****

0- Ensembles, applications et logique

On sera amené à manipuler des ensembles vu comme "événements" d'où les quelques rappels utiles suivants.

1- Exemples d'ensembles et notations. — $\mathbb{N}$ est l'ensemble des entiers naturels, $\mathbb{R}$ celui des nombres réels, $\mathbb{C}$ celui des nombres complexes. Un seul ensemble a la propriété de ne pas avoir d'éléments, on l'appelle *l'ensemble vide* et il est noté $\emptyset$. Des ensembles peuvent être décrits par une liste, par exemple, il est raisonnable de noter $L = \{a, b, c, \cdots, z\}$ l'ensemble des lettres de l'alphabet (latin) et l'on notera couramment $[\![m, n]\!] = \{m, m+1, \cdots, n\}$ l'ensemble des entiers entre m et n.

2- Opérations. — A et B étant deux ensembles, *l'intersection* notée $A \cap B = \{x | x \in A \text{ et } x \in B\}$ et *la réunion* notée $A \cup B = \{x | x \in A \text{ ou } x \in B\}$. Il est particulièrement important de ne pas faire de confusion entre "et" et "ou".

De manière générale si $(A_i)_{i\in I}$ est une famille d'ensembles indexée par I (on a fréquemment $I = [\![1, n]\!]$ ou $I = \mathbb{N}$) on note :

$\cap_{i\in I} A_i = \{x | x \in A_i,\ \forall i \in I\}$ *l'intersection des* A_i et $\cup_{i\in I} A_i = \{x | x \in A_i |\ \exists i \in I\}$ *la réunion des* A_i.

Exemples : si A est l'ensemble des entiers pairs et B l'ensemble des multiples de 3 alors $A \cap B$ est l'ensemble des multiples de 6. Si $A_n =]0, 1/n + 1[\subset \mathbb{R}$ et $B_n =]0, n[\subset \mathbb{R}$ alors $\cap_{n\in\mathbb{N}} A_n = \emptyset$ et $\cup_{n\in\mathbb{N}} B_n =]0, \infty[$.

Propriété : si A, B, C sont trois ensembles $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

2- Inclusion, passage au complémentaire. — L'inclusion de A dans B est notée $A \subset B$. On a $A \subset B \iff x \in A \Rightarrow x \in B$ d'où le *principe de double inclusion* : $A \subset B$ et $B \subset A \iff A = B$.

Si $A \subset E$ le *complémentaire* de A dans E noté $E \setminus A = \{x \in E | x \notin A\}$ et quand il n'y a de confusion possible c'est à dire si tous les ensembles considérés sont des parties d'un même ensemble E, la partie $E \setminus A$ est notée plus simplement A^c.

Propriété : si A, B sont deux ensembles inclus dans E, on a

a) $(A^c)^c = A$, *b)* $(A \cap B)^c = A^c \cup B^c$ et *c)* $(A \cup B)^c = A^c \cap B^c$.

et de manière générale si $(A_i)_{i\in I}$ est une famille d'ensembles inclus dans E indexée par I on a :

$(\cap_{i\in I} A_i)^c = \cup_{i\in I} A_i^c$ et $(\cup_{i\in I} A_i)^c = \cap_{i\in I} A_i^c$.

3- Relation avec la logique. — Soient P et Q des propriétés à un argument c'est à dire dont on peut tester la validité sur un seul objet à la fois. Par exemple, P pourrait être la propriété d'être un nombre entier pair. Soit E un ensemble, alors $A = \{x \in E | x \text{ satisfait } P\}$ et $B = \{x \in E | x \text{ satisfait } Q\}$ sont des parties bien définies de E et on a les :

Propriétés :

a) $(P \Rightarrow Q) \iff A \subset B$.

b) $(P \iff Q) \iff A = B$.

c) P et Q sont incompatibles $\iff A \cap B = \emptyset$.

d) $A^c = \{x \in E | x \text{ satisfait nonP}\}$

e) La contraposition $(P \Rightarrow Q) \iff (nonQ \Rightarrow nonP)$ équivaut à $A \subset B \iff B^c \subset A^c$.

4- Applications entre ensembles. — Soit $f : E \to F$ l'application qui à $x \in E$ fait correspondre $f(x) \in F$.

Image directe : Si $A \subset E$ on note $f(A) = \{y \in F | \exists x \in A, y = f(x)\}$ *l'image (directe)* de A par f.
Propriétés :

a) $A \subset B \Rightarrow f(A) \subset f(B)$

b) $f(A \cap B) \subset f(A) \cap f(B)$ mais il n'y a pas toujours égalité.

c) $f(A \cup B) = f(A) \cup f(B)$.

Image réciproque : Si $A \subset F$ on note $f^{-1}(A) = \{x \in E | f(x) \in A\}$ *l'image réciproque* de A par f.

Propriétés :

a) $A \subset B \Rightarrow f^{-1}(A) \subset f^{-1}(B)$

b) $f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B)$.

c) $f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B)$.

Propriétés possibles de l'application f :

a) f injective $\iff$ $f(x) = f(x') \Rightarrow x = x'$

b) f surjective $\iff$ $f(E) = F$.

c) f bijective $\iff$ f est injective et surjective.

1- Dénombrements

1- Equipotence. — Pour savoir s'il y a autant de garçons que de filles dans une classe, on peut compter les deux groupes séparément et comparer les résultats. On peut aussi demander aux étudiants de se mettre par couple (mixte) et on se rendra alors compte s'il y a égalité ou non. En termes mathématiques, on cherche à construire une *bijection* entre l'ensemble F des filles et l'ensemble G des garçons.

Définition. — *Deux ensembles F et G sont équipotents si et seulement si il existe une bijection de F dans G.*

Un ensemble E est dit *fini* s'il est en bijection avec $[\![1, n]\!]$ pour un certain $n \in \mathbb{N}$. L'entier n est le *cardinal* ou le *nombre d'éléments* de E noté $\operatorname{Card} E$ ou plus commodément $|E|$. Clairement, deux ensembles finis E et F sont équipotents si et seulement si $\operatorname{Card} E = \operatorname{Card} F$. On a aussi la proposition qui serait encore vraie sans supposer les ensembles finis :

Proposition. — *Deux ensembles finis E et F sont équipotents si et seulement si il existe une injection $f : E \to F$ et une surjection $g : E \to F$.*

Remarque. — Historiquement, le mot "Cardinal" a été employé comme synonyme du mot "Puissance" qui est encore utilisé pour parler du cardinal d'un ensemble infini. Ceci explique le terme "équipotent"= "qui a même puissance" = "qui a même cardinal".

A partir de maintenant pour un ensemble fini E on va utiliser la notation $|E|$ déjà indiquée de préférence à $\operatorname{Card} E$.

2- Réunion disjointe. — Si A et B sont des ensembles *disjoints* (c'est-à-dire tels que $A \cap B = \emptyset$), alors $|A \cup B| = |A| + |B|$. Plus généralement, si $A_1, \ldots, A_n$ sont des ensembles finis *deux à deux disjoints*, alors $|A_1 \cup \ldots \cup A_n| = \sum_{k=1}^{n} |A_k|$ (Principe d'addition).

Notation : si $A_1, \ldots, A_n$ sont des ensembles deux à deux disjoints, dans ce cas et *uniquement* dans ce cas, on notera $A_1 + \ldots + A_n = \sum_{k=1}^{n} A_k$ la réunion des A_i. Avec cette notation conventionnelle, l'égalité précédente s'écrit $|\sum_{k=1}^{n} A_k| = \sum_{k=1}^{n} |A_k|$.

3- Réunion quelconque. — Si A, B, C sont des ensembles finis, alors :

$$|A \cup B| = |A| + |B| - |A \cap B| \quad (*)$$

formule déjà vue dans le secondaire dont on déduit :

$|A \cup B \cup C| = |A \cup B| + |C| - |(A \cup B) \cap C| = |A| + |B| + |C| - |A \cap B| - |B \cap C| - |A \cap C| + |A \cap B \cap C|$.

Plus généralement, par récurrence et toujours grâce à la formule $(*)$, si $A_1, \ldots, A_n$ sont des ensembles finis, on a la *formule de Poincaré* :

$$\begin{aligned} |A_1 \cup \ldots \cup A_n| &= \sum_{p=1}^{n} (-1)^{p-1} \sum_{1 \leq k_1 < \ldots < k_p \leq n} |A_{k_1} \cap \ldots \cap A_{k_p}| \\ &= \sum_{k=1}^{n} |A_k| - \sum_{1 \leq i < j \leq n} |A_i \cap A_j| + \cdots + (-1)^{n-1} |A_1 \cap \ldots \cap A_n|. \end{aligned}$$

4- Principe des bergers. — *Dans un troupeau de moutons, il y a quatre fois plus de pattes que de moutons.* Voici pourquoi : 1) chaque mouton a quatre pattes 2) chaque patte appartient à un seul mouton.

Numérotons les moutons de 1 à m et notons E_k l'ensemble des pattes du $k^{\text{ième}}$ mouton. Puisque les $E_1, \ldots, E_m$ sont disjoints deux à deux, le nombre total de pattes est : $|E_1 + \ldots + E_m| = \sum_{k=1}^{m} |E_k| = \sum_{k=1}^{m} 4 = 4m$.

Formulation en termes de choix. — Pour choisir une patte, on commence par choisir un mouton puis on choisit une de ses pattes. Le second choix n'est pas indépendant du premier, mais le *nombre de possibilités* pour le second choix est indépendant du premier choix effectué.

Ce ne serait plus le cas du même troupeau auquel on aurait ajouté n individus à 5 pattes numérotés de 1 à n. En effet, si l'on note alors F_k l'ensemble des pattes du $k^{\text{ième}}$ mouton à 5 pattes, le nombre total de pattes serait cette fois ci : $|E_1 + \ldots + E_m + F_1 + \ldots + F_n| = \sum_{k=1}^{m} |E_k| + \sum_{k=1}^{n} |F_k| = \sum_{k=1}^{m} 4 + \sum_{k=1}^{n} 5 = 4m + 5n$.

5- Produit cartésien. — Le produit cartésien de deux ensembles A et B est l'ensemble $A \times B = \{(a, b) | a \in A, b \in B\}$ autrement dit l'ensemble des *couples* (a, b) d'éléments de A et de B.

Propriété : *Si A et B sont finis, on a* $|A \times B| = |A| \times |B|$. (Principe de multiplication)

En effet, pour tout $b \in B$, notons $E_b = A \times \{b\} = \{(a, b) | a \in A\}$. On a $A \times B = \cup_{b \in B} E_b$ et donc, comme les E_b sont disjoints et ont tous $|A|$ éléments, il vient : $|A \times B| = |\sum_{b \in B} E_b| = \sum_{b \in B} |E_b| = \sum_{b \in B} |A| = |A| \times |B|$.

Exemple. — Dans un jeu de cartes ordinaire, une carte peut être vue comme un couple $(a, b) \in A \times B$ où :

$A = \{1, 2, \ldots, 10, V, D, R\}$ (ensemble des hauteurs) et $B = \{\clubsuit, \diamondsuit, \heartsuit, \spadesuit\}$ (ensemble des couleurs)

Il y a $13 \times 4 = 52$ cartes.

Formulation en termes de choix. — Choisir une carte revient à choisir indépendamment : une hauteur parmi 13 et une couleur parmi 4. Le nombre de choix possibles est donc 13×4.

6- Nombre d'applications. — Soient E et F deux ensembles finis, de cardinaux respectifs p et n. On note F^E l'ensemble des applications de E dans F. Alors : $\text{Card}\, F^E = (\text{Card}\, F)^{\text{Card}\, E} = n^p$.

En effet, soient $x_1, \ldots, x_p$ les p éléments de E. Choisir une application $f : E \to F$ équivaut à choisir le p-uplet $(f(x_1), \ldots, f(x_p))$ arbitrairement dans $F^p = F \times \ldots \times F$ (produit cartésien de F par lui-même p fois) ou encore à choisir indépendamment les images $f(x_1), \ldots, f(x_p)$. Donc $\text{Card}\, F^E = \text{Card}\,(F^p) = n \times \ldots \times n = n^p$.

Exemple. — *Tirages successifs avec remise* : Une urne contient n boules. On effectue p tirages successifs, en remettant la boule dans l'urne après chaque tirage. Un tirage peut être vu comme un p-uplet d'éléments de F (où F est l'ensemble des boules) $(y_1, \ldots, y_p)$ où y_k est la $k^{\text{ième}}$boule tirée. Il y a n^p tirages possibles.

7- Nombre d'arrangements. — Que se passe-t-il si on effectue p *tirages successifs sans remise*? Il devient alors impossible de tirer deux fois de suite la même boule. Le résultat de l'expérience peut alors être vu comme un p-uplet d'éléments de F *distincts* ou encore comme une *injection* de $\{1, 2, \ldots, p\}$ dans F.

Les résultats de chaque tirage ne sont pas indépendants.

- si $p > n$, il n'est pas possible d'effectuer p tirages sans remise.
- si $p \leq n$, il y a n choix possibles au 1^{er}tirage, $n-1$ choix possibles au $2^{\text{ième}}$tirage, ..., $n-p+1$ choix possibles au p-ième tirage.

En appliquant successivement le principe des bergers, il y a donc $n(n-1)\ldots(n-p+1)$ façons d'effectuer p tirages successifs sans remise dans une urne contenant initialement n boules.

DÉFINITION. — *On appelle arrangement de p éléments pris dans F tout p-uplet formé d'éléments distincts de F.*

PROPRIÉTÉ. — *Le nombre d'arrangements de p éléments pris dans un ensemble à n éléments est* $A_n^p = n(n-1)\ldots(n-p+1)$. *Si $p \leq n$,* $A_n^p = \dfrac{n!}{(n-p)!}$ *et si $p > n$,* $A_n^p = 0$.

Cas particulier où $p = n$. — On tire alors toutes les boules et seul l'ordre dans lequel on les tire compte. Dans ce cas, on compte en fait le nombre d'*injections* de $\{1, 2, \ldots, n\}$ dans F. Mais comme une *injection* entre ensembles finis ayant même nombre d'éléments n est une *bijection*, on compte le nombre de *bijections* de $\{1, 2, \ldots, n\}$ dans F. Ce nombre est $A_n^n = n!$ qui est aussi le nombre de *permutations* de l'ensemble F c'est-à-dire le cardinal de l'ensemble $\mathcal{S}(F)$ de toutes les bijections (ou permutations) de F. L'ensemble $\mathcal{S}(F)$ a une structure de groupe pour la composition des bijections qui le constituent, on l'appelle le *groupe symétrique* de F.

8- Nombre de combinaisons. — Si on tire *simultanément* p boules d'une urne contenant n boules, le résultat est une partie ou *combinaison* de p éléments de F où F est l'ensemble des boules. Le nombre de ces combinaisons est noté $\binom{n}{p}$ (ou C_n^p).

Pour choisir une combinaison (*i.e.* une partie) à p éléments de F, on peut effectuer p tirages successifs sans remise puis oublier l'ordre d'apparition des boules tirées. Il y a alors $p!$ façons d'obtenir une combinaison donnée. En appliquant le principe des bergers à l'envers, on obtient $A_n^p = p!\binom{n}{p}$, d'où : $\binom{n}{p} = \dfrac{n(n-1)\dots(n-p+1)}{p(p-1)\dots 1}$ et $\binom{n}{p} = \dfrac{n!}{(n-p)!p!}$ si $p \leq n$ et $\binom{n}{p} = 0$ si $p > n$.

Pour effectuer p *tirages successifs sans remise*, on peut procéder de la façon suivante :

– on effectue le tirage *simultané* de p boules et on met les boules selectionnées dans une *autre* urne : $\binom{n}{p}$ possibilités.

– puis on effectue dans cette nouvelle urne p tirages *successifs* sans remise des p boules : $p!$ possibilités.

D'après le principe des bergers, on a donc $A_n^p = \binom{n}{p} \times p!$

Propriétés des $\binom{n}{p}$. — Fixons un élément b de F et séparons les parties de F à p éléments en deux catégories disjointes :

– celles qui ne contiennent pas b; ce sont les parties à p éléments de $F \setminus \{b\}$ et il y en a $\binom{n-1}{p}$.

– celles qui contiennent b; il y en a autant que de parties à $p-1$ éléments de $F \setminus \{b\}$ c'est-à-dire $\binom{n-1}{p-1}$.

On a ainsi $\binom{n}{p} = \binom{n-1}{p} + \binom{n-1}{p-1}$. Cette formule de récurrence permet de calculer les coefficients $\binom{n}{p}$ grâce au triangle de Pascal.

• Pour tout p on a $\binom{n}{n-p} = \binom{n}{p}$

• Formule du binôme (c'est pour cela que les $\binom{n}{p}$ sont aussi appelés *coefficients du binôme*) :

Pour a, b réels ou complexes et $n \in \mathbb{N}$

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k$$

que l'on démontre par récurrence sur n.

9- Nombre de parties d'un ensemble. — Soit F un ensemble à n éléments. Alors F possède C_n^p parties à p éléments pour tout $p \in \{0, \dots, n\}$. Le nombre total de parties de F est donc

$$\sum_{p=0}^{n} \binom{n}{p} = \sum_{p=0}^{n} C_n^p \, 1^p \, 1^{n-p} = (1+1)^n.$$

On peut montrer ce résultat plus simplement. A toute partie A de F associons sa *fonction indicatrice* (*i.e.* qui répond par "oui" (=1) ou "non" (=0) à la question " x est-il dans A?") :
$\mathbb{1}_A : F \to \{0,1\}$ est définie par : $\mathbb{1}_A(x) = 1$ si $x \in A$ et $\mathbb{1}_A(x) = 0$ si $x \notin A$. Alors l'application $A \mapsto \mathbb{1}_A$ est une bijection de l'ensemble des parties $\mathcal{P}(F)$ dans $\{0,1\}^F$. Donc $\operatorname{Card} \mathcal{P}(F) = \operatorname{Card} \{0,1\}^F = 2^n$

Résumé. — Si on tire p boules d'urne contenant n boules

– successivement avec remise : n^p p-uplets.

– successivement sans remise : A_n^p arrangements.

– simultanément : $\binom{n}{p}$ combinaisons.

2- Définition d'un espace probabilisé $(\Omega, \mathcal{T}, P)$ **fini**. — Soit Ω un ensemble fini et $\mathcal{T} \subset \mathcal{P}(\Omega)$ un ensemble de parties de Ω.

Tribu : On dit que $\mathcal{T}$ est une *tribu* si $\mathcal{T}$ vérifie les trois propriétés

(i) $\Omega \in \mathcal{T}$

(ii) Si $A \in \mathcal{T}$ alors $A^c \in \mathcal{T}$

(iii) Pour toute *suite* $A_n \in \mathcal{T}$ on a $\cup_{n \in \mathbb{N}} A_n \in \mathcal{T}$ (et $\cap_{n \in \mathbb{N}} A_n \in \mathcal{T}$ grâce à (ii)).

Dans le cas qui nous intéresse pour l'instant où Ω est fini, il suffit de vérifier la propriété (iii) pour les suites *finies* c'est-à-dire de vérifier que $\mathcal{T}$ est une *algèbre* de parties.

L'exemple fondamental et le plus utilisé est la tribu $\mathcal{T} = \mathcal{P}(\Omega)$ de toutes les parties de Ω.

Image réciproque d'une tribu : Si $\mathcal{T}$ est une tribu de parties de Ω et $f : \Omega' \to \Omega$ une application alors $\mathcal{T}' = f^{-1}(\mathcal{T}) = \{f^{-1}(A) | A \in \mathcal{T}\}$ est une tribu de parties de Ω' appelée *tribu image réciproque* par f de la tribu $\mathcal{T}$.

Une tribu est définie par une partition : Une suite de parties $\Omega_1, \Omega_2, \ldots, \Omega_k$ de Ω est une *partition de* Ω si $\Omega = \Omega_1 + \Omega_2, + \cdots + \Omega_k$ (Ω est réunion disjointe des $\Omega_1, \Omega_2, \ldots, \Omega_k$).

On peut vérifier que, étant donné une telle partition, l'ensemble de parties $\mathcal{T} = \{\sum_1^s \Omega_{i_l}, 1 \leq i_1 < \ldots < i_s \leq k\}$ est une tribu qu'on dit *engendrée* par la partition $\Omega_1, \Omega_2, \ldots, \Omega_k$. On peut observer que si l'on définit l'application $f : \Omega \to [\![1, k]\!]$ par $f(\omega) = i$ si $\omega \in \Omega_i$, on a $\mathcal{T} = f^{-1}(\mathcal{P}([\![1, k]\!])$.

Réciproquement, il est facile de vérifier que si $\mathcal{T}$ est une est une tribu de parties de l'ensemble fini Ω, les parties $A_\omega = \cap_{\omega \in A, A \in \mathcal{T}} A$ avec $\omega \in \Omega$ sont des éléments de $\mathcal{T}$ formant une partition engendrant la tribu $\mathcal{T}$.

Espace probabilisable : Une tribu $\mathcal{T}$ de parties de Ω étant fixée, le couple $(\Omega, \mathcal{T})$ est un *espace probabilisable*. Les éléments de $\mathcal{T}$ en sont les *événements*. Si $\mathcal{T}$ est engendrée par la partition $\Omega_1, \Omega_2, \ldots, \Omega_k$ de Ω, les Ω_i sont appelés *événements élémentaires*. Dans le cas fréquent où $\mathcal{T} = \mathcal{P}(\Omega)$, les événements élémentaires sont les singletons $\{\omega\}$ où $\omega \in \Omega$. On a aussi les

DÉFINITIONS. — *Si* $(\Omega, \mathcal{T})$ *est un espace probabilisable fini, alors*

a) Ω *est l'événement certain;* $\emptyset$ *est l'événement impossible.*

b) A^c *(le complémentaire de* A *dans* Ω*) est l'événement contraire de* A*.*

c) Deux événements A *et* B *sont dits incompatibles lorsque* $A \cap B = \emptyset$*.*

d) Si $\omega \in A$ *on dit que l'éventualité (ou le résultat d'expérience)* ω *réalise l'événement* A*.*

Espace probabilisé : Une probabilité P sur l'espace probabilisable $(\Omega, \mathcal{T})$ est une application $P : \mathcal{T} \to [0, 1]$ vérifiant

(i) $P(\Omega) = 1$

(ii) Si $A, B \in \mathcal{T}$ sont deux événements *disjoints* on a $P(A + B) = P(A) + P(B)$.

On dit que $(\Omega, \mathcal{T}, P)$ est un *espace probabilisé fini*. Dans le cas le plus habituel où $\mathcal{T} = \mathcal{P}(\Omega)$, on omettra l'indication de la tribu et on notera de manière concise l'espace probabilisé (Ω, P).

Exemple. — L'application $P_0 : \mathcal{P}(\Omega) \to \mathbb{R}^+$ définie par $P_0(A) = |A|/|\Omega|$ est une probabilité sur Ω appelée *probabilité uniforme* sur Ω (on dit aussi qu'on est alors dans le cas d'équiprobabilité)

PROPRIÉTÉS. — *a)* $P(\emptyset) = 0$*;* $P(A^c) = 1 - P(A)$ *et donc* $P(A) \leq 1$*.*

b) Si $A \subset B$ *alors* $P(B) = P(A) + P(B \setminus A) \geq P(A)$

c) $P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B)$

d) Soit un entier $n > 0$ *et* $A_1, \ldots, A_n$ *des événements, on a* $P(A_1 \cup \ldots \cup A_n) \leq \sum_{k=1}^n P(A_k)$ *et il y a égalité lorsque les événements sont deux à deux disjoints* $(P(A_1 + \ldots + A_n) = P(A_1) + \ldots + P(A_n))$*.*

e) On a la formule de Poincaré $P(A_1 \cup \ldots \cup A_n) = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \ldots < i_k \leq n} P(A_{i_1} \cap \ldots \cap A_{i_k})$*.*

Si on note $\Omega_1, \Omega_2, \ldots, \Omega_k$ les événements élémentaires (qui engendrent $\mathcal{T}$) on a le

COROLLAIRE. — *Pour tout événement* A*,* $P(A) = \sum_{\Omega_l \subset A} P(\{\Omega_l\})$*.*

On voit donc qu'une probabilité sur un ensemble fini est entièrement déterminée par la probabilité des événements élémentaires.

Inversement, une suite $(p_l)_{1 \leq l \leq k}$ de réels *positifs de somme 1* détermine une (et une seule) probabilité telle que $P(\{\Omega_l\}) = p_l$ par la formule $P(A) = \sum_{\Omega_l \subset A} p_l$.

A partir de maintenant, on considèrera, sauf mention explicite du contraire, le cas où la tribu d'événements est $\mathcal{P}(\Omega)$ et les événements élémentaires les singletons $\{\omega\}$ et les probabilités élémentaires $P(\{\omega\}) = p_\omega, \omega \in \Omega$.

Exemple. — Un dé pipé, $\Omega = \{1,2,3,4,5,6\}$ et (par exemple) $p_1 = p_2 = p_3 = 1/4$ et $p_4 = p_5 = p_6 = 1/12$.

Le cas de l'équiprobabilité. — Il y a *équiprobabilité* lorsque $P(\{\omega\})$ ne dépend pas de $\omega \in \Omega$. Dans ce cas, pour tout $\omega \in \Omega$ on a $P(\{\omega\}) = 1/|\Omega|$ et P est la *probabilité uniforme* sur Ω.

Modélisation. — Une expérience aléatoire (de alea = nom latin féminin qui signifie dé, sort ou hasard) est une expérience dont on ne peut prévoir le résultat pour diverses raisons (manque d'informations, extrême dépendance par rapport aux conditions initiales). Lorsqu'une expérience a un ensemble fini de résultats possibles Ω, on peut tenter une modélisation, c'est-à-dire tenter de justifier l'existence d'une probabilité P qui semble bien rendre compte des chances de réalisations de résultats élémentaires $\omega \in \Omega$.

Exemple du jeu de l'oie. — Dans ce jeu, on lance 2 dés et on avance d'un nombre de cases égal à la somme des deux chiffres obtenus. L'espace des résultats possibles de l'expérience aléatoire est donc $\Omega = \{2, 3, \dots, 12\}$. Il reste à déterminer $P(\{\omega\})$ pour $\omega \in \Omega$. La seule chose que l'on puisse dire avec conviction, c'est que si les dés ne sont pas pipés (ce que l'on suppose), l'espace $\Omega' = \{1, \dots 6\}^2$ des couples (i,j) de chiffres obtenus sur le 1er et le 2ème dé est pourvu de la probabilité uniforme $P_0(\{(i,j)\}) = 1/36$. Et donc si $\omega \in \Omega$, l'événement élémentaire "obtenir la somme ω" est le même que l'événement E_ω de Ω' suivant : $E_\omega = \{(i,j) | i + j = \omega, 1 \leq i, j \leq 6\}$ et on vérifie facilement que $P_0(E_{14-\omega}) = P_0(E_\omega) = (\omega - 1)/36$ si $\omega \leq 7$. Par conséquent il est entièrement justifié que la probabilité P à prendre sur Ω est donnée par : $P(\{\omega\}) = P_0(E_\omega)$ comme ci-dessus.

3- Variables aléatoires sur un espace probabilisé fini. — On se place sur un espace probabilisé fini $(\Omega, \mathcal{T}, P)$. Intuitivement, une variable aléatoire (abréviation v.a.) est une grandeur, un caractère ... dépendant du hasard.

DÉFINITION. — *Soit un ensemble $E \neq \emptyset$. Une variable aléatoire X (sur $(\Omega, \mathcal{T}, P)$) à valeurs dans E est une application de Ω dans E telle que pour tout sous-ensemble $A \subset E$ on a $X^{-1}(A) \in \mathcal{T}$. Quand la tribu $\mathcal{T} = \mathcal{P}(\Omega)$, cette condition est toujours remplie. Dans le cas fréquent où $E \subset \mathbb{R}$, on dit que X est une variable aléatoire réelle (v.a.r.).*

Les événements du type $X^{-1}(A)$ (noté aussi $[X \in A]$) forment une sous-tribu $\mathcal{T}'$ de $\mathcal{T}$ constituée par les événements descriptibles à l'aide de la seule variable X ou encore *les événements relatifs à la variable X*.

Exemple. — Soit Ω une population de souris que l'on munit de la probabilité uniforme P_0 ce qui correspond à l'expérience aléatoire consistant à tirer au hasard (de manire équiprobable) une souris dans la population Ω. Si ω est une des souris de Ω, on définit $N(\omega) \in \mathbb{N}$ comme étant le nombre de souris nées le même jour que la souris ω; $S(\omega) \in \{F, M\}$ son sexe et $T(\omega) \in \mathbb{R}^+$ son poids à la naissance en grammes. Les applications ainsi définies sont des variables aléatoires.

Loi d'une variable aléatoire. — Si X est une v.a. sur $(\Omega, \mathcal{T}, P)$ à valeurs dans E, alors X ne peut prendre qu'un nombre fini de valeurs différentes puisque Ω est fini ($|X(\Omega)| \leq |\Omega| < \infty$). Quitte à restreindre l'ensemble d'arrivée E (c'est-à-dire remplacer E par $X(\Omega)$), on peut supposer que E est fini. Pour toute partie A de E, l'image réciproque de A par X c'est-à-dire $X^{-1}(A) = \{\omega \in \Omega | X(\omega) \in A\} \in \mathcal{T}$ est un événement noté souvent $[X \in A]$.

On pose $P_X(A) = P(X^{-1}(A)) = P[X \in A]$ qu'on lit "probabilité que X appartienne à A".

PROPOSITION. — *L'application $P_X : \mathcal{P}(E) \to \mathbb{R}^+$ ainsi définie est une probabilité sur E appelée probabilité image de P par X ou encore loi de X.*

Remarques. — a) Si l'on ne s'intéresse qu'à la v.a. X, il suffit de connaître la loi de X. L'espace probabilisé $(\Omega, \mathcal{T}, P)$ n'a pas d'importance car les calculs peuvent être faits dans l'espace probabilisé (E, P_X).

b) Pour connaître la loi de X, il suffit de connaître pour tout $x \in E$ les probabilités $P_X(\{x\}) = P(X^{-1}(\{x\})) = P(\{\omega \in \Omega | X(\omega) = x\})$. L'événement $X^{-1}(\{x\})) = \{\omega \in \Omega | X(\omega) = x\}$ est aussi noté en abrégé $[X = x]$ et sa probabilité est notée $P[X = x]$ (lire "probabilité pour que X soit égale à x").

Exemple du jeu de l'oie. — On prend $\Omega' = \{1, \dots 6\}^2$ et P_0 la loi uniforme sur Ω'. Pour $\omega = (\omega_1, \omega_2) \in \Omega'$, on pose $X_1(\omega) = \omega_1$ (chiffre fourni par le premier dé); $X_2(\omega) = \omega_2$ (chiffre fourni par le second dé) et $S(\omega) = \omega_1 + \omega_2$ (somme des deux chiffres). Quelle est la loi de la v.a. S? La v.a. S est à valeurs dans $\Omega = \{2, 3, \dots, 12\}$. Déterminer la loi P_S revient à déterminer pour tout $s \in \Omega$ les nombres

$P_S(\{s\}) = P_0[S = s] = P_0\{(\omega_1, \omega_2) \in \Omega' | \omega_1 + \omega_2 = s\} = (\mathrm{Card}\,\{(\omega_1, \omega_2) \in \Omega' | \omega_1 + \omega_2 = s\})/36.$

Or $[S = 2] = \{(1, 1)\}, [S = 3] = \{(1, 2), (2, 1)\}, [S = 4] = \{(1, 3), (2, 2), (3, 1)\}, \dots, [S = 12] = \{(6, 6)\}.$

Donc $P_0[S = 14 - s] = P_0[S = s] = (s - 1)/36$ si $2 \leq s \leq 7$ ce qui détermine la loi de S.

L'espace $\mathcal{V}$ des variables aléatoires réelles sur $(\Omega, \mathcal{T}, P)$. — Décrivons l'ensemble des v.a.r. sur un espace probabilisé fini $(\Omega, \mathcal{T}, P)$. Supposons que la tribu $\mathcal{T}$ est engendrée par la partition $\Omega = \Omega_1 + \Omega_2, + \cdots + \Omega_k$, les Ω_i sont les événements élémentaires, c'est-à-dire les plus petits éléments non vides de $\mathcal{T}$. Soit X une v.a. sur $(\Omega, \mathcal{T}, P)$ et $i \in [\![1, k]\!]$ on choisit $\omega \in \Omega_i$, on a alors $\omega \in X^{-1}(\{X(\omega)\}) \in \mathcal{T}$ et donc $\Omega_i \subset X^{-1}(\{X(\omega)\}) \iff X$ est constante sur Ω_i. Il est facile de voir réciproquement qu'une application $X : \Omega \to E$ constante sur chaque événement élémentaire Ω_i est une variable aléatoire. D'où la

Proposition. — *Une v.a.r. X sur $(\Omega, \mathcal{T}, P)$ est une fonction à valeurs réelles constante sur chaque événement élémentaire $\Omega_i, i = 1, \dots, k$ ou ce qui revient au même qui peut s'écrire $X = \sum_{i=1}^k x_i \mathbb{1}_{\Omega_i}$. Dans le cas particulier où $\mathcal{T} = \mathcal{P}(\Omega)$ on a $X = \sum_{\omega \in \Omega} X(\omega) \mathbb{1}_{\{\omega\}} = \sum_{x \in X(\Omega)} x \mathbb{1}_{[X = x]}$.*

On en déduit que l'ensemble $\mathcal{V}$ des v.a.r. sur $(\Omega, \mathcal{T}, P)$ forment un espace vectoriel de dimension k ayant pour base les fonctions indicatrices $\mathbb{1}_{\Omega_i}$ des événements élémentaires. Dans le cas $\mathcal{T} = \mathcal{P}(\Omega)$, ce sont les indicatrices $\mathbb{1}_{\{\omega\}}$ des singletons. L'espace $\mathcal{V}$ a même une structure d'algèbre : si X et Y sont des v.a.r. sur $(\Omega, \mathcal{T}, P)$ on peut définir la v.a.r. produit XY par $(XY)(\omega) = X(\omega)Y(\omega)$ pour tout $\omega \in \Omega$ (vérifier qu'on obtient bien ainsi une v.a.r.).

Si X est une v.a.r. et f une fonction de $\mathbb{R}$ dans $\mathbb{R}$ définie en tout point de $X(\Omega)$, on définit une nouvelle v.a.r. $Y = f(X)$ par $f(X)(\omega) = f(X(\omega))$ pour tout $\omega \in \Omega$. Par exemple si X est une variable positive, on peut définir $\sqrt{X}$ la v.a. racine carrée de X.

Remarque. — *Si $A_1, \dots, A_n$ sont des événements, la somme $\mathbb{1}_{A_1} + \dots + \mathbb{1}_{A_n}$ représente le nombre de ces événements qui se réalisent.*

Exemples de lois. — Si on tire au hasard une boule d'une urne remplie de n boules numérotées de 1 à n et que l'on note X le numéro de la boule tirée, on a $P[X = k] = 1/n$ si $1 \leq k \leq n$ et $P[X = k] = 0$ sinon. On dit que X *suit la loi uniforme* sur $\{1, 2, \dots n\}$.

Soient $n \in \mathbb{N}^*$ et $p \in]0, 1[$. On dit que X *suit la loi binomiale de paramètres* (n, p) notée $B(n, p)$ si $P[X = k] = \binom{n}{k} p^k (1 - p)^{n-k}$ (vérifier que c'est bien une loi). Par exemple, si on tire au hasard (hypothèse d'équiprobabilité) un sous-ensemble A d'un ensemble E à n éléments et que l'on note $X = |A|$, la loi de X est la binomiale $B(n, 1/2)$ (le vérifier).

On joue n fois à pile ou face avec une pièce bien équilibrée et on note N le nombre de face obtenu avant le premier pile. On conviendra que $N = n$ si pile n'apparaît pas. On vérifie que la loi de N est donnée par $P[N = k] = (1/2)^{k+1}$ pour $0 \leq k \leq n - 1$ et $P[N = n] = (1/2)^n$. Cette loi est *la loi géométrique tronquée.*

4- Espérance. Variance. — Dans tout ce paragraphe, X et Y sont des v.a.r. sur un espace probabilisé fini $(\Omega, \mathcal{T}, P)$. On suppose que la tribu $\mathcal{T}$ est engendrée par la partition $\Omega = \Omega_1 + \Omega_2, + \cdots + \Omega_k$ (les Ω_i sont les événements élémentaires).

Définition. — *On appelle espérance ou moyenne de la variable X le réel*

$$\mathbb{E}(X) = \sum_{i=1}^{k} P(\Omega_i) x_i \text{ avec } x_i = X(\omega), \forall \omega \in \Omega_i \text{ pour } i = 1, \dots, k$$

Dans le cas où $\mathcal{T} = \mathcal{P}(\Omega)$ les singletons sont les événements élémentaires et la formule s'écrit

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} P(\{\omega\}) X(\omega).$$

Remarques. — a) Les coefficients $p_i = P(\Omega_i)$ étant positifs de somme 1, l'espérance de X est donc une moyenne pondérée des valeurs de X, la valeur x_i prise par X sur Ω_i ayant le poids p_i.

b) Si Ω_i est tel que $P(\Omega_i) = 0$, la valeur x_i prise par X sur Ω_i n'a pas d'influence sur le calcul de $\mathbb{E}(X)$.

Cas particuliers. — a) Pour tout événement $A \subset \Omega$, l'espérance de l'indicatrice $\mathbb{1}_A$ est $\mathbb{E}(\mathbb{1}_A) = P(A)$.
b) Si X est *presque sûrement égale à la constante a* (noté $X = a$ p.s.) c'est-à-dire si $P[X = a] = 1$ (l'événement $[X = a]$ est certain) alors $\mathbb{E}(X) = a$.

Propriétés. — *a) Croissance : si $X \leq Y$ (c'est-à-dire si $X(\omega) \leq Y(\omega)$ pour tout $\omega \in \Omega$, alors $\mathbb{E}(X) \leq \mathbb{E}(Y)$. En particulier, si $0 \leq Y$ alors $0 \leq \mathbb{E}(Y)$ (i.e. l'espérance d'une v.a.r. positive est positive)*

b) Linéarité : pour tout $\lambda \in \mathbb{R}$ on a $\mathbb{E}(X + \lambda Y) = \mathbb{E}(X) + \lambda\mathbb{E}(Y)$; l'application $\mathbb{E} : \mathcal{V} \to \mathbb{R}$ est donc une forme linéaire de l'espace vectoriel $\mathcal{V}$ des v.a.r. définies sur $(\Omega, \mathcal{T}, P)$.

Conséquences. — a) Comme $X = \sum_{x \in X(\Omega)} x\mathbb{1}_{[X=x]}$, on a $\mathbb{E}(X) = \sum_{x \in X(\Omega)} xP[X = x]$. Pour calculer $\mathbb{E}(X)$, il suffit donc de connaître les $P[X = x]$ pour $x \in X(\Omega)$ l'ensemble des valeurs possibles de X, c'est-à-dire connaître la loi de X.
b) Si f est une fonction définie sur $X(\Omega)$, on obtient ainsi la formule $\mathbb{E}(f(X)) = \sum_{x \in X(\Omega)} f(x)P[X = x]$.

Inégalité de Markov. — *Soit X une v.a.r. positive ($X \geq 0$). Alors pour tout $a > 0$ en comparant X à $Y = a\mathbb{1}_{[X \geq a]}$ on obtient l'inégalité de Markov*

$$P[X \geq a] \leq \frac{\mathbb{E}(X)}{a}$$

En particulier, si $\mathbb{E}(X) = 0$ alors $P[X = 0] = 1$ (i.e. une v.a.r. positive d'espérance nulle est p.s. nulle).

Polynôme générateur d'une v.a. entière. — On se restreint ici au cas où X et Y sont des v.a. à valeurs entières (v.a.e.) c'est-à-dire à valeurs dans $\mathbb{N}$ toujours avec Ω fini.

Proposition-Définition. — *On pose pour tout t réel $G_X(t) = \mathbb{E}(t^X)$. La fonction $G_X(t)$ est un polynôme appelé polynôme générateur de la loi de X.*

En effet, si $p_k = P[X = k], 0 \leq k \leq N$ où X est à valeurs dans $[\![0, N]\!]$, on a $G_X(t) = \mathbb{E}(t^X) = \sum_{k=0}^{N} p_k t^k$.

Propriétés. — *On a*

1) $G_X(1) = 1$ et $G'_X(1) = \mathbb{E}(X)$.

2) Si Y est aussi une v.a. entière et $G_X = G_Y$ alors les v.a. X et Y ont même loi.

Covariance.Variance. Ecart-type. — X et Y désignent des v.a.r. sur un espace probabilisé fini $(\Omega, \mathcal{T}, P)$.

Définitions. — *La covariance de (X, Y) est le réel $\mathrm{Cov}(X, Y) = \mathbb{E}((X - m_X)(Y - m_Y))$ où $m_X = \mathbb{E}(X)$ et $m_Y = \mathbb{E}(Y)$. On appelle variance ou écart quadratique moyen de la variable X le réel positif $\mathrm{Var}(X) = \mathrm{Cov}(X, X) = \mathbb{E}[(X - m_X)^2]$. L'écart-type de X est la quantité $\sigma(X) = \sqrt{\mathrm{Var}(X)}$.*

Interprétation. — $X - m_X$ est l'écart de X à sa moyenne. Donc $\mathrm{Var}(X)$ est la moyenne des carrés (*i.e.* la moyenne quadratique) des écarts de X à sa moyenne. Cette quantité mesure donc la dispersion de la variable X. Elle est plus commode à manipuler que la moyenne des valeurs absolues des écarts à la moyenne $\mathbb{E}[|X - m_X|]$. On observera que $\mathbb{E}[X - m_X] = 0$: les écarts (algébriques) à la moyenne se compensent.

Propriété. — *Si $\mathcal{V}$ est l'espace des v.a.r. sur $(\Omega, \mathcal{T}, P)$, l'application $(X, Y) \in \mathcal{V} \times \mathcal{V} \mapsto \mathrm{Cov}(X, Y) \in \mathbb{R}$ est une forme bilinéaire, $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$ est la forme quadratique (positive) qui lui est associée. Sa signature est $(k - 1, 0)$ si $\dim \mathcal{V} = k$.*

Formules pratiques. — $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ *et donc*

a) $\mathrm{Var}(X)$ ne dépend que de la loi de X

b) $\mathbb{E}[X]^2 \leq \mathbb{E}[X^2]$.

On a pour tous réels a, b : $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$ et $\sigma(aX + b) = |a|\sigma(X)$

Conditions d'annulation. — *Il y a équivalence entre*

1) $\mathrm{Var}(X) = 0$

2) $P[X = m_X] = 1$

3) Il existe un réel a tel que $P[X = a] = 1$ (i.e. la v.a. X est p.s. égale à a).

Inégalité de Bienayme-Tchebicheff. — *Pour tout réel* $a > 0$, *en appliquant l'inégalité de Markov à la v.a.r. positive* $(X - m_X)^2$ *et au nombre* a^2, *il vient :* $P[(X - m_X)^2 \geq a^2] \leq \dfrac{\mathrm{Var}(X)}{a^2}$ *et comme* $[(X - m_X)^2 \geq a^2] = [|X - m_X| \geq a]$ *on a :* $P[|X - m_X| \geq a] \leq \dfrac{\sigma(X)^2}{a^2}$.
On remarque que cette inégalité n'est intéressante que pour $a > \sigma(X)$.

Indépendance. — On dit que deux événements A et B de $(\Omega, \mathcal{T}, P)$ sont *indépendants* si $P(A \cap B) = P(A)P(B)$; deux v.a. X et Y à valeurs respectives dans E et F sont *indépendantes* si pour tous $C \subset E$ et $D \subset F$ les événements $A = [X \in C]$ et $B = [Y \in D]$ sont indépendants.

Propriétés. — *Soient* X *et* Y *deux v.a.r.* *indépendantes* *alors :*
a) $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ *et* $\mathrm{Cov}(X, Y) = 0$.
b) $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$
c) Si X, Y *sont des v.a. entières leurs polynômes générateurs vérifient* $G_{X+Y} = G_X G_Y$.

5 - Probabilités conditionnelles, événements indépendants. — On se place sur un espace probabilisé fini $(\Omega, \mathcal{T}, \mathcal{P})$, mais les définitions que l'on va donner sont générales. Soit A un événement tel que $P(A) > 0$. Si on sait que l'événement A est réalisé, on souhaite définir une probabilité P_A qui tienne compte de cette information. Il est naturel de demander que la probabilité $P_A(B)$ soit proportionelle à $P(A \cap B)$ et que $P_A(A) = 1$ vaille 1. On pose donc la

Définition. — *Si* A *et* B *sont deux événements tels que* $P(A) > 0$, *la quantité*

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

s'appelle la probabilité de B *sachant* A. *(Rappel : les événements* A *et* B *de* $(\Omega, \mathcal{T}, P)$ *sont indépendants si* $P(B|A) = P(B)$ *c'est-à-dire si l'information sur la réalisation de* A *ne change pas la probabilité de celle de* B*)*

Propriétés immédiates. — *a)* P_A *est une probabilité sur* $(\Omega, \mathcal{T})$.
b) Si $B \subset A$, $P_A(B) = \dfrac{P(B)}{P(A)}$.
c) Si $A \subset B$, $P_A(B) = 1$.
d) Si $P(B) > 0$, $P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$.

Attention! La notation $P(B|A)$ est assez trompeuse : elle peut laisser penser qu'elle représente la probabilité d'un événement qui serait $B|A$, mais il n'y a évidemment aucun événement de Ω qui soit "B sachant A".

Cas particulier. — Si P est la probabilité uniforme sur Ω, alors pour tout $B \in \mathcal{P}(A)$, on a $P_A(B) = \dfrac{P(B)}{P(A)} = \dfrac{|B|}{|A|}$ et donc la restriction de P_A à $\mathcal{P}(A)$ est la probabilité uniforme sur A.

Exemple. — Dans une population, 5% des individus ont une maladie donnée. On dispose d'un test tel que : 80% des individus malades ont un test positif et 90% des individus sains ont un test négatif. Est-ce que le test est efficace?
a) Modélisation : Ω est l'ensemble des individus et P la loi uniforme. M est l'ensemble des individus malades et T celui des individus dont le test est positif.
Les hypothèses se traduisent par $P(M) = 1/20$ $P(T|M) = 4/5$ $P(T^c|M^c) = 9/10$.
b) Le test est d'autant plus efficace que la probabilité $P(M|T)$ est proche de 1. Calculons donc cette probabilité conditionnelle : $P(M|T) = P(M \cap T)/P(T) = P(M)P(T|M)/P(T)$. Et on a $P(T) = P((M \cap T) \cup (M^c \cap T)) = P(M)P(T|M) + P(M^c)P(T|M^c) = 4/100 + 19/200 = 27/200$ d'où $P(M|T) = P(M)P(T|M)/P(T) = 8/27 < 1/3$. On peut dire que ce test n'est pas très efficace!

Deux formules utiles (à savoir retrouver). — On suppose que Ω est la réunion d'événements $A_1, \ldots, A_n$ *deux à deux disjoints* de probabilités non nulles. Alors pour tout événement B on a :
a) $P(B) = \sum_{k=1}^n P(A_k \cap B) = \sum_{k=1}^n P(A_k)P_{A_k}(B)$ (formule des probabilités totales)
b) Si $P(B) \neq 0$ on a $P(A_k|B) = \dfrac{P(A_k)P_{A_k}(B)}{\sum_{k=1}^n P(A_k)P_{A_k}(B)}$ (probabilités des causes ou de Bayes)

Evénements indépendants. — Deux événements A et B de $(\Omega, \mathcal{T}, P)$ sont *indépendants* si et seulement si $P(A \cap B) = P(A)P(B)$.

Remarques : a) Si $P(A) = 0$ ou si $P(A) = 1$ alors pour tout B, A et B sont indépendants.

b) Si $P(A) > 0$, les événements A et B sont indépendants si et seulement si $P(B|A) = P(B)$.

c) Si A et B sont indépendants, alors A et B^c sont indépendants de même que A^c et B^c.

Définition-Proposition. — *Des événements $A_1, \ldots, A_n$ de $(\Omega, \mathcal{T}, P)$ sont dits indépendants (dans leur ensemble) si et seulement si pour toute partie I de $\{1, \ldots, n\}$ on a $P(\cap_{i \in I} A_i) = \prod_{i \in I} P(A_i)$. Les événements $A_1^c, A_2 \ldots, A_n$ sont alors à leur tour indépendants.*

Attention! L'indépendance deux à deux ne garantit pas l'indépendance (dans leur ensemble) de trois événements ou plus comme le montre l'

Exemple. — On lance 2 fois une pièce équilibrée. Pour $i = 1, 2$, on note $A_i =$"obtenir pile au coup i" et $A_3 =$ "obtenir la même chose aux 2 coups". On a $P(A_1) = P(A_2) = P(A_3) = 1/2$ et $P(A_1 \cap A_2) = P(A_2 \cap A_3) = P(A_3 \cap A_1) = P(A_1 \cap A_2 \cap A_3) = 1/4$ car $A_1 \cap A_2 = A_2 \cap A_3 = A_3 \cap A_1$. Donc les 3 événements A_1, A_2, A_3 sont indépendants 2 à 2 mais pas dans leur ensemble.

Loi de l'hérédité. — Un caractère d'un individu tel que la couleur des yeux *Bruns* (en fait foncés) ou *bleus* (en fait clairs) est déterminé par la présence d'une paire (non ordonné) de gènes BB, Bb ou bb appelée *génotype*. Pour qu'un individu ait les yeux bleus (caractère récessif), il faut qu'il soit bb; il a les yeux Bruns (caractère dominant) s'il est BB ou Bb.

On considère une population avec des proportions respectives $u, 2v, w$ d'individus de génotypes BB, Bb, bb. Si l'on choisit au hasard un couple (homme, femme) et que l'on note (abusivement) $\{BB, Bb\}, \{BB, bb\}, \{Bb, bb\}, \{BB\}, \{Bb\}, \{bb\}$ les événements "l'ensemble des génotypes du couple choisi est (respectivement) $\{BB, Bb\}, \{BB, bb\}, \{Bb, bb\}, \{BB\}, \{Bb\}$ ou $\{bb\}$", on a

$$P(\{BB, Bb\}) = 4uv, P(\{BB, bb\}) = 2uw, P(\{Bb, bb\}) = 4vw, P(\{BB\}) = u^2, P(\{Bb\}) = 4v^2, P(\{bb\}) = w^2$$

Si on suppose que chaque couple a un rejeton et que l'on note BB', Bb', bb' les événements "le génotype du rejeton du couple choisi est (respectivement) BB, Bb ou bb" et que ce génotype est constitué d'un des 2 gènes de l'homme tiré au hasard et d'un des 2 gènes de la femme tiré au hasard, on a

$$P(BB') = P(\{BB, Bb\})P(BB'|\{BB, Bb\}) + P(\{BB\})P(BB'|\{BB\}) + P(\{Bb\})P(BB'|\{Bb\}) =$$
$$4uv/2 + u^2 + 4v^2/4 = (u + v)^2$$

de même (en échangeant le rôle de B et b) $P(bb') = 4wv/2 + w^2 + 4v^2/4 = (v + w)^2$ et
$P(Bb') = P(\{BB, Bb\})P(Bb'|\{BB, Bb\}) + P(\{BB, bb\})P(Bb'|\{BB, bb\}) + P(\{Bb, bb\})P(Bb'|\{Bb, bb\}) +$
$P(\{Bb\})P(Bb'|\{Bb\}) = 4uv/2 + 2uw + 4vw/2 + 4v^2/2 = 2(u + v)(v + w)$ d'où la

Loi de l'hérédité. — *Soit une population avec des proportions respectives $u, 2v, w$ d'individus de génotypes BB, Bb, bb. Aux générations suivantes ces proportions seront* $(u + v)^2, 2(u + v)(v + w), (v + w)^2$.

Une autre façon de voir ceci serait de mettre tous les gènes de la population dans une urne et d'en tirer indépendamment deux. Les proportions de gènes B et b sont respectivement $u + v$ et $v + w$. L'événement BB'="tirer 2 fois B" et donc $P(BB') = (u + v)^2$, de même $P(bb') = (v + w)^2$. L'événement Bb'="tirer d'abord B puis b" + "tirer d'abord b puis B"et donc $P(Bb') = 2(u + v)(v + w)$. C.Q.F.D.

*****

Problème de Ruine : Deux joueurs jouent à pile ou face avec une pièce bien équilibrée. Le 1er dispose d'une fortune $a \in \mathbb{N}$ et le 2ème d'une fortune $b \in \mathbb{N}$. Ils jouent jusqu'à la ruine de l'un des deux joueurs (et peut-être une infinité de fois...). Le joueur gagnant est celui qui a ruiné l'autre. On note $p(a, b)$ la probabilité que le 1er joueur a de gagner. On se propose dans ce qui suit de déterminer cette probabilité en admettant que l'événement E="un des joueurs ruine l'autre" a une probabilité $P(E) = 1$.

1) Que vaut $p(a, b)$ en fonction de $p(b, a)$? Que vaut $p(a, a)$?

2) Déterminer $p(a, 0)$ et $p(0, b)$ si $a, b > 0$.

3) Montrer que $p(a, b) = \frac{1}{2}p(a - 1, b + 1) + \frac{1}{2}p(a + 1, b - 1)$ si $a, b > 0$.

4) On fixe la somme $s = a + b > 0$. Par une récurrence sur a, déduire de 3) la valeur de $p(a, s - a)$ pour $0 \leq a \leq s$. En déduire la formule recherchée $p(a, b) = \dfrac{a}{a + b}$ si a ou $b > 0$.

*****

6- Variables aléatoires à une infinité de valeurs, espaces probabilisés infinis. — Certaines expériences aléatoires ont une infinité de résultats possibles et ne peuvent être modélisées au moyen d'espaces probabilisés finis. De même, certaines variables aléatoires très naturelles peuvent prendre une infinité de valeurs et ne peuvent donc pas être définies sur un espace probabilisé fini. Nous allons commencer par deux exemples de telles variables à valeurs entières.

1- Deux exemples de v.a.e. à une infinité de valeurs :

A) **Loi géométrique (ou de Pascal) de paramètre** p : on joue à pile ou face autant de fois qu'on le veut avec une pièce qui a la probabilité p $(0 < p < 1)$ de tomber sur pile. Les lancers sont indépendants (dans leur ensemble). Soit X l'ordre du lancer où "pile" apparait pour la première fois. On a défini ainsi une variable aléatoire entière (v.a.e.) X. Il nous reste à établir la loi de X, c'est à dire à déterminer les probabilités $P[X = n]$. Pour tout $k \in \mathbb{N}^*$, soit l'événement F_k = "obtenir face au kième lancer". On peut écrire $[X = n] = F_1 \cap \ldots \cap F_{n-1} \cap F_n^c$ et donc en posant $q = 1 - p$, comme les F_k sont indépendants (dans leur ensemble) et $P(F_k) = q$ pour tout $k \in \mathbb{N}^*$, on a $P[X = n] = P(F_1) \ldots P(F_{n-1})P(F_n^c) = pq^{n-1}$. Observons que $\displaystyle\sum_1^{\infty} pq^{n-1} = \frac{p}{1-q} = 1$ ce qui signifie que X prend presque sûrement une valeur finie (il n'y a aucune chance que n'apparaisse jamais pile). On dit que X suit la loi géométrique de paramètre p.

Si l'on considère que l'expérience aléatoire que l'on fait ici a pour résultat la valeur de X, on est conduit à construire l'espace de probabilité $\Omega = \mathbb{N}^*$ avec $P(\{n\}) = pq^{n-1}$. Il reste à vérifier que si pour tout $A \in \mathcal{P}(\Omega)$ on pose $P(A) = \sum_{n \in A} pq^{n-1}$ on définit bien une probabilité P sur Ω (le faire). On a ainsi construit le premier espace probabilisé infini de ce cours.

B) **Loi de Poisson de paramètre** λ : on s'intéresse à compter le nombre de fois N qu'un événement pouvant se répéter survient dans un intervalle de temps donné. Par exemple le nombre N d'étoiles nouvelles qui apparaitront dans une partie du ciel donnée au cours de l'année prochaine (c'est ce type d'événement qui a conduit Poisson a élaborer la loi qui porte son nom). On dispose de données historiques qui permettent de dire qu'en moyenne il apparaît $\lambda > 0$ étoiles par unité de temps. A partir des deux postulats simples qui vont suivre on va établir la loi de la variable aléatoire N.

Postulat 1 : Au cours d'un intervalle de temps assez petit $[t, t + \Delta t[$ il apparaît au plus une étoile avec une probabilité $\lambda \Delta t$.

Postulat 2 : Ce qui se passe après le temps t est indépendant de ce qui se passe avant.

Notons $N(t)$ le nombre d'étoiles apparues au temps t de l'année prochaine ($t = 0$ étant le début de l'année). La v.a.e. dont on cherche la loi est $N = N(1)$ où 1 est le temps d'une année. Pour $t < s$, soit l'événement $A_{t,s}$ = "il apparaît une étoile nouvelle dans l'intervalle de temps $[t, s[$".

A partir du postulat 1, on a pour tout $n \in \mathbb{N}$ et Δt petit

$$[N(t + \Delta t) = n] = [N(t) = n] \cap A^c_{t,t+\Delta t} + [N(t) = n - 1] \cap A_{t,t+\Delta t}$$

la réunion étant disjointe, on a donc

$$P([N(t + \Delta t) = n]) = P([N(t) = n] \cap A^c_{t,t+\Delta t}) + P([N(t) = n - 1] \cap A_{t,t+\Delta t}).$$

Posons $p_n(t) = P([N(t) = n])$. Du postulat 2, il vient l'équation

$$p_n(t + \Delta t) = p_n(t)(1 - \lambda \Delta t) + p_{n-1}(t)\lambda \Delta t.$$

En divisant par Δt puis en faisant tendre Δt vers 0, il vient la suite d'équations différentielles

$$p'_n(t) + \lambda p_n(t) = \lambda p_{n-1}(t)$$

que l'on résout par récurrence en partant de $p_{-1}(t) = 0 \; \forall t$ (évidemment) et $p_0(0) = 1$ et $p_n(0) = 0 \; \forall n > 0$ pour obtenir

$$p_n(t) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}$$

d'où la loi de N

$$P([N = n]) = e^{-\lambda} \frac{\lambda^n}{n!}$$

appelée *loi de Poisson de paramètre* λ et notée $\mathcal{P}(\lambda)$.

Loi binomiale et Loi de Poisson : on va retrouver la loi de Poisson d'une autre façon qui consiste à l'approcher par une loi binomiale.

Tout d'abord, décrivons de manière précise comment apparaît la loi binomiale quand on étudie le nombre S de succès à une épreuve (jet d'une pièce, arrivée d'un appel à un standard téléphonique etc...) répétée n fois indépendamment avec une probabilité de succès p.

Si l'on note E_k^1="la kième épreuve réussit" et E_k^0="la kième épreuve échoue", on a $P(E_k^1) = p$ et $P(E_k^0) = q = 1 - p$ si $1 \leq k \leq n$. L'événement $[S = s]$="parmi les n épreuves il y a eu s succès" peut s'écrire $[S = s] = \sum_{(\varepsilon_k)\in A_s} E_1^{\varepsilon_1} \cap \ldots \cap E_n^{\varepsilon_n}$ où A_s est l'ensemble des n-uplets de 0 et de 1 comportant exactement s fois 1. Les épreuves E_k étant indépendantes dans leur ensemble, on a pour $(\varepsilon_k) \in A_s$, $P(E_1^{\varepsilon_1}\cap\ldots\cap E_n^{\varepsilon_n}) = P(E_1^{\varepsilon_1})\ldots P(E_n^{\varepsilon_n}) = p^s q^{n-s}$ et donc $P[S = s] = \sum_{(\varepsilon_k)\in A_s} p^s q^{n-s}$ et comme $\mathrm{Card}\, A_s = C_n^s$, la loi de S est $P[S = s] = C_n^s p^s q^{n-s}$ pour $0 \le s \le n$ et 0 sinon. C'est la loi binomiale $B(n, p)$ de paramètres n et p.

Etudions maintenant à nouveau le nombre N d'étoiles nouvelles qui apparaitront au cours de l'année prochaine en partant toujours de la connaissance statistique qu'en moyenne il apparaît $\lambda > 0$ étoiles dans une unité de temps (ici une année). Soit n un entier (qui sera amené à tendre vers l'infini). Partageons l'année en les n intervalles $[k/n, (k + 1)/n[$ où $0 \le k \le n - 1$. Si n est assez grand, on peut approximativement estimer que le postulat 1 s'applique : dans un intervalle de temps $[k/n, (k + 1)/n[$ il apparaît au plus une étoile avec une probabilité $p = \lambda/n$. On continue à appliquer le postulat 2 qui garantit que les événements E_k^1="il apparaît une étoile dans l'intervalle de temps $[k/n, (k+1)/n[$" sont indépendants. La loi de N est donc approximativement celle d'une variable N_n de loi binomiale $B(n, \lambda/n)$ si n est assez grand.

En fait, nous allons voir que, à s fixé, les $P[N_n = s]$ tendent vers une limite p_s que l'on peut considérer comme étant la "vraie" loi de N.

Rappelons que le polynôme générateur de N_n est $G_n(t) = (1 + \lambda(t - 1)/n)^n$ qui à t fixé tend vers $G(t) = e^{\lambda(t-1)}$ quand $n \to \infty$ (exercice classique sur les équivalents à savoir faire). De même, la $s$ième dérivée $G_n^{(s)}(t) = n\ldots(n - s + 1)(\lambda/n)^s(1 + \lambda(t - 1)/n)^{n-s}$ tend vers $G^{(s)}(t) = \lambda^s e^{\lambda(t-1)}$ quand $n \to \infty$.

Par définition de $G_n(t)$ on a $P[N_n = s] = C_n^s p^s q^{n-s} = G_n^{(s)}(0)/s!$ qui à s fixé tend donc vers $G^{(s)}(0)/s! = \lambda^s e^{-\lambda}/s!$ quand $n \to \infty$. On retrouve que la loi de N est la loi $\mathcal{P}(\lambda)$ de Poisson de paramètre λ :

$$P[N = s] = e^{-\lambda}\frac{\lambda^s}{s!}.$$

On dit que les variables binomiales N_n tendent en loi vers N. On peut aussi tirer la conséquence pratique suivante de ce phénomène de convergence en loi : si une variable X suit une loi binomiale $B(n, p)$ avec n très grand et p assez petit de sorte que $\lambda = np$ soit de taille raisonnable, on peut pour les calculs considérer que X suit la loi de Poisson $\mathcal{P}(\lambda)$.

Exemple. — Dans la population française, on estime que 1% de la population environ possède une tortue (pourcentage faible). On note X le nombre de résultats positifs à la question "Avez vous une tortue chez vous ?" posée à un échantillon de 100 personnes tirées au hasard. La loi de X est en toute rigueur la loi binomiale $B(100, 1/100)$ mais on ne fait pas de grosse erreur en considérant qu'on peut la remplacer par la loi de Poisson $\mathcal{P}(1)$.

2- Loi de probabilité d'une v.a.e.. — En général, se donner la loi d'une v.a. entière (v.a.e) X revient à déterminer la suite $p_k = P[X = k] \ge 0$. Cette suite vérifie $\sum_0^\infty p_k = 1$. On observe, comme dans les deux exemples vus précédemment, que l'on ne précise pas ici l'espace Ω de définition de X (il y en a pourtant un, mais peu importe...). On s'attache uniquement à déterminer les chances que la quantité entière aléatoire X a de valoir tel ou tel entier. La loi de X est en fait une *probabilité sur* $\mathbb{N}$. Pour pouvoir préciser cette notion, il est nécessaire d'introduire la

Définition. — *Soit une suite de réels $u_k \ge 0$ telle que $\sum_0^\infty u_k$ est convergente et $A \in \mathcal{P}(\mathbb{N})$ un sous-ensemble d'entiers. On définit le réel $\sum_{k\in A} u_k$ par*

$$\sum_{k\in A} u_k = \lim_{n\to\infty} \sum_{k\in A\cap[0,n]} u_k.$$

Proposition-Définition. — *Soit une suite de réels $p_k \ge 0$ telle que $\sum_0^\infty p_k = 1$ et $A \in \mathcal{P}(\mathbb{N})$ un sous-ensemble d'entiers. La formule $P(A) = \sum_{k\in A} p_k$ définit une probabilité sur $\mathbb{N}$ dans le sens que (comme dans le cas d'un espace fini) P est une application de l'ensemble $\mathcal{P}(\mathbb{N})$ dans $\mathbb{R}^+$ telle que $P(\mathbb{N}) = 1$ et pour toutes parties disjointes A et B de $\mathbb{N}$ on a $P(A + B) = P(A) + P(B)$.*
De plus, pour toute suite A_n, $n \in \mathbb{N}$ de parties disjointes de $\mathbb{N}$ on a $P(\sum_{n\in\mathbb{N}} A_n) = \sum_0^\infty P(A_n)$.

Espérance. — L'espérance d'une v.a.e. X dont la loi est donnée par la suite p_k est la somme de la série à termes positifs $\mathbb{E}(X) = \sum_{k\in\mathbb{N}} kP[X = k] = \sum_0^\infty kp_k$ (par convention $+\infty$ si elle est divergente).
La **variance** de X est donnée par $\mathrm{Var}X = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.

Exemple. — Soit la suite $p_0 = p_1 = 1/4$ et $p_k = \dfrac{1}{k(k-1)} - \dfrac{1}{k(k+1)}$ pour $k \ge 2$. Il est facile de constater que $\displaystyle\sum_0^n p_k = 1 - \frac{1}{n(n+1)}$ de sorte que $\displaystyle\sum_0^\infty p_k = 1$. Il y a donc une variable aléatoire entière X de loi $P[X = k] = p_k$.

On a $E(X) = \sum_0^\infty kp_k = 1/4 + \sum_2^\infty (\frac{1}{k-1} - \frac{1}{k+1}) = 7/4$. On vérifie que la variance $\mathrm{Var}X = \infty$ sur cet exemple.

Série génératrice. — La série génératrice d'une v.a.e. X dont la loi est donnée par la suite p_k est la série entière $\sum_0^\infty p_k t^k$. Elle est normalement convergente sur $[-1, 1]$ (pourquoi ?). Sa somme $G_X(t) = \sum_0^\infty p_k t^k$ pour $t \in [-1, 1]$ est appelé la *fonction génératrice* de X.

Propriétés : 1) Si Y est une autre v.a.e. et $G_X = G_Y$ alors les v.a.e. X et Y ont même loi.

2) Si X et Y sont indépendantes $G_{X+Y} = G_X G_Y$

3) Si $\mathbb{E}(X) < \infty$ alors G_X est dérivable sur $[-1, 1]$ et $\mathbb{E}(X) = G'_X(1)$.

4) Si $\mathbb{E}(X(X-1)) < \infty$ alors G_X est deux fois dérivable sur $[-1, 1]$ et $\mathbb{E}(X(X-1)) = G''_X(1)$;
de plus $\mathrm{Var}(X) = G''_X(1) + G'_X(1) - (G'_X(1))^2$.

3- Loi de probabilité d'une v.a.r. qui a une suite de valeurs possibles (variable aléatoire discrète).
Si X est une v.a.e. et $f : \mathbb{N} \to \mathbb{R}$ une suite de réels, la v.a.r. $Y = f(X)$ est particulière, l'ensemble de ses valeurs possibles Ω est constitué des éléments d'une suite $\Omega = \{f(k), k \in \mathbb{N}\}$. Si Ω est infini, on peut toujours réécrire $\Omega = \{y_k, k \in \mathbb{N}\}$ avec $y_k \neq y_l$ si $k \neq l$. Se donner la loi de Y revient à déterminer la suite $p_k = P[Y = y_k] \geq 0$. La loi de Y est en fait une *probabilité sur* Ω. Ceci généralise à un ensemble Ω *dénombrable* (i.e. en bijection avec $\mathbb{N}$) ce qui a été fait pour $\mathbb{N}$ lui-même.

PROPOSITION-DÉFINITION. — *Soit une suite de réels $p_k \geq 0$ telle que $\sum_0^\infty p_k = 1$ et $A \in \mathcal{P}(\Omega)$. La formule $P(A) = \sum_{y_k \in A} p_k$ définit une probabilité sur Ω dans le sens que (comme dans le cas d'un espace fini) P est une application de l'ensemble $\mathcal{P}(\Omega)$ dans $\mathbb{R}^+$ telle que $P(\Omega) = 1$ et pour toutes parties disjointes A et B de Ω on a $P(A + B) = P(A) + P(B)$.*
De plus, pour toute suite A_n, $n \in \mathbb{N}$ de parties disjointes Ω on a $P(\sum_{n \in \mathbb{N}} A_n) = \sum_0^\infty P(A_n)$.

DÉFINITION. — *Les v.a.r. qui ont une suite de valeurs possibles sont appelées variables aléatoires discrètes (v.a.d.). Soit X une telle v.a.d. et $\Omega = \{x_k, k \in \mathbb{N}\}$ son ensemble de valeurs possibles. Sa loi est donnée par la suite $p_k = P[X = x_k]$.*

DÉFINITIONS. — *1) L'espérance $\mathbb{E}(X)$ de X dont la loi est donnée par la suite p_k est la somme de la série $\mathbb{E}(X) = \sum_{k \in \mathbb{N}} x_k P[X = x_k] = \sum_0^\infty x_k p_k$ si elle est absolument convergente et elle n'est pas définie sinon.*
2) La variance de X, si elle existe, est donnée par $\mathrm{Var}X = \mathbb{E}(X - \mathbb{E}(X))^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.

On a encore (à vérifier) comme dans le cas des v.a. à un nombre fini de valeurs les

PROPRIÉTÉS. — *Soient X et Y deux v.a. discrètes indépendantes alors :*

a) $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ si les moyennes $\mathbb{E}(X)$ et $\mathbb{E}(Y)$ existent.

b) $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ si les variances $\mathrm{Var}(X)$ et $\mathrm{Var}(Y)$ existent.

Résumé sur les séries utile pour les probabilités. — Rappelons qu'une suite réelle est une application $n \mapsto u_n$ de $[\![a, +\infty[\![\subset \mathbb{N}$ dans $\mathbb{R}$.

Définition. — *A toute suite réelle (u_n), on associe la suite (S_n) des sommes partielles $S_n = \sum_{k=0}^{n} u_k$. On a donc $S_0 = u_0,\ S_1 = u_0 + u_1,\ S_2 = u_0 + u_1 + u_2, \ldots,\ S_n = u_0 + u_1 + \ldots + u_n, \ldots$. Etudier la suite (S_n) des sommes partielles revient étudier la série $\sum u_n$ de terme général u_n.*

Si la suite $(S_n)_{n\in\mathbf{N}}$ converge vers S on dit que la série $\sum u_n$ converge et on note $S = \sum_{n=0}^{\infty} u_n$, appelée *somme de la série*. Si la suite $(S_n)_{n\in\mathbf{N}}$ diverge on dit que la série $\sum u_n$ diverge.

Exemples :

• $u_n = 1$ pour tout n, alors $S_k = k + 1 \to \infty$ et donc la série diverge : $\sum_{n=0}^{\infty} u_n$ n'existe pas.

• $u_n = (-1)^n$, $S_0 = 1$, $S_1 = 0$, $S_2 = 1$, $\ldots$ la suite $(S_n)_{n\in\mathbf{N}}$ admet deux sous suites ayant deux limites différentes. Donc la série diverge.

• Série géométrique : c'est une série associée à une suite $u_n = r.u_{n-1} = cr^n$ d'où $S_n = c\dfrac{1 - r^{n+1}}{1 - r}$ si $r \neq 1$. La série converge et a pour somme $\dfrac{c}{1-r}$ ssi $|r| < 1$. On retiendra donc qu'une *série géométrique de raison r avec $|r| < 1$ a pour somme le terme initial sur $1 - r$.*

• Série "téléscopique" : une série associée à une suite $(u_n)_{n\geq 0}$ vérifiant $u_n = v_{n+1} - v_n$ où $(v_n)_{n\geq 0}$ est une suite donnée. On a alors le phénomène "téléscopique" : $S_n = v_1 - v_0 + v_2 - v_1 + v_3 - v_2 + \cdots + v_{n+1} - v_n = v_{n+1} - v_0$. Donc à constante près, la suite S_n est la suite v_n.

• Série de l'exponentielle : $\sum \dfrac{t^n}{n!}$. La formule de Taylor pour la fonction exponentielle e^t en $t = 0$ donne à l'ordre n : $e^t = \sum_{k=0}^{n} \dfrac{t^k}{k!} + e^{c_n}\dfrac{t^{n+1}}{(n+1)!}$ où $c_n \in [0, t]$ dépend de t mais on a avec $S_n = \sum_{k=0}^{n} \dfrac{t^k}{k!}$ la majoration $|e^t - S_n| \leq e^{|t|}\dfrac{t^{n+1}}{(n+1)!}$ qui à t fixé tend vers 0. On obtient que pour tout réel t fixé, la série $\sum \dfrac{t^n}{n!}$ converge et on a $\sum_{k=0}^{\infty} \dfrac{t^k}{k!} = e^t$.

Propriété : Si la série $\sum_{n\geq 0} u_n$ converge, il est nécessaire (il faut) que $\lim_{n\to\infty} u_n = 0$.

Preuve : si S_k tend vers S alors S_{k-1} aussi et donc $u_k = S_k - S_{k-1}$ tend vers $l - l = 0$.

⚠La réciproque est fausse!!

Contre-exemple : $u_n = \ln \dfrac{n+1}{n} = \ln(n + 1) - \ln n$ définie pour $n \geq 2$.

On a $u_n = \ln \dfrac{n+1}{n} \to \ln 1 = 0$ et pourtant $S_n = \ln(n + 1)$ diverge.

Critère : Si, à partir d'un certain rang N, on a $0 \leq u_n \leq v_n$ alors

$\sum v_n$ converge $\Rightarrow \sum u_n$ converge et $\sum u_n$ diverge $\Rightarrow \sum v_n$ diverge.

Preuve : $S_n - S_{N-1} = u_N + u_{N+1} + \ldots + u_n \leq v_N + v_{N+1} + \ldots + v_n = \tilde{S}_n - \tilde{S}_{N-1}$ si $\tilde{S}_n$ est la somme partielle des termes de la suite $(v_n)_{n\geq 0}$. Si $\sum_{n\in\mathbf{N}} v_n$ converge alors $\tilde{S}_n$est majorée et donc S_n est croissante majorée donc convergente et $\sum_{n\in\mathbf{N}} u_n$ converge .

Application (fonction génératrice) : Si p_n est une suite de nombres positifs telle que la série $\sum p_n$ converge et a pour somme $\sum_{n=0}^{\infty} p_n = 1$ alors pour tout $t \in [0, 1]$ la fonction $g(t) = \sum_{n=0}^{\infty} p_n t^n$ est bien définie et $g(1) = 1$.

Ajoutons que l'on peut montrer (théorie des séries entières) que la fonction $g(t)$ est toujours dérivable pour $t \in [0, 1[$ et $g'(t) = \sum_{n=0}^{\infty} np_n t^{n-1}$ et g est dérivable sur $[0, 1]$ ssi $\sum_{n=0}^{\infty} np_n$ est convergente et alors cette somme vaut $g'(1)$.

Enfin, $g(t)$ est toujours deux fois dérivable pour $t \in [0, 1[$ et $g''(t) = \sum_{n=0}^{\infty} n(n-1)p_n t^{n-2}$ et g est deux fois dérivable sur $[0, 1]$ ssi $\sum_{n=0}^{\infty} n(n-1)p_n$ est convergente et alors cette somme vaut $g''(1)$.

7- Loi de probabilité d'une v.a.r. : cas général. — Certaines variables aléatoires réelles peuvent non seulement prendre une infinité de valeurs mais leurs valeurs possibles peuvent constituer un intervalle $I \neq \emptyset$ de réels. Nous allons décrire un exemple naturel de v.a.r. donné comme le *temps d'attente* d'un événement aléatoire.

Loi exponentielle de paramètre λ : On revient au problème de l'apparition d'étoiles nouvelles dans le ciel, mais on s'intéresse cette fois-ci au temps $T > 0$ qu'il faudra attendre pour voir la première apparition d'une étoile sachant qu'en moyenne il apparaît $\lambda > 0$ étoiles par unité de temps. Nous avions noté $N(t)$ le nombre d'étoiles apparues au temps t et montré en adoptant deux postulats naturels qu'à t fixé, la v.a.e. $N(t)$ suivait la loi de Poisson $\mathcal{P}(\lambda t)$, c'est-à-dire $P[N(t) = k] = e^{-\lambda t} \frac{(\lambda t)^k}{k!}$.

Nous allons à partir de là donner une expression de la probabilité de l'événement
$[T \leq t_0] =$ "il a fallu attendre moins de t_0 pour voir l'apparition d'une première étoile nouvelle".
On a $[T \leq t_0] = [N(t_0) > 0] = \sum_{k=1}^{\infty} [N(t_0) = k]$ et donc

$$P[T \leq t_0] = P[N(t_0) > 0] = \sum_{k=1}^{\infty} P[N(t_0) = k] = \sum_{k=1}^{\infty} e^{-\lambda t_0} \frac{(\lambda t_0)^k}{k!} = 1 - e^{-\lambda t_0}.$$

La fonction croissante (au sens large) $F(t) = P[T \leq t] = 1 - e^{-\lambda t}$ sur l'ensemble $I = [0, \infty[$ des valeurs possibles de T et valant 0 ailleurs est la *fonction de répartition* de la v.a.r. T. Elle suffit à calculer toutes les probabilités d'événements exprimables uniquement à l'aide de T, elle détermine donc la *loi* de T. Par exemple, si $0 < a < b$, $P[a < T \leq b] = F(b) - F(a)$; on a donc $P[T = b] \leq F(b) - F(t)$ pour tout $t < b$ et donc comme $F(t) = 1 - e^{-\lambda t}$ est continue en b, $P[T = b] = 0$ pour tout b. Cette dernière propriété caractérise les v.a.r. continues (par rapport aux v.a.r. discrètes).

Fonction de répartition. — En général, se donner la loi d'une v.a.r. X revient à se donner une fonction *croissante* au sens large F_X, continue à droite *i.e.* $\lim_{t \to x^+} F_X(t) = F(x^+) = F(x)$ et telle que $\lim_{x \to -\infty} F_X(x) = 0$ et $\lim_{x \to +\infty} F_X(x) = 1$.

La fonction F_X est appelée la *fonction de répartition* de la v.a.r. X.

Si F_X est continue sur $\mathbb{R}$, on dit que la v.a.r. est *continue*.

Si X est une variable entière dont la loi est donnée par la suite p_k c'est en particulier une v.a.r. et sa fonction de répartition s'écrit $F_X(t) = P[X \leq t] = \sum_{k=0}^{\infty} p_k \mathbb{1}_{\{x \leq k\}}(t) = \sum_{k=0}^{[t]} p_k$ où $[t]$ est la partie entière du réel t. Cette fonction est une fonction en escalier qui en général a des discontinuités aux entiers donc ce n'est pas du tout une variable continue!

Evénements réels. — On appellera *événement réel* toute partie $A = \sum_1^n I_k$ où les I_k sont des intervalles de $\mathbb{R}$ *deux à deux disjoints*. L'ensemble des événements réels sera noté $\mathcal{E}$.

L'ensemble $\mathcal{E} \subset \mathcal{P}(\mathbb{R})$ est une *algèbre* de parties de $\mathbb{R}$ c'est-à-dire vérifie les trois propriétés :

(i) $\mathbb{R}$ et $\emptyset \in \mathcal{E}$ (la partie pleine et la partie vide sont des événements).

(ii) Si $A \in \mathcal{E}$ alors $A^c \in \mathcal{E}$ (le contraire d'un événement est un événement).

(iii) Si A et $B \in \mathcal{E}$ alors $A \cup B \in \mathcal{E}$ (et $A \cap B \in \mathcal{E}$ aussi grâce à (ii))

On a ainsi une catégorie d'ensembles de réels que l'on peut appeler des événements et on est en mesure de définir ce que peut être une probabilité sur $\mathbb{R}$:

Construction d'une probabilité sur $\mathbb{R}$. — Soit une fonction *croissante* (au sens large) et *continue à droite* $F : \mathbb{R} \to [0, 1]$ telle que $\lim_{x \to -\infty} F(x) = 0$ et $\lim_{x \to +\infty} F(x) = 1$.

On vérifie qu'on définit bien une application $P : \mathcal{E} \to [0, 1]$ de la façon suivante :

1) On pose d'abord pour tout $a \in \mathbb{R}$:

$P(] - \infty, a]) = 1 - P(]a, +\infty]) = F(a)$ et $P(] - \infty, a[) = 1 - P([a, +\infty]) = F(a^-)$

2) Si I et J sont des intervalles du type précédent et $I \subset J$ on pose $P(J - I) = P(J) - P(I)$. On obtient ainsi pour tous réels a, b avec $a \leq b$:

$$P(]a, b]) = F(b) - F(a) \ ; \ P([a, b]) = F(b) - F(a^-) \ ; \ P([a, b[) = F(b^-) - F(a^-) \ ; \ P(]a, b[) = F(b^-) - F(a)$$

3) Les $P(I)$ étant ainsi déterminés pour tout intervalle I, on pose enfin $P(\sum_1^n I_k) = \sum_1^n P(I_k)$ si I_k $(k = 1, \ldots, n)$ sont des intervalles *disjoints*.

L'application $P : \mathcal{E} \to [0, 1]$ ainsi définie est la *probabilité sur* $\mathbb{R}$ *dont la fonction de répartition est* $F(x) = P(] - \infty, x])$

Comme dans le cas d'un espace fini et d'un espace dénombrable, P vérifie les deux propriétés fondamentales d'une probabilité :

$P(\mathbb{R}) = 1$ et pour toutes parties *disjointes* $A \in \mathcal{E}$ et $B \in \mathcal{E}$ on a $P(A + B) = P(A) + P(B)$.

Espérance (cas d'une v.a.r. positive). — Soit X une v.a.r. positive et F_X sa fonction de répartition. On définit l'espérance de X par

$$\mathbb{E}(X) = \int_0^{+\infty} (1 - F_X(x))dx$$

expression qui peut être finie ou $+\infty$ (par convention) selon que l'intégrale généralisée converge ou diverge.

Par exemple, si $X = \text{Cte} = c \geq 0$ alors $F_X(x) = 0$ si $x < c$ et $F_X(x) = 1$ si $x \geq c$ et on a $\mathbb{E}(X) = \int_0^c dx = c$ (la moyenne de la constante c est la constante c).

Dans le cas particulier où X est une v.a.e. dont la loi est donnée par la suite p_k on a alors $F_X(x) = P[X \leq x] = \sum_{k=0}^{[x]} p_k$ et donc

$$\mathbb{E}(X) = \int_0^{\infty} (\sum_{k=[x]+1}^{\infty} p_k)dx = \sum_{l=0}^{\infty}(\sum_{k=l+1}^{\infty} p_k) = \sum_{k=1}^{\infty} kp_k$$

qui coïncide bien avec l'expression déjà donnée dans ce cas particulier.

Exemple. — La v.a.r. $T > 0$ vue précédemment dont la fonction de répartition est $F(t) = 1 - e^{-\lambda t}$ (loi exponentielle de paramètre λ) a pour moyenne $\mathbb{E}(T) = \int_0^\infty e^{-\lambda t}dt = 1/\lambda$.

Inégalité de Markov. — *Soit X une v.a.r. ≥ 0. Alors pour tout $a > 0$ on a $P[X \geq a] \leq \frac{\mathbb{E}(X)}{a}$. En particulier, si $\mathbb{E}(X) = 0$ alors $P[X = 0] = 1$ (i.e. une v.a.r. positive d'espérance nulle est p.s. nulle).*

Définitions. — *Soit X une v.a.r. On dit que X est majorée par $M \in \mathbb{R}$ que l'on note $X \leq M$ si $P[X \leq M] = F_X(M) = 1$ et X est bornée s'il existe $M \in \mathbb{R}^+$ tel que $|X| \leq M$.*

La v.a.r. X dite minorée par $m \in \mathbb{R}$ (noté $X \geq m$) si $-X$ est majorée par $-m$. Pour tout entier n, on pose $X_n = \inf(X, n)$ qui est naturellement une v.a.r. majorée par n.

Proposition 1. — *Soit X et Y des v.a.r. ≥ 0 et $M \geq 0$ une constante.*

1) Si $X \leq Y$ alors $\mathbb{E}(X) \leq \mathbb{E}(Y)$, en particulier si $X \leq M$ alors $\mathbb{E}(X) \leq M$.

2) Dans tous les cas $\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$.

3) Pour tout réel $\lambda \geq 0$, $\mathbb{E}(\lambda X) = \lambda\mathbb{E}(X)$.

Soit X une v.a.r. $0 \leq X < 1$. Notons $Y_n = \sum_{k=1}^n \frac{k-1}{n} 1\!\!1_{(k-1)/n \leq X < k/n}$ et $Z_n = \sum_{k=1}^n \frac{k}{n} 1\!\!1_{(k-1)/n \leq X < k/n}$ on a $0 \leq Y_n \leq X \leq Z_n$ donc $\mathbb{E}(Y_n) \leq \mathbb{E}(X) \leq \mathbb{E}(Z_n)$ et les v.a.r. finies Y_n et Z_n qui encadrent X vérifient $Z_n - Y_n \leq 1/n$ d'où $\mathbb{E}(Z_n) - \mathbb{E}(Y_n) \leq 1/n$ et donc la suite de v.a.r. finies Y_n est telle que $\mathbb{E}(X) - \mathbb{E}(Y_n) \leq 1/n$.

Grâce ces remarques et à la proposition 1, on peut obtenir facilement de la linéarité de l'espérance $\mathbb{E}$ sur les v.a.r. finies la

Proposition 2. — *Soit X et Y des v.a.r. ≥ 0 alors $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$.*

Pour définir l'espérance d'une v.a.r. X *de signe variable*, il est commode d'introduire les deux v.a.r. positives $X^+ = \sup(X, 0)$ (*la partie positive* de X) et $X^- = \sup(-X, 0)$ (*la partie négative* de X).

On a $X = X^+ - X^-$ et on peut maintenant donner la définition générale :

Espérance (cas général). — Soit X une v.a.r. on définit l'espérance $\mathbb{E}(X)$ (parfois appelée moyenne de X et notée m_X) par

$$\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-) \text{ si cette différence a un sens.}$$

on dit qu'elle est finie si $\mathbb{E}(X^+)$ et $\mathbb{E}(X^-)$ le sont tous les deux.

Grâce aux propositions 1 et 2 on montre facilement la

Proposition 3. — *Si les v.a.r. X et Y ont des espérances finies alors pour tout $\lambda \in \mathbb{R}$ il en est de même de λX et de $X + Y$ et on a $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ et $\mathbb{E}(\lambda X) = \lambda\mathbb{E}(X)$.*

Autrement dit les v.a.r. d'espérance finie forment un espace vectoriel $\mathcal{V}$ et l'espérance $\mathbb{E}$ est une application linéaire $\mathbb{E} : \mathcal{V} \to \mathbb{R}$.

Variables à densité. — Une v.a.r. X continue dont la fonction de répartition F_X est non seulement continue mais dérivable sauf en des points isolés est dite "à densité" on note alors $f_X = F'_X$ la *densité* de X là où elle est définie, on a alors, si l'intégrale est convergente :

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} xf_X(x)dx$$

Pour voir ceci, on peut clairement se ramener au cas où $X \geq 0$, on a alors en faisant une intégration par parties :

$$\mathbb{E}(X) = \int_0^{+\infty} (1 - F_X(x))dx = [x(1 - F_X(x))]_0^{+\infty} + \int_0^{+\infty} x f_X(x)dx = \lim_{a\to\infty} a(1 - F_X(a)) + \int_0^{+\infty} x f_X(x)dx$$

Si $\mathbb{E}(X) < +\infty$, la limite $L = \lim_{a\to\infty} a(1 - F_X(a))$ existe. On a $L = 0$, sinon on aurait $1 - F_X(x) \sim \frac{L}{x}$ en l'infini avec $L > 0$ ce qui contredirait la convergence de $\mathbb{E}(X) = \int_0^{+\infty}(1 - F_X(x))dx$.

On vérifie, et nous l'admettrons, la

Propriété. — *Si une v.a.r. X a pour densité f_X et g est dérivable alors la v.a.r. $Y = g(X)$ a pour moyenne $\mathbb{E}(Y) = \int_{-\infty}^{+\infty} g(x) f_X(x)dx$ si l'intégrale est convergente.*

Exemples de variables à densité. — La v.a.r. T vue précédemment dont la loi est exponentielle de paramètre λ a une fonction de répartition continue $F(t) = (1 - e^{-\lambda t})\mathbb{1}_{\mathbb{R}^+}$ dérivable sauf en 0 elle a donc pour densité sur $\mathbb{R}^*$ la fonction f définie par $f(t) = F'(t) = \lambda e^{-\lambda t}$.

La condition pour qu'une $f(x)$ (localement intégrable) soit la densité d'une v.a.r. est qu'elle soit positive et que $\int_{-\infty}^{+\infty} f(x)dx = 1$ (en particulier l'intégrale est convergente).

Loi uniforme sur l'intervalle $[a, b]$ **(**$a < b$**)** : $f(x) = \frac{1}{b-a}\mathbb{1}_{[a,b]}(x)$ est la densité de la loi uniforme sur $[a, b]$.

Loi de Cauchy : $f(x) = \frac{1}{\pi(1+x^2)}$ est la densité de la loi de Cauchy.

Loi normale ou gaussienne $N(m, \sigma^2)$: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$ est la densité de la loi normale de moyenne m et de variance σ^2.

Variance. — Soit X une v.a.r. telle que $\mathbb{E}(X) = m_X$ est finie, la *variance* de X est l'expression $\mathrm{Var}(X) = \mathbb{E}([X - m_X]^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 \geq 0$ qui est $< +\infty$ si et seulement si $\mathbb{E}(X)^2 < +\infty$.

L'*écart type* est $\sigma(X) = \sqrt{\mathrm{Var}(X)}$.

On a pour tous réels a, b : $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$ et $\sigma(aX + b) = |a|\sigma(X)$

Conditions d'annulation. — *Il y a équivalence entre*

1) $\mathrm{Var}(X) = 0$

2) $P[X = m_X] = 1$

3) Il existe un réel a tel que $P[X = a] = 1$ (i.e. la v.a. X est p.s. égale à a).

Indépendance. — Soient X et Y deux v.a.r. indépendantes alors :

a) $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$

b) $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

Inégalité de Bienayme-Tchebicheff. — *Soit X une v.a.r. ayant moyenne m_X finie et un réel $a > 0$. On a l'inégalité $P[|X - m_X| \geq a] \leq \frac{\sigma(X)^2}{a^2}$.*

Résumé sur les intégrales généralisées utiles pour les proba :

Soit $f : \mathbb{R} \to \mathbb{R}$ localement intégrable, c'est-à-dire telle que $\int_a^x f(t)dt$ existe pour tous a et x. On dit alors que l'intégrale généralisée $\int_a^\infty f(t)dt$ est convergente si $S = \lim_{x\to\infty} \int_a^x f(t)dt$ existe et S est alors la valeur de l'intégrale généralisée $\int_a^\infty f(t)dt$. De même, $\int_{-\infty}^a f(t)dt = \lim_{x\to-\infty} \int_x^a f(t)dt$ si cette limite existe. Et l'intégrale généralisée $\int_{-\infty}^\infty f(t)dt$ est convergente ssi $\int_{-\infty}^a f(t)dt$ et $\int_a^\infty f(t)dt$ sont toutes deux convergentes et on a alors $\int_{-\infty}^\infty f(t)dt = \int_{-\infty}^a f(t)dt + \int_a^\infty f(t)dt$.

Exemple de Riemann : Soit f définie par $f(x) = x^\alpha$ si $x \geq 1$ et $f(x) = 0$ si $x \leq 1$. On a clairement $\int_{-\infty}^1 f(t)dt = 0$ et si $\alpha \neq -1$, $\int_1^\infty f(t)dt = \lim_{x\to\infty} \int_1^x t^\alpha dt = \lim_{x\to\infty} \left[\frac{t^{\alpha+1}}{\alpha+1}\right]_1^x$ existe ssi $\alpha < -1$ et vaut $-\frac{1}{\alpha+1}$. Si $\alpha = -1$, $\int_1^\infty f(t)dt = \lim_{x\to\infty}[\ln t]_1^x = +\infty$ l'intégrale généralisée diverge. Et donc au total, l'intégrale généralisée $\int_{-\infty}^\infty f(t)dt$ est convergente ssi $\alpha < -1$ et vaut alors $-\frac{1}{\alpha+1}$.

Autres exemples :

1) Soit $f(x) = xe^{-x}\mathbb{1}_{]0,+\infty[}(x)$ alors
$\int_{-\infty}^\infty f(t)dt = \int_0^\infty f(t)dt = \lim_{x\to\infty} \int_0^x te^{-t}dt = \lim_{x\to\infty}([-te^{-t}]_0^x + \int_0^x e^{-t}dt) = 1$.

2) Soit $f(x) = 1/(x + 1)(x + 2)$ alors
$\int_0^\infty f(t)dt = \lim_{x\to\infty} \int_0^x (1/(t + 1) - 1/(t + 2))dt = \lim_{x\to\infty}[\ln(t + 1) - \ln(t + 2)]_0^x = \ln 2$.

Convergences. — On dit qu'une suite de v.a.r. Z_n converge en loi vers la v.a.r. Z si $F_{Z_n}(z) \to F_Z(z)$ pour tout $z \in \mathbb{R}$ ce que l'on note $Z_n \xrightarrow{\mathcal{L}} Z$ (on a par exemple vu que si Z_n est de loi binomiale $B(n, \lambda/n)$ alors $Z_n \xrightarrow{\mathcal{L}} Z$ où Z suit la loi de Poisson $\mathcal{P}(\lambda)$).

La suite Z_n converge en probabilité vers Z noté $Z_n \xrightarrow{P} Z$ si pour tout $\varepsilon > 0$, $P[|Z_n - Z| \geq \varepsilon] \to 0$.

La suite Z_n converge presque sûrement vers Z noté $Z_n \xrightarrow{p.s.} Z$ si $P[Z_n \to Z] = 1$.

On peut montrer que $Z_n \xrightarrow{p.s.} Z \Rightarrow Z_n \xrightarrow{P} Z \Rightarrow Z_n \xrightarrow{\mathcal{L}} Z$ (la convergence p.s. est plus forte que la convergence en probabilité et en loi).

Loi faible des grands nombres. — *Soit X_n une suite de v.a.r. indépendantes de moyenne m et de variance σ^2, alors $\frac{1}{n}\sum_{k=1}^n X_k = \bar{X}_n \xrightarrow{P} m$.*

Preuve : $\mathbb{E}(\bar{X}_n) = m$ grâce à la linéarité de l'espérance et $\mathrm{Var}(\bar{X}_n) = \sigma^2/n$ grâce à l'indépendance des X_n. D'après l'inégalité de Bienayme-Tchebicheff appliquée à $\bar{X}_n$ à $\varepsilon > 0$ fixé, $P[|\bar{X}_n - m| \geq \varepsilon] \leq \frac{\sigma^2}{n\varepsilon^2} \to 0\cdot$

Loi forte des grands nombres. — *Soit X_n une suite de v.a.r. indépendantes de moyenne m et de variance σ^2, alors $\frac{1}{n}\sum_{k=1}^n X_k = \bar{X}_n \xrightarrow{p.s.} m$.*

Théorème central limite. — *Sous les mêmes hypothèses $\frac{1}{\sqrt{n}\sigma}(\sum_{k=1}^n X_k - nm) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$.*

8- Couple de v.a.r. Loi conjointe. Covariance. — Dire que deux v.a.r. X et Y sont *indépendantes* revient à dire que $P[X \leq x \text{ et } Y \leq y] = P[X \leq x]P[Y \leq y] = F_X(x)F_Y(y)$ (les évènements décrits à l'aide de X sont indépendants de ceux décrits à l'aide de Y). Mais en général, deux v.a.r. quelconques X et Y ne sont pas indépendantes et donc la fonction de deux variables $F_{X,Y}(x,y) = P[X \leq x \text{ et } Y \leq y]$ appelée *fonction de répartition* du couple (X,Y) n'est pas le produit de $F_X(x)$ et de $F_Y(y)$.

Se donner la *loi conjointe* des v.a.r. X et Y ou la *loi du couple* (X,Y) revient à se donner la fonction de répartition $F_{X,Y} : \mathbb{R}^2 \to [0,1]$. Cette fonction contient toute l'information probabiliste sur le couple (X,Y), en particulier les lois de X et de Y appelées *lois marginales* du couple, on a en effet : $F_X(x) = \lim_{y\to+\infty} F_{X,Y}(x,y)$ et $F_Y(y) = \lim_{x\to+\infty} F_{X,Y}(x,y)$.

Couples de variables discrètes. — Dans le cas d'un couple (X,Y) de v.a.r. *discrètes*, par exemple à valeurs entières, si on note $\{(x_k, y_l), k, l \in \mathbb{N}\}$ l'ensemble des valeurs possibles du couple, il est clair que se donner $F_{X,Y}$ revient à se donner les probabilités $p_{k,l} = P[X = x_k \text{ et } Y = y_l]$ pour tous les $(k,l) \in \mathbb{N}^2$. Connaissant la loi du couple (X,Y), on retrouve les lois marginales, c'est-à-dire les suites $p_k = P[X = x_k]$ et $q_l = P[Y = y_l]$, par les formules $p_k = \sum_{l=0}^{+\infty} p_{k,l}$ et $q_l = \sum_{k=0}^{+\infty} p_{k,l}$.

Soit $g : \mathbb{R}^2 \to \mathbb{R}$ une fonction. Si on pose $Z = g(X,Y)$ et si $\mathbb{E}(Z)$ est définie, on a la formule
$$E(Z) = \sum_{k=0}^{+\infty}\sum_{l=0}^{+\infty} g(x_k, y_l)p_{k,l} = \sum_{l=0}^{+\infty}\sum_{k=0}^{+\infty} g(x_k, y_l)p_{k,l}$$

Couples à densité. — Si la fonction de répartition $F_{X,Y} : \mathbb{R}^2 \to [0,1]$ d'un couple (X,Y) de v.a.r. est continue, on dit que le couple est continu, si de plus les dérivées partielles de $F_{X,Y}(x,y)$ d'ordre 2 existent sauf pour des valeurs isolées de x et de y, on dit que le couple (X,Y) a une *densité* qui est la fonction $f_{X,Y}(x,y) = \dfrac{\partial^2 F_{X,Y}}{\partial x \partial y}(x,y) = \dfrac{\partial^2 F_{X,Y}}{\partial y \partial x}(x,y)$ là où elle est définie.

Soit $A \subset \mathbb{R}^2$ un domaine (*i.e.* un rectangle, un disque, plus généralement l'intérieur d'une courbe fermée) on a alors
$$P[(X,Y) \in A] = \int\int_A f_{X,Y}(x,y)\,dxdy.$$
Soit $g : \mathbb{R}^2 \to \mathbb{R}$ une fonction ayant des dérivées partielles. Si on pose $Z = g(X,Y)$ et si $\mathbb{E}(Z)$ est définie, on a la formule
$$E(Z) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f_{X,Y}(x,y)\,dxdy.$$
Les variables X et Y sont alors à densités qui sont respectivement $f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dy$ et $f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dx$ (densités marginales du couple). Dire que les v.a. d'un couple (X,Y) à densité sont indépendantes revient à dire que $f_{X,Y}(x,y) = f_X(x)f_Y(y)$. La somme $S = X + Y$ est à densité et on a $f_S(s) = \int_{-\infty}^{+\infty} f_{X,Y}(x, s-x)\,dx$.

Exercice : Soit (X,Y) couple de v.a.r. indépendantes de loi normale $N(0,1)$ (*i.e.* de densités respectives $f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ et $f_Y(y) = \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}$); soient R et Θ les coordonnées polaires correspondantes. Montrer que R a pour densité $re^{-\frac{r^2}{2}}\mathbb{1}_{\mathbb{R}^+}(r)$ et que Θ suit la loi uniforme sur $[0, 2\pi]$.

Couple (X, Y) **de loi uniforme sur un domaine** $D \subset \mathbb{R}^2$: c'est le cas où la densité est $f_{X,Y}(x,y) = \dfrac{\mathbb{1}_D(x,y)}{\text{Aire}(D)}$ (on suppose bien sûr $\text{Aire}(D) \neq 0$). Le couple (X, Y) représente alors les coordonnées d'un point "choisi au hasard de manière uniforme dans le domaine D". On peut vérifier que les coordonnées X et Y du couple sont indépendantes $\iff D = [a, b] \times [c, d]$ (*i.e.* un rectangle horizontal de $\mathbb{R}^2$).

Dans ce cas particulier, pour un domaine $A \subset \mathbb{R}^2$, on a $P[(X, Y) \in A] = \dfrac{\text{Aire}(A \cap D)}{\text{Aire}(D)}$.

Méthode de Monte-Carlo : C'est le nom donné à la méthode probabiliste de calcul des aires. Par exemple, on tire au hasard n fois (tirages indépendants) un point du carré $D = [-1, 1]^2$ et on compte le nombre de fois N_n où ce point est dans le disque $A = \{(x, y)|x^2 + y^2 \leq 1\}$. La loi forte des grands nombres nous dit qu'il est presque sûr que $\lim_{n\to\infty} N_n = P[(X, Y) \in A] = \pi/4$ où (X, Y) est le couple de densité $\dfrac{\mathbb{1}_D(x,y)}{4}$. D'où une méthode pour approcher π (avec une certaine probabilité d'erreur) !

Covariance. — Soit X et Y deux v.a.r. telles que $\mathbb{E}(X) = m_X$ et $\mathbb{E}(Y) = m_Y$ sont finies ainsi que $\text{Var}(X)$ et $\text{Var}(Y)$, la *covariance* de X et Y est le nombre

$$\text{Cov}(X, Y) = \mathbb{E}([X - m_X][Y - m_Y]) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

qui peut être de signe quelconque.

PROPRIÉTÉS. — *1)* $\text{Var}(X) = \text{Cov}(X, X)$.
2) $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$.
3) $|\text{Cov}(X, Y)| \leq \sigma(X)\sigma(Y)$ *(Inégalité de Cauchy-Schwarz)*

Corrélation. — Si les variables X et Y ne sont pas presque sûrement égales à une constante (*i.e.* $\sigma(X) \neq 0$ et $\sigma(Y) \neq 0$), on définit leur *coefficient de corrélation* $\rho(X, Y)$ par

$$\rho(X, Y) = \frac{|\text{Cov}(X, Y)|}{\sigma(X)\sigma(Y)} \in [-1, 1]$$

PROPRIÉTÉS. — *1)* X, Y *indépendantes* $\Rightarrow \text{Cov}(X, Y) = \rho(X, Y) = 0$ *(Attention : réciproque fausse!)*
2) $\rho(X, Y) = \pm 1 \iff \exists a \in \mathbb{R}^*$ et $b \in \mathbb{R}$ *tels que* $P[Y = aX + b] = 1$.

Loi conditionnelle (couples de variables discrètes). — Soit un couple (X, Y) de v.a.r. *discrètes* (par exemple à valeurs entières). On note encore $\{(x_k, y_k), k \in \mathbb{N}\}$ l'ensemble des valeurs possibles du couple et $p_{k,l} = P[X = x_k \text{ et } Y = y_l]$ sa loi. Il est naturel de considérer *la loi conditionnelle de* X : il s'agit de la donnée pour chaque y_l tel que $q_l = P[Y = y_l] \neq 0$ de la suite
$P[X = x_k|Y = y_l] = \frac{p_{k,l}}{q_l}$ qui est la loi d'une v.a.r. discrète X_l ; la donnée des espérances (si elles existent) $\mathbb{E}(X_l)$ est l'*espérance conditionnelle de* X connaissant Y et est notée $\mathbb{E}(X|Y = y_l)$. Si X et Y sont indépendantes $\mathbb{E}(X|Y = y_l) = \mathbb{E}(X)$ pour tout l.

Propriété. — Soit $f(y_l) = \mathbb{E}(X|Y = y_l)$ et $Z = f(Y)$ alors

$$\mathbb{E}(Z) = \sum_{l=0}^{\infty} \mathbb{E}(X|Y = y_l)q_l = \mathbb{E}(X).$$

Par exemple, si le couple (X, Y) représente la taille en centimètre $X \in [30, 250]$ et le poids en kilogramme $Y \in [1, 200]$ d'un individu tiré au hasard dans la population française, $\mathbb{E}(X|Y = l)$ représente la moyenne des tailles des individus de poids l. C'est évidemment dans ce cas une fonction croissante de l : les deux variables sont fortement corrélées (*i.e.* pas indépendantes du tout).

Corollaire. — On suppose de plus que X est une variable aléatoire entière. Notons $G_{X|Y=y_l}(t)$ la fonction génératrice de la loi de X sachant $Y = y_l$. Alors

$$G_X(t) = \sum_{l=0}^{\infty} q_l G_{X|Y=y_l}(t).$$

Appendice : Espaces probabilisés généraux. Variables aléatoires et couples de v.a.

On a vu comment on pouvait définir sur $\mathbb{N}$ (resp. sur l'ensemble des éléments Ω d'une suite $x_{k,k\in\mathbb{N}}$) une probabilité P à partir de la donnée d'une suite $p_{k,k\in\mathbb{N}}$ telle que $\sum p_k = 1$, l'ensemble des évènements étant $\mathcal{P}(\mathbb{N})$ (resp. $\mathcal{P}(\Omega)$). Ceci suffisait à définir convenablement la loi et l'espérance d'une v.a. à valeurs entières ou même à valeurs dans une suite de nombre réels.

On a vu aussi comment on pouvait définir sur $\mathbb{R}$ une notion de probabilité P à partir de la donnée d'une fonction de répartition $F : \mathbb{R} \to [0,1]$, l'ensemble des événements étant l'algèbre des événements réels $\mathcal{E} \subset \mathcal{P}(\mathbb{R})$. Cela suffit encore à définir la loi et l'espérance d'une v.a.r.

Il manque à cette présentation plusieurs choses que nous avons admises implicitement :

a) Préciser un espace de probabilité sur lequel sont définies les v.a. considérées.
b) Pouvoir dans tous les cas considérer des réunions ou des intersections d'une suite d'événements.
c) Pouvoir considérer des couples de v.a. définies sur un même espace de probabilité.

Ce qui suit est le résumé d'une présentation de la théorie des probabilités qui répond à ces contraintes.

Définition générale d'un espace probabilisé $(\Omega, \mathcal{T}, P)$: Soit Ω un ensemble et $\mathcal{T} \subset \mathcal{P}(\Omega)$ un ensemble de parties de Ω.

Tribu : On dit que $\mathcal{T}$ est une *tribu* ou une *σ-algèbre* si $\mathcal{T}$ vérifie les trois propriétés

(i) Ω et $\emptyset \in \mathcal{T}$ (ii) Si $A \in \mathcal{T}$ alors $A^c \in \mathcal{T}$

(iii) Pour toute *suite* $A_n \in \mathcal{T}$ on a $\cup_{n\in\mathbb{N}} A_n \in \mathcal{T}$ (et $\cap_{n\in\mathbb{N}} A_n \in \mathcal{T}$ grâce à (ii)).

Si la propriété (iii) n'était vraie que pour les suites *finies*, $\mathcal{T}$ ne serait qu'une *algèbre*. Une tribu ou σ-algèbre est donc une algèbre de parties qui a en plus la propriété des réunions de suites (d'où le σ ajouté...). Si $\Omega = \mathbb{R}$, l'algèbre $\mathcal{E}$ des événements réels que nous avons considérée auparavant n'est pas une tribu. Pour tout Ω, $\mathcal{P}(\Omega)$ est une tribu.

Espace probabilisable : Une tribu $\mathcal{T}$ étant fixée, le couple $(\Omega, \mathcal{T})$ est un *espace probabilisable*. Les éléments de $\mathcal{T}$ en sont les *événements*.

Espace probabilisé : Une probabilité P sur l'espace probabilisable $(\Omega, \mathcal{T})$ est une application $P : \mathcal{T} \to [0,1]$ vérifiant

(i) $P(\Omega) = 1$ (ii) Si $A \in \mathcal{T}$ alors $P(A^c) = 1 - P(A)$

(iii) Pour toute *suite* d'événements $A_n \in \mathcal{T}$ *disjoints* on a $P(\sum_{n\in\mathbb{N}} A_n) = \sum_0^\infty P(A_n)$.

Cette dernière propriété est la σ-additivité (plus forte que l'additivité finie).

Une probabilité P étant fixée sur l'espace probabilisable $(\Omega, \mathcal{T})$, le triplet $(\Omega, \mathcal{T}, P)$ est un *espace probabilisé* (ou encore un *espace de probabilité*).

L'espace probabilisable $(\mathbb{R}, \mathcal{B})$: Nous l'avons dit, l'algèbre $\mathcal{E}$ des événements réels qui semblait adaptée au concept de v.a.r. n'est pas une tribu, on doit donc ajouter à $\mathcal{E}$ de nouveaux événements pour avoir une tribu, c'est

La tribu borélienne $\mathcal{B}$: La plus petite tribu de parties $\mathbb{R}$ contenant les événements réels $\mathcal{E}$ (réunions disjointes finies d'intervalles) est notée $\mathcal{B}$ et appelée la *tribu borélienne*.

Une probabilité $P : \mathcal{B} \to [0,1]$ détermine une fonction de répartition $F(t) = P(]-\infty, t])$. Réciproquement on a le

Théorème. — *Etant donnée une fonction de répartition F, il existe une unique probabilité $P : \mathcal{B} \to [0,1]$ telle que $F(t) = P(]-\infty, t])$.*

Dans le cours, nous avons facilement défini $P(A)$ pour $A \in \mathcal{E}$, la difficulté est d'étendre P à la tribu $\mathcal{B}$ tout entière.

Variable aléatoire : Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé. Une fonction $X : \Omega \to \mathbb{R}$ est appelée variable aléatoire si pour tout $B \in \mathcal{B}$, $[X \in B] \in \mathcal{T}$.

La loi P_X de X est une probabilité sur $(\mathbb{R}, \mathcal{B})$ définie par $P_X(B) = P[X \in B]$ pour tout $B \in \mathcal{B}$.

Les v.a.r. définies sur un même espace de probabilité $(\Omega, \mathcal{T}, P)$ forment un espace vectoriel $\mathcal{V}$.

Espace probabilisé produit : Soit $(\Omega_1, \mathcal{T}_1)$ et $(\Omega_2, \mathcal{T}_2)$ deux espaces probabilisables. On note $\Omega = \Omega_1 \times \Omega_2$ et $\mathcal{T} \subset \mathcal{P}(\Omega)$ la plus petite tribu $\mathcal{T}$ (ou tribu engendrée par) contenant les $A_1 \times A_2$, $(A_1, A_2) \in \mathcal{T}_1 \times \mathcal{T}_2$. L'espace probabilisable $(\Omega, \mathcal{T})$ est le *produit des espaces probabilisables* $(\Omega_1, \mathcal{T}_1)$ et $(\Omega_2, \mathcal{T}_2)$. On peut aussi noter $\mathcal{T} = \mathcal{T}_1 \otimes \mathcal{T}_2$ et l'appeler la *tribu produit*.

Si P_1 et P_2 sont des probabilités respectivement sur $(\Omega_1, \mathcal{T}_1)$ et $(\Omega_2, \mathcal{T}_2)$, il existe (on peut le démontrer) une unique probabilité P sur $(\Omega, \mathcal{T})$, vérifiant $P(A_1 \times A_2) = P_1(A_1)P_2(A_2)$ pour tout $(A_1, A_2) \in \mathcal{T}_1 \times \mathcal{T}_2$.

On appelle P la *probabilité produit* et on la note $P = P_1 \otimes P_2$. L'espace probabilisé $(\Omega, \mathcal{T}, P)$ est l'*espace probabilisé produit* des espaces $(\Omega_1, \mathcal{T}_1, P_1)$ et $(\Omega_2, \mathcal{T}_2, P_2)$.

Couple de variables aléatoires Soient $(\Omega_i, \mathcal{T}_i, P_i)$ $i = 1, 2$ deux espaces probabilisés et $X_i : \Omega_i \to \mathbb{R}$ $i = 1, 2$ deux variables aléatoires. Le couple $X = (X_1, X_2)$ est aussi appelé *vecteur aléatoire*; c'est une application $X : \Omega \to \mathbb{R}^2$ où $\Omega = \Omega_1 \times \Omega_2$ qui a la propriété que si $B \in \mathcal{B} \otimes \mathcal{B}$ alors $[X \in B] \in \mathcal{T} = \mathcal{T}_1 \otimes \mathcal{T}_2$.

La loi P_X du couple (ou du vecteur aléatoire) X est une probabilité sur l'espace probabilisable $(\mathbb{R}^2, \mathcal{B} \otimes \mathcal{B})$ (*i.e.* le plan muni de la tribu borélienne) définie par $P_X(B) = P[X \in B]$ pour tout $B \in \mathcal{B} \otimes \mathcal{B}$.