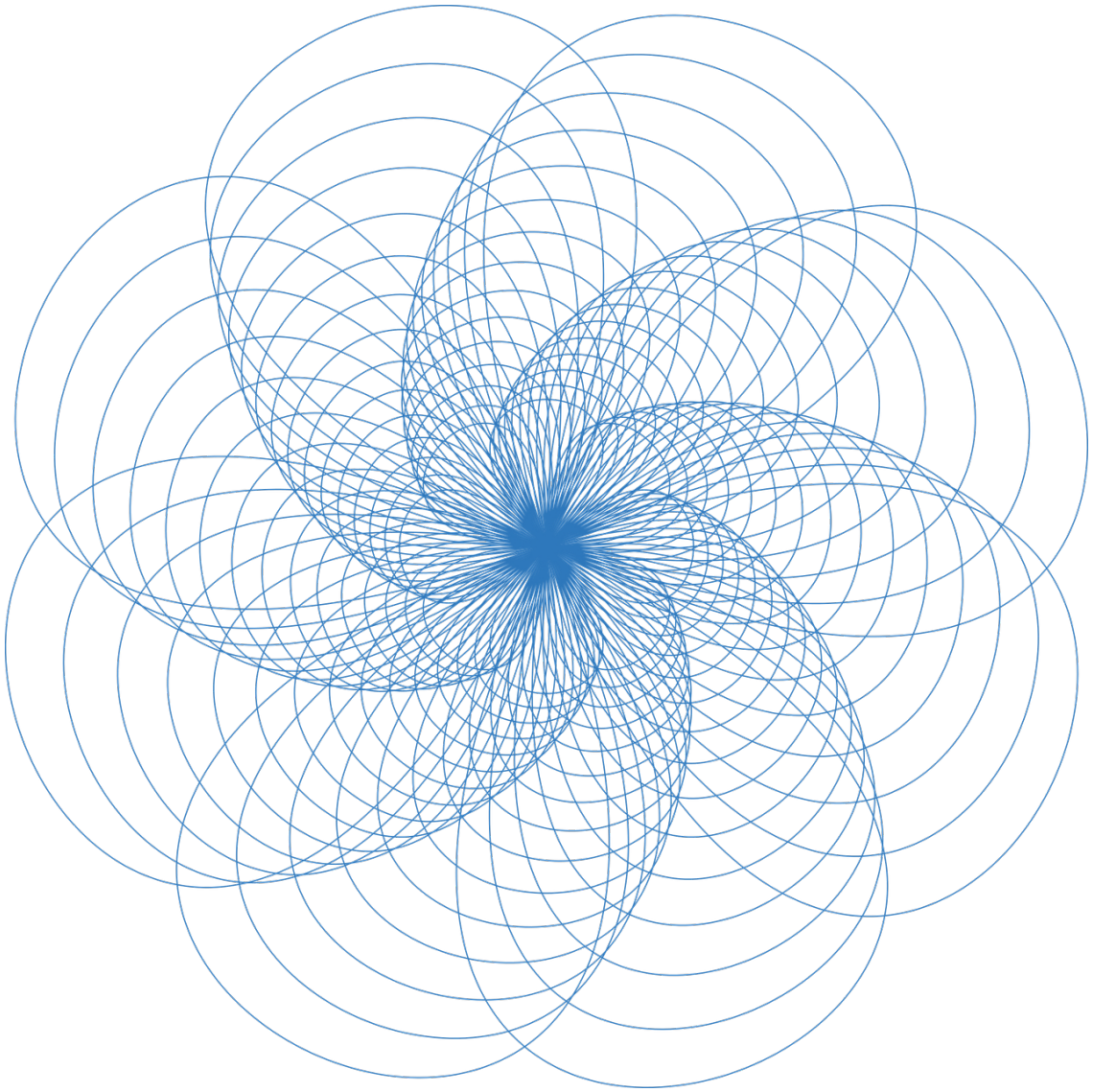
Mathématiques

Licence



Université Jean-François Champollion
Année 2022 - 2025

Table des matières

I — Raisonnements

Soit $\mathscr{P}$ une proposition et n un entier naturel.

Disjonctions & Conjonctions :

Si $\mathscr{P}$ est une disjonction de la forme $\mathscr{A} \vee \mathscr{B}$, il suffit alors de supposer **l'une des deux propriétés fausse** et de montrer que l'autre est vraie.

Si $\mathscr{P}$ est une conjonction de la forme $\mathscr{A} \wedge \mathscr{B}$, il faut simplement prouver $\mathscr{A}$ et $\mathscr{B}$.

Raisonnements par l'absurde :

Raisonner par l'absurde revient à utiliser le principe du **tiers exclu**, ie l'axiome qui affirme que la proposition ci-dessous est toujours vraie:

$$\mathscr{P} \vee \neg\mathscr{P}$$

Donc si on veut prouver $\mathscr{P}$, on peut alors simplement montrer que $\neg\mathscr{P} \implies \bot$ avec "$\bot$" comme notation d'une contradiction logique. Alors on peut conclure d'après l'axiome du tiers exclu que $\mathscr{P}$ est vraie.

Raisonnement par Analyse / Synthèse :

Le raisonnement par Analyse / Synthèse permet de déterminer **l'ensemble des solutions d'un problème**, il s'effectue en deux étapes, tout d'abord l'étape d'analyse suppose qu'une telle solution existe, alors on circonscrit son existence à des propriétés connues qu'elle vérifie nécessairement. Cette étape permet de "cerner" les solutions en question. Si les propriétés sont assez contraignantes, alors on peut même prouver **l'unicité**, ie l'ensemble des solutions se réduit à un singleton.

Puis lors de l'étape de synthèse, on considère un objet vérifiant les propriétés qu'on a utilisé lors de l'étape d'analyse, et on **vérifie** que cet objet est bien une solution au problème initial. C'est lors de cette étape qu'on prouve bien **l'existence** de solutions. Si aucun des objets circonscrits par l'analyse ne conviennent, le problème n'a alors pas de solutions.

Implications & Équivalences :

Si $\mathscr{P}$ est une implication de la forme $\mathscr{A} \implies \mathscr{B}$, on a les équivalences suivantes:

$$\mathscr{P} \iff \neg\mathscr{A} \vee \mathscr{B} \iff \neg\mathscr{B} \implies \neg\mathscr{A}$$

Aussi en raisonnant **par l'absurde**, il suffit alors de prouver:

$$\mathscr{A} \wedge \neg\mathscr{B} \implies \bot$$

Il est important de noter que l'implication **n'est pas une opération associative**, en effet, soit une propriété de la forme:

$$\mathscr{A}_1 \implies \mathscr{A}_2 \implies \mathscr{A}_3$$

Alors de manière générale, on a:

$$\mathscr{A}_1 \implies (\mathscr{A}_2 \implies \mathscr{A}_3) \not\iff (\mathscr{A}_1 \implies \mathscr{A}_2) \implies \mathscr{A}_3$$

Prouver une équivalence revient à prouver une **double implication** dans la majorité des cas.

Cas particulier : Si $\mathscr{P}$ est de la forme $A_1 \iff A_2 \iff \cdots \iff \mathscr{A}_{n-1} \iff \mathscr{A}_n$, il suffit alors de montrer:

$$\mathscr{A}_1 \implies \mathscr{A}_2 \implies \cdots \implies \mathscr{A}_{n-1} \implies \mathscr{A}_n \implies \mathscr{A}_1$$

Ainsi pour toute paire de $\mathscr{A}_i$, on a bien double implication entre les deux membres et donc la chaîne d'équivalence est démontrée.

Raisonnements par récurrence :

Soit $\mathscr{P}$ une propriété dépendante de n qu'on veut démontrer sur $[\![\alpha\ ;\ +\infty]\!]$, soit k en entier fixé supérieur à α, démontrer $\mathscr{P}$ par récurrence simple revient à utiliser **l'axiome de récurrence** (issu de la construction de $\mathbb{N}$) ci-dessous:

$$\Big[\mathscr{P}_\alpha \wedge \big[\mathscr{P}_k \implies \mathscr{P}_{k+1}\big]\Big] \implies \forall n \in \mathbb{N}\ ;\ \mathscr{P}_n$$

Si la propriété à prouver est plus complexe, on peut avoir besoin de récurrences d'une autre type, en effet si $\mathscr{P}$ dépend **des deux rangs précédents**, et on utilise alors une récurrence à deux pas qui s'exprime:

$$\Big[\mathscr{P}_\alpha \ \wedge\ \mathscr{P}_{\alpha+1} \wedge \big[\mathscr{P}_{k-1} \wedge \mathscr{P}_k \implies \mathscr{P}_{k+1}\big]\Big] \implies \forall n \in \mathbb{N}\ ;\ \mathscr{P}_n$$

Enfin pour le cas limite, si $\mathscr{P}$ dépends **d'exactement tout les rangs précédents**, alors on peut utiliser une récurrence forte qui s'exprime:

$$\Big[\mathscr{P}_\alpha \wedge\big[\mathscr{P}_{\alpha+1} \ \wedge\ \ldots\ \mathscr{P}_{k-1} \ \wedge\ \mathscr{P}_k \implies \mathscr{P}_{k+1}\big]\Big] \implies \forall n \in \mathbb{N}\ ;\ \mathscr{P}_n$$

Un dernier type de récurrence appelé **récurrence limitée** permet simplement d'utiliser la récurrence sur un intervalle entier fini, et donc on initialise et on prouve l'hérédité avec la contrainte de cette intervalle.

Remarque sur la récurrence forte :

Une telle récurrence forte ne nécessitera qu'une **unique** initialisation pour compléter l'hérédité.

D'un point de vue heuristique, il peut arriver d'engager une récurrence forte sur un problème qui n'aurait nécessité qu'une récurrence à p pas.
Ce cas précis reviendra alors, lors de l'étape d'hérédité, à **ne pas utiliser l'ensemble de l'hypothèse de récurrence**, et alors il faudra modifier le nombre d'initialisation à réaliser et l'intervalle de notre hypothèse de récurrence.

Admettons que $\mathscr{P}_\alpha$ soit vraie, supposons qu'elle soit vraie sur $[\![\alpha\ ;\ k]\!]$. Alors, on doit montrer que la propriété est vraie au rang $k+1$.

Alors, selon **le plus petit rang** nécessaire à compléter l'hérédité, on a:

Si on a besoin de $\mathscr{P}_k$ alors **on se ramène à une récurrence simple.**
Si on a besoin de $\mathscr{P}_{k-1}$ alors **on se ramène à une récurrence double.**
Si on a besoin de $\mathscr{P}_{k-2}$ alors **on se ramène à une récurrence triple.**
.....

Et donc, les initialisations et l'intervalle de notre hypothèse de récurrence changeront en conséquence et on remarque alors que si le plus petit rang nécessaire est $\mathscr{P}_{k-p}$, alors on se ramène nécessairement à une **récurrence à p pas**, avec p initialisations et l'hypothèse de récurrence qui commence à $\alpha+p$.

Une récurrence forte n'est alors qu'une récurrence qui nécessite des hypothèses sur **tout les rangs précédents.**

Récurrences imbriquées :

Soit $\mathscr{P}_{n,m}$ une propriété qui dépend **de deux variables entières**, alors on pourrait prouver $\mathscr{P}_{n,0}$ par récurrence et alors cela constituerait l'initialisation d'une récurrence imbriquée qui supposerait par exemple $\mathscr{P}_{n,k}$ vraie pour prouver $\mathscr{P}_{n,k+1}$.

I — Ensembles

Soit E un ensemble de parties. On munit cet ensemble des opérations élémentaires d'union, d'intersection ainsi que de complémentation définies pour toutes familles d'ensembles $(A_i)_I$ par:

- **Union:** $x \in \bigcup_i A_i \iff \exists i \in I \ ; \ x \in A_i$
- **Intersection:** $x \in \bigcap_i A_i \iff \forall i \in I \ ; \ x \in A_i$
- **Complémentaire:** $x \in A^c \iff x \notin A$

Ces opérations sont compatibles entre elles au sens où elle sont distributives l'une sur l'autre et associatives (pour une suite d'un seul type d'opération, si il y a mélange d'intersections et d'unions, on n'a pas associativité).

Inclusion :

On définit une **relation d'ordre** sur l'ensemble des parties de E appelée inclusion, elle est **réflexive, transitive et antisymétrique**. Si l'inclusion est stricte, on parle de **sous-ensemble propre**. En particulier, les opérations élémentaires préservent l'inclusion, en effet si $F \subseteq G$, on a:

$$F \cap X \subseteq G \cap X$$
$$F \cup X \subseteq G \cup X$$

Néanmoins la complémentation inverse l'inclusion:

$$F^c \subseteq E^c$$

Complémentaire et différence :

On peut montrer deux propriétés fondamentales du complémentaire apellées **lois de De Morgan** qui nous donnent:

$$(E \cap F)^c = E^c \cup F^c$$
$$(E \cup F)^c = E^c \cap F^c$$

On peut aussi raffiner la notion de complémentaire en définissant **la différence ensembliste** pour tout partie $F, G \subseteq E$, on pose:

$$F \setminus G = F \cap G^c$$

Cas particulier : On peut aussi définir l'opération de **différence symétrique** notée Δ qui permet d'obtenir tout les éléments qui appartiennent exactement à un seul des deux ensembles:

$$E \Delta F = (E \cup F) \setminus (E \cap F)$$

Produit cartésien :

Soit n une entier naturel, le produit cartésien des ensembles E_1, E_2, ... , E_{n-1}, E_n est l'ensemble des n-uplets de la forme $(e_1, e_2, \ldots, e_{n-1}, e_n)$ avec $e_i \in E_i$ pour $i \in [\![1 \ ; \ n]\!]$. Il y a **unicité** de ces n-uplets.
Plus formellement, on note:

$$\prod_{i=1}^{n} E_i = \left\{ (e_1, e_2, \ldots, e_{n-1}, e_n) \ ; \ e_1 \in E_1, e_2 \in E_2, \ldots, e_n \in E_n \right\}$$

Le produit cartésien est distributif sur l'union et l'intersection ie si on note $\star$ une de ces deux opérations, on a:

$$A \times (B \star C) = (A \times B) \star (A \times C)$$

Cardinalité :

Supposons que E et F tout deux inclus dans X et ayant **un nombre fini d'éléments**. On a alors différentes propriétés:

$$\begin{aligned} |E \cup F| &= |E| + |F| - |E \cap F| \\ |E \times F| &= |E| \times |F| \\ |E^c| &= |X| - |E| \\ |\mathcal{P}_E| &= 2^{|E|} \end{aligned}$$

Partitions et recouvrements:

Soit $(P_i)_{i\in\mathbb{N}}$ une famille de parties **non vides et deux à deux disjointes** de E.
On dit que (P_i) est une **partition** de E si et seulement si:

$$\bigcup_{i\in\mathbb{N}} P_i = E$$

On remarque immédiatement deux partitions singulières:

- La famille contenant uniquement E qu'on appelle **partition grossière**.
- La famille contenant tout les singletons de E qui est la partition **la plus fine**.

On peut donc intuitivement parler de **finesse** d'une partition, en regard de la taille des parties de la famille.

On peut généraliser le concept de partitition à celui de **recouvrement**, alors E ne nécessite que d'être contenu par l'union des (P_i).

Algèbre de Boole :

On peut montrer que l'ensemble **ordonné** des parties de E muni de l'union, l'intersection, le complémentaires forment une **Algèbre de Boole**.

Cela signifie que la structure $(\mathcal{P}(E), \cup, \cap, X^c)$ vérifie les axiomes suivants:

- Les deux opérations binaires sont **associatives, commutatives et distributives l'une sur l'autre.**
- Les deux opérations binaires sont **idempotentes.**
- **L'élément neutre** pour l'union est l'ensemble vide, et pour l'intersection l'ensemble E.
- **L'élément absorbant** pour l'union est l'ensemble E, et pour l'intersection l'ensemble vide.

- Le complémentaire est **involutif**.
- L'intersection d'un élément et de son complémentaire est **vide**.
- L'union d'un élément et de son complémentaire est **l'ensemble tout entier**.
- Les **lois de De Morgan** sont vérifiées.

De manière analogue, en considérant $\{0, 1\}$ comme les valeurs de vérité d'une proposition, on a:

La structure $(\{0, 1\}, \vee, \wedge, \neg)$ est aussi une algèbre de Boole.

Cette structure est à la base de la logique formelle et vérifie les même axiomes que l'algèbre de l'ensemble des parties d'un ensemble.

I — Relations

Une **relation** entre des objets d'un ensemble est une propriété que vérifient ces objets **entre eux**.
Les relations sont des objets **fondamentaux** en mathématiques, elles sont entre autres des objets primitifs de la théorie des ensembles.

On appelle **arité** le nombre d'éléments mis en jeu par la relation.
Par exemple une relation d'arité 2 est appelée **relation binaire** et met en jeu deux éléments. On définit ainsi le cas général de relation **n-aire** qui met en jeu n éléments $x_1, x_2, \ldots, x_{n-1}, x_n$ et on note:

$$\mathscr{R}(x_1, x_2, \ldots, x_{n-1}, x_n)$$

Par abus de langage, on appelle **classe** un ensemble d'ensembles.
Formellement une classe n'est pas un ensemble mais un élément primitif de la théorie ZFC, mais ici on verra qu'on appelle classe des objets qui **sont** des ensembles.

Zoologie :

Il existe un grand nombre de relations très connues et élémentaires, par exemple:

- La relation d'appartenance à un ensemble
- La relation d'égalité
- La relation d'ordre
- La relation d'inclusion
- La relation de congruence
- La relation de parallélisme de deux droites du plan

On peut remarque que la relation d'appartenance à un ensemble est une relation binaire fondamentale, à la base de la théorie des ensembles.

Relations binaires :

Soit $x, y, z \in E$, une relation entre deux éléments peut vérifier plusieurs propriétés remarquables:

- **Réflexivité :** $\mathscr{R}(x, x)$
- **Irréflexivité :** $\not\mathscr{R}(x, x)$
- **Symétrie :** $\mathscr{R}(x, y) \implies \mathscr{R}(y, x)$
- **Antisymétrie :** $\mathscr{R}(x, y) \wedge \mathscr{R}(y, x) \implies x = y$

Elle peut aussi être **transitive**:

$$\mathscr{R}(x, y) \wedge \mathscr{R}(y, z) \implies \mathscr{R}(x, z)$$

On appelle aussi relation **totale** une relation telles si pour toute paire d'éléments, on a $\mathscr{R}(x, y) \vee \mathscr{R}(y, x)$.

Relations d'ordre :

Une **relation d'ordre** est une relation **réflexive, antisymétrique et transitive**. Elle induit un ordre sur l'ensemble E, qui peut potentiellement être **total**.

Des relations d'ordre très connues sont la relation $\leq$ sur les ensembles de nombres ou la relation $\subseteq$ sur l'ensemble des parties de E.

On appelle relation de **préordre** toute relation relation d'ordre qui n'est pas antisymétrique. Intuitivement, une relation de préordre est une relation d'ordre à "équivalence près" des éléments.

Relations d'équivalence :

Une **relation d'équivalence** est une relation **réflexive, symétrique et transitive**. Intuitivement, elle met en relation les éléments des ensembles qui sont "similaires".

Des relations d'équivalence très connues sont la relation $=$ et $\equiv$ sur les ensembles de nombres, ou encore la relation $\sim$ sur l'ensemble des fonctions.

Classes d'équivalence :

Soit $(E, \sim)$ un ensemble muni d'une relation d'équivalence.
Les **classes d'équivalence** de E par rapport à la relation $\sim$ sont alors les parties de E contenant des éléments en relation.

Soit $x \in E$, on définit alors la **classe d'équivalence** de x et on note $[x]$ l'ensemble:

$$[x] := \Big\{ \alpha \in E \ ; \ \alpha \sim x \Big\}$$

D'après les propriété de la relation, on a alors:

$$\boxed{x \sim y \iff [x] = [y]}$$

Et on appelle **représentant** de $[x]$ tout élément qui appartient à $[x]$.

Ensembles quotient :

L'ensemble des classes d'équivalence de E forme alors une **partition** de E, et on appelle **ensemble quotient**, ou encore **ensemble quotienté par la relation d'équivalence** l'ensemble:

$$E/\sim := \Big\{ [x] \in \mathcal{P}(E) \ ; \ x \in E \Big\}$$

C'est alors un ensemble de classes d'équivalences par rapport à la relation $\sim$.

Travailler avec l'ensemble quotient revient alors à ne pas distinguer les éléments équivalents entre eux.
On peut aussi créer des **structures** quotient, il suffit alors de quotienter une structure algébrique de telle sorte que les propriétés de structure soient conservées.

Quelques exemples connus de structures quotient:

- **L'anneau** $\mathbb{Z}/n\mathbb{Z} := (\mathbb{Z}/\sim, +, \times)$ **pour la relation** $a \sim b \iff a \equiv b[n]$
- **Le corps** $\mathbb{Q} := ((\mathbb{Z}\,;\, \mathbb{Z}\setminus\{0\})/\sim, +, \times)$ **pour la relation** $(a,b) \sim (c,d) \iff ad = bc$

I — Fonctions & Applications

On appelle **fonction** ou **application** des cas particulier de relation entre deux ensembles, soit f, g deux fonctions telles que:

$$\begin{aligned} f : E &\longrightarrow F & \qquad g : G &\longrightarrow H \\ x &\longmapsto f(x) & \qquad x &\longmapsto g(x) \end{aligned}$$

On note D_f le sous-ensemble de E tel que $f(x)$ existe, alors f est une **application** si et seulement si $E = D_f$ et on note alors $\mathscr{F}(E, F)$ l'ensemble des **applications** de E vers F.

Si $F \subseteq G$, alors on définit **la composée** $g \circ f$ par la fonction $h : x \in E \longmapsto g(f(x)) \in H$

Cas des suites :

Une suite à valeurs dans E n'est alors qu'un cas particulier en la forme d'une fonction $u : \mathbb{N} \longrightarrow E$, ce sont des objets d'étude trés importants en analyse et notamment en topologie. Dans le cas des suites on peut définir la notion de **suite extraite**, car si u_n est une suite dans E et k_n est **suite d'entiers croissante**, alors on définit une suite extraite de u_n par:

$$u \circ k : \mathbb{N} \longrightarrow E$$

C'est simplement les termes de la suite u_n dont on ne choisit que les termes d'indices donnés par k_n.

Graphe :

On définit **le graphe** de f comme suit:

$$G_f := \Big\{ (x, f(x)) \in E \times F \ ; \ x \in E \Big\}$$

Intuitivement, c'est l'ensemble des couples d'éléments, des points, qui caractérise uniquement la fonction.

Restrictions & Prolongements :

On note $f|_A$ la restriction de **l'ensemble de départ** de f à une partie A de E.
On note $f|^B$ la restriction de **l'ensemble d'arrivée** de f à une partie B de F.

Soit $x \in D_f$, on appelle **prolongement** de f, l'application g telle que $D_f \subset D_g$ et $g(x) = f(x)$

Image directe :

Une fonction induit canoniquement une autre fonction sur l'ensemble des parties, notée aussi f, qui à chaque partie associe la partie **image directe**, ie pour toute partie A de E on définit:

$$f(A) := \Big\{ f(x) \ ; \ x \in A \Big\}$$

L'image directe est compatible avec **certaines opérations ensemblistes**, plus précisément:

- **Intersection:** $f(A \cap B) = f(A) \cap f(B)$
- **Union:** $f(A \cup B) \subset f(A) \cup f(B)$

Image Réciproque :

Toute fonction induit aussi une autre fonction sur l'ensemble des parties qui à chaque partie associe la partie **image réciproque**, ie pour toute partie A de E on définit:

$$f^{-1}(B) := \Big\{ x \in A \ ; \ f(x) \in B \Big\}$$

L'image réciproque est compatible avec **toutes les opérations ensemblistes**, plus précisément:

- **Intersection:** $f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B)$
- **Union:** $f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B)$

Injections :

L'application f est dite **injective** si et seulement si:

$$\forall x_1, x_2 \in E^2 \ ; \ f(x_1) = f(x_2) \implies x_1 = x_2$$

Avoir une injection de $E \longrightarrow F$ permet **d'identifier une partie de F à E**. Réciproquement, la non-injectivité représente le fait que la fonction détruit de l'information [1], ou alors que l'espace d'arrivée est trop petit.

Si on considère une composée $g \circ f$ injective alors on peut montrer que f est nécessairement injective. L'injectivité est stable par composition.

Surjections :

L'application f est dite **surjective** si et seulement si:

$$\forall y \in F \ , \ \exists x \in E \ ; \ f(x) = y$$

Avoir une surjection de $E \longrightarrow F$ permet **d'identifier une partie de E à F**. Réciproquement, la non-surjectivité indique que la fonction transporte trop peu d'informations, ou que l'espace d'arrivée est trop grand.

Si on considère une composée $g \circ f$ injective alors on peut montrer que g est nécessairement surjective. La surjectivité est stable par composition.

Bijections :

L'application f est bijective si et seulement si elle est surjective et injective. Dans ce cas, **une application réciproque** g existe et elle vérifie:

$$\begin{cases} f \circ g &= Id_F \\ g \circ f &= Id_E \end{cases}$$

Réciproquement, si il existe une application g telle que f soit inversible à gauche et à droite par g, alors f est bijective. Aussi la bijectivité est stable par composition.

Equipotence & Cardinalités :

On peut étendre la notion de cardinalité d'un ensemble fini via la notion d'application, en particulier on dira que pour tout ensembles A, B, alors:

- Si il existe une injection $f : E \longrightarrow F$, alors on a nécessairement $|E| \leq |F|$
- Si il existe une surjection $f : E \longrightarrow F$, alors on a nécessairement $|E| \geq |F|$

Si il existe une bijection de E vers F, alors on dit que ces ensembles sont **équipotents**, et on a:

$$|E| = |F|$$

Cette définition du cardinal par les bijections permet de parler de cardinal d'un ensemble dans le cas **infini**. En particulier:

- Si il existe une bijection entre $\mathbb{N}$ et E, on dit que E est **dénombrable**[2] et on note $|E| = \aleph_0$
- Si il existe une bijection de $\mathbb{R}$ dans E, alors on dit que E est **indénombrable** et on note $|E| = \aleph_1$

Dans notre cadre théorique (ZFC), il n'existe aucun ensemble dont le cardinal se situerait entre $\aleph_0$ et $\aleph_1$, c'est **l'hypothèse du continu**.

[1] Dans le sens où si deux valeurs différentes ont la même image, on ne peut plus les distinguer à l'arrivée.
[2] En fait, une injection suffit car on considérera par la suite que les ensembles finis sont dénombrables.

I — Dénombrement

Soit E un ensemble, on dit que E est **fini** si il existe une bijection de $[\![1\ ;\ n]\!]$ sur E.

On considère maintenant que E est fini, dénombrer E consiste à déterminer sa cardinalité. Informellement il s'agit souvent de compter le nombres **d'issues possibles** d'une situation donnée, on dispose alors de trois grands modèles, les **listes**, les **arrangements** et les **combinaisons**.

Listes

On appelle **liste** à p éléments de E un p-uplet constitué d'éléments de E, c'est à dire **un élément du produit cartésien** E^p on remarque alors la propriété:

Dans une liste, l'ordre compte et les répétitions sont possibles

On peut alors montrer que le nombre d'applications d'un ensemble à p éléments dans un ensemble à n éléménts est p^n

Arrangements

On appelle **arrangement** tout liste à p éléments **distincts** de E, on remarque alors:

Dans un arrangement, l'ordre compte mais les répétitions sont impossibles

On note alors A_n^p le nombre d'arrangements de p éléménts d'un ensemble à n éléments et on a:

$$A_n^p = \frac{n!}{(n-p)!}$$

Et on peut alors montrer que le nombre d'applications **injectives** d'un ensemble à p éléments dans un ensemble à n éléménts est A_n^p.

Un arrangement de la forme A_n^n est apellée une **permutation** de E qui est simplement donnée par $n!$, c'est aussi le nombre de **bijections** de E dans E.

Combinaisons

On appelle **combinaison** de p éléments tout **partie** de E à p éléments, on remarque alors:

Dans une combinaison, l'ordre ne compte pas et les répétitions sont impossibles

On apelle alors **coefficient binomial** et on note $\binom{n}{p}$ le nombre de parties à p éléménts d'un ensemble à n éléments et on a:

$$\binom{n}{p} = \frac{n!}{p!(n-p)!}$$

On peut remarquer que le nombre de parties à p éléménts de E est exactement le nombre d'arrangements à p éléménts de E auquel on retire toutes les permutations des p éléments choisis, ce qui revient exactement à **retirer la contrainte d'ordre**.

Propriétés du coefficient binomial

Le coefficient binomial possède plusieurs propriétés intéréssantes, on peut tout d'abord remarquer une **symétrie** évidente mais aussi:

Formule de Pascal: $\binom{n}{p} = \binom{n-1}{p} + \binom{n-1}{p-1}$ **Formule du capitaine:** $p\binom{n}{p} = n\binom{n-1}{p-1}$

La **formule de Pascal** se comprends si on considère un élément fixé de l'ensemble et qu'on dénombreux tout ceux qui le contiennent, et les autres, ie:

Le nombre de parties à p éléments est exactement la somme du nombre de parties qui ne contiennent pas un certain x et du nombre de partie qui contiennent ce x.

La **formule du capitaine** se comprends si on considère le choix d'une équipe sportive de p joueurs (dont un capitaine) parmi un groupe de n candidats:

Choisir une équipe de p joueurs puis un capitaine parmi les p joueurs revient à choisir un capitaine parmi les n candidats, puis les $p-1$ joueurs restants.

Enfin on a aussi:

$$\sum_{p=0}^{n} \binom{n}{p} = 2^n$$

Le cardinal de l'ensemble des parties d'un ensemble à n éléments est donc exactement la somme des parties qui ont respectivement $1, 2, \ldots, n$ éléments.

Généralisation

On peut remarquer que le coefficient binomial est le nombre de partitions en deux parties de E telles que le cardinal de la première soit p. Par exemple si on considère les partitions de $E := \{1, 2, 3\}$ en deux parties dont la première ait 1 élément, on remarque qu'il y a 3 telles partitions:

$$P = \big(\{1\}, \{2, 3\}\big) \text{ ou } \big(\{2\}, \{1, 2\}\big) \text{ ou } \big(\{3\}, \{1, 2\}\big)$$

On peut alors généraliser cette idée et définir le **coefficient multinomial** $\binom{n}{k_1,\ldots,k_p}$ qui sera le nombre de partitions en p parties telles que la p-ième partie soit de cardinal k_p avec la somme des k_p **qui soit égale au cardinal total**:

$$\boxed{\binom{n}{k_1, \ldots, k_p} = \frac{n!}{k_1! k_2! \ldots k_p!}}$$

Pour fixer les idées on remarque que si $p = 2$ on a bien notre coefficient binomial usuel[1][2]:

$$\binom{n}{k_1, k_2} = \binom{n}{k_1, n-k_1} = \binom{n}{k_1} = \frac{n!}{k_1!(n-k_1)} = \frac{n!}{k_1! k_2!}$$

On peut alors utiliser ce coefficient multinomial, pour compter le nombre d'anagramme d'un mot de n lettres avec m lettres distinctes répétées k_m fois, ou encore le nombre de façon de mettre n objets dans m boites qui peuvent en contenir k_m.
Par exemple, le nombre d'anagrammes de MISSISSIPI est donné par $\binom{11}{1,4,4,1} = \frac{11!}{4!4!} = 34650$

On peut même pour définir la **formule du multinôme de Newton** qui généralise celle du binôme:

$$(x_1 + x_2 + \ldots + x_p)^n = \sum_{k_1+k_2+\ldots+k_p=n} \binom{n}{k_1, k_2 \ldots, k_n} x_1^{k_1} x_2^{k_2} \ldots x_p^{k_p}$$

[1] La première égalité vient de la contrainte sur la somme des k_p.
[2] La seconde égalité se comprends par symétrie, compter le nombre de partitions en deux parties dont la première contient k_1 éléments revient à compter le nombre de parties à k_1 éléments et le reste sera nécessairement dans la seconde partie.

I — Arithmétique élémentaire

Dans ce chapitre on énonce quelques définitions et propriétés arithmétiques simples dans $\mathbb{Z}$, qui seront généralisées plus tard dans le chapitre d'algèbre au cas général. Dans cet ensemble on peut définir une relation d'ordre de divisibilité définie par:

$$a|b \iff \exists k \in \mathbb{Z}\ ;\ b = ak$$

On dira alors que b est un multiple de a et que a divise b. On notera $\mathcal{D}(n)$ l'ensemble des diviseurs de n et $n\mathbb{Z}$ l'ensemble de ses multiples. Une première propriété trés utile de cette relation et que si a divise b, c alors pour tout $n, m \in \mathbb{Z}$ on a:

$$a|nb + mc$$

Division Euclidienne :

Soit $a, b \in \mathbb{Z} \times \mathbb{Z}^*$, on peut montrer qu'il existe un unique couple $(q, r) \in \mathbb{Z} \times \mathbb{N}$ avec $r < |b|$ tel que:

$$a = bq + r$$

On appelle alors cette décomposition **la division euclidienne** de a par b. La preuve se fait par l'exibition de l'algorithme bien connu.

Plus grand diviseur commun :

Soit $a, b \in \mathbb{Z}$ non simultanément nuls, alors le pgcd est l'entier $a \wedge b$ qui vérifie:

$$a \wedge b := \max\left\{n \in \mathbb{N}\ ;\ n|a \text{ et } n|b\right\}$$

Alors on l'appelle **plus grand diviseur commun** de a et de b et on le note $a \wedge b$. On peut alors monter plusieurs propriétés de cette quantité:

- **Maximalité:** Si d est un diviseur commun de a, b alors $d|a \wedge b$.
- **Réduction:** On peut réduire le pgcd par division euclidienne, ie $a \wedge b = b \wedge r$.

On peut alors définir la notion de deux entier n, m **premiers entre eux** par le fait que $n \wedge m = 1$.

Plus petit commun multiple:

Soit $a, b \in \mathbb{Z}$ non simultanément nuls, alors le ppcm est l'entier $a \vee b$ qui vérifie:

$$a \vee b := \min\left\{n \in \mathbb{N}\ ;\ a|n \text{ et } b|n\right\}$$

Alors on l'appelle **plus petit commun multiple** de a et de b et on le note $a \vee b$. On peut alors monter plusieurs propriétés de cette quantité:

- **Minimalité:** Si d est un multiple commun de a, b alors c'est un multiple de $a \vee b$.

Identité de Bézout :

Soit $a, b \in \mathbb{Z}^2$, on peut montrer par une extension de l'algorithme d'Euclide appelé **algorithme d'Euclide étendu**[1] qu'il existe deux entiers $u, v \in \mathbb{Z}^2$ tels que:

$$au + bv = a \wedge b$$

Il existe donc une combinaison linéaire (à coefficients entiers) de a, b qui donne leur PGCD.

[1] En effet remontrer l'algorithme par subsitution permet d'écrire le dernier reste non nul, le pgcd, comme combinaison linéaire des deux nombres de départ.

Lemme de Gauss :

Soit 3 entiers $a, b, c \in \mathbb{Z}$, alors grâce à l'identité de Bézout, on peut montrer le **lemme de Gauss**:

$$\begin{cases} a \mid bc \\ a \wedge b = 1 \end{cases} \implies a \mid c$$

Nombres premiers :

On appelle nombres premiers tout nombre différent de 1 qui n'admet aucun diviseur. Il existe une infinité de nombres premiers et on a le **théorème de décomposition**:

$$\forall n \in \mathbb{Z}\ ;\ n = \prod_{p \in \mathcal{P}} p^{v_p(n)}$$

Où $v_p(n) = \max\left\{k \in \mathbb{N}\ ;\ p^k | n\right\}$ est apellée **valuation p-adique** de n.

Indicatrice d'Euler :

En algèbre, il sera utile de connaître **le nombre d'entiers inférieurs à n et premiers avec n**, pour ceci on définit **la fonction indicatrice d'Euler** par:

$$\begin{aligned} \varphi : \mathbb{N} &\longrightarrow \mathbb{N} \\ n &\longmapsto n \prod_{p \mid n} \left(1 - \frac{1}{p}\right) \end{aligned}$$

Le produit se faisant sur tout les diviseurs premiers distincts de n. L'utilité de cette fonction vient de la propriété suivante que justement $\phi(n)$ est exactement le nombre d'entiers inférieurs à n et premiers avec n.

Exemple: $\varphi(30) = \varphi(2 \times 3 \times 5) = 30 \times \left(1 - \frac{1}{2}\right)\left(1 - \frac{1}{3}\right)\left(1 - \frac{1}{5}\right) = 30 \times \frac{1}{2} \times \frac{2}{3} \times \frac{4}{5} = 8$

XII — Espaces probabilisés

Le domaine des probabilités chercher à modéliser des **expériences aléatoires** ie des expériences dont toutes les **issues** possibles sont connues à priori mais dont le résultat peut varier lorsqu'on la répète (lancer de dés, tirage dans une urne...).

Le cadre de la théorie de la mesure, nous permet de formaliser la théorie axiomatique des probabilités, ainsi que les différents objets en jeu, en particulier, on considère un espace mesurable $(\Omega, \mathcal{A})$ muni d'une mesure $\mathbb{P}$ à valeurs dans $[0\,;\,1]$ et telle que $\mathbb{P}(\Omega) = 1$. On apelle une telle mesure **loi de probabilité**.

Le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ est alors apellé **espace probabilisé**. Dans ce cadre l'ensemble Ω des issues possibles de l'expérience est apellé **univers**, les parties mesurables sont apellées **évenements** et deux parties mesurables disjointes seront dites **incompatibles**.

Espace probabilisé discret et continu

La mesure de l'espace doit être égale à 1 donc en particulier, on doit avoir $\int_\Omega d\mathbb{P} = 1$, ceci étant dit, on peut alors distinguer deux grands cas d'espaces probabilisés:

- Le cas où les parties non-négligeables de Ω sont **dénombrables**, alors par applications de la relation de Chasles et l'invisibilité des parties négligeables, on obtient que:

$$\int_\Omega d\mathbb{P} = \int_{\bigcup x_n} d\mathbb{P} = \sum_n \mathbb{P}(x_n)$$

  On remarque alors que la loi de probabilité est entièrement déterminée par la probabilité **d'évenements élémentaires** d'une certaine famille (x_n) dont la série vaut 1. On appele cette famille **distribution** de $\mathbb{P}$ et de tels espaces **espaces probabilisés discrets**.

- Le cas où elles ne le sont pas et que $\Omega \subseteq \mathbb{R}^n$, alors il est toujours possible de définir la **fonction de répartition** de la mesure de probabilité par la fonction suivante qui caractérise la loi:

$$F : x \longrightarrow \mathbb{P}(]-\infty\,;\,x])$$

  Si de plus les non-boréliens sont négligeables pour $\mathbb{P}$ (on dit aussi que $\mathbb{P}$ est **absolument continue** par rapport à la mesure de Lebesgue) alors on peut montrer que $\mathbb{P}$ admet une densité, c'est à dire une fonction intégrable f dont l'intégrale vaut 1 et qui caractérise alors la probabilité:

$$\mathbb{P}(A) = \int_A f(x)dx$$

  On dira alors que ces lois sont des **lois à densité**. Si la dernière condition n'est pas vérifiée, on dira alors que la loi est **mixte ou singulière**.

Espace probabilisé produit

Si on se donne une famille de n espaces probabilisés $(\Omega_i, \mathscr{A}_i, \mathbb{P}_i)$, on peut alors conformément à la théorie de la mesure définir l'espace produit $(\prod \Omega_i, \mathscr{A}_\otimes, \mathbb{P}_\otimes)$ avec la tribu et la loi produit. Dans tout la suite on considère simplement le cas $n = 2$ pour simplifier.
Selon le cas discret ou a densité, on a alors que la loi est caractérisée par:

- **Cas dénombrable:**

$$\mathbb{P}_\otimes(A) = \sum_{\mathbb{N}\times\mathbb{N}} \mathbb{P}_\otimes(x_n, y_m)$$

- **Cas absolument continu:**

$$\mathbb{P}_\otimes(A) = \int_{\Omega_1\times\Omega_2} f(x,y)d\mathbb{P}_\otimes$$

Lois marginales

Sachant la loi produit $\mathbb{P}$, une question intéressante est alors de determiner les lois **marginales** des espaces composantes, on montre alors qu'on a:

- **Dans le cas dénombrable:**
$$\mathbb{P}_1(\{k\}) = \sum_{\mathbb{N}} \mathbb{P}(\{k\}, y_m)$$
- **Dans le cas absolument continu:**
$$f_1(x) = \int_{\Omega_1} f(x,y)dy$$

Malheureusement on peut montrer que la donnée des lois marginales ne caractérise pas la loi produit, en effet les lois marginales dans le cas fini par exemple correspondent aux sommes des lignes ou colonnes du tableau des probabilités et deux sommes peuvent être égales sans que les valeurs individuelles soient toutes égales.

Exemples

Plusieurs exemples de différentes natures:

- **Discret fini:** Si on cherche à modéliser 3 tirages successifs à pile ou face avec une pièce non truquée, on peut modéliser cette expérience par l'espace probabilisé suivant:
$$\left(\left\{(r_1,r_2,r_3)\ ;\ (r_i) \in \{P,F\}\right\}, \mathcal{P}(\Omega), \mathbb{P}(A) = \frac{|A|}{|\Omega|}\right)$$
- **Discret infini:** Si on cherche à modéliser le nombres de visiteurs qui se présentent dans un musée, on peut alors modéliser ce phénomène par un espace probabilisé dénombrable et une probabilité rapidement décroissante, ie on pose par exemple:
$$\left(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mathbb{P}(\{n\}) = \frac{1}{2^{n+1}}\right)$$
Alors ceci est bien un espace probabilisé discret infini.
- **Absolument continu:** Si on cherche à modéliser un tir de fléchette sur le disque unité $D \subseteq \mathbb{R}^2$où la probabilité suit une densité uniforme, alors on peut poser:
$$\left(D, \mathcal{B}(D), \mathbb{P}(D) = \frac{1}{\pi}\int_A d\mu = \frac{\mu(A)}{\pi}\right)$$
- **Mixte:** Si on cherche à modéliser une loterie où l'on tire un nombre de $[0\,;\,10]$ avec $\mathbb{P}(\{0\}) = 0.1$ qui correspond au jackpot et densité uniforme pour le reste des nombres. Alors on a naturellement une structure d'espace probabilisé mixte.
- **Espace produit:** Si on cherche à modéliser le choix uniforme d'un point dans $[\![1\ ;\ n]\!]^2$, on modèlise ceci par l'espace produit:
$$([\![1\ ;\ n]\!]^2, \mathcal{P}([\![1\ ;\ n]\!]^2), \mathbb{P}(\{(k,l)\}) = \frac{1}{n^2})$$
Alors les distributions marginales sont facilement $\mathbb{P}_1(\{k\}) = \mathbb{P}_2(\{k\}) = \frac{1}{n}$, on a en fait le tableau des probabilités décrit par la matrice de taille n suivante:
$$\begin{pmatrix} \frac{1}{n^2} & \cdots & \frac{1}{n^2} \\ \vdots & \ddots & \vdots \\ \frac{1}{n^2} & \cdots & \frac{1}{n^2} \end{pmatrix}$$
Les colonnes (resp. lignes) correspondant aux probabilités $\mathbb{P}_1(\{k\})$ (resp. $\mathbb{P}_2(\{k\})$)

XII — Probabilités conditionnelles

Lorsque l'on dispose d'informations sur le résultat d'une expérience donnée, il est possible d'affiner nos prédictions.
Soit X un évenement qui n'est pas négligeable, alors on définit l'application:

$$\begin{aligned}\mathbb{P}(\cdot|X) : \mathcal{P}(\Omega) &\longrightarrow [0\,;\,1] \\ A &\longmapsto \frac{\mathbb{P}(A \cap X)}{\mathbb{P}(X)}\end{aligned}$$

On peut montrer que c'est une mesure de probabilité sur Ω et on l'appelle **probabilité de A sachant X**. De la symétrie de l'intersection on peut alors en déduire la **formule de Bayes** qui permet alors **d'inverser le conditionnement**:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A)\mathbb{P}(B|A)}{\mathbb{P}(B)}$$

Formule des probabilités composées

On en déduit directement la **formule des probabilités composées**:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A) = \mathbb{P}(B)\mathbb{P}(A|B)$$

Qui se généralise pour une famille finie d'évenements $(A_n)_{n\in I}$ d'intersection non nulle:

$$\mathbb{P}(A_1 \cap \ldots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}_{A_1}(A_2)\mathbb{P}_{A_1\cap A_2}(A_3)\ldots\mathbb{P}_{A_1\cap\ldots\cap A_{n-1}}(A_n)$$

Formule des probabilités totales

On considère une partition $(A_n)_n$ de Ω en évenements disjoints (on appelle une telle partition **systême complet d'évenements**), alors on peut montrer la **formule des probabilités totales**:

$$\mathbb{P}(B) = \sum_{k=1}^{n} \mathbb{P}(A_k)\mathbb{P}_{A_k}(B)$$

Indépendance

On dit que deux évenements A, B sont **indépendants** si et seulement si la donnée de la réalisation d'un des évenements n'influence pas l'autre, ie:

$$\mathbb{P}(A|B) = \mathbb{P}(A)$$

Ou encore par la formule conditionelle:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

Si deux évenements sont indépendants, alors n'importe quelle paire de $A, B, \overline{A}, \overline{B}$ est indépendante.

XII — Lois usuelles

Dans ce chapitre, on énumère les lois usuelles en probabilité et leurs cas d'utilisation. Comme vu précédement, on distingue le cas discret et absolument continu. Dans tout la suite on considère un espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$.

Lois discretes usuelles

On appelle **épreuve de Bernoulli** est une expérience aléatoire qui n'a que deux issues, usuellement nommées **succés et échec**.

On dit que la loi est **uniforme** si on a la distribution:

$$\forall A \in \mathscr{A} \ ; \ \mathbb{P}(A) = \frac{|A|}{|E|}$$

On dit que la loi est **binomiale** de paramêtres n, p si $\Omega = \{1, \dots, n\}$ et si on a la distribution:

$$\forall k \leq n \ ; \ \mathbb{P}(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}$$

On dit que la loi est **géométrique** de paramêtre p si $\Omega = \mathbb{N}^*$ et si on a la distribution:

$$\forall k \geq 1 \ ; \ \mathbb{P}(\{k\}) = p(1-p)^{k-1}$$

On dit que la loi est **hypergéométrique** de paramêtres (p, n, N) si $\Omega = \mathbb{N}^*$ et si on a la distribution:

$$\forall k \geq 1 \ ; \ \mathbb{P}(\{k\}) = \frac{\binom{pN}{k}\binom{(1-p)N}{n-k}}{\binom{N}{n}}$$

On dit que la loi est **de Poisson** de paramêtres λ si $\Omega = \mathbb{N}^*$ et si on a la distribution:

$$\forall k \geq 1 \ ; \ \mathbb{P}(\{k\}) = \frac{\lambda^k}{k!} e^{-\lambda}$$

La **loi binomiale** est utilisée pour déterminer la probabilité d'obtenir exactement k succés aprés n itérations d'une épreuve de Bernoulli.
La **loi géométrique** est utilisée pour déterminer la probabilité d'un temps d'attente k avant le le premier succés d'une épreuve de Bernoulli.
La **loi hypergéométrique** est utilisée pour déterminer la probabilité d'obtenir k succés aprés n itérations d'une épreuve de tirage sans remise dans une urne contenant N boules, dont pN boules gagantes, et $(1-p)N$ boules perdantes, avec la contrainte que pN soit un entier.
La **loi de Poisson** est utilisée pour déterminer le nombre d'événements se produisant dans un intervalle de temps fixé, si ces événements se produisent avec une fréquence moyenne connue, et indépendamment du temps écoulé depuis l'événement précédent[1].

[1] C'est une loi qui s'obtient asymptotiquement à partir d'une loi binomiale de paramêtres $T, \frac{\lambda}{T}$ en faisant tendre T vers l'infini.

Lois à densité usuelles

On dit que la loi est **uniforme** si sa densité f est constante sur un intervalle $[a\,;\,b]$ et nulle en dehors.
On dit que la loi est **exponentielle** de paramètre λ si on a la densité:

$$f(x) = \lambda \exp(-\lambda x)\ ;\ x \geq 0$$

On dit que la loi est **normale** dee paramêtres[1] μ, σ si on a la densité:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

La **loi exponentielle** est utilisée pour modéliser le temps d'attente d'un phénomène sans mémoire, en particulier, c'est l'analogue continue de la loi géométrique[2].
La **loi normale** est fondamentale en probabilités du fait de son omniprésence dans les sciences expérimentales, en effet, un théorème fondamental montrera que la somme d'une suite de variables aléatoires (comprendre expériences) convergera vers une certaine loi normale. Elle est donc d'importance capitale en statistiques.

[1] Ces paramêtres correspondent alors à l'éspérance et l'écart type de la loi.
[2] Elle s'obtient asymptotiquement à partir d'une loi géométrique de paramètre λT en faisant tendre T vers 0.

XII — Variables aléatoires

Trés souvent, il se trouve que l'espace probabilisé de l'expérience est inconnu, trop grand ou trop complexe, on considérera alors simplement son existence et on étudiera celui ci via des fonctions définies sur cet espace, appelées **variables aléatoires**. Ces fonctions induiront un nouvel espace probabilisé correspondant à notre expérience précise (souvent un espace probabilisé numérique).

On considère alors souvent $(\Omega, \mathscr{A}, \mathbb{P})$ comme un espace probabilisé abstrait et on l'oublie même complétement trés souvent. Par exemple:

- On considère une expérience aléatoire qui tire au hasard un gateau dans une chaîne de fabrication, on peut alors définir une variable aléatoire sur l'espace probabilisé naturellement défini qui à chaque évenement associe le volume du gateau, son taux de sucre, le nombre de raisins secs ... Et faire alors des suppositions sur la loi de ces variables aléatoires par exemple on pourra supposer que le taux de sucre d'un gateau choisi au hasard suit une loi normale.
- Si on considère un groupe de N personnes vivant un épisode épidémique, alors il est trés compliqué de modéliser l'état épidémique du groupe à un instant donné du fait des différentes intéractions et dépendances, on préfère alors éudier des espaces probabilisés plus simples induits par des variables aléatoires comme le nombre de personnes infectées, le temps mis par l'épidémie pour atteindre une certaine taille etc ..

Définition

On dira que X est une **variable aléatoire** de $(\Omega_1, \mathscr{A}, \mathbb{P})$ vers un espace mesurable $(\Omega_2, \mathscr{B})$ si et seulement si c'est une **fonction mesurable** sur cet espace. Elle définit alors une loi naturelle sur Ω_2 définie par la **mesure image**:

$$\begin{aligned} \mathbb{P}_X : \mathscr{B} &\longrightarrow [0\,;\,1] \\ B &\longmapsto \mathbb{P}(X^{-1}(B)) \end{aligned}$$

On note alors plus simplement $\mathbb{P}_X(B) = \mathbb{P}(X \in B)$. Trés souvent, on considérera $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ comme espace d'arrivée et donc la variable aléatoire sera dite **réelle** et définira une loi sur les boréliens.

Cas réel

Dans le cas de variables aléatoires **réelles** on peut aussi créer une notation qui s'applique si B est un intervalle et on a:

$$X^{-1}\big(]a\,;\,b[\big) = \Big\{\omega \in \Omega\,;\ a < X(\omega) < b\Big\} \overset{\text{notation}}{=} (a < X < b)$$

Propriétés

La famille $((X = a))_{a \in X(\Omega)}$ est un **système complet d'évenements**, en effet car si on considère une issue $\omega \in \Omega$, on a:

$$\begin{cases} X(\omega) = x \implies \omega \in (X = x) \\ X(\omega) \neq x \implies \omega \notin (X = x) \end{cases}$$

On peut donc partitionner les éléments de Ω selon leur image par X

On peut aussi noter que si f est une application mesurable, alors $f \circ X$ **est une variable aléatoire** sur les espaces correspondants.

Indépendance

On dira alors que deux variables aléatoires X, Y sont indépendantes si et seulement si pour tout couple x, y, les évenements correspondants sont indépendants:

$$\mathbb{P}(X = x \cap Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$$

Par ailleurs, si X, Y sont deux variables aléatoires, il existe une mesure de la dépendance (corrélation) de deux variables aléatoires appellée **covariance** définie dans la dernière partie.

Cas des vecteurs aléatoires

Dans le cas où la variable aléatoire $X = (X_1, \ldots, X_n)$ est à valeurs dans $\mathbb{R}^n$, alors elle définit une loi produit (appelée dans ce cadre **loi conjointe** de X) sur les boréliens et on l'appelle **vecteur aléatoire**. En particulier les lois marginales sont alors les lois des variables composantes X_i. Par exemple si on prends un vecteur aléatoire choisissant uniformément un point dans $[\![1\ ;\ n]\!]^2$, alors on a que:

$$\mathbb{P}(X = (k, l)) = \frac{1}{n^2}$$

Et les loi marginales sont données par:

$$\mathbb{P}(X_1 = k) = \sum_{i=1}^{n} \mathbb{P}(X = (k, i))$$

On retrouve alors les même lois marginales que dans l'exemple analogue sans variable aléatoire, en particulier les lois conjointe et marginales sont exactement les lois produits et marginales sur $\mathbb{R} \times \mathbb{R}$.

Intégrabilité et formule de transfert

On se donne une variable aléatoire réelle intégrable par rapport à la mesure $\mathbb{P}$ ie telle que:

$$\int_\Omega X(\omega) d\mathbb{P} < \infty$$

Alors on peut montrer l'identité suivante par les propriétés de la mesure image $d\mathbb{P}_X$:

$$\int_\Omega X(\omega) d\mathbb{P} = \int_{\mathbb{R}} x d\mathbb{P}_X$$

Et même plus généralement, on a le **théorème de transfert** pour tout fonction ϕ telle qu'une des intégrale ait un sens:

$$\int_\Omega \phi(X(\omega)) d\mathbb{P} = \int_{\mathbb{R}} \phi(x) d\mathbb{P}_X$$

Et les intégrales du membre de droite se calculent souvent facilement, par exemple dans les deux cas classiques:

- **Cas dénombrable:** On a

$$\int_{\mathbb{R}} x d\mathbb{P}_X = \sum_{\mathbb{N}} \int_{y_n} x d\mathbb{P}_X = \sum_{\mathbb{N}} y_n \mathbb{P}(X = y_n)$$

- **Cas absolument continu:** On a

$$\int_{\mathbb{R}} x d\mathbb{P}_X = \int_{\mathbb{R}} x f(x) dx$$

XII — Indicateurs

Dans tout la suite, on considère $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ un espace probabilisé fini et X, Y deux variables aléatoires **intégrables** pour la mesure $\mathbb{P}$.

On appelle **indicateur de position** un nombre réel permettant de situer les valeurs d'une série statistique, par exemple l'esperance et la médiane sont des indicateurs de position.
On appelle **indicateur de dispersion** un nombre réel permettant de mesurer la variabilité des valeurs d'une série statistique autour d'une valeur (généralement autour de la moyenne), par exemple la variance, l'écart-type ou l'écart interquartile sont des indicateurs de dispersion.

Esperance

L'espérance mathématique correspond à la moyenne théorique du résultat qu'on peut espérer avoir en répétant une expérience aléatoire un grand nombre de fois, c'est **la moyenne des valeurs de la variable aléatoire, pondérées par leur probabilités respectives**, ou c'est aussi le centre de masse de la densité, on définit alors celle ci par:

$$\mathbb{E}(X) := \int_\Omega X d\mathbb{P}$$

L'esperance prends alors la forme d'une somme pondérée dans le cas discret, ou d'une intégrale pondéree dans le cas absolument continue. Elle existe toujours dans le cas d'une variable aléatoire **finie** mais ce n'est pas le cas en général, et il faut alors étudier l'intégrabilité de la variable aléatoire.

L'espérance possède plusieurs propriétés remarquables, elle est **linéaire et croissante** et l'espérance d'une constante est cette constante.
Mais en général, l'espérance **n'est pas multiplicative**, c'est néanmoins le cas quand les deux variables aléatoires sont **indépendantes**.

Variables centrées

On rappelle qu'on a le théorème de transfert donc $\mathbb{E}(\phi(X))$ existe si $\phi(X)$ est intégrable. Aussi, on appelle **variable centrée** une variable aléatoire d'espérance nulle. On peut alors centrer une variable aléatoire par la translation $X' = X - \mathbb{E}(X)$.

Moments d'ordre k

On généralise cette définition et on définit le **moment d'ordre k** de la variable X, si il existe par la quantité suivante:

$$m_k(X) = \mathbb{E}(X^k)$$

On remarque alors que cette quantité existe si et seulement si X^k est intégrable. De manière générale, on comprends facilement qu'on a la propriété suivante (où la disjonction dépends de la discrétude ou non de X):

$$X \text{ admet un moment d'ordre } k \iff X \in L^k(\Omega) \text{ ou } \ell^k(\Omega)$$

Variance

On manque alors d'informations sur les valeurs de X, elles peuvent tout aussi bien rester toujours très proches de $\mathbb{E}(X)$, ou s'en éloigner beaucoup, on a donc besoin de mesurer la distance moyenne entre X et $\mathbb{E}(X)$, qui serait alors $\mathbb{E}(|X - \mathbb{E}(X)|)$.

Cette formule est techniquement impraticable du fait de la valeur absolue, on utile donc **la moyenne des carrés des distances** entre X et $\mathbb{E}(X)$, et on définit alors la variance:

$$\mathbb{V}(X) := \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

La deuxième expression apellée **formule de Koenig-Huygens** se déduit facilement de la première par les propriétés de l'espérance et on en déduit qu'une variable aléatoire admet une variance si et seulement si elle admet un moment d'ordre 2.
On voit directement que la variance est **positive** (ou nulle si X est constante presque partout). Elle vérifie aussi les propriétés suivantes:

- **Invariante par translation** $\mathbb{V}(X + a) = \mathbb{V}(X)$
- **Quadratique** $\mathbb{V}(\lambda X) = \lambda^2 \mathbb{V}(X)$

Ecart type

On peut alors définir **l'écart-type** de la variable X qui est défini par $\sigma(X) = \sqrt{\mathbb{V}(X)}$

Enfin, on appelle **variable réduite** une variable aléatoire d'écart type 1 et on peut alors définir la **variable centrée réduite** associée à X:

$$X^* = \frac{X - \mathbb{E}(X)}{\sigma(X)}$$

Covariance

Dans le cas d'un couple aléatoire et contrairement à l'espérance, le concept de variance perds son sens. Plutôt que de chercher un écart par rapport à la moyenne, on va préférer chercher **un écart moyen entre les deux variables**[1] qui se définit par:

$$\mathrm{Cov}(X, Y) = \mathbb{E}\big[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\big] = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

Gràce à la covariance on peut aussi définir la variance d'une somme:

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\mathrm{Cov}(X, Y)$$

La covariance de X, Y n'existe bien sûr que si X, Y et XY sont intégrables.

Propriétés de la covariance

La covariance vérifie plusieurs propriétés intéressantes:

- Si les deux variables aléatoires sont indépendantes, alors on a $\mathrm{Cov}(X, Y) = 0$, la reciproque étant **fausse**.
- Si X admet une variance alors on retrouve celle ci comme $\mathrm{Cov}(X, X)$

On peut même montrer que la covariance est **bilinéaire, symétrique et positive**. Informellement c'est un *"pseudo produit scalaire"* sur les variables aléatoires, néanmoins suffisament proche du produit scalaire pour avoir **l'inégalité de Cauchy-Schwartz**:

$$|\mathrm{Cov}(X, Y)| \leq \sigma(X)\sigma(Y)$$

Plus précisément, considérons l'espace des variables aléatoires centrées réduites, alors c'est **un espace pre-hilbertien**, et la covariance définit son produit scalaire, l'écart type définit alors la **norme** associée et on peut définir le coefficient de corrélaition linéaire par:

$$\varrho_{X,Y} = \frac{\mathrm{Cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

Qui s'intérpréterait alors comme *l'angle* entre les variable aléatoires.

[1] On remarque que la variance n'est alors que la covariance de X avec elle-même. Aussi si les deux variables aléatoires sont indépendantes, dans ce cas l'espérance est multiplicative et on a $\mathrm{Cov}(X, Y) = 0$, la reciproque étant **fausse**.

Espérance & variances usuelles

Il est intéressant de considérer les différents indicateurs de position et de dispersion des lois usuelles[1]

Lois	Espérance	Variance
$X \sim \mathcal{U}(E)$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$
$X \sim \mathcal{B}(n,p)$	$n \cdot p$	$n \cdot p(1-p)$
$X \sim \mathcal{G}(p)$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
$X \sim \mathcal{H}(p,n,N)$	$n \cdot p$	$n \cdot p(1-p)\frac{N-n}{N-1}$

Inégalités

On souhaite majorer la probabilité d'avoir des valeurs "extrèmes", ie éloignées de l'espérance, alors si X est une variable aléatoire positive presque partout et $\alpha \in \mathbb{R}^{+*}$ on peut montrer **l'inégalité de Markov**:

$$\mathbb{P}(X \geq \alpha) \leq \frac{\mathbb{E}(X)}{\alpha}$$

Si la variable aléatoire admet une variance, on a alors **l'inégalité de Bienaymé-Tchebychev**:

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \alpha) \leq \frac{\mathbb{V}(X)}{\alpha^2}$$

Cette dernière est un cas particulier de la première mais est en général *plus fine* que la majoration donnée par l'inégalité de Markov.

[1] Pour la loi uniforme on considère X à valeurs dans $[\![1\ ;\ n]\!]$

XII — Convergences Stochastiques

On peut alors tenter d'appliquer les résultat sur les espaces L^p à la théorie des probabilité, en particulier étudier des intégrales de variables aléatoires, des espérances, etc .. revient à étudier la finitude d'un norme dans $L^p(\Omega)$ en particulier pour tout $p \in [1\,;\infty]$, on en déduis que les normes p s'appliquent aux variables aléatoires, ie on a:

$$\begin{cases} \|X\|_p := \left(\int_\Omega |X|^p d\mathbb{P}\right)^{\frac{1}{p}} = \left(\mathbb{E}(|X|^p)\right)^{\frac{1}{p}} \\ \|X\|_\infty := \sup \operatorname{ess}\{|X|\} \end{cases}$$

Où ici X est supposée à densité ou à distribution $f(x)$.

Cette approche sera alors trés fructueuse, en effet par l'étude et la définition de différent *modes de convergences* bien choisis sur des suites de variables aléatoires (X_n), on pourra alors démontrer les grands théorèmes probabilistes.

Convergence en loi

On dira que X_n **converge en loi** vers X si et seulement si pour tout $x \in \mathbb{R}$, la loi de X_n est abitrairement proche de la loi de X, ie si on a:

$$\mathbb{P}(X_n \leq x) \longrightarrow \mathbb{P}(X \leq x)$$

On notera alors:

$$(X_n) \xrightarrow{\mathcal{L}} X$$

Convergence en probabilité

On dira que X_n **converge en probabilité** vers X si et seulement si pour tout $\varepsilon > 0$, la probabilité que $(X_n)_n$ s'éloigne de X tends vers 0, ie si on a:

$$\mathbb{P}(|X_n - X| > \varepsilon) \longrightarrow 0$$

On notera alors:

$$(X_n) \xrightarrow{\mathcal{P}} X$$

Convergence presque partout

On dira que X_n **converge presque partout** vers X si et seulement si elle converge sauf sur un domaine de mesure nul, ie:

$$\mathbb{P}\left(\left\{\omega \in \Omega\,;\ \lim_{n\to+\infty} X_n(\omega) \neq X(\omega)\right\}\right) = 0$$

On notera alors:

$$(X_n) \longrightarrow X\ p.p.$$

Convergence L^p

On dira qu'une suite (X_n) converge en norme p vers une variable aléatoire X si et seulement si:

$$\lim_{n\to\infty} \|X_n - X\|_p = \lim_{n\to\infty} \left(\mathbb{E}(|X_n - X|^p)\right)^{\frac{1}{p}} = 0$$

Relations entre les convergences

On peut montrer les différentes implications suivantes:

$$\textbf{Convergence p.p.} \implies \textbf{Convergence en probabilité} \implies \textbf{Convergence en loi}$$

Et pour $p > q \geq 1$

$$\textbf{Convergence } L^p \implies \textbf{Convergence } L^q \implies \textbf{Convergence en probabilité}$$

XII — Théorèmes limites

Munis de nos nouveaux outils et modes de convergences des suites de variables aléatoires, on peut alors énoncer et démontrer les grands théorèmes probabilistes.

Lois des grands nombres

On se donne une suite de variables aléatoires indépendantes et identiquement distribuées (communément noté iid) (X_n) admettant une espérance μ, et on pose une variable aléatoire appellée **moyenne**[1] **empirique**:

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$$

Alors on peut montrer la loi **faible** des grands nombre, ie:

$$\overline{X}_n \xrightarrow{\mathcal{P}} \mu$$

Pour les memes hypothèses que précédemment la loi **forte** des grands nombres nous assure une convergence plus forte, ie on a:

$$\overline{X}_n \xrightarrow{p.s.} \mu$$

Théorème central limite

On se donne une suite de variables aléatoires iid (X_n) admettant une espérance μ et un écart-type σ finis, alors on considère à nouveau la moyenne empirique:

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$$

Mais cette fois on considère cette variable sous sa forme **centrée réduite**, qu'on notera $\overline{X}^*_n$, alors le théorème central limite nous donne que:

$$\overline{X}^*_n \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

[1] Elle correspond à la moyenne faite sur les réalisations d'une expérience par exemple.