11. Probabilités – Variables aléatoires discrètes

I - Préliminaires techniques

1) Ensembles dénombrables

Définition : un ensemble est dit *dénombrable* si et seulement s'il est en bijection avec $\mathbb{N}$ (on dit qu'il est *équipotent à* $\mathbb{N}$).

Premiers exemples

1) $\mathbb{N}^*$ est dénombrable ; l'ensemble des entiers pairs (*resp.* impairs) est dénombrable.

2) $\mathbb{Z}$ est dénombrable : on obtient une bijection f de $\mathbb{N}$ dans $\mathbb{Z}$ en posant pour tout q dans $\mathbb{N}$ $f(2q) = q$ et $f(2q+1) = -q-1$.

NB : pour montrer qu'une application est bijective, le plus simple est parfois d'exhiber sa bijection réciproque ! Sinon, on peut montrer que tout élément de l'ensemble d'arrivée admet un unique antécédent (souvent par analyse-synthèse, ce qui rejoint la première idée...). On peut aussi montrer qu'elle est surjective et injective.
Pour montrer qu'un ensemble est dénombrable, il existe des résultats plus généraux.

Lemme : toute partie infinie de $\mathbb{N}$ est dénombrable.

Théorème : toute partie infinie d'un ensemble dénombrable est dénombrable.

Corollaire : un ensemble E est *fini ou dénombrable* (on dit aussi *au plus dénombrable*) si et seulement s'il est en bijection avec une partie de $\mathbb{N}$ (*i.e.* si et seulement s'il existe une injection de E dans $\mathbb{N}$ ou encore si et seulement s'il existe une surjection de $\mathbb{N}$ dans E).

Corollaire : un ensemble non vide est fini ou dénombrable si et seulement s'il peut être décrit en extension sous la forme $\{x_n,\ n \in \mathbb{N}\}$.

Lemme : $\mathbb{N}^2$ est dénombrable.

Théorème : le produit cartésien de deux ensembles dénombrables est dénombrable.

Applications : pour tout m dans $\mathbb{N}^*$, $\mathbb{N}^m$ est dénombrable ; $\mathbb{Q}$ est dénombrable.

NB : $\mathbb{R}$ n'est pas dénombrable ; $\{0,1\}^{\mathbb{N}}$ non plus ! On peut montrer qu'ils sont équipotents.

2) Compléments sur les séries (*hors programme, résultats admis*)

a) Permutation des termes d'une série absolument convergente

Si la série numérique $\sum u_n$ est absolument convergente, alors pour toute permutation σ de $\mathbb{N}$ (*i.e.* bijection de $\mathbb{N}$ dans $\mathbb{N}$), $\sum u_{\sigma(n)}$ est également absolument convergente et a même somme.

Ce résultat permet de donner un sens à une somme indexée par un ensemble dénombrable, dès que la série des termes indexée selon un ordre arbitraire est absolument convergente.

b) Sommation par paquets

Soit $\sum u_n$ une série absolument convergente, I un ensemble fini ou dénombrable et $(P_i)_{i\in I}$ une partition de $\mathbb{N}$. Alors, pour tout i dans I, $S_i = \sum\limits_{n\in P_i} u_n$ a un sens (somme finie ou somme d'une série absolument convergente) ; $\sum\limits_{i\in I} S_i$ a également un sens et l'on a

$$\sum_{n=0}^{\infty} u_n = \sum_{i\in I} S_i = \sum_{i\in I}\left(\sum_{n\in P_i} u_n\right)$$

(d'où l'appellation de *sommation par paquets*).

Comme au paragraphe précédent, ce résultat se généralise à une somme indexée par un ensemble dénombrable autre que $\mathbb{N}$.

c) Suites doubles

On appelle *suite double* une suite $(u_{p,q})$ indexée par les couples (p,q) de $\mathbb{N}^2$.

<u>**Cas des suites doubles à termes dans $\mathbb{R}^+$**</u>

Si les $u_{p,q}$ sont dans $\mathbb{R}^+$, les assertions suivantes sont équivalentes :

- pour tout p dans $\mathbb{N}$, la série $\sum\limits_{q\geq 0} u_{p,q}$ converge, de somme σ_p, et la série $\sum\limits_{p\geq 0} \sigma_p$ converge
- pour tout q dans $\mathbb{N}$, la série $\sum\limits_{p\geq 0} u_{p,q}$ converge, de somme τ_q, et la série $\sum\limits_{q\geq 0} \tau_q$ converge
- la série de terme général $\sum\limits_{p+q=n} u_{p,q}$ converge
- la suite de terme général $\sum\limits_{0\leq p,q\leq N} u_{p,q}$ converge

Lorsque ces assertions sont vraies, on a

$$\sum_{p=0}^{\infty}\left(\sum_{q=0}^{\infty} u_{p,q}\right) = \sum_{q=0}^{\infty}\left(\sum_{p=0}^{\infty} u_{p,q}\right) = \sum_{n=0}^{\infty}\left(\sum_{p+q=n} u_{p,q}\right) = \lim_{N\to\infty} \sum_{0\leq p,q\leq N} u_{p,q}.$$

On dit alors que la suite double $(u_{p,q})$ est *sommable* et la valeur commune ci-dessus est notée $\sum\limits_{(p,q)\in\mathbb{N}^2} u_{p,q}$.

Les deux premières écritures se généralisent au cas de l'indexation par un produit cartésien de deux ensembles dénombrables, on parle alors de *famille sommable* $(u_{i,j})_{(i,j)\in I\times J}$.

<u>**Cas des suites doubles à termes dans $\mathbb{C}$**</u>

Si les $u_{p,q}$ sont dans $\mathbb{C}$, la suite double $(u_{p,q})$ est dite sommable si et seulement si $(|u_{p,q}|)$ l'est.
Auquel cas les quatre expressions ci-dessus ont un sens et ont la même valeur, toujours notée $\sum\limits_{(p,q)\in\mathbb{N}^2} u_{p,q}$.

II - Espaces probabilisés

1) Notion de tribu – Notations et vocabulaire

Lorsqu'on veut généraliser la notion de probabilité à un univers Ω infini, on s'aperçoit que l'ensemble des événements ne peut pas toujours être $\mathcal{P}(\Omega)$ tout entier (par exemple si les calculs de probabilités impliquent des calculs d'intégrales).

Définition : si Ω est un ensemble, on appelle *tribu sur* Ω toute partie $\mathcal{A}$ de $\mathcal{P}(\Omega)$ telle que :

- $\Omega \in \mathcal{A}$
- <u>*$\mathcal{A}$ est stable par passage au complémentaire*</u> : pour tout A de $\mathcal{A}$, $\overline{A} = \Omega\backslash A \in \mathcal{A}$
- <u>*$\mathcal{A}$ est stable par union dénombrable*</u> : pour toute suite $(A_n)_{n\in\mathbb{N}}$ d'éléments de $\mathcal{A}$, $\bigcup\limits_{n=0}^{\infty} A_n \in \mathcal{A}$.

Un *espace probabilisable* est un couple $(\Omega, \mathcal{A})$ où $\mathcal{A}$ est une tribu sur l'ensemble Ω.

<u>*Exemples*</u> : $\mathcal{A} = \{\varnothing, \Omega\}$ est une tribu sur Ω, *la tribu grossière sur* Ω ; $\mathcal{P}(\Omega)$ est une tribu sur Ω.

Ces notions sont utilisés pour modéliser une *expérience aléatoire*.
Les éléments de Ω sont *les issues* (ou *réalisations*, ou encore *résultats possibles*).
Les éléments de $\mathcal{A}$ sont *les événements* (ce sont des parties de Ω, c'est-à-dire des ensembles de résultats possibles). Les événements élémentaires sont ceux de la forme $\{\omega\}$ où $\omega \in \Omega$.
$\varnothing$ est *l'événement impossible*, Ω *l'événement certain* ; pour $A \in \mathcal{A}$, $\overline{A}$ est *l'événement contraire de* A.
Pour $(A,B) \in \mathcal{A}^2$, $A\cup B$ est appelé *événement "A ou B"*, $A\cap B$ est appelé *événement "A et B"*.
On dit que A et B sont des *événements incompatibles* si et seulement si $A\cap B = \varnothing$, que A *implique* B si et seulement si $A \subset B$.

Un système complet fini d'événements est une famille finie $(A_1, \ldots, A_m)$ d'événements tels que :

- les A_n sont incompatibles deux à deux : $i \neq j \Rightarrow A_i \cap A_j = \varnothing$
- l'union des A_n est l'univers tout entier : $\bigcup\limits_{n=1}^{m} A_n = \Omega$.

Un système complet dénombrable d'événements est une suite $(A_n)_{n\in\mathbb{N}}$ d'événements tels que :

- les A_n sont incompatibles deux à deux : $i \neq j \Rightarrow A_i \cap A_j = \varnothing$
- l'union des A_n est l'univers tout entier : $\bigcup\limits_{n=0}^{\infty} A_n = \Omega$.

NB : c'est la version probabiliste des partitions ; on parle communément de *système complet d'événements* dans la mesure où la notation permet de distinguer les deux cas précédents.

2) Notion d'espace probabilisé

Définition : si Ω est un ensemble et $\mathcal{A}$ une tribu sur Ω, on appelle *probabilité sur* $(\Omega, \mathcal{A})$ une application $P : \mathcal{A} \to [0, 1]$ telle que :

- $P(\Omega) = 1$
- <u>σ-additivité</u> : pour toute suite $(A_n)_{n\in\mathbb{N}}$ d'événements incompatibles 2 à 2,

$$P\left(\bigcup_{n=0}^{\infty} A_n\right) = \sum_{n=0}^{\infty} P(A_n).$$

Le triplet $(\Omega, \mathcal{A}, P)$ est dit *espace probabilisé.*

- un événement de probabilité 1 est dit *presque sûr* ou *presque certain.*
- un événement de probabilité 0 est dit *négligeable* ou *presque impossible.*

NB : la propriété de σ-additivité contient la **convergence** de la série $\sum P(A_n)$, pour toute suite $(A_n)_{n\in\mathbb{N}}$ d'événements incompatibles 2 à 2.
Elle est bien entendu vraie aussi pour toute famille **finie** d'événements incompatibles 2 à 2 (qu'il suffit de compléter par des $\varnothing$ pour obtenir une suite, sans pour autant changer la réunion ni la somme des probabilités, voir § **4)**). On parle alors d'*additivité.*

3) Germe de probabilité sur $\left(\Omega, \mathcal{P}(\Omega)\right)$ lorsque Ω est au plus dénombrable

Théorème : si Ω est au plus dénombrable et si $p : \Omega \to [0, 1]$ vérifie $\sum\limits_{\omega\in\Omega} p(\omega) = 1$, alors il existe une unique probabilité P sur $\left(\Omega, \mathcal{P}(\Omega)\right)$ telle que $P(\{\omega\}) = p(\omega)$. P est définie par :

$$\forall A \in \mathcal{P}(\Omega) \quad P(A) = \sum_{\omega\in A} p(\omega).$$

<u>*Dém.*</u> La preuve — banale dans le cas fini — est hors programme en PSI dans le cas dénombrable ; elle utilise la notion de famille sommable, cf. **I-2)**.

<u>Le cas d'**équiprobabilité** sur un univers fini</u> :

Si Ω est un ensemble **fini non vide** et $\mathcal{A} = \mathcal{P}(\Omega)$, $P : A \mapsto \dfrac{\#A}{\#\Omega}$ est une probabilité sur $(\Omega, \mathcal{A})$, dite *probabilité uniforme sur* Ω ; c'est un cas particulier du théorème précédent, caractérisé par le fait que tous les événements élémentaires ont même probabilité (cela est impossible sur un univers infini !). Dans ce cas, la probabilité d'un événement est le quotient du nombre de cas “favorables” par le nombre de cas possibles.

<u>*Exemples*</u> : parmi les situations classiques d'équiprobabilité, citons le lancer d'une pièce équilibrée, d'un dé non pipé, le tirage “au hasard” d'un carte dans un jeu, d'une boule indiscernable des autres dans une urne. . .

4) Premières propriétés

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé.

- Pour toute suite $(A_n)_{n\in\mathbb{N}}$ d'événements, $\overline{\bigcup_{n=0}^{\infty} A_n} = \bigcap_{n=0}^{\infty} \overline{A_n}$ et $\overline{\bigcap_{n=0}^{\infty} A_n} = \bigcup_{n=0}^{\infty} \overline{A_n}$;
- $\mathcal{A}$ est stable par intersection dénombrable.
- $\mathcal{A}$ est stable par union finie et par intersection finie.
- $\mathcal{A}$ est stable par *différence* ($A\backslash B = A \cap \overline{B}$: A et non B) et par *différence symétrique* :
$$A\Delta B = (A\backslash B) \cup (B\backslash A) = (A \cup B) \setminus (A \cap B) \text{ (ou **exclusif** : soit } A\text{, soit } B)$$
- Si $A_0, \ldots, A_m$ sont des événements incompatibles 2 à 2, $P\left(\bigcup_{n=0}^{m} A_n\right) = \sum_{n=0}^{m} P(A_n)$.
- $P(\varnothing) = 0$ et $\forall A \in \mathcal{A} \quad P\left(\overline{A}\right) = 1 - P(A)$.
- Pour A, B dans $\mathcal{A}$ tels que $A \subset B$, $P(B\backslash A) = P(B) - P(A)$.
- $\forall (A,B) \in \mathcal{A}^2 \quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- *Croissance* : pour A, B dans $\mathcal{A}$ tels que $A \subset B$, $P(A) \leq P(B)$.
- *Continuité croissante* : si $(A_n)_{n\in\mathbb{N}}$ est une suite croissante d'événements ($\forall n \quad A_n \subset A_{n+1}$), alors
$$\lim_{n\to\infty} P(A_n) = P\left(\bigcup_{n=0}^{\infty} A_n\right).$$
- *Continuité décroissante* : si $(A_n)_{n\in\mathbb{N}}$ est une suite décroissante d'événements ($\forall n \quad A_{n+1} \subset A_n$),
$$\lim_{n\to\infty} P(A_n) = P\left(\bigcap_{n=0}^{\infty} A_n\right).$$
- *Sous-additivité* : si $(A_n)_{n\in\mathbb{N}}$ est une suite d'événements, alors
$$P\left(\bigcup_{n=0}^{\infty} A_n\right) \leq \sum_{n=0}^{\infty} P(A_n)$$
(où cette dernière somme vaut par convention $+\infty$ si la série diverge...).

  Pour une famille finie $A_0, \ldots, A_m$ d'événements, on a $P\left(\bigcup_{n=0}^{m} A_n\right) \leq \sum_{n=0}^{m} P(A_n)$.

5) Répétition une infinité de fois d'une expérience à deux issues

De nombreux contextes variés correspondent à ce modèle, typiquement : "on lance une infinité de fois une pièce de monnaie" (c'est bien sûr une expérience virtuelle, même pour une infinité dénombrable !).

L'univers Ω est alors, $\{\mathbf{P}, \mathbf{F}\}^{\mathbb{N}^*}$ (où **P** désigne "pile" et **F** "face" et où l'on numérote les lancers 1, 2,...). Ω n'est pas dénombrable et cela pose des problèmes de définition d'une tribu et d'une probabilité "raisonnables". Par exemple, dans le cas d'une pièce équilibrée, on souhaite que la probabilité d'un événement où l'on impose le résultat d'exactement n lancers soit $\frac{1}{2^n}$; alors, si l'on avait pour tribu $\mathcal{P}(\Omega)$, pour respecter la continuité décroissante, il faudrait que la probabilité de tout événement élémentaire soit nulle... On ne peut donc pas généraliser la construction du § **3)** au cas non dénombrable. Hélas — même si c'est possible — il est délicat (et hors programme !) de construire une tribu qui permette de modéliser convenablement la situation.

C'est pourquoi, en pratique, on n'explicite pas la tribu utilisée et l'on se contente d'opérations sur des événements "raisonnables" comme ceux évoqués ci-dessus.

Le lecteur scrupuleux remarquera que c'est souvent le cas dans les problèmes de probabilité.

Par exemple, soit l'événement A : "obtenir pile à tous les lancers". Ainsi $A = \bigcap_{n=1}^{\infty} A_n$ où A_n est l'événement "les n premiers lancers donnent pile", de sorte que $P(A_n) = \frac{1}{2^n}$. Il est clair que la suite (A_n) est décroissante et donc $P(A) = \lim_{n\to\infty} P(A_n) = 0$: A est un événement presque impossible (on s'en doutait !). $\overline{A}$ ("obtenir au moins une fois face") est donc presque certain.

6) Conditionnement

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé.

a) Probabilité conditionnelle

Théorème et définition : soient A et B deux événements, tels que $P(A) > 0$.

On appelle *probabilité de B sachant A* le réel
$$P_A(B) = \frac{P(A \cap B)}{P(A)}, \text{ noté aussi } P(B|A).$$
L'application P_A est une probabilité sur $(\Omega, \mathcal{A})$, parfois appelée *probabilité conditionnée à A*.

NB : on peut "justifier" cette appellation par analogie avec le calcul de la fréquence de réalisation de B sachant que A est réalisé, sur un grand nombre N de répétitions de l'expérience (approche "fréquentiste" des probabilités) :
$$\frac{n_{A\cap B}}{n_A} = \frac{\frac{n_{A\cap B}}{N}}{\frac{n_A}{N}} \sim \frac{P(A \cap B)}{P(A)}.$$

Exemple : je lance 2 dés (cubiques) équilibrés, en les maintenant cachés.
Quelle est la probabilité qu'il y ait au moins un 6 ?
Je découvre l'un des dés qui montre un 5, quelle est la probabilité que l'autre soit un 6 ?

Convention importante

Dans toute la suite on convient que $P(B|A)\, P(A) = 0$ lorsque $P(A) = 0$ (alors que $P(B|A)$ n'a pas de sens *a priori*).

b) Formule des probabilités composées

Théorème : pour tous événements A, B on a : $P(A \cap B) = P(A)\, P(B|A)$.

Cette formule absolument banale est très utile en pratique ! Voir par exemple l'expérience emblématique de tirage **sans remise** : on extrait successivement sans remise deux boules d'une urne contenant 7 boules blanches et 4 noires ; quelle est la probabilité d'obtenir deux boules noires ? On cherche ainsi la probabilité de l'événement $N_2 \cap N_1$, où N_k est l'événement "la k-ième boule est noire". Alors, en supposant les tirages équiprobables :
$$P(N_1 \cap N_2) = P(N_1)\, P(N_2|N_1) = \frac{4}{11} \cdot \frac{3}{10} = \frac{6}{55} \approx 0,11.$$

Une récurrence immédiate donne la version générale :

Théorème : pour toute famille finie d'événements $(A_1, \dots, A_m)$,
$$P\left(\bigcap_{n=1}^{m} A_n\right) = P(A_1)\, P(A_2|A_1)\, P(A_3|A_1 \cap A_2) \cdots P(A_m|A_1 \cap \cdots \cap A_{m-1})$$
(ce produit étant nul dès que l'une des intersections $A_1 \cap \ldots \cap A_k$ a une probabilité nulle !).

Exemple : dans le cas précédent, la probabilité d'obtenir 4 boules noires est $\frac{4}{11} \cdot \frac{3}{10} \cdot \frac{2}{9} \cdot \frac{1}{8}$ et celle d'obtenir au moins 5 boules noires vaut 0 !

c) Formule des probabilités totales

Pour rassembler les énoncés correspondant aux cas fini et dénombrable, on note $I = [\![1, m]\!]$ ou $I = \mathbb{N}$.

Définition : $(A_n)_{n\in I}$ est un système *quasi complet* (ou *presque complet*) d'événements si et seulement si les A_n sont incompatibles deux à deux et $P\left(\bigcup_{n\in I} A_n\right) = 1$.

Théorème : soit $(A_n)_{n\in I}$ un système complet (*resp.* quasi complet) d'événements ; alors
$$\forall B \in \mathcal{A} \quad P(B) = \sum_{n\in I} P(B|A_n)\, P(A_n) \quad \textit{(formule des probabilités totales)}$$
(avec toujours la convention officielle : $P(B|A_n)\, P(A_n) = 0$ lorsque $P(A_n) = 0$).
Là encore, dans le cas dénombrable, le résultat comprend la convergence de la série !

NB : cette formule permet de calculer la probabilité d'un événement, connaissant ses probabilités conditionnées aux événements d'un système complet (ou quasi complet), ce qui est une situation assez répandue.
Appliquée à plusieurs événements, elle conduit souvent à des égalités matricielles.

Exemple 1 (toujours le même) : en utilisant le système complet d'événements $\left(N_1, \overline{N_1}\right)$ j'obtiens

$$P(N_2) = P(N_2|N_1)\,P(N_1) + P\left(N_2|\overline{N_1}\right)P\left(\overline{N_1}\right) = \frac{3}{10}\cdot\frac{4}{11} + \frac{4}{10}\cdot\frac{7}{11} = \frac{4}{11}.$$

Exemple 2 : on considère un groupe de $2q$ personnes ; le nombre k de femmes vérifie $q \leq k \leq 2q$ et les différentes valeurs de k sont équiprobables. On choisit une personne au hasard. Quelle est la probabilité que ce soit une femme ?

d) Formule de Bayes

De nouveau des formules banales, compte tenu des résultats précédents, mais formules très utiles aussi, car permettant de “remonter dans le temps” puisqu'elles peuvent donner une probabilité conditionnelle dans l'ordre inverse de l'ordre “chronologique” des expériences... De son temps, la formule de Bayes s'appelait “théorème de probabilité des causes".

Théorème : si A, B sont deux événements de probabilité non nulle, alors

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} \quad \textit{(formule de Bayes)}$$

Et son corollaire compte tenu de la formule des probabilités totales :

Théorème : soit $(A_n)_{n\in I}$ un système complet (*resp.* quasi complet) d'événements.
Si B est un événement de probabilité non nulle, alors

$$\forall i \in I \quad P(A_i|B) = \frac{P(B|A_i)\,P(A_i)}{\sum\limits_{n\in I} P(B|A_n)\,P(A_n)} \quad \textit{(formule de Bayes)}$$

Exemple 1 (suite de l'exemple 2 du **c)**) : la personne choisie étant une femme, quelle est la probabilité que le groupe soit exclusivement composé de femmes ?

Exemple 2 : on dispose de 100 dés dont 80 sont équilibrés et 20 sont pipés au point qu'ils donnent un 6 avec une probabilité de 1/2. On lance un dé choisi au hasard et on obtient 6. Quelle est la probabilité que ce dé soit pipé ?

7) Indépendance

Soit toujours $(\Omega, \mathcal{A}, P)$ un espace probabilisé.

Définition : deux événements A et B sont *indépendants* si et seulement si $P(A\cap B) = P(A)\,P(B)$ (on remarque que cette relation est symétrique...).

NB : dans le cas où $P(A) \neq 0$, A et B sont indépendants si et seulement si $P(B|A) = P(B)$, ce qui semble intuitivement raisonnable. De même, si $P(A) \neq 1$, A et B sont indépendants si et seulement si $P\left(B|\overline{A}\right) = P(B)$.

Il faut bien comprendre que l'on ne **démontre** pas que deux événements sont indépendants par du “baratin” sur la réalisation de telle ou telle expérience, mais en vérifiant l'égalité ci-dessus.

En revanche, c'est dans la définition même de la probabilité que l'on peut faire en sorte que deux événements intuitivement indépendants le soient formellement.

Définition : soit $(A_n)_{1\leq n\leq m}$ une famille finie d'événements.

* les A_n, $1 \leq n \leq m$, sont *indépendants deux à deux* si et seulement si

$$\forall (i,j) \in [\![1,m]\!]^2 \quad i \neq j \Rightarrow P(A_i \cap A_j) = P(A_i)\,P(A_j)\,.$$

* les A_n, $1 \leq n \leq m$, sont *mutuellement indépendants* si et seulement si, pour tout sous-ensemble I de $[\![1,m]\!]$,

$$P\left(\bigcap_{i\in I} A_i\right) = \prod_{i\in I} P(A_i)\,.$$

NB : bien entendu, l'indépendance mutuelle implique l'indépendance 2 à 2, mais la réciproque est fausse. Voir l'exemple ci-après.

Exemple : on lance deux fois une pièce équilibrée. Soient les événements A : "les deux résultats sont différents", B : "obtenir face au premier lancer", C : "obtenir pile au second lancer". Montrer que A, B et C sont indépendants deux à deux mais pas mutuellement indépendants.

Propriété : si les A_n, $1 \leq n \leq m$, sont mutuellement indépendants, alors pour tout sous-ensemble I de $[\![1, m]\!]$ et tout j de $[\![1, m]\!]$ n'appartenant pas à I, A_j et $\bigcap_{i \in I} A_i$ sont indépendants.

Propriété : si A et B sont indépendants, alors A et $\overline{B}$ sont indépendants, de même que $\overline{A}$ et $\overline{B}$.

Propriété : si les A_n, $1 \leq n \leq m$, sont mutuellement indépendants, alors toute famille $(B_n)_{1 \leq n \leq m}$ où, pour tout n, $B_n = A_n$ ou $B_n = \overline{A_n}$ est aussi une famille d'événements mutuellement indépendants.

III - Variables aléatoires discrètes

1) Généralités

a) Définitions et notations

Une *variable aléatoire discrète* X sur un espace probabilisable $(\Omega, \mathcal{A})$ est une application de Ω dans un ensemble E, telle que :

1) l'ensemble $X(\Omega)$ des valeurs prises par X est fini ou dénombrable ;

2) pour tout **élément** x de $X(\Omega)$, l'ensemble des antécédents de x est un événement, autrement dit

$$\forall x \in X(\Omega) \quad X^{-1}(\{x\}) \in \mathcal{A}.$$

NB : la propriété **2** est évidemment vérifiée lorsque $\mathcal{A} = \mathcal{P}(\Omega)$!

Si $(\Omega, \mathcal{A}, P)$ est un espace probabilisé, une variable aléatoire discrète sur $(\Omega, \mathcal{A})$ est aussi appelée variable aléatoire discrète sur $(\Omega, \mathcal{A}, P)$.

Lorsque $E \subset \mathbb{R}$, on parle de *variable aléatoire réelle discrète.*
Lorsque $E \subset \mathbb{R}^d$, on parle de *vecteur aléatoire réel discret.*

Propriété : pour toute **partie** U de $X(\Omega)$, $X^{-1}(U)$ est un événement.

Notations : pour x **élément** de $X(\Omega)$, l'événement $X^{-1}(\{x\})$ est aussi noté $(X = x)$ ou $\{X = x\}$; pour U **partie** de $X(\Omega)$, l'événement $X^{-1}(U)$ est aussi noté $(X \in U)$ ou $\{X \in U\}$.

Exemple 1 (cas fini où $\mathcal{A} = \mathcal{P}(\Omega)$) : on lance deux dés et l'on associe à chaque issue la somme des résultats des deux dés.

Exemple 2 (cas où $X(\Omega)$ est dénombrable) : **temps d'attente**

Dans le cas (classique) de la répétition infinie d'une expérience, ladite expérience pouvant se solder par la survenue d'un événement A. On appelle *temps d'attente de l'événement* A la variable aléatoire T qui, à chaque suite de résultats, associe le rang de la première occurrence de A, avec la convention que ce rang vaut ∞ si A ne se produit jamais. Ainsi T prend ses valeurs dans $\mathbb{N}^* \cup \{\infty\}$, qui est bien dénombrable. Il est beaucoup plus délicat de démontrer la propriété **2** de la définition, d'autant qu'on n'a déjà pas (en PSI) les outils pour expliciter une tribu "raisonnable" !
Normalement, les énoncés **diront** que ce type d'application est une variable aléatoire.

Exemple 3 (cas général) : **variable aléatoire indicatrice**

Si $A \in \mathcal{A}$, l'application $\mathbf{1}_A : \Omega \to \{0, 1\}$ qui à ω associe 1 si $\omega \in A$, 0 sinon, est une variable aléatoire discrète sur $(\Omega, \mathcal{A})$.

Variables aléatoires sur $\big(\Omega, \mathcal{P}(\Omega)\big)$ lorsque Ω est fini ou dénombrable

Dans ce cas toute application de Ω dans E est une variable aléatoire discrète !

b) Loi d'une variable aléatoire discrète

Théorème et définition : soit une variable aléatoire discrète X sur un espace probabilisé $(\Omega, \mathcal{A}, P)$.
La loi de X est l'application $P_X : \begin{array}{ccc} \mathcal{P}\left(X(\Omega)\right) & \to & [0,1] \\ U & \mapsto & P(X \in U) \end{array}$.
P_X est ainsi une probabilité sur $\left(X(\Omega), \mathcal{P}(X(\Omega))\right)$.
La loi de X est entièrement déterminée par la famille $\left(P(X = x)\right)_{x \in X(\Omega)}$.

La loi de X est appelée par certains auteurs “loi de probabilité de X”.
On appelle souvent “loi de X” la famille $\left(P(X = x)\right)_{x \in X(\Omega)}$; on remarquera que la donnée de cette famille comprend l'ensemble $X(\Omega)$ et les valeurs des $P(X = x)$ pour tous les x de $X(\Omega)$.

Propriété : soit I un ensemble fini ou dénombrable. Si X, variable aléatoire discrète sur $(\Omega, \mathcal{A}, P)$, prend ses valeurs dans $\{x_i,\ i \in I\}$ (où les x_i sont distincts 2 à 2), alors les $(X = x_i)$, $i \in I$, forment un système complet d'événements. En particulier, on a

$$\sum_{i \in I} P(X = x_i) = 1.$$

Le théorème suivant fournit en quelque sorte une réciproque à cette propriété banale.

Théorème : soit I un ensemble fini ou dénombrable. Si $X : \Omega \to E$ prend ses valeurs dans $\{x_i,\ i \in I\}$ (où les x_i sont distincts 2 à 2) et si $(p_i)_{i \in I}$ est une famille de réels de $[0,1]$ telle que $\sum_{i \in I} p_i = 1$, alors il existe un espace probabilisé $(\Omega, \mathcal{A}, P)$ tel que

$$\forall i \in I \quad P(X = x_i) = p_i.$$

Dém. Banale pour I fini, hors programme pour I dénombrable.

Ce théorème a pour avantage (appréciable) de permettre d'étudier une variable aléatoire définie par sa loi, sans avoir à expliciter l'espace probabilisé correspondant.

c) Fonction de répartition d'une variable aléatoire réelle discrète

Définition : soit une variable aléatoire **réelle** discrète X sur un espace probabilisé $(\Omega, \mathcal{A}, P)$.
La fonction de répartition de X est l'application $F_X : \begin{array}{ccc} \mathbb{R} & \to & [0,1] \\ x & \mapsto & P(X \leq x) \end{array}$.

Propriétés : soit F_X la fonction de répartition d'une variable aléatoire réelle discrète X.

a) F_X est croissante sur $\mathbb{R}$

b) $\lim\limits_{-\infty} F_X = 0$ et $\lim\limits_{+\infty} F_X = 1$

c) Si X prend ses valeurs dans $\{x_i,\ i \in I\}$, où I est fini ou dénombrable et les x_i distincts 2 à 2, alors

$$\forall x \in \mathbb{R} \quad F_X(x) = \sum_{i \in I_x} P(X = x_i) \quad \text{où} \quad I_x = \{i \in I \ /\ x_i \leq x\}$$

NB : dans le cas d'une variable aléatoire à valeurs dans $\mathbb{N}$ (cas très fréquent) on a

$$\forall x < 0 \quad F_X(x) = 0 \quad \text{et} \quad \forall k \in \mathbb{N} \quad \forall x \in [k, k+1[\quad F_X(x) = \sum_{i=0}^{k} P(X = i).$$

On peut penser que, dans le cas discret, la fonction de répartition n'apporte rien puisqu'elle se déduit de la loi de X. Mais, dans le cas continu (hors programme), c'est l'outil essentiel…

De plus, dans le cas d'une variable aléatoire à valeurs dans $\mathbb{N}$, la fonction de répartition se calcule parfois naturellement et l'on en **déduit** la loi de X :

$$\forall n \in \mathbb{N} \quad P(X = n) = F_X(n) - F_X(n-1).$$

Exemple : une urne contient N boules indiscernables numérotées de 1 à N ($N \geq 3$). On en extrait simultanément 3 et l'on note X le plus grand des 3 numéros obtenus. Alors $X(\Omega) = [\![3, N]\!]$ et

$$\forall n \in [\![3, N]\!] \quad F_X(n) = P(X \leq n) = \frac{\binom{n}{3}}{\binom{N}{3}} \quad \text{d'où} \quad P(X = n) = \frac{\binom{n}{3} - \binom{n-1}{3}}{\binom{N}{3}}.$$

2) Quelques lois usuelles

a) Loi uniforme

Soit E un ensemble fini non vide.
Une variable aléatoire X suit *la loi uniforme sur* E si et seulement si $X(\Omega) = E$ et

$$\forall x \in E \quad P(X = x) = \frac{1}{\#E}.$$

On écrit alors $X \hookrightarrow \mathcal{U}(E)$.

Cette loi correspond à l'équiprobabilité d'obtenir les différentes valeurs possibles.

b) Loi de Bernoulli *(pas de "i" avant les deux "l")*

Soit $p \in [0,1]$.
Une variable aléatoire X suit *la loi de Bernoulli de paramètre* p si et seulement si $X(\Omega) = \{0,1\}$ et

$$P(X = 1) = p \quad \text{et} \quad P(X = 0) = 1 - p.$$

On écrit alors $X \hookrightarrow \mathcal{B}(p)$.

C'est le modèle des expériences à deux résultats, souvent appelés *succès* $(X = 1)$ et *échec* $(X = 0)$. Une telle expérience est dite *expérience de Bernoulli.*

Noter que l'univers "ambiant" ne se réduit pas nécessairement à deux issues, comme dans le jeu de "pile ou face". Le succès peut par exemple être "obtenir un 6 en lançant un dé", auquel cas l'événement $(X = 1)$ est $\{6\}$, tandis que $(X = 0) = \{1,2,3,4,5\}$.

c) Loi binomiale *(pas d'accent circonflexe)*

Soient $n \in \mathbb{N}^*$ et $p \in [0,1]$.
Une variable aléatoire X suit *la loi binomiale de paramètres* n, p si et seulement si $X(\Omega) = [\![0,n]\!]$ et

$$\forall k \in [\![0,n]\!] \quad P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

On écrit alors $X \hookrightarrow \mathcal{B}(n,p)$.

C'est le modèle du nombre de succès lors de la réalisation de n expériences de Bernoulli indépendantes, de même paramètre p (pour $n = 1$, on retrouve la loi de Bernoulli $\mathcal{B}(p)$!).

Par exemple : lancers consécutifs d'une même pièce, d'un même dé (ou de plusieurs, mais ayant les mêmes caractéristiques !) ; tirages dans une urne **avec remise**. . . .

d) Loi géométrique

Soit $p \in]0,1[$.
Une variable aléatoire X suit *la loi géométrique de paramètre* p si et seulement si $X(\Omega) = \mathbb{N}^*$ et

$$\forall k \in \mathbb{N}^* \quad P(X = k) = p(1-p)^{k-1}.$$

On écrit alors $X \hookrightarrow \mathcal{G}(p)$.

C'est le modèle du rang du premier succès lors de la réalisation d'une infinité d'expériences de Bernoulli indépendantes, de même paramètre p. On jette alors un voile pudique sur le cas où ledit premier succès ne surviendrait jamais, ce qui est presque impossible (on dit que X est définie *presque sûrement* plutôt que d'ajouter ∞ parmi les valeurs possibles, cela pour les calculs ultérieurs : cf. **V** et **VI**).

Propriété : si $X \hookrightarrow \mathcal{G}(p)$, alors, pour tout $(n,k) \in \mathbb{N}^2$, $P(X > n + k \mid X > n) = P(X > k)$.

On dit que la loi géométrique est une loi *sans mémoire.*

Cette propriété est caractéristique des lois géométriques :

Théorème : soit X une variable aléatoire à valeurs dans $\mathbb{N}^*$, telle que : $\forall k \in \mathbb{N}^* \quad P(X = k) > 0$.

Si l'on a :

$$\forall (n,k) \in \mathbb{N}^2, P(X > n + k \mid X > n) = P(X > k),$$

alors X suit une loi géométrique.

e) Loi de Poisson *(avec une majuscule)*

Soit $\lambda \in \mathbb{R}^{+*}$.
Une variable aléatoire X suit *la loi de Poisson de paramètre* λ si et seulement si $X(\Omega) = \mathbb{N}$ et

$$\forall k \in \mathbb{N} \quad P(X = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}.$$

On écrit alors $X \hookrightarrow \mathcal{P}(\lambda)$.

Cette loi est "abstraite" au sens où elle ne découle pas d'un calcul de probabilité concret, contrairement aux précédentes. Elle est appelée *"loi des événements rares"* car elle est apparue à partir d'un modèle considérant un événement se produisant à un rythme "raisonnable", précisément un nombre constant de fois par unité de temps (par exemple, connexion à un serveur web, arrivée à une caisse de supermarché, décès suite à une ruade de cheval dans la cavalerie prussienne,...).

On s'intéresse à la variable aléatoire X donnant le nombre de réalisations pendant une durée T. On découpe T en n intervalles de temps d'amplitude T/n et, par hypothèse, la probabilité de survenue de l'événement durant chacun de ces intervalles est de la forme $p = \alpha T/n$.
Comme l'intervalle $[0, T]$ est l'enchaînement de n intervalles de même amplitude T/n, on considère que la loi binomiale de paramètres n, p s'applique et, pour $k \in [\![0, n]\!]$ fixé :

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{k!} \cdot \left(\frac{\alpha T}{n}\right)^k \cdot \exp\left((n-k)\ln\left(1 - \frac{\alpha T}{n}\right)\right)$$

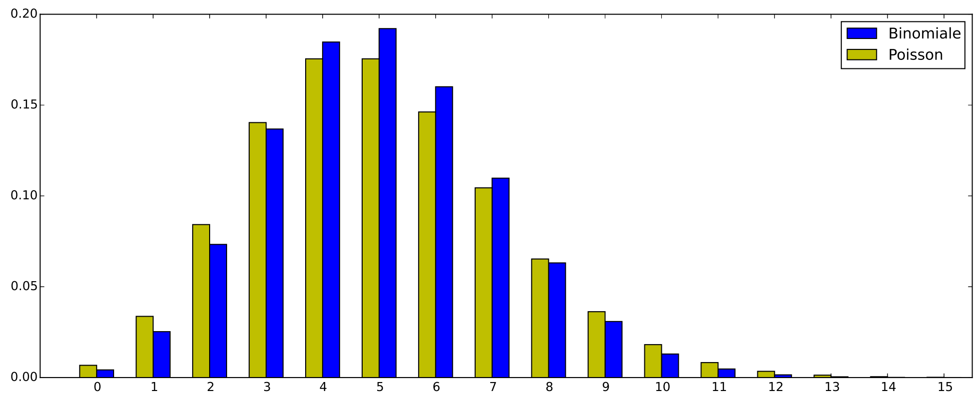
d'où (k est fixé !)

$$P(X = k) \underset{n\to\infty}{\longrightarrow} \frac{(\alpha T)^k}{k!} \exp(-\alpha T)$$

où l'on reconnaît la loi de Poisson de paramètre $\lambda = \alpha T$.

On peut ainsi espérer approcher la loi binomiale $\mathcal{B}(n, p)$ par la loi de Poisson $\mathcal{P}(np)$, beaucoup plus simple à calculer. L'expérience montre que la précision est satisfaisante pourvu que n soit "assez grand" et np "pas trop grand", une condition (empirique) souvent proposée étant $\left(n \geq 30 \quad \text{et} \quad np \leq 5\right)$.

Voici les graphes pour $n = 30$ et $np = 5$ (limité à $[0, 15]$) :



Théorème : soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires telle que, pour tout n, $X_n \hookrightarrow \mathcal{B}(n, p_n)$ où $np_n \underset{n\to\infty}{\longrightarrow} \lambda \in \mathbb{R}$. Alors

$$\forall k \in \mathbb{N} \quad P(X_n = k) \underset{n\to\infty}{\longrightarrow} e^{-\lambda} \cdot \frac{\lambda^k}{k!}.$$

3) Fonction d'une variable aléatoire discrète

Soit X une variable aléatoire discrète sur l'espace probabilisé $(\Omega, \mathcal{A}, P)$, à valeurs dans un ensemble E.
Si φ est une fonction de E dans un ensemble F, on note (abusivement) $\varphi(X)$ la fonction $\varphi \circ X$.

Propriété : $\varphi(X)$ est une variable aléatoire discrète sur $(\Omega, \mathcal{A}, P)$.

NB : la loi de probabilité de $\varphi(X)$ n'est pas toujours facile à expliciter, lorsque φ n'est pas injective.

<u>*Exemple*</u> : soit $\varphi : x \mapsto x^2$ et $X \hookrightarrow \mathcal{U}(E)$ où $E = \{-1, 0, 1, 2\}$. Alors $X^2(E) = \{0, 1, 4\}$ et

$P\left(X^2 = 0\right) = P(X = 0) = \dfrac{1}{4}$; $P\left(X^2 = 4\right) = P(X = 2) = \dfrac{1}{4}$; $P\left(X^2 = 1\right) = P(X = -1) + P(X = 1) = \dfrac{1}{2}$.

IV - Couple de variables aléatoires discrètes

1) Généralités

Soient X et Y deux variables aléatoires discrètes sur un même espace probabilisé $(\Omega, \mathcal{A}, P)$, à valeurs respectivement dans les ensembles E et F. On dispose alors sur $(\Omega, \mathcal{A}, P)$ du vecteur aléatoire discret $Z = (X, Y) : \omega \mapsto (X(\omega), Y(\omega))$. En effet :

- $Z(\Omega) \subset X(\Omega) \times Y(\Omega)$ donc $Z(\Omega)$ est fini ou dénombrable
- pour $(x, y) \in Z(\Omega)$, $Z^{-1}\left(\{(x, y)\}\right) = X^{-1}(\{x\}) \cap Y^{-1}(\{y\}) \in \mathcal{A}$.

Pour alléger, l'événement $((X, Y) = (x, y))$ est noté $(X = x, Y = y)$.

Définition : *la loi conjointe* (ou *loi mutuelle*) *de* X *et de* Y est la loi du couple (X, Y) :

$$\forall (x, y) \in X(\Omega) \times Y(\Omega) \quad P(X = x, Y = y) = P\big((X = x) \cap (Y = y)\big).$$

les lois marginales du couple (X, Y) sont la loi de X et la loi de Y :

$$\forall x \in X(\Omega) \quad P(X = x) = \sum_{y \in Y(\Omega)} P(X = x, Y = y)$$

et

$$\forall y \in Y(\Omega) \quad P(Y = y) = \sum_{x \in X(\Omega)} P(X = x, Y = y).$$

Attention ! Les lois marginales se déduisent de la loi conjointe, mais elles ne permettent pas de la déterminer.

Exemple : une urne contient 3 boules blanches et 4 noires ; on en extrait deux boules et l'on note X_1 la variable aléatoire donnant 1 (*resp.* 0) si la première boule est blanche (*resp.* noire). On définit de même X_2 pour la deuxième boule et l'on représente dans un tableau la loi conjointe de (X_1, X_2) dans les deux cas avec et sans remise :

avec remise

X_1 \ X_2	0	1	loi de X_1
0	$\frac{16}{49}$	$\frac{12}{49}$	$\frac{4}{7}$
1	$\frac{12}{49}$	$\frac{9}{49}$	$\frac{3}{7}$
loi de X_2	$\frac{4}{7}$	$\frac{3}{7}$	1

sans remise

X_1 \ X_2	0	1	loi de X_1
0	$\frac{2}{7}$	$\frac{2}{7}$	$\frac{4}{7}$
1	$\frac{2}{7}$	$\frac{1}{7}$	$\frac{3}{7}$
loi de X_2	$\frac{4}{7}$	$\frac{3}{7}$	1

Définition : pour $x \in X(\Omega)$ tel que $P(X = x) > 0$, on appelle *loi conditionnelle de* Y *sachant* $(X = x)$ la loi de Y relativement à la probabilité $P_{(X=x)}$, c'est-à-dire l'application

$$\begin{array}{ccl} Y(\Omega) & \to & [0, 1] \\ y & \mapsto & P_{(X=x)}(Y = y) = \dfrac{P(X = x, Y = y)}{P(X = x)} \end{array}$$

On définit de même la loi conditionnelle de X sachant $(Y = y)$ lorsque $P(Y = y) > 0$.

2) Somme de variables aléatoires discrètes

Théorème : toute somme finie de variables aléatoires réelles discrètes est une variable aléatoire réelle discrète.

Dém. Pour deux variables aléatoires X et Y, on considère le vecteur (X, Y), auquel on applique la fonction $\varphi : (X, Y) \mapsto X + Y$ (cf. **III-3)**). Le cas général s'en déduit par récurrence.

NB : comme le produit par un scalaire est banal, on en déduit que l'ensemble des variables aléatoires réelles discrètes est un $\mathbb{R}$-espace vectoriel.

3) Variables aléatoires discrètes indépendantes

Définition : X et Y sont dites *indépendantes* si et seulement si

$$\forall (x,y) \in X(\Omega) \times Y(\Omega) \quad P(X=x, Y=y) = P(X=x)\,P(Y=y)$$

(autrement dit, pour tout x tel que $P(X=x) > 0$, la loi conditionnelle de Y sachant $(X=x)$ coïncide avec la loi de Y).

Théorème : si X et Y sont indépendantes, alors

$$\forall (A,B) \in \mathcal{P}(X(\Omega)) \times \mathcal{P}(Y(\Omega)) \quad P(X \in A, Y \in B) = P(X \in A)\,P(Y \in B)\,.$$

Dém. Hors programme dans le cas dénombrable...

Théorème : si X et Y sont indépendantes et f et g des fonctions définies respectivement sur $X(\Omega)$ et $Y(\Omega)$, alors $f(X)$ et $g(Y)$ sont indépendantes.

Définition : soit $(X_i)_{i\in I}$ une famille finie de variables aléatoires discrètes. On dit que *les X_i, $i \in I$, sont mutuellement indépendantes* si et seulement si, pour toute famille $(x_i)_{i\in I}$ de $\prod\limits_{i\in I} X_i(\Omega)$, les événements $\big((X_i = x_i)\big)_{i\in I}$ sont mutuellement indépendants.

Lorsque I est dénombrable, on dit que *les X_i, $i \in I$, sont mutuellement indépendantes* si et seulement si toute sous-famille finie est formée de variables aléatoires discrètes mutuellement indépendantes.

NB : la vérification de cette propriété serait certes fastidieuse, mais en général c'est le contexte qui **fournit** des variables aléatoires mutuellement indépendantes, par exemple les variables de Bernoulli associées à des lancers successifs d'une pièce...

Exemple fondamental

Si $X_1, \dots, X_n$ sont mutuellement indépendantes et suivent chacune la loi de Bernoulli $\mathcal{B}(p)$, alors $X_1 + \cdots + X_n$ suit la loi binomiale $\mathcal{B}(n,p)$.

Lemme des coalitions : le théorème précédent se généralise ; supposons pour simplifier que $(X_k)_{1\le k\le n}$ est une famille finie de variables aléatoires réelles discrètes mutuellement indépendantes, que p est un entier de $[\![1, n-1]\!]$ et que f (*resp.* g) est une fonction de $\mathbb{R}^p$ (*resp.* $\mathbb{R}^{n-p}$) dans $\mathbb{R}$.
Alors les variables aléatoires réelles discrètes $f(X_1, \dots, X_p)$ et $g(X_{p+1}, \dots X_n)$ sont indépendantes.

V - Espérance et variance

1) Espérance d'une variable aléatoire réelle discrète

Définition : soit X une variable aléatoire réelle discrète à valeurs dans un ensemble dénombrable $\{x_n,\ n \in \mathbb{N}\}$ (où les x_n sont distincts 2 à 2).
On dit que *X est d'espérance finie* si et seulement si la série $\sum x_n P(X = x_n)$ est absolument convergente ; si c'est le cas, on appelle *espérance de X* le réel

$$E(X) = \sum_{n=0}^{\infty} x_n P(X = x_n)\,.$$

On l'admet en PSI, mais l'hypothèse de convergence **absolue** permet de montrer que cette notion ne dépend pas de l'ordre dans lequel on a "numéroté" les x_n...

Plus généralement, on dit que X *admet un moment d'ordre m* si et seulement si X^m est d'espérance finie, auquel cas le réel $E(X^m)$ est dit *moment d'ordre m de X*.

Dans le cas où X prend un nombre fini de valeurs, elle est d'espérance finie et l'on retrouve la définition de 1re année :

$$E(X) = \sum_{x\in X(\Omega)} xP(X = x) \left(= \sum_{\omega\in\Omega} X(\omega)\,P(\{\omega\}) \text{ si } \Omega \text{ est fini} \right).$$

Comme la somme des "coefficients" $P(X = x_n)$ vaut 1, on retrouve aussi l'idée de *moyenne pondérée* des valeurs prises par X (l'espérance est un indicateur de *position*).

Exemple 1 : on lance deux dés et l'on désigne par X la variable aléatoire qui donne la somme des résultats des deux dés. On trouve $E(X) = 7$, résultat prévisible pour raisons de symétrie...

Exemple 2 : d'après le théorème du **III-1)b)** *in fine*, il existe une variable aléatoire réelle discrète X d'image $\mathbb{N}^*$ telle que : $\forall n \in \mathbb{N}^* \quad P(X = n) = \dfrac{1}{n(n+1)}$ (en effet $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 1$).
Cela fournit un exemple de variable aléatoire qui n'est pas d'espérance finie.

Exemple 3 : si X est bornée, elle est d'espérance finie (ne pas oublier que $\sum_{n=0}^{\infty} P(X = x_n) = 1$).

Théorème : soit X à valeurs dans $\mathbb{N}$; X est d'espérance finie si et seulement si la série $\sum P(X \geq n)$ converge, auquel cas
$$E(X) = \sum_{n=1}^{\infty} P(X \geq n).$$

2) Propriétés de l'espérance

a) Une idée utile pour les calculs d'espérance

Voici une piste pour démontrer plusieurs résultats parmi ceux qui suivent. Démonstrations hors programme car il s'agit d'utiliser abondamment des sommations par paquets et des séries doubles.

L'idée majeure est que l'on n'a pas besoin d'expliciter la loi d'une variable aléatoire discrète Z pour calculer son espérance : il suffit de connaître une partition $(A_k)_{k\in K}$ de Ω (où K est fini ou dénombrable) telle que Z prenne une valeur constante, disons c_k, sur chacun des A_k. On a alors le résultat suivant : Z est d'espérance finie si et seulement si $\sum_{k\in K} |c_k| P(A_k)$ converge et si c'est le cas $E(Z) = \sum_{k\in K} c_k P(A_k)$.

Idée de la dém. Pour tout z de $Z(\Omega)$, on note $K_z = \{k \in K \ / \ c_k = z\}$, de sorte que $(K_z)_{z\in Z(\Omega)}$ est une partition de K et que
$$\forall z \in Z(\Omega) \quad |z| P(Z = z) = \sum_{k\in K_z} |c_k| P(A_k) \quad \text{et} \quad zP(Z = z) = \sum_{k\in K_z} c_k P(A_k)$$

b) Théorème du transfert

Soient X une variable aléatoire réelle discrète à valeurs dans $\{x_n,\ n \in \mathbb{N}\}$ (où les x_n sont distincts 2 à 2) et f une application à valeurs réelles, dont l'ensemble de définition contient $\{x_n,\ n \in \mathbb{N}\}$.
Alors $f(X)$ est d'espérance finie si et seulement si la série $\sum f(x_n) P(X = x_n)$ est absolument convergente, auquel cas
$$E\big(f(X)\big) = \sum_{n=0}^{\infty} f(x_n) P(X = x_n).$$
NB : résultat similaire dans le cas où X prend un nombre fini de valeurs (sans soucis de convergence !).

c) Linéarité de l'espérance

Soient X et Y deux variables aléatoires réelles discrètes sur un même espace probabilisé et $(a, b) \in \mathbb{R}^2$.
Si X et Y sont d'espérance finie, alors $aX + bY$ est d'espérance finie et
$$E(aX + bY) = aE(X) + bE(Y).$$

Corollaire : si X est d'espérance finie, alors $E\big(X - E(X)\big) = 0$ (variable aléatoire *centrée*).

d) Positivité et croissance de l'espérance

Positivité soit X une variable aléatoire réelle discrète à valeurs dans $\mathbb{R}^+$.
Si X est d'espérance finie, alors $E(X) \geq 0$.

Croissance soient X et Y deux variables aléatoires réelles discrètes sur un même espace probabilisé, telles que $X \leq Y$ (*i.e.* $X(\omega) \leq Y(\omega)$ pour tout $\omega \in \Omega$).
Si X et Y sont d'espérance finie, alors $E(X) \leq E(Y)$.

e) Espérance du produit de deux variables aléatoires réelles indépendantes

Soient X et Y deux variables aléatoires réelles discrètes sur un même espace probabilisé.
Si X et Y sont indépendantes et d'espérance finie, alors XY est d'espérance finie et
$$E(XY) = E(X)E(Y).$$

Attention ! Réciproque fausse.

3) Variance et écart type

a) Généralités

Théorème et définition : soit X une variable aléatoire réelle discrète.

Si X^2 est d'espérance finie, alors X est également d'espérance finie et l'on appelle *variance de* X le réel

$$V(X) = E\left((X - E(X))^2\right) = E\left(X^2\right) - E(X)^2.$$

Dans ce cas *l'écart type de* X est le réel $\sigma(X) = \sqrt{V(X)}$.

NB : la variance et surtout l'écart type (qui s'exprime dans la même "unité" que X) sont des indicateurs de la *dispersion* des valeurs prises par X de part et d'autre de l'espérance.

Propriété : si X^2 est d'espérance finie et $(a, b) \in \mathbb{R}^2$, alors $(aX + b)^2$ est d'espérance finie et

$$V(aX + b) = a^2 V(X) \quad ; \quad \sigma(aX + b) = |a|\,\sigma(X).$$

4) Trois inégalités fondamentales

a) Inégalité de Markov

Soit X une variable aléatoire réelle discrète à valeurs dans $\mathbb{R}^+$, d'espérance finie. Alors

$$\forall \alpha > 0 \quad P(X \geq \alpha) \leq \frac{E(X)}{\alpha}.$$

b) Inégalité de Bienaymé-Tchebychev

Soit X une variable aléatoire réelle discrète telle que X^2 soit d'espérance finie. Alors

$$\forall \alpha > 0 \quad P\left(\left|X - E(X)\right| \geq \alpha\right) \leq \frac{V(X)}{\alpha^2}.$$

NB : ces deux dernières majorations ont le mérite d'être simples, mais sont en général assez brutales.

c) Inégalité de Cauchy-Schwarz

Soient X et Y deux variables aléatoires réelles discrètes sur un même espace probabilisé telles que X^2 et Y^2 soient d'espérance finie.

Alors XY est d'espérance finie et

$$|E(XY)| \leq \sqrt{E(X^2)}\sqrt{E(Y^2)}.$$

Noter que sous les hypothèses précédentes, si besoin, $E(XY)$ se calcule grâce à l'idée du **a)** : si $X(\Omega) = \{x_i,\ i \in I\}$ et $Y(\Omega) = \{y_j,\ j \in J\}$, alors

$$E(XY) = \sum_{(i,j) \in I \times J} x_i y_j P(X = x_i, Y = y_j).$$

5) Covariance et coefficient de corrélation

Soient X et Y deux variables aléatoires réelles discrètes sur un même espace probabilisé telles que X^2 et Y^2 soient d'espérance finie.

Alors $\big(X - E(X)\big)\big(Y - E(Y)\big)$ est d'espérance finie et on appelle *covariance de* X *et* Y le réel

$$\operatorname{Cov}(X, Y) = E\Big(\big(X - E(X)\big)\big(Y - E(Y)\big)\Big) = E(XY) - E(X)E(Y)$$

Ainsi, si X et Y sont indépendantes, alors $\operatorname{Cov}(X, Y) = 0$ (réciproque fausse...).

Si X et Y ont une variance non nulle, on appelle *coefficient de corrélation de* X *et* Y le réel

$$\rho(X, Y) = \frac{\operatorname{Cov}(X, Y)}{\sigma(X)\,\sigma(Y)}.$$

L'inégalité de Cauchy-Schwarz, appliquée aux variables centrées $X - E(X)$ et $Y - E(Y)$, donne

$$|\rho(X, Y)| \leq 1.$$

L'appellation "coefficient de corrélation" vient du fait que, dans le cas fini, une valeur de $|\rho(X,Y)|$ proche de 1 indique que le nuage de points formé par les valeurs de (X,Y) est "voisin d'une droite". Cf. la *méthode des moindres carrés* vue en TD et le cas d'égalité dans l'inégalité de Cauchy-Schwarz.

6) Variance d'une somme finie de variables aléatoires réelles discrètes

Soient $X_1,\ldots,X_n$ des variables aléatoires réelles discrètes dont le carré est d'espérance finie. On a

$$V\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} V(X_k) + 2 \sum_{1\leq i<j\leq n} \mathrm{Cov}(X_i,X_j).$$

Si en outre les X_k sont indépendantes deux à deux, alors

$$V\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} V(X_k)$$

Exemple : si $X \hookrightarrow \mathcal{B}(p)$, alors $E(X) = E\left(X^2\right) = p$ et $V(X) = p(1-p)$. On obtient ainsi que, si $X \hookrightarrow \mathcal{B}(n,p)$, alors $E(X) = np$ et $V(X) = np(1-p)$.

VI - Fonction génératrice d'une variable aléatoire à valeurs dans $\mathbb{N}$

Dans toute cette section, X (*resp.* Y) désigne une variable aléatoire à valeurs dans $\mathbb{N}$ (donc réelle discrète !) sur un espace probabilisé $(\Omega,\mathcal{A},P)$.

1) Notion de fonction génératrice

La série entière $\sum P(X=n)\,t^n$ a un rayon de convergence au moins égal à 1.

Sa fonction somme $G_X : t \mapsto E\left(t^X\right) = \sum_{n=0}^{\infty} P(X=n)\,t^n$ est *la fonction génératrice de X*.

G_X est continue sur $[-1,1]$ (convergence normale !), $\mathcal{C}^\infty$ sur $]-1,1[$ et

$$\forall n \in \mathbb{N} \quad P(X=n) = \frac{G_X^{(n)}(0)}{n!}.$$

Par conséquent, on dit que *la loi de X est caractérisée par* G_X.

2) Fonction génératrice et espérance

X est d'espérance finie si et seulement si G_X est dérivable en 1, auquel cas

$$E(X) = G_X'(1).$$

NB : c'est toujours le cas si le rayon de convergence est strictement supérieur à 1 !

3) Fonction génératrice et variance

X^2 est d'espérance finie si et seulement si G_X est deux fois dérivable en 1, auquel cas

$$G_X''(1) = E\big(X(X-1)\big) = E\left(X^2\right) - E(X)$$

d'où

$$V(X) = G_X''(1) + G_X'(1) - G_X'(1)^2.$$

NB : c'est toujours le cas si le rayon de convergence est strictement supérieur à 1 !

4) Fonction génératrice d'une somme en cas d'indépendance

Si X et Y sont indépendantes, alors :

$$\forall t \in\,]-1,1[\quad G_{X+Y}(t) = G_X(t)\cdot G_Y(t).$$

NB : cette égalité peut être valide en dehors de $]-1,1[$ selon les rayons de convergence, mais $]-1,1[$ suffit pour identifier les coefficients des séries entières.

5) Lois usuelles

a) Loi de Bernoulli

Si $X \hookrightarrow \mathcal{B}(p)$, $G_X(t) = 1 - p + pt = 1 + p(t-1)$. On retrouve

$$E(X) = p \quad \text{et} \quad V(X) = 0 + p - p^2 = p(1-p).$$

b) Loi binomiale

Si $X \hookrightarrow \mathcal{B}(n,p)$,

$$\begin{aligned} G_X(t) &= (1-p+pt)^n = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} t^k \\ &= \big(1 + p(t-1)\big)^n \underset{t\to 1}{=} 1 + np(t-1) + \frac{n(n-1)}{2} p^2 (t-1)^2 + o\Big((t-1)^2\Big) \end{aligned}$$

On retrouve

$$E(X) = np \quad \text{et} \quad V(X) = n(n-1)p^2 + np - (np)^2 = np(1-p).$$

Conséquence : si $X \hookrightarrow \mathcal{B}(n,p)$, $Y \hookrightarrow \mathcal{B}(m,p)$ (avec le même p !) et si X et Y sont indépendantes, alors $X + Y \hookrightarrow \mathcal{B}(n+m,p)$.

c) Loi géométrique

Si $X \hookrightarrow \mathcal{G}(p)$ ($p \in]0,1[$), $G_X(t) = \sum_{n=1}^{\infty} p(1-p)^{n-1} t^n = \dfrac{pt}{1-(1-p)t}$ (RCV $\dfrac{1}{1-p} > 1$).
Il en résulte que X^2 est d'espérance finie et que :

$$E(X) = \frac{1}{p} \quad \text{et} \quad V(X) = \frac{1-p}{p^2}.$$

d) Loi de Poisson

Si $X \hookrightarrow \mathcal{P}(\lambda)$ ($\lambda > 0$), $G_X(t) = \sum_{n=0}^{\infty} e^{-\lambda} \cdot \dfrac{\lambda^n}{n!} \cdot t^n = e^{\lambda(t-1)}$ (RCV $+\infty$).
Il en résulte que X^2 est d'espérance finie et que :

$$E(X) = \lambda \quad \text{et} \quad V(X) = \lambda.$$

Conséquence : si $X \hookrightarrow \mathcal{P}(\lambda)$, $Y \hookrightarrow \mathcal{P}(\mu)$ et si X et Y sont indépendantes, alors $X + Y \hookrightarrow \mathcal{P}(\lambda+\mu)$.

VII - La loi faible des grands nombres

Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles discrètes, deux à deux indépendantes, de même loi, telles que les X_n^2 soient d'espérance finie.

On note m et σ l'espérance et l'écart type communs aux X_k et $S_n = \sum_{k=1}^{n} X_k$.

Alors on a, pour tout $\varepsilon > 0$,

$$P\left(\left|\frac{1}{n}S_n - m\right| \geq \varepsilon\right) \underset{n\to\infty}{\longrightarrow} 0$$

avec, plus précisément, pour tout n dans $\mathbb{N}^*$,

$$P\left(\left|\frac{1}{n}S_n - m\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

Ce résultat justifie d'une certaine façon l'approche *fréquentiste* des probabilités : si les X_k suivent une même loi de Bernoulli de paramètre m, $\dfrac{1}{n}S_n$ est la fréquence d'obtention d'un succès durant n répétitions de l'expérience...