici Card (Ω)= 2^3=8

La variable aléatoire X associe à chacun de ces lancers, le nombre de 'Piles' correspondants. Ce nombre est aléatoire car il dépend du résultat de l'expérience.

	$\xrightarrow{X}$
ω_1= {PPP}	$X(\omega_1)=3$
ω_2= {PPF}	$X(\omega_2)=2$
ω_3= {PFP}	$X(\omega_3)=2$
ω_4= {PFF}	$X(\omega_4)=1$
ω_5= {FPP}	$X(\omega_5)=2$
ω_6= {FPF}	$X(\omega_6)=1$
ω_7= {FFP}	$X(\omega_7)=1$
ω_8= {FFF}	$X(\omega_8)=0$

$$X(\Omega)=\{0,1,2,3\}$$

A chacune des valeurs prises par X va s'associer une probabilité. C'est la probabilité de l'image réciproque par X.
Par exemple : Quelle est la probabilité pour que X (le nombre de 'Pile') soit égale à 2 ?
L'union des événements élémentaires ω_2 , ω_3 , ω_5 ,défini par l'ensemble $\{\omega_2\cup\omega_3\cup\omega_5\}$ correspond au résultat X=2 de l'expérience.

Dans cet exemple , nous sommes dans un cas d'équiprobabilité et chacun des événements élémentaires ω_i de Ω a la même probabilité de se réaliser.
Nous pouvons ainsi définir les probabilités suivantes :

$$P(\omega_1)=P(X=3)=\frac{1}{8}$$

$$P(\omega_2\cup\omega_3\cup\omega_5)=P(X=2)=\frac{1}{8}+\frac{1}{8}+\frac{1}{8}=\frac{3}{8}$$

$$P(\omega_4\cup\omega_6\cup\omega_7)=P(X=1)=\frac{1}{8}+\frac{1}{8}+\frac{1}{8}=\frac{3}{8}$$

$$P(\omega_8)=P(X=0)=\frac{1}{8}$$

On remarque que :

$$P(X=x_i)=p_i\leq 1$$

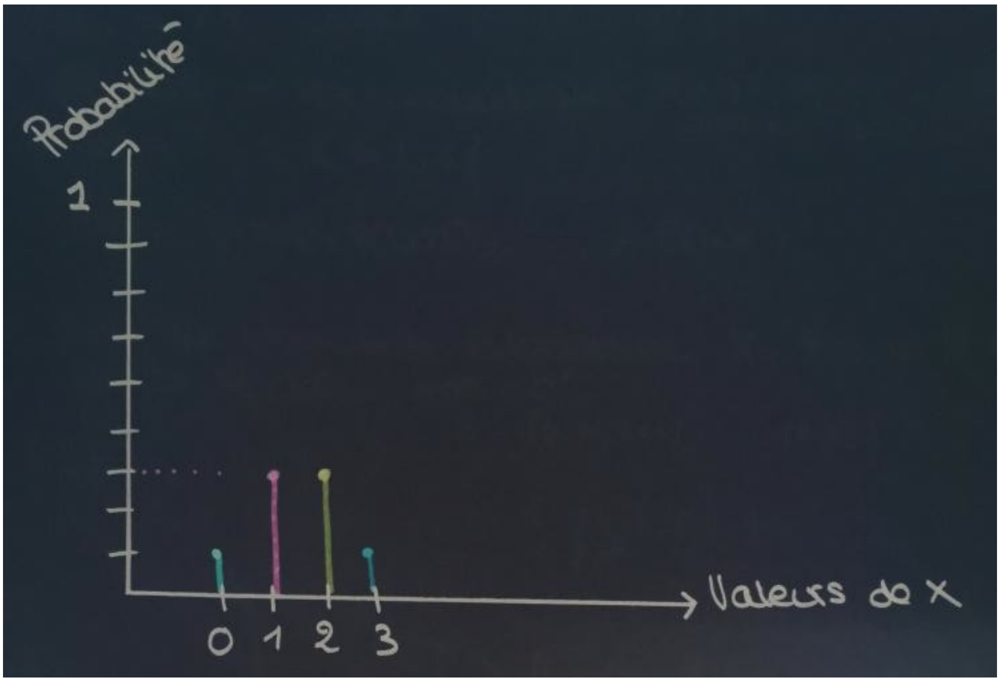
$$\sum_i P(X=x_i)=\sum_i p_i=1$$

L'ensemble des couples (x_i, p_i) forme la **loi de probabilité de la variable aléatoire X** :

- $X(\Omega)=\ldots.$
- $\forall\, x_k\in X(\Omega), P(X=X_k)=\ldots$

La loi de probabilité est le premier outil dont nous disposons pour décrire notre variable aléatoire.

Nous pouvons représenter cette distribution par une fonction de masse grâce à un diagramme en bâton :



La variable étant une variable discrète, nous ne pouvons pas relier les points entre eux.

Nous allons construire des mesures nous permettant de caractériser notre distribution :

L'Espérance mathématique (ou la valeur moyenne de notre variable aléatoire)
On appelle Espérance mathématique la quantité $E(x)$ définie par :

$$E(x)=\sum_{i}^{n}(x_i p_i), \forall x_i \in X(\Omega) \text{ et tel que } 1 \leq i \leq n$$

Elle désigne la valeur prise par X à laquelle on peut s'attendre en général au cours de l'expérience (sans être ni optimiste ni pessimiste).

Nous pouvons aussi définir l'espérance mathématique comme un barycentre, qui n'est rien d'autre qu'une moyenne pondérée d'un certains nombres de points, et dont la somme des pondérations fait 1.

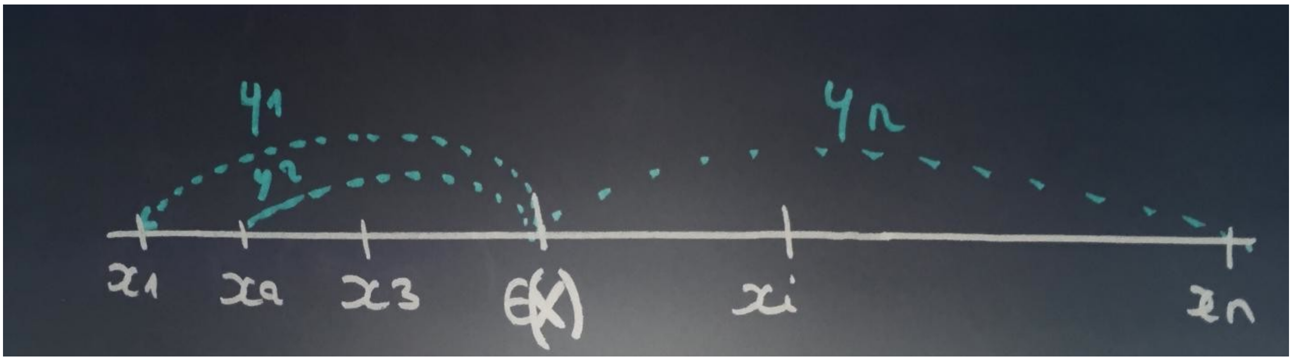
Il existe un cas pour lequel l'espérance mathématique ne peut pas se calculer : c'est le fameux « paradoxe de Saint-Petersbourg ».

- <u>**Propriétés de l'Espérance**</u>
- Si b constante alors $E[b]=b$
- Monotonie :Si X, Y v.a tel que $X <= Y$ alors $E[X] \leq E[Y]$
- Linéarité : pour deux v.a $X, Y \in \Omega$ et pour deux nombre $a, b \in \mathbb{R}$ alors
  $E[aX+bY]=aE[X]+bE[Y]$
- Si X, Y **v.a indépendantes** alors $E[XY]=E[X]E[Y]$
  Si X, Y **v.a non indépendantes** alors, en général $E[XY] \neq E[X]E[Y]$

La Variance et l'Écart-type (la dispersion) de notre variable aléatoire autour de sa moyenne.
Plus les grandes valeurs de la variance sont nombreuse, plus les valeurs prises par notre variable aléatoire s'étalent de part et d'autre de la moyenne.

Nous allons construire notre indicateur de dispersion.

Pour commencer nous définissons Y , variable aléatoire tel que $y_i = x_i - E(X)$.
Chacun des y_i est la mesure de la distance entre x_i et $E(X)$: Y est la variable aléatoire qui renseigne sur la dispersion des valeurs de la variable X autour de sa moyenne (ou les écarts). Nous pouvons l'illustrer de la façon suivante :



Nous souhaitons agréger nos y_i pour construire notre indicateur de dispersion.

Notre outil d'agrégation est l'espérance : $E(Y)=\sum_{i}^{n} y_i p_i$.

Cependant, en observant notre schéma, un problème se pose :

si $x_i > E(X)$ alors $y_i > 0$

si $x_i < E(X)$ alors $y_i < 0$

C’est à dire que lorsque nous allons sommer nos y_i nous aurons des valeurs positives et négatives: les mesures de distance d’ un x_i « à gauche » de son espérance annuleront les mesures de distance d’un x_i situé « à droite » de l’espérance.

Nous devons donc traiter Y de manière à supprimer cette information.
Par exemple, nous pourrions redéfinir y_i tel que $y_i = |x_i - E(X)|$.
Cependant, les valeurs absolues ne sont pas des outils facilement manipulables.

Nous conservons notre définition : Y tel que $y_i = x_i - E(X)$, mesure de la distance entre x_i et $E(X)$.
Nous posons alors Y^2 tel que $y_i^2 = (x_i - E(X))^2$ carré de la mesure de la distance entre x_i et $E(X)$.

Les carrés étant strictement positifs, nous pouvons les agréger en utilisant l’espérance, sans perte d’informations.

$E(Y^2)$ est donc l’espérance de la distance au carré entre x_i et $E(X)$

$\sqrt{E(Y^2)}$ est donc l’espérance de la distance entre x_i et $E(X)$: notre indicateur de dispersion initialement recherché.

Nous posons notre mesure de dispersion : $y_i = x_i - E(X)$
Nous élevons cette mesure au carré : $y_i^2 = [x_i - E(X)]^2$
Nous développons l’expression obtenue : $y_i^2 = x_i^2 - 2x_i E(X) + E(X)^2$
Nous calculons l’espérance de cette expression :

$$E[Y^2] = \sum_i^n p_i (x_i^2 - 2x_i E(X) + E(X)^2)$$

$$rappel : \sum (a x_i + b y_i) = a \sum x_i + b \sum y_i$$
$$rappel : \sum k x_i = k \sum x_i \,\forall\, k\ constant$$

$$E[Y^2] = \sum_i^n p_i (x_i^2) - 2E[X] \sum_i^n p_i x_i + E[X]^2 \sum_i^n p_i$$

$$rappel : \sum_i^n p_i = 1 \quad \text{et} \quad \sum_i^n x_i p_i = E[X]$$

$$E[Y^2] = \sum_i^n p_i (x_i^2) - 2E[X]^2 + E[X]^2$$

$$E[Y^2] = E[X^2] - E[X]^2$$

$E(Y^2)$ se nomme la variance et $\sqrt{E(Y^2)}$ se nomme l'écart quadratique moyen ou écart-type.
La variance de X est notée : $Var(X)$ et son écart-type : $\sigma(X)$.

Nous avons donc :

$$Var(X)=E(X^2)-E(X)^2$$
$$\sigma(X)=\sigma_X=\sqrt{E(X^2)-E(X)^2}$$

Propriétés de la Variance :

- $Var(X)\geq 0$
- $Var(kX)\geq k^2 Var(X)\forall k\, constant$
- $Var(k)=0\forall k\, constant$
- Pour X,Y **v.a indépendantes** $Var(X+Y)=Var(X)+Var(Y)$

Démonstration

$$\begin{aligned} V(X+Y) &= E((X+Y)^2)-(E(X+Y))^2 \\ &= E(X^2+2XY+Y^2)-(E(X)+E(Y))^2 \\ &= E(X^2)-(E(X))^2+E(Y^2)-(E(Y))^2-2[E(XY)-E(X)E(Y)] \\ &= V(X)+V(Y) \end{aligned}$$

car puisque X et Y sont indépendantes $E(XY)=E(X)E(Y)$.

Pour X,Y v.a $Var(X+Y)=Var(X)+Var(Y)+2Cov(X,Y)$

En reprenant l'expérience étudiée précédemment, nous pouvons construire le tableau suivant qui définit la **loi uniforme discrète** de notre variable aléatoire X :

Nous vous invitons à vous renseigner sur les Moments des variables aléatoires discrètes pour découvrir l'espérance et la variance en tant qu'application linéaire.

En reprenant l'expérience étudiée précédemment, nous pouvons construire le tableau suivant qui définit la **loi uniforme** de notre variable aléatoire X :

Mesures	x_k	0	1	2	3	TOTAL
Probabilité	$P(X=x_k)$	1/8	3/8	3/8	1/8	1
Espérance	$x_k \times P(X=x_k)$	0	3/8	6/8	3/8	3/2
Variance	$x_k^2 \times P(X=x_k)$ $-(x_k \times P(X=x_k))^2$	0	3/8-9/64	12/8-36/64	9/8-9/64	3/4
Écart-type						√3/2

Les valeurs de X s'écartent en moyenne de l'espérance d'environ 0.9 'Piles'.

La loi uniforme n'est rien d'autre que la loi propre à notre variable X.

Nos ancêtres mathématiciens nous ont facilité la tâche en créant des **lois usuelles de probabilités** applicables à n'importe quel phénomène se rapprochant de l'expérience relative à la loi utilisée.
Ainsi nous pouvons utiliser directement les formules de calcul et les représentations associées à ces lois.

2.a. Lois usuelles des variables aléatoires discrètes finies

Épreuve de Bernoulli

Notation
$\mathcal{B}(p)$ avec p le paramètre du succès

Ensemble des valeurs
$X(\Omega)= [0,1]$

Loi

- $P(X=1)$ = probabilité du succès = p
- $P(X=0) = 1-p$
- $E(X) = p$
- $E(X^2) = p$
- $Var(X) = p(1-p)$

Utilisation

- Réponse Oui/Non d'un sondage

$$X = \begin{cases} 0 \text{ si Non} \\ 1 \text{ si Oui} \end{cases} \text{, p inconnu à priori : } X \sim \mathcal{B}(p)$$

- Succès dans un jeu binaire :

$$X = \begin{cases} 1 \text{ si Pile} \\ 0 \text{ si Face} \end{cases} \text{, } p=P(X=1) = 1/2 \text{ : } X \sim \mathcal{B}(1/2)$$

Loi Binomiale

Notation
Bin(n,p) pour n ε ℕ*, et p ε [0,1]

Ensemble des valeurs
X(Ω)= {0,….,n]

Loi

- P (X=k) = $\binom{n}{k}$. p^k . $(1-p)^{n-k}$, avec 0<k<n
- E(X) = np
- Var(X) = np(1-p)

Théorème
Si X et Y sont deux variables aléatoires suivants des lois Binomiales :
Soient X~ *Bin*(n,p) et Y~ *Bin*(m,p) alors X+Y~ *Bin*(n+m,p)

Utilisation

La loi Binomiale est applicable dans les cas d'événements indépendants

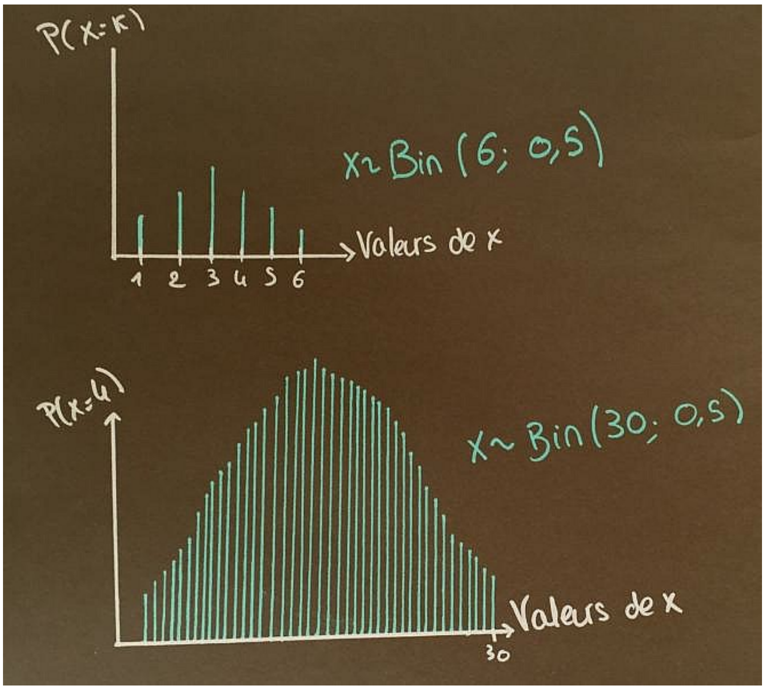
- **Nombre de succès** dans une épreuve de Bernoulli répétée n fois de façon **indépendante**

> **Exemple** **[Bin]** : Nous lançons 9 fois un dés , Soit X= nombre de 3 obtenus
> X ~ *Bin*(9, $\frac{1}{6}$)
> P(X=2)= 0.28

- Comptage d'un caractère dans un **tirage avec remise**

> **Exemple** **[Bin]** : Nous avons un lot de 1000 pantalons dont 10 % défectueux.
> Nous tirons 15 pantalons avec remise.
> Soit X le nombre de pantalons défectueux.
> X ~ *Bin*(15, $\frac{1}{100}$)

Représentation de la fonction de masse en histogramme



2.b Lois usuelles des variables aléatoires discrètes infinies

Loi géométrique

Notation
G(p) avec p le paramètre du succès

Ensemble des valeurs
X(Ω)= {0,….,n] ε ℕ

Loi

- P (X=k) = $p(1-p)^k$, avec 0<k<n
- E(X) = $\frac{1-p}{p}$
- Var(X) = $\frac{1-p}{p^2}$

Utilisation

- **Nombre d’échecs avant d’avoir 1 succès.** Elle décrit une situation de répétition d’épreuves de Bernouilli indépendantes à observer avant d’avoir un succès.

Loi Binomiale négative

Notation

Neg*Bin*(n,p) avec p le paramètre du succès et n le nombre de succès

Ensemble des valeurs

X(Ω)= {1….,n] ε ℕ*

Loi

- $P(X=k) = \binom{n-1}{k-1} p^n (1-p)^k$
- E(X) = n(1-p)/p
- Var(X) =n(1-p)/p^2

Utilisation

- **Nombre d'échecs avant d'avoir n succès.** Elle décrit une situation de répétition d'épreuves de Bernouilli indépendantes à observer avant d'avoir n succès.
  Nous pouvons utiliser cette loi lorsque l'on souhaite définir le temps d'attente avant d'avoir n succès suite à X=k épreuves indépendantes.

> **Exemple [NegBin]** :
> Nous lançons une pièce de monnaie (truquée) dont la probabilité d'obtenir pile est p=0.3.
> Nous notons X le nombre d'échecs précédant le 4ème pile.
> Alors X ~ Neg*Bin*(4, 0.3).

Nous pouvons donc définir la loi géométrique comme une loi de Bernoulli négative avec n=1.

Loi de Poisson

Notation

P(λ) pour λ ε ℝ⁺, représentant le nombre moyen de réalisations.

Ensemble des valeurs

X(Ω)= {0,….,n] ε ℕ

Loi

- $P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}$
- E(X) = λ
- Var(X) = λ

Théorème

Soient X~ P(λ) et Y~ P(λ') alors (X+Y)~P(λ+λ').

Utilisation

- Nombre de clients entrant dans un magasin entre 14h et 18h
- Nombre de bus passant à 1 arrêt de bus en 35 minutes
- Nombre de requêtes sur un serveur de minuit à 6 heures

Approximation

En pratique lorsque l'on a un « grand nombre » d'événements qui suivent une loi binomiale et qu'on connaît la moyenne λ, on peut utiliser une loi de Poisson.

Lorsque $n \to +\infty$: $\lambda = np$.
L'approximation de la loi Binominale B(n,p) par une loi de Poisson P(λ) avec λ=np n'est valide que lorsque les événements sont rares et suffisamment nombreux, c'est-à-dire :

- $p \leq 0.1$
- $np < 10$
- $n > 50$

3. Quand est-il pour une variable aléatoire continue ?

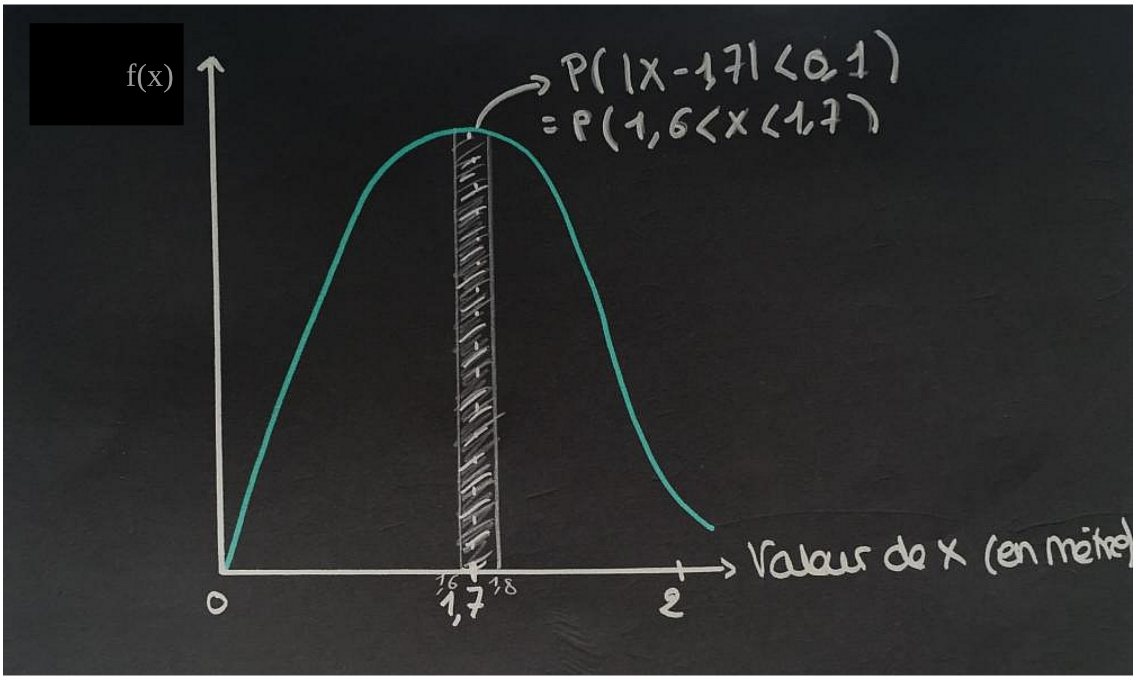
En pratique, nous cherchons souvent des modèles correspondant à des résultats continus.
Une variable aléatoire continue ne peut pas prendre une seule valeur exacte mais se positionne plutôt sur un intervalle de valeurs. Par exemple les variables faisant référence à l'age, la durée sont des variables aléatoires continues.
Soit X une variable aléatoire continue : $P(X=x_i)= p_i = 0$ alors que $P(X \in [a,b]) \neq 0$.
Dans ces conditions il est impossible de définir la loi de probabilité pour l'ensemble des couples (x_i, p_i). On introduit alors la notion de **fonction de répartition** de la variable aléatoire X :

$$F(x)= P(X\leq x)$$

Supposons une variable aléatoire continue dont les valeurs varient de -∞ à +∞. Pour toutes les valeurs de X la fonction F(x) possède une dérivée f(x) telle que :

$$F(x_i)=P(X\leq x_i) = \int_{-\infty}^{xi} f(x)\,dx$$

Nous pouvons représenter les variables aléatoires continues par la représentation graphique d'une fonction appelée **fonction de densité.**
Soit X la variable aléatoire continue : « Taille des élèves de la promotion Capgemini-Simplon ».
Nous pouvons tracer la courbe suivante :



Nous comprenons que l'aire sous la courbe de densité n'est rien d'autre que notre fonction de répartition, soit la probabilité que la valeur de X soit comprise dans l'intervalle souhaité.

Nous comprenons également que calculer P(X=a) revient à calculer l'aire d'une ligne :

$$P(X=a) = 0$$

Les axiomes de probabilités nous rappelle que l'aire totale sous la courbe est égale à 1 :

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1$$

La fonction de densité respecte les conditions suivantes :

- f est continue
- $\int_{-\infty}^{+\infty} f(x)\,dx = 1$
- $f(x) \geq 0$

2.a. Lois usuelle Uniforme continue

La loi Uniforme des variables aléatoires continues est applicable lorsque l'on est dans un tirage borné I = [a,b] et que l'on est dans un cas d'équiprobabilité.

Notation
U(a,b)

Ensemble des valeurs
X(Ω)= [a,….,b] $\in$ $\mathbb{N}$

Nous avons notre fonction de répartition F(x)= $\int_a^b f(x)\,dx$

Loi

- f(x) = $\begin{cases} \dfrac{1}{b-a} & \text{quand } a \le x \le b \\ 0 \text{ sinon} & \end{cases}$
- F(x) = $\begin{cases} \dfrac{x-a}{b-a} & \text{quand } a \le x \le b \\ 0 \text{ sinon} & \end{cases}$
- E(X) = $\dfrac{a+b}{2}$
- Var(X)= $\dfrac{(b-a)^2}{12}$

Utilisation

- Soit X = « tirage d'un nombre aléatoire »
  X~ U(0,100)

Petit aparté Geek : Loi applicable à la fonction random dans Python

Propriété
- Symétrique par rapport à E(X)

2.b. Lois usuelle Exponentielle

<u>Notation</u>

E (λ) avec $\lambda > 0$

<u>Ensemble des valeurs</u>

X(Ω)= [0…., +∞] ε ℕ

<u>Loi</u>

- $f(x|\lambda) = \begin{cases} \lambda\, e^{-\lambda x} \text{ si } x \geq 0 \\ 0 \text{ sinon} \end{cases}$

- $F(x) = \begin{cases} 1\text{-}e^{-\lambda x} \text{ si } x \geq 0 \\ 0 \text{ sinon} \end{cases}$

- $E(X) = \dfrac{1}{\lambda}$

- $Var(X)= \dfrac{1}{\lambda^2}$

<u>Utilisation</u>

- Décroissance d'une quantité (au cours du temps)
- Fiabilité des systèmes : Durée de vie de composants électroniques par exemple
- Tremblement de terre
- Désintégration d'un noyau radioactif

<u>Propriétés</u>

- Soit X une variable aléatoire qui suit une loi exponentielle de paramètre 0,1.
  $P(1\leq X \leq 3) = P(X \leq 3) - P(X <1) = 1- e^{-0,1\times3} - (1^{- e^{-0,1\times1}}) = e^{-0,1} - e^{-0,3} \approx 0,164$

- Durée de vie sans vieillissement : Soit X une variable aléatoire qui suit une loi exponentielle de paramètre λ .
  Alors, pour tout réel t et h positifs, on a : $P(X \geq t + h \mid X \geq t) = P(X \geq h)$.

  La durée de vie, d'un composant électrique par exemple, sur une période h ne dépend pas de l'âge t à partir duquel on considère cet événement. La probabilité que le phénomène dure au moins t+h heures (par exemple) sachant qu'il a déjà duré t heures sera la même que la probabilité de durer h heures à partir de sa mise en fonction initiale.

2.c. Lois usuelle Normale

La distribution normale est plus souvent utilisée lors des analyses. C'est la fameuse courbe en cloche.

Notation

N(μ,σ) avec μ la moyenne de l'échantillon et σ son écart-type

Ensemble des valeurs

X(Ω)= [-∞, +∞] ∈ ℝ

Loi

- $f(x|\mu,\sigma) = \begin{cases} \frac{1}{\sqrt{2\Pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} & \text{si } x \geq 0 \\ 0 \text{ sinon} & \end{cases}$
- $E(X) = \mu$
- $Var(X) = \sigma^2$

L'Espérance nous renseigne sur le point où la cloche est centrée et là où elle est à son maximum et l'écart-type indique la largeur de la cloche.

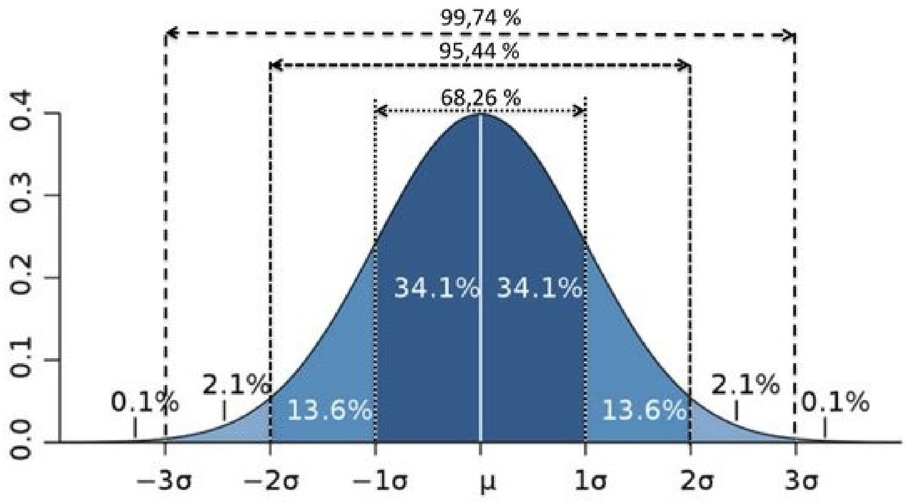
Utilisation

- Lorsqu'une expérience aléatoire est répétée suffisamment de fois, la variable aléatoire (qu'elle soit discrète ou continue) suit une loi Normale de paramètre (μ, σ).

Propriétés

- Symétrique par rapport à μ
- Soient X_1, X_2 deux variables aléatoires indépendantes qui suivent une loi normale : $X_1+X_2 \sim N(\mu_1+\mu_2, \sigma_1+\sigma_2)$
- Lorsque μ=0 et σ=1 on parle de loi Normale centrée réduite. Soit X une variable aléatoire. $X \sim N(\mu, \sigma)$. Alors $U = \frac{X-\mu}{\sigma} \sim N(0,1)$.

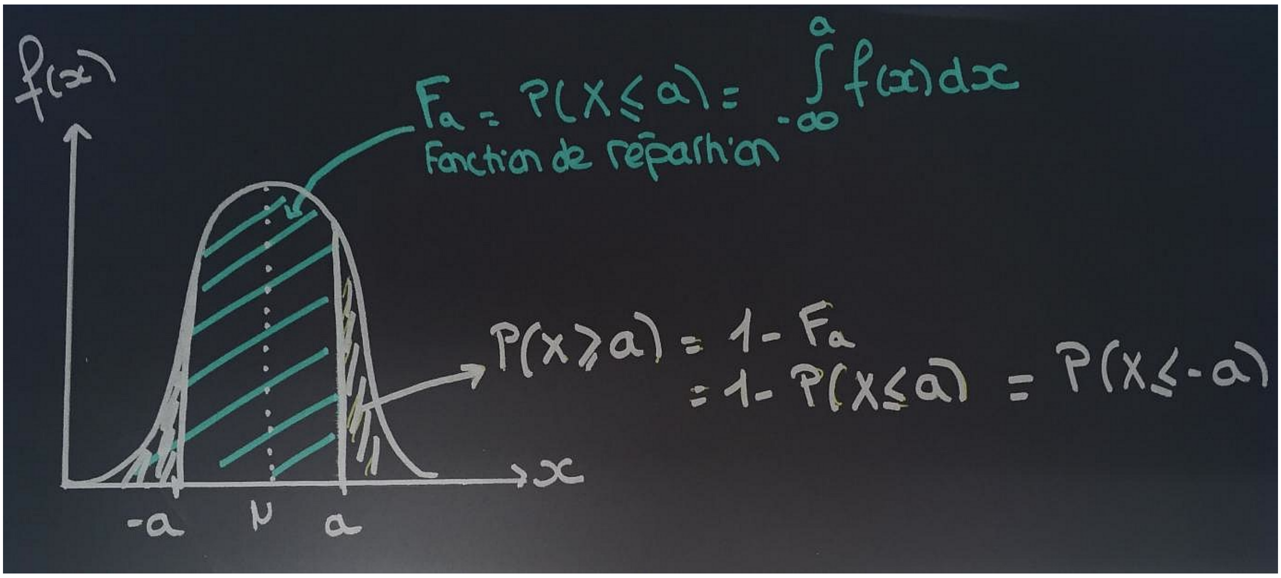
<u>Représentation graphique de la fonction de densité de la loi normale</u>



L’axe des abscisses est formé par les valeurs x_i prises par notre variable aléatoire X.
L’axe des ordonnées est formé par les valeurs prises par la fonction de densité f.

Nous comprenons que 68 % des informations sont comprises entre -σ et + σ. Et 99 % informations se trouvent entre -3σ et +3σ .

<u>Autre représentation graphique de la fonction de densité de la loi normale</u>



Les variables ayant des valeurs extrêmes sont appelés des outliers. Les outliers sont définis par l’ensemble des valeurs { $\mu \pm 2 * \sigma$ }.

La Loi normale est un des plus grands outils couramment utilisé en statistiques et probabilité.
Il a été prouvé que lorsqu’une expérience a été réalisée un certain nombre de fois alors on peut approximer la distribution de la variable aléatoire étudiée par une loi Normale.

Nous vous invitons à vous renseigner sur la loi des grands nombres.

3. Le Théorème Central Limite...Mais encore ?

Le théorème central limite est une théorie statistique qui va nous permettre d'agréger plusieurs variables aléatoires, et de les traiter « comme » une seule variable suivant une loi normale.

Plus précisément, le TCL établit la **convergence en loi** de la somme **d'une suite de variables aléatoires** vers **une loi normale**.

Théorème:

Soit $X_1, X_2, \ldots X_n$ une suite de variables aléatoires, **indépendantes et identiquement distribuées,** et définies sur le même espace de probabilité Ω. Nous pouvons donc dire que chacune de ces variables suit une loi D quelconque, d'espérance μ et d'écart type σ.

$$\{X_1, X_2, \ldots X_n\} \sim D(\mu, \sigma)$$

Considérons $S_n = X_1 + X_2 + \ldots + X_n$ la somme des variables aléatoires.

Rappel : $E[X+Y] = E[X] + E[Y]$ (propriété de l'espérance)

et $V[X+Y] = E[X] + E[Y] \forall X, Y \, indépendant$ (propriété de la variance)

Nous en déduisons que $E[S_n] = E[X_1] + E[X_2] + \ldots + E[X_n] = \mu + \mu + \ldots + \mu = n\mu$

et que $Var[S_n] = Var[X_1] + Var[X_2] + \ldots + Var[X_n] = \sigma^2 + \sigma^2 + \ldots + \sigma^2 = n\sigma^2$ donc que l'écart-type de S_n est égal à $\sqrt{n} \times \sigma$

Le TCL nous permet de dire que, **pour n assez grand**, $S_n \overset{L}{\sim} \mathrm{N}(n\mu, \sqrt{n} \times \sigma)$

n assez grand si n>30 pour des v.a continues.

Si les variables aléatoires considérées suivent une loi de Bernoulli, alors il suffit que n.p > 5 et n.(1-p) > 5 pour approximer la loi de Bernoulli par une Loi Normale.

Afin d'exploiter ce théorème, nous allons centrer (soustraction de l'espérance $n\mu$) et réduire (division par l'écart type $\sqrt{n} \times \sigma$) notre variable S_n .

Nous posons $Z_n = \dfrac{S_n - n\mu}{\sigma\sqrt{n}}$ notre variable centre réduite dérivé de S_n

On multiplie le numérateur et le dénominateur par n, ce qui revient à multiplier par 1.

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} \times \frac{n}{n} = \frac{S_n - n\mu}{n} \times \frac{n}{\sigma\sqrt{n}} = \left(\frac{S_n}{n} - \frac{n\mu}{n} \right) \times \frac{n}{\sigma\sqrt{n}} \quad (1)$$

or, $S_n = X_1 + X_2 + \ldots + X_n$ donc $\dfrac{S_n}{n} = \bar{X}$ (2)

et $\dfrac{n\mu}{n} = \mu$ (3)

Par ailleurs on peut écrire $\frac{n}{\sigma\sqrt{n}} = \frac{n^{1}}{1} \times \frac{1}{\sigma} \times \frac{n^{-1/2}}{1} = \frac{\sqrt{n}}{\sigma}$ (4)

On injecte (2), (3) et (4) dans (1)

Nous obtenons : $Z_n = (\bar{X} - \mu) \times \frac{\sqrt{(n)}}{\sigma}$

Rappel: $a \times \frac{c}{d} = \frac{a}{\frac{d}{c}}$

Nous obtenons : $Z_n = \frac{(\bar{X} - \mu)}{\sigma/\sqrt{n}}$

$E[Z_n] = 0$ car $E\left[\frac{(\bar{X}-\mu)}{\sigma/\sqrt{n}}\right] = \frac{1}{\sigma/\sqrt{n}} E[\bar{X} - \mu] = \frac{1}{\sigma/\sqrt{n}} E[0]$ puisque $\bar{X} = \mu$

Rappel : $Var(\bar{X}) = Var\left[\frac{S_n}{n}\right] = \frac{1}{n^2} Var[S_n] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$ car $Var[aX + k] = a^2 Var[X] \forall a, k \, constant$

$Var[Z_n] = 1$ car $Var\left[\frac{(\bar{X}-\mu)}{\sigma/\sqrt{n}}\right] = \frac{1}{(\sigma/\sqrt{n})^2} Var[\bar{X} - \mu] = \frac{1}{\sigma^2/n} Var[\bar{X}] = \frac{\sigma^2/n}{\sigma^2/n} = 1$

Conclusion : $Z_n \overset{L}{\sim} \mathrm{N}(0{,}1)$

C'est sympa toutes ces équations, mais concrètement ça va nous servir à quoi ?

Nous venons de démontrer que nous pouvons, à l'aide de quelques manipulations simples (la réduction et la centralisation) approximer la loi de probabilité d'une suite de variable i.i.d à une loi Normale d'Espérance 0 et d'Écart type 1.

Les moyennes d'un grand nombre d'échantillon (n) suivent une loi normale (de moyenne μ, et d'écart-type $\frac{\sigma}{\sqrt{n}}$) même si ceux-ci suivent individuellement une autre loi de probabilité.

Le TCL permet aussi de montrer qu'au-delà d'un certain effectif, la plupart des lois peuvent être approchées par une loi normale sous condition d'indépendance.

Au lieu de devoir calculer une par une les probabilités d'une suite de variables qui nous intéresse, il nous suffira de consulter une table de loi Normale, où toutes ces valeurs ont déjà été calculé.

__Applications : Saturation de standard__

On s'intéresse à un standard téléphonique

On note :

- N le nombre de lignes installées.
- X le nombre de personnes téléphonant au même instant *t*.
- La probabilité qu'une personne appelle à l'instant *t* est de 1/10.

On cherche à connaître N tel que la probabilité que X soit supérieure à N soit inférieur à 2,5 % : $P(X \geq N) \leq 0.025$

En supposant les appels indépendants , et en sachant que X suit une loi binomiale de paramètre n = 300 et p = 1/10, d'espérance $E(X) = 30$ et d'écart type $\sigma_X = \sqrt{(27)}$.

On pose $X_{cr} = \frac{(X-30)}{\sqrt{(27)}}$

Par le théorème central limite on peut considérer que X_{cr} converge en loi vers $N = (0,1)$.

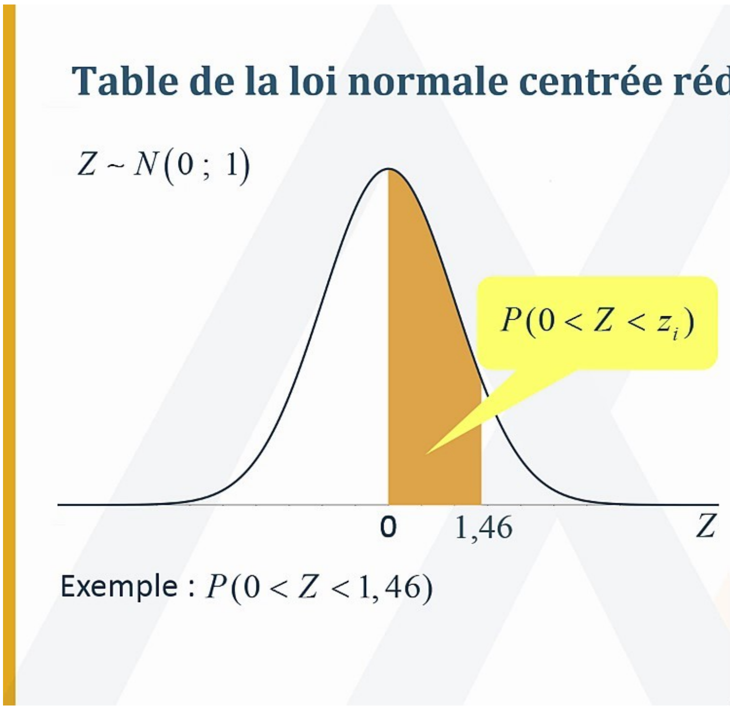
On note ϕ la fonction de répartition de la loi normale centrée réduite .

On a :

$$P(X \geq N) \leq 0.025 \Leftrightarrow P\left(X_{cr} \geq \frac{(N-30)}{\sqrt{(27)}}\right) \leq 0.025$$

$$\Leftrightarrow 1 - \phi\left(\frac{(N-30)}{\sqrt{(27)}}\right) \leq 0.025$$

$$\Leftrightarrow \phi\left(\frac{(N-30)}{\sqrt{(27)}}\right) \geq 0.975 = \phi(1.96)$$



z_i	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000
4,0	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

$$\Leftrightarrow \frac{(N-30)}{\sqrt{(27)}} \geq 1.96$$

$$\Leftrightarrow N \geq 41$$

Conclusion : il faut installer au moins 41 lignes pour éviter la saturation.

A présent vous avez les bases pour vous attaquer au monde de la statistique, outil fondamental dans le Machine Learning !