Probabilités

par Emmanuel AMIOT

14 janvier 2021

Introduction et motivation

On étend les notions vues en Sup à des v.a. prenant une infinité de valeurs, mais une infinité *dénombrable*. En effet, si on essaye de faire une probabilité sur $\mathbb{R}$ (par exemple) en définissant la probabilité de « tomber » (évènement...) dans l'intervalle $[a, b]$ par $\mathcal{P}([a, b]) = \int_a^b p(t)\,dt$, où p serait une fonction positive, intégrable, telle que $\int_{\mathbb{R}} p = 1$ (probabilités totales), on n'a pas de pb avec $\mathcal{P}([a, c]) = \mathcal{P}([a, b]) + \mathcal{P}([b, c])$ pour $a < b < c$ ou encore avec $\mathcal{P}([a, b] \cap [c, d])$ (et encore, à condition que la probabilité d'un singleton soit nulle...). Mais quand on commence à prendre des unions et intersections d'intervalles de toutes les façons, on se demande un peu à quoi ressemble le résultat, qui n'est pas sans rappeler les propriétés des ouverts/fermés en Topologie... On attendra donc un peu plus de maturité pour examiner les tribus de Boréliens ! (L3/M1).

Exemple de base : on joue à pile ou face sans jamais s'arrêter. Si la probabilité de « pile » est p et si $q = 1 - p$ alors la probabilité que **le premier jet faisant pile ait le rang** n est : $\mathcal{P}(X = n) = pq^{n-1}$. *On a ainsi défini une v.a. à valeurs dans* $\mathbb{N}$ (cette loi de probabilité s'appelle une *loi géométrique*).

Systématisons les notions nécessaires pour étudier de telles lois. On a besoin d'une axiomatique suffisante pour traiter le cas des univers dénombrables, sans trop s'écarter du modèle fini étudié en Sup, mais qui vous prépare aussi à étudier plus tard des univers non dénombrables (intervalles, parties de $\mathbb{R}$, ensembles abstraits). On prend donc la formalisation de Kolmogorov :

Une variable aléatoire X va d'un univers infini Ω (ce qui arrive, qui se produit) dans un ensemble éventuellement infini E (les « cas possibles ») *pour nous* E *sera dénombrable*. Pour pouvoir mesurer des quantités comme $\mathcal{P}(X = x_i)$ ou $\mathcal{P}(X > x_i)$, on est amené à regarder les parties de ω de la forme $X^{-1}(\{x_i\}), X^{-1}(]x_i, +\infty[)$ et à imposer un certain nombre de conditions à ces parties : c'est la notion de **tribu** ci-dessous.

1 Tribus

Il s'agit de reprendre les notions vues en Sup mais avec une infinité de valeurs possibles pour une v.a. X.

DÉFINITION 1. *On appelle* **tribu** *sur une ensemble Ω (supposé dénombrable dans notre programme) une famille (non vide) $\mathcal{A}$ de parties de Ω, qui soit stable par intersections et réunions dénombrables, et par passage au complémentaire. Le couple $(\Omega, \mathcal{A})$ est un* **espace probabilisable**.

NB : en particulier une tribu est stable par intersections et réunions **finies**.

EXERCICE 1. *Montrer qu'on peut retirer l'hypothèse « stable par intersection dénombrable ». Montrer qu'une tribu doit contenir Ω et $\emptyset$.*

EXERCICE 2. *Quelle est la plus petite tribu de l'univers « la classe de MP » qui contienne « les filles » et « les 5/2 » ?*

Exemple :
— $\Omega = \mathbb{N}, \mathcal{A} = P(\mathbb{N})$. Ou à l'extrême opposé, $\mathcal{A} = \{\emptyset, \mathbb{N}\}$.
— L'ensemble des parties de $\mathbb{N}$ qui sont finies ou bien dont le complémentaire est fini est-il une tribu ?

DÉFINITION 2. *On appelle* **probabilité** *sur* $(\Omega, \mathcal{A})$ *une application* $\mathcal{P} : \mathcal{A} \to [0,1]$ *telle que* $\mathcal{P}(\Omega) = 1$ *et*

$$\mathcal{P}\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \sum_{i=0}^{\infty} \mathcal{P}(A_i)$$

dès que les A_i *sont 2 à 2 disjoints.* $(\Omega, \mathcal{A}, \mathcal{P})$ *est appelé espace probabilisé.*

DÉFINITION 3. *On appelle* **variable aléatoire discrète** *définie sur* Ω *une application* $X : \Omega \to E$ *telle que l'image de* X *soit (finie ou) dénombrable et telle que les parties*

$$X^{-1}(\{x\}) = \{\omega \in \Omega \mid X(\omega) = x\}$$

soient éléments de la tribu $\mathcal{A}$*, pour tout* $x \in E$*.*
On peut alors définir la loi de X *par* $\mathcal{P}_X(X = x) = \mathcal{P}(X^{-1}(\{x\}))$ *et plus généralement* $\mathcal{P}_X(A) = \sum_{a\in A} \mathcal{P}(X^{-1}(\{a\}))$*.*

Généralement $E \subset \mathbb{R}$ et on définit naturellement $\mathcal{P}(X \leqslant x)$ et autres valeurs similaires. Noter que par définition, $X^{-1}(A)$ est élément de la tribu pour toute partie A de $\mathbb{N}$ (réunion dénombrable des $X^{-1}(a)$ pour $a \in A$). On voit que cette condition de stabilité par réunion dénombrable n'est pas gratuite...
NB : cette définition semble sans problème, mais elle repose en fait sur la sommabilité de la famille des $\mathcal{P}(X^{-1}(\{a\}))$ (ce sont des réels positifs et leur somme sur tous les a, qui se calcule par le théorème de sommation par paquets, est majorée par 1 : cette somme est donc parfaitement définie, ayant la même valeur quelque soit la façon de calculer). Se méfier des évidences...

Exemple : La **loi géométrique de paramètre** p est $\mathcal{G}(p)$ définie par

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

On a vu qu'elle modélise un jeu de pile ou face infini (X est la première occurrence du tirage pile), ou une séquence infinie de tirages avec remise, ou...
$\mathcal{G}(p)$ est parfaitement définie sur $\mathcal{P}(\mathbb{N}^*)$, puisqu'on a

$$\forall A \subset \mathbb{N}^* \ \mathcal{P}(A) = \sum_{k\in A} p \times (1-p)^{k-1}$$

et c'est bien une probabilité puisque

$$\sum_{k\in\mathbb{N}^*} \mathcal{P}(X = k) = \sum_{k\in\mathbb{N}^*} p \times (1-p)^{k-1} = p \times \frac{1}{1-(1-p)} = 1.$$

Plus généralement, la probabilité que la pièce tombe sur pile *après* le k^e tirage est

$$\mathcal{P}(X > k) = \sum_{i>k} p \times (1-p)^{i-1} = p \times \frac{(1-p)^k}{1-(1-p)} = (1-p)^k$$

ce qui est cohérent, puisque c'est aussi l'évènement « la pièce tombe sur face lors des k premiers tirages ».

EXERCICE 3. *Trouver d'autres exemples concrets qui se modélisent par une loi géométrique.*

REMARQUE 1. *Plus généralement, on constate qu'il est nécessaire, et suffisant, de connaître les probabilités* $p_k = \mathcal{P}(X = k)$ *pour en déduire les probabilités de toute partie de* $\mathbb{N}$ *(ou* $\mathbb{Z}$*).*
Réciproquement, pour toute famille dénombrable de réels positifs $(p_k)_{k\in\mathbb{N}}$ *telle que* $\sum_{k\geqslant 0} p_k = 1$*, on peut fabriquer un espace probabilisé tel que* $p_k = \mathcal{P}(\{k\})$ *(ce fait est admis).*

REMARQUE 2. *Imaginons une tribu infinie non nécessairement dénombrable.*

Comme $A_n = \{\{x\} \in \mathcal{A} \mid \mathcal{P}(\{x\}) \geqslant 1/n\}$, l'ensemble des "atomes de probabilité $\geqslant 1/n$" est nécessairement fini (et même de cardinal $\leqslant n$ car $A_n \in \mathcal{A}$ et $\mathcal{P}(A_n) \geqslant \mathrm{Card}(A_n) \times 1/n$), cela prouve que l'ensemble des atomes de la tribu qui ont une probabilité non nulle est dénombrable, puisque c'est $\bigcup A_n$.

Ceci permet de pressentir les difficultés théoriques de la définition de probabilités sur des ensembles infinis non dénombrables... même si vous avez vu (sans aucune définition rigoureuse !) la distribution normale sur $\mathbb{R}$ en terminale ! C'est pourquoi après d'âpres discussions il a été décidé de se contenter des probabilités discrètes dans notre programme.

REMARQUE 3. *Dans la mesure où l'on convient que tout ensemble fini est* au plus *dénombrable, toutes les lois étudiées en Sup sont de droit des probabilités discrètes (Uniforme, Bernoulli, Binomiale...)*

Reprenons la loi binômiale $\mathcal{B}(n,p)$: $\mathcal{P}(X_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$. C'est une loi sur l'intervalle d'entiers $[\![0,n]\!]$; elle modélise le nombre de réussites de n épreuves suivant une même loi de Bernoulli de paramètre p – en langage clair, quelle est la probabilité de faire k fois pile si vous faites n parties de pile ou face ?

EXERCICE 4. *Préciser $\Omega, \mathcal{A}$. Retrouver l'espérance mentionnée ci-dessous.*

On se souvient que son espérance $E(X_n) = \sum k\,\mathcal{P}(X_n = k)$ vaut $n\,p$.[1] Essayons de la prolonger en une probabilité sur $\mathbb{N}$, en conservant cette valeur de l'espérance.

Hypothèse : $n\,p \to m > 0$ **quand** $n \to +\infty$**,** $k \in \mathbb{N}$ **est fixé.**

On rappelle la formule de Stirling : $n! \sim \sqrt{2\pi n}\; n^n\, e^{-n}$, qui nous donne ici

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \sim \frac{1}{k!}\sqrt{\frac{2\pi n}{2\pi(n-k)}}n^n(n-k)^{k-n}e^{-k} \sim \frac{1}{k!}\left(1-\frac{k}{n}\right)^{-n} n^k e^{-k} \sim \frac{n^k}{k!}$$

car $\left(1-\frac{k}{n}\right)^{-n} = \exp(-n\ln(1-\frac{k}{n})) = \exp(+k + o(1))$.

Par ailleurs (rigoureusement il faudrait tout écrire en exponentielles),

$$p^k(1-p)^{n-k} \sim \left(\frac{m}{n}\right)^k \left(1-\frac{m}{n}\right)^n \left(1-\frac{m}{n}\right)^{-k} \sim \frac{m^k}{n^k}e^{-m}$$

car $\left(1-\frac{m}{n}\right)^{-k} \to 1$, l'exposant étant fixe. On a donc une limite à cette probabilité quand $n \to +\infty$, qui est égale à $e^{-m}\dfrac{m^k}{k!}$. Cela mérite un nom :

DÉFINITION 4. *On appelle* **loi de Poisson de paramètre** m *la loi* $\mathcal{P}(m)$ *définie par*

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

C'est une probabilité sur $\mathbb{N}$, car $\sum\limits_{k \geqslant 0} \dfrac{m^k}{k!} = e^m$. En anticipant sur sa définition, on a bien son espérance qui est égale à son paramètre :

$$\sum_{k\geqslant 0} k e^{-m}\frac{m^k}{k!} = \sum_{k \geqslant 1} e^{-m}\frac{m.m^{k-1}}{(k-1)!} = m\sum_{k\geqslant 1} e^{-m}\frac{m^{k-1}}{(k-1)!} = m$$

Exemple *:* Vous allez à la pêche, et vous pêchez – longuement... Le nombre de truites que vous ramenez de votre journée au bord de la Têt peut raisonnablement être modélisé par une loi de Poisson : elle est adaptée aux *évènements rares*. Nous verrons plus bas comment trouver

1. Savoir le retrouver par le calcul direct.

expérimentalement son paramètre. Déjà la probabilité de *faire bredouille* vaut e^{-m}. Donc si vous ne ramenez rien une fois sur deux, m vaut environ $\ln 2 \approx 0.69$ autrement dit sur 100 journées de pêche, vous pouvez espérer ramener un total de 69 poissons...

DÉFINITION 5. *Si X est une v.a. discrète sur Ω à valeurs dans E et si f est une fonction de E dans F, on définit comme dans le cas fini la v.a. $f(X)$, dite* **variable image**, *par*

$$\mathcal{P}(f(X) = y) = \mathcal{P}(X \in f^{-1}(\{y\})).$$

Cependant, cette définition peut être problématique dans certains cas (les $f^{-1}(\{y\})$ sont-ils bien éléments de la tribu de X?) et nous ne l'utiliserons que dans des cas robustes, e.g. $E = F = \mathbb{N}$.

REMARQUE 4. *Quel est l'univers, la tribu, les évènements correspondant à cette nouvelle loi ? En notant (arbitrairement) par 0 ou 1 le fait de prendre un poisson, l'espace est $\Omega = \{0,1\}^{\mathbb{N}}$ mais on s'intéresse à des $A_k = \{x \in \Omega \mid \sum_n x_n = k\}$: l'ensemble des suites de 0 et de 1 qui contiennent exactement k fois la valeur 1. Pour la loi géométrique, on considère plutôt des parties de Ω du genre $B_k = \{(0,0,0,0\ldots,0,1,*,*,*,\ldots\}$, i.e. toutes les suites qui commencent par le même début : $k-1$ zéros suivis d'un un, puis n'importe quoi. Ce n'est pas du luxe que d'admettre (conformément au programme) qu'il existe une tribu sur $\Omega = \{0,1\}^{\mathbb{N}} \approx \{P,F\}^{\mathbb{N}}$ telle que les projections de cette tribu, i.e. les applications $p_n : (x_0, x_1, \ldots x_n, \ldots) \mapsto x_n$ de $\Omega \to \mathbb{R}$ soient des lois identiques de Bernoulli indépendantes (i.e. $\mathcal{P}(x_n = 1) = p, \mathcal{P}(x_n = 0) = 1-p$). En effet, l' ensemble Ω n'est pas dénombrable et sort du contexte étudié, bien que le résultat puisse paraître intuitif (il n'est en fait pas* **du tout** *évident).*

Pour le même prix, on admettra qu'il existe des espaces probabilisés correspondant à une suite (dénombrable) de v.a. indépendantes de même loi donnée sur un même ensemble fini.

2 Propriétés des probabilités

2.1 Propriétés ensemblistes

Les propriétés mettant en jeu un nombre *fini* d'évènements restent vraies :

— $A \subset B \Rightarrow \mathcal{P}(A) \leqslant \mathcal{P}(B)$
— $\mathcal{P}(A \cup B) = \mathcal{P}(A) + \mathcal{P}(B) - \mathcal{P}(A \cap B)$
— $\mathcal{P}(A \setminus B) = \mathcal{P}(A) - \mathcal{P}(A \cap B)$
— $\mathcal{P}(\Omega) = 1, \mathcal{P}(\emptyset) = 0, \mathcal{P}(\overline{A}) = 1 - \mathcal{P}(A)$.

NB : $\overline{A}$ désigne le complémentaire de A dans Ω et non pas son adhérence. [2]

EXERCICE 5. *Prouver ces relations.*

Maintenant, on peut considérer des suites monotones pour l'inclusion dans la tribu $\mathcal{A}$:

THÉORÈME DE CONTINUITÉ CROISSANTE. *Soit (A_n) une famille* **croissante** *d'éléments de $\mathcal{A}$, i.e. telle que $\forall n, A_n \subset A_{n+1}$. Alors*

$$\mathcal{P}(\bigcup_{n\in\mathbb{N}} A_n) = \lim_{n\to+\infty} \mathcal{P}(A_n)$$

Démonstration. Posons $A_\infty = \bigcup_{n\in\mathbb{N}} A_n$ pour simplifier les notations. Déjà la suite des $\mathcal{P}(A_n)$ est croissante et majorée (par $\mathcal{P}(A_\infty)$ puisque $A_\infty \supset A_n \forall n$), donc converge vers une limite $\ell \leqslant \mathcal{P}(A_\infty)$. Posons $B_n = A_{n+1} \setminus A_n = \{x \in \Omega \mid x \in A_{n+1}, x \notin A_n\}$. $B_0 = A_0$. Alors les B_n sont deux à deux disjoints, et $B_0 \cup B_1 \cup \ldots B_n = A_{n+1}$, d'où

$$\lim_{n\to+\infty} \mathcal{P}(A_n) = \sum_{n\geqslant 0} \mathcal{P}(B_n) = \mathcal{P}(\bigcup B_n) = \mathcal{P}(\bigcup A_n) = \mathcal{P}(A_\infty)$$

– cela rappelle le passage des suites croissantes aux séries à termes positifs... ♦

2. Vu qu'on a recommencé à parler d'images inverses, cela peut prêter à confusion, mais surtout quand Ω est un evn ce qui n'est pas prévu dans le cadre de notre programme.

Par passage au complémentaire, on a

THÉORÈME DE CONTINUITÉ DÉCROISSANTE.
Soit (A_n) *une famille d'éléments de* $\mathcal{A}$ *telle que* $\forall n, A_{n+1} \subset A_n$. *Alors*

$$\mathcal{P}(\bigcap_{n\in\mathbb{N}} A_n) = \lim_{n\to+\infty} \mathcal{P}(A_n)$$

Un Corollaire important du Thm. 2.1 est

PROPOSITION. *Pour toute suite d'évènements* $B_n \in \mathcal{A}$ *on a*

$$\mathcal{P}(\bigcup_{n\in\mathbb{N}} B_n) \leqslant \sum_{n\in\mathbb{N}} \mathcal{P}(B_n)$$

où le terme de droite peut être infini (mais pas celui de gauche).

Démonstration. Poser $A_n = B_0 \cup \cdots \cup B_n$ pour se ramener au calcul du Thm. 2.1. ♦

DÉFINITION 6. *Un évènement* $A \in \mathcal{A}$ *est* négligeable *(resp.* presque sûr*) si* $\mathcal{P}(A) = 0$ *(resp.* $\mathcal{P}(A) = 1$*).*

Par exemple, il est presque sûr que la pièce finira un jour par tomber sur pile ; ou que vous trouverez un examinateur sympathique à l'oral de l'X, ou que vous aller rencontrer le prince charmant (ou la princesse) UN samedi soir (on parle de répéter une infinité de fois l'épreuve. . .).

EXERCICE 6. *On considère une séquence infinie de pile ou face, la pièce ayant à chaque fois la probabilité* p *de tomber sur pile. Montrer que l'évènement ≪ on finira bien par tomber sur pile ≫ est presque sûr.*
Variante : on considère une série numérique prise au hasard. Quelle est la probabilité qu'elle soit à termes tous positifs ? Qu'elle soit alternée ?

La dernière proposition entraîne la première des deux assertions suivantes (l'autre par passage aux complémentaires) :

PROPOSITION. *Toute réunion [dénombrable] d'évènements négligeable est négligeable ; toute intersection [dénombrable] d'évènements presque sûrs est presque sûre.*

Ces notions sont donc invariantes par intersection (resp. réunion). La moralité est que l'on peut **négliger les évènements négligeables (! !)**, puisque aucun calcul ne pourra les combiner pour faire quelque chose de significatif, contrairement au principe ternaire de Raymond Devos :
≪ *Bon, une fois rien, c'est rien. Deux fois rien, c'est pas beaucoup, d'accord, mais trois fois rien. . . pour trois fois rien on peut déjà acheter quelque chose ! Et pour pas cher !*[3] ≫
En pratique (et notamment pour les formules de probabilités conditionnelles du paragraphe suivant) on ne considère tout simplement pas les évènements de probabilité nulle (à quoi bon ? !).
Attention néanmoins qu'un évènement presque sûr n'est pas un évènement certain (il est presque sûr qu'une série numérique "au hasard" n'est pas alternée. [4] Pourtant vous en avez rencontré !).

2.2 Probabilités liées et conditionnelles

On révise les no(ta)tions vues en Sup. Seule la formule de Bayes peut subir une modification éventuelle.

3. Bon, maintenant si vous multipliez trois fois rien par trois fois rien, rien multiplié par rien égale rien, trois multiplié par trois égal neuf, ça fait rien de neuf !
4. Avec des arguments hors-programmes, on démontre qu'une série $\sum \varepsilon_n/n$ où $\varepsilon_n = \pm 1$ à une chance sur deux, est presque sûrement convergente. Pourtant vous connaissez la série harmonique, qui est bien certainement divergente !

DÉFINITION 7. *La* probabilité de A sachant B *(probabilité conditionnelle) est définie par*

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

L'application $\mathcal{P}_B$ est une probabilité, définie sur la même tribu $\mathcal{A}$ que P.

NB : ceci n'a de sens que si B n'est pas négligeable, bien sûr.

Exemple : On considère une loi géométrique Y, montrons qu'elle est **sans mémoire** :

$$\mathcal{P}(Y = n + k \mid Y > n) = \mathcal{P}(Y = k).$$

En effet, $\mathcal{P}(Y = n + k) = p\, q^{n+k-1}$ et $\mathcal{P}(Y > n) = q^n$ (comme on l'a vu, c'est la probabilité que les n premiers lancers donnent face). Or

$$\mathcal{P}(Y = n + k \text{ et } Y > n) = \mathcal{P}(Y = n + k) \text{ (certes !)}$$

D'où $\mathcal{P}(Y = n + k \mid Y > n) = \dfrac{\mathcal{P}(Y = n + k \text{ et } Y > n)}{\mathcal{P}(Y > n)} = \dfrac{p\, q^{n+k-1}}{q^n} = q^{k-1} = \mathcal{P}(Y = k)$.

EXERCICE 7. *Considérons la loi $\mathcal{G}(p)$ associée à une séquence infinie de pile ou face, avec $q = 1 - p$. Soit $A = 3\mathbb{N}$ (« le premier tirage pile se produit à un rang multiple de 3 »), et $B = 2\mathbb{N}$ (« le premier pile se produit à un rang pair »). Vérifier que*

$$\mathcal{P}(A) = \frac{q^2}{1 + q + q^2} \qquad \mathcal{P}(B) = \frac{q}{1 + q} \qquad \mathcal{P}(A \cap B) = \frac{q^5}{1 + q + \dots q^5}$$

et en déduire $\mathcal{P}_B(A) = \dfrac{q^4}{1 + q^2 + q^4}$. Quelle est la probabilité de l'ensemble $6\mathbb{N}$?

Par définition on a immédiatement la *formule des Probabilités Composées* :

PROPOSITION.

$$\mathcal{P}(A \cap B) = \mathcal{P}_B(A)\mathcal{P}(B) = \mathcal{P}_A(B)\mathcal{P}(A) \quad \textit{i.e.} \quad \mathcal{P}_B(A) = \mathcal{P}_A(B) \times \frac{\mathcal{P}(A)}{\mathcal{P}(B)}$$

(cf. `https://xkcd.com/1236/`) qui rappelle la

DÉFINITION 8. A *et* B *sont indépendants ssi* $\mathcal{P}(A \cap B) = \mathcal{P}(A)\mathcal{P}(B)$

cela donne une caractérisation équivalente de l'indépendance :

$$\mathcal{P}(A) = \mathcal{P}_B(A) = \mathcal{P}_{\overline{B}}(A)$$

(A n'en a rien à carrer, de savoir si B est réalisé ou pas. . .)

REMARQUE 5. *Si A et B sont indépendants il en est de même de $\overline{A}$ et B (ou $\overline{B}$). En revanche A et $\overline{A}$ ne sont* jamais *indépendants (sauf si l'un des deux évènements est presque sûr).*

REMARQUE 6. *La différence entre $\mathcal{P}(A\cap B)$ et $\mathcal{P}(A)\mathcal{P}(B)$ est bornée (forcément !...) et on peut même être plus précis :*

Si on pose $\mathcal{P}(A\cap B)=x, \mathcal{P}(A\cap \overline{B})=y, \mathcal{P}(\overline{A}\cap B)=z$ et $\mathcal{P}(\overline{A}\cap\overline{B})=t$, on a

- $\mathcal{P}(A)=x+y$;
- $\mathcal{P}(B)=x+z$;
- $x,y,z,t\geqslant 0$ *et* $x+y+z+t=1$.

Cette dernière condition peut s'écrire $x+y+z\leqslant 1$ et $x,y,z\geqslant 0$ ce qui définit un « coin ». On a

$$\begin{aligned}\Delta=\mathcal{P}(A\cap B)-\mathcal{P}(A)\mathcal{P}(B)&=x-(x+y)(x+z)\\&=x-x^2-(xy+xz+yz)\\&=(x-x^2-xy-xz)-yz=x(1-x-y-z)-yz=xt-yz\end{aligned}$$

On a donc d'une part

$$\Delta\leqslant x-x^2\leqslant 1/4 \quad \textit{car } x\in[0,1]$$

et ce cas est atteint ($x=1/2=t, y=z=0$). D'autre part,

$$\Delta\geqslant -yz\geqslant -y(1-y)\geqslant -1/4$$

pour des raisons similaires. Ce cas est aussi atteint. Finalement,

$$\boxed{\left|\mathcal{P}(A\cap B)-\mathcal{P}(A)\mathcal{P}(B)\right|\leqslant 1/4.}$$

On peut voir les variations de la fonction Δ sur le domaine concerné sur la figure 1.

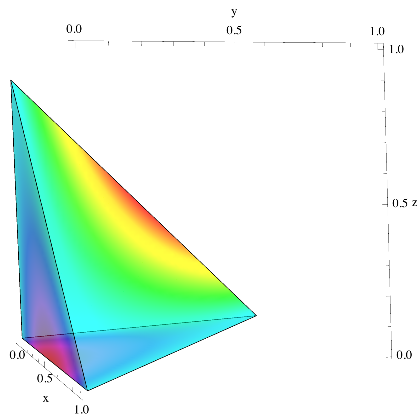


FIGURE 1 – Variations de Δ sur le trièdre $x+y+z\leqslant 1$ et $x,y,z\geqslant 0$.

PROPOSITION (FORMULE DES PROBABILITÉS TOTALES).

Si les B_i forment une famille **complète**, *i.e. ce sont des évènements non négligeables, d'intersection négligeable 2 à 2 et de réunion presque sûre i.e.*

$$\sum\mathcal{P}(B_i)=1 \qquad \mathcal{P}(B_i\cap B_j)=0 \ \textit{ pour } i\neq j, \quad \mathcal{P}(B_i)>0\ \forall i,$$

alors on peut retrouver la probabilité de A connaissant ses probabilités sachant les B_i :

$$\mathcal{P}(A)=\sum_i \mathcal{P}_{B_i}(A)\mathcal{P}(B_i).$$

Évident puisque $\mathcal{P}(A)=\mathcal{P}(\bigcup A\cap B_i)=\sum\mathcal{P}(A\cap B_i)$. Typiquement les B_i sont une partition de Ω, mais la définition est plus générale (on parle aussi de « famille exhaustive ») : c'est une partition, mais à des rognures de probabilité nulle près !

Exemple classique : on suppose que Y suit une loi de Poisson de paramètre λ, et que X sachant Y suit une loi binomiale de paramètre (p, Y) (typiquement, Y pourrait être la loi décrivant le nombre total de poissons que l'on pêche, et, sachant que l'on a pêché n poissons, la probabilité d'avoir pris k brochets parmi ces n poissons serait $\mathcal{P}(X = k \mid Y = n) = \binom{n}{k}p^k(1-p)^{n-k}$).
Alors on calcule la loi de X grâce à la formule précédente :

$$\begin{aligned}
\mathcal{P}(X = k) &= \sum_{n\in\mathbb{N}} \mathcal{P}(X = k \mid Y = n) \times \mathcal{P}(Y = n) \\
&= \sum_{n\geqslant k} \binom{n}{k} p^k(1-p)^{n-k} \times e^{-\lambda}\frac{\lambda^n}{n!} = e^{-\lambda} \sum_{n\geqslant k} \frac{p^k(1-p)^{n-k}\lambda^n}{k!(n-k)!} \\
&= e^{-\lambda}\frac{p^k\lambda^k}{k!} \sum_{n-k\geqslant 0} \frac{(1-p)^{n-k}\lambda^{n-k}}{(n-k)!} = \frac{e^{-\lambda}(p\lambda)^k}{k!} e^{(1-p)\lambda} = \frac{e^{-p\lambda}(p\lambda)^k}{k!}
\end{aligned}$$

et donc $X \sim P(\lambda p)$, ce qui est conforme à l'intuition (ça arrive).
La formule de Bayes vue en Sup reste vraie, et se prolonge éventuellement à des sommes infinies : comme on a $\mathcal{P}(A \mid B) = \dfrac{\mathcal{P}(B \mid A)\mathcal{P}(A)}{\mathcal{P}(B)}$ d'après la formule des Probabilités Composées [5], on déduit de la formule des probabilités totales que

THÉORÈME DE BAYES (VERSION SPé).
Si les A_j forment une famille exhaustive (par exemple une partition de Ω), alors

$$\mathcal{P}(A_i \mid B) = \frac{\mathcal{P}(B \mid A_i)\mathcal{P}(A_i)}{\sum_j \mathcal{P}(B \mid A_j) \times \mathcal{P}(A_j)} \quad \textit{i.e.} \quad \mathcal{P}_B(A_i) = \frac{\mathcal{P}_{A_i}(B)\mathcal{P}(A_i)}{\sum_j \mathcal{P}_{A_j}(B) \times \mathcal{P}(A_j)}$$

Notez que la série $\sum_j \mathcal{P}_{A_j}(B) \times \mathcal{P}(A_j)$ est toujours convergente. Cette formule permet d'inverser l'information, passant des $\mathcal{P}_{A_j}(B)$ aux $\mathcal{P}_B(A_i)$. C'est en ce sens qu'on parle parfois de formule de *probabilités a posteriori* – mais une telle interprétation est dangereuse, en cas de doute pensez toujours à la définition (tout part d'une tribu c'est à dire d'une axiomatique de l'univers considéré).
Même dans le cadre des v.a. sur une tribu infinie, on utilise souvent la forme binaire de la formule de Bayes :

$$\mathcal{P}_B(A) = \frac{\mathcal{P}_A(B)\mathcal{P}(A)}{\mathcal{P}_A(B)\mathcal{P}(A) + \mathcal{P}_{\overline{A}}(B)\mathcal{P}(\overline{A})}$$

Exemple : http ://xkcd.com/795 Un bel exemple (mortel !) de probabilité conditionnelle.

Exemple : Une cause célèbre : Sally Clark a été accusée de meurtre sur ses deux enfants par le procureur Roy Meadow au prétexte que, si la probabilité d'une mort subite du nourrisson (MSN) est de 1/8543 alors celle d'avoir *deux* décès par MSN serait de $\frac{1}{8543} \times \frac{1}{8543} \approx \frac{1}{73\,000\,000}$. Le procureur en a déduit qu'il était très improbable que la mère soit innocente. Condamnée deux fois, elle fut relaxée en appel en 2003 mais elle est morte 4 ans après d'un coma éthylique. . .
À part l'erreur sur l'indépendance des deux décès, [6] voyez-vous l'énorme faute de raisonnement ? Si on appelle I l'évènement « Sally est Innocente », et M l'évènement « deux enfants décèdent coup sur coup », le procureur a calculé (mal) $\mathcal{P}(M|I)$. Alors que la question était d'évaluer $\mathcal{P}(I|M)$! En fait on a

$$\mathcal{P}(I|M) = \frac{\mathcal{P}(M|I)\mathcal{P}(I)}{\mathcal{P}(M|I)\mathcal{P}(I) + \mathcal{P}(M|\bar{I})\mathcal{P}(\bar{I})}$$

où $\mathcal{P}(\bar{I})$ – la probabilité qu'un parent tue *deux* de ses enfants successivement – est minuscule (un statisticien l'a évaluée à 1/8,7 milliards). De ce fait, $\mathcal{P}(I|M)$ est bien plus proche de 1 que de 0.

5. Certains appellent déjà cette expression « formule de Bayes ».
6. Si une MSN se produit il est bien possible que ce soit pour des raisons génétiques/physiologiques qui tiennent aussi pour les enfants suivants. . .

2.3 Indépendance

2.3.1 Famille d'évènements indépendants

On a déjà rappelé la définition de deux évènements indépendants. Cette définition se généralise à un nombre quelconque :

DÉFINITION 9. *La famille des* $(A_i)_{i\in I} \in \Omega$ *est indépendante (ou : les évènements sont* mutuellement *indépendants) si*

$$\mathcal{P}(\bigcap_{i\in J} A_i) = \prod_{i\in J} \mathcal{P}(A_i)$$

pour toute sous-famille J finie *de* I.

Noter que $\prod_{i\in I} \mathcal{P}(A_i) = \operatorname{Inf} \prod_{i\in J} \mathcal{P}(A_i)$, J finie $\subset I$, a toujours un sens (suite décroissante et minorée par 0).

Par ailleurs, l'indépendance de la famille prouve l'indépendance des évènements 2 à 2, mais cette dernière propriété est plus faible que l'indépendance mutuelle. Un parallèle : une famille de vecteurs libre, c'est plus fort que la condition "non colinéaires (deux à deux)".

Exemple *:* Soient deux entiers p, q premiers entre eux et n un multiple de p, q. Alors dans $\mathbb{Z}/n\,\mathbb{Z}$, les évènements « x est multiple de p » et « x est multiple de q » sont indépendants. Ceci permet, à n fixé, de définir (par projection dans $\mathbb{Z}/n\mathbb{Z}$) la probabilité qu'un entier $x \in \mathbb{Z}$ soit multiple d'un facteur m de n, cette probabilité étant égale à $1/m$.

EXERCICE 8. *Retrouver ainsi l'expression de la valeur de la fonction indicatrice d'Euler* $\phi(n)$ *en fonction des facteurs premiers de* n.

2.3.2 V.A. indépendantes

Un couple de v.a. discrètes définit une v.a. discrète (sur l'espace probabilisé produit, qui est dénombrable car produit de deux espaces dénombrables, et qui hérite automatiquement de la structure d'espace probabilisé). Généralement X et Y prennent leurs valeurs sur la même tribu, ce qui simplifie les notations.

DÉFINITION 10. *La loi conjointe du couple* (X, Y) *est définie par les valeurs*

$$\mathcal{P}_{ij} = \mathcal{P}(X = i, Y = j) = \mathcal{P}(\{X = i\} \cap \{Y = j\}).$$

La loi marginale de X *(resp. de* Y*) vérifie*

$$\mathcal{P}(X = i) = \sum_j p_{ij} \qquad \mathcal{P}(Y = j) = \sum_i p_{ij}.$$

NB : on trouve les lois marginales en sommant les lignes ou les colonnes du tableau de la loi conjointe.

À noter que les séries écrites sont forcément convergentes, il n'y a donc pas de difficulté à cette généralisation de la définition déjà vue en Sup.

En sommant une partie des p_{ij} on peut généraliser aux v.a. ce que l'on a dit sur les éléments d'une tribu :

DÉFINITION 11. *La loi conditionnelle de « Y sachant que $X = x$ » est définie (pour* $\mathcal{P}(X = x) \neq 0$*) par* $\mathcal{P}(Y \mid X = x)$*, probabilité définie par* $\mathcal{P}(Y \in I \mid X = x) = \dfrac{\sum_{i\in I} p_{x,i}}{\mathcal{P}(X = x)}$ *ou de façon atomique par* $\mathcal{P}(Y = y \mid X = x) = \dfrac{p_{x,y}}{\mathcal{P}(X = x)} = \dfrac{p_{x,y}}{\sum_j p_{x,y_j}}$.

C'est une probabilité.

On définit naturellement

$$\mathcal{P}(Y \mid X \geqslant x) = \sum_{k \geqslant x} \mathcal{P}(Y \mid X = k)$$

et de façon similaire $\mathcal{P}(Y \mid X \leqslant x)$ ou $\mathcal{P}(Y \mid X > x)$.
Comme dans le cas fini, on a

DÉFINITION 12. *Deux v.a.* X *et* Y *sont indépendantes quand*

$$\mathcal{P}(Y = y_i \textit{ et } X = x_j) = \mathcal{P}(X = x_j).\mathcal{P}(Y = y_i)$$

(la loi conjointe vaut le produit des lois marginales)

C'est **seulement** dans le cas de lois indépendantes qu'on peut récupérer la loi conjointe à partir des lois marginales !

Exemple : (CCP 2016)
On considère un couple de lois (X, Y) à valeurs dans $\mathbb{N}^2$ telles que

$$\mathcal{P}(X = i, Y = j) = \frac{i+j}{2^{i+j+3}}$$

En sommant par paquets, on a bien

$$\sum_{(i,j)\in\mathbb{N}^2} \frac{i+j}{2^{i+j+3}} = \sum_{n\geqslant 0} \sum_{i+j=n} \frac{i+j}{2^{i+j+3}} = \sum_{n\geqslant 0} (n+1)\frac{n}{2^{n+3}} = 1$$

compte tenu des relations :

$$\forall x \in]-1,1[\quad \sum x^n = \frac{1}{1-x} \qquad \sum n x^{n-1} = \frac{1}{(1-x)^2} \qquad \sum (n+1)n x^{n-1} = \frac{2}{(1-x)^3}$$

en appliquant la dernière pour $x = 1/2$.
On vérifie alors que les deux lois marginales X, Y sont non indépendantes, car

$$\mathcal{P}(X = i) = \sum_{j\in\mathbb{N}} \frac{i+j}{2^{i+j+3}} = \frac{i}{2^i} \sum_{j\in\mathbb{N}} \frac{1}{2^{j+3}} + \frac{1}{2^i} \sum_{j\in\mathbb{N}} \frac{j}{2^{j+3}} = \frac{i}{2^i}\frac{1}{4} + \frac{1}{2^i}\frac{1}{4} = \frac{i+1}{2^{i+2}}$$

et donc $0 = \mathcal{P}((X, Y) = (0, 0)) \neq \mathcal{P}(X = 0) \times \mathcal{P}(Y = 0) = \frac{1}{16}$ par exemple.[7]
On peut aussi s'amuser à calculer des probabilités conditionnelles : par exemple

$$\mathcal{P}(X = i \mid Y = j) = \frac{\mathcal{P}(X = i, Y = j)}{\mathcal{P}(Y = j)} = \frac{\frac{i+j}{2^{i+j+3}}}{\frac{j+1}{2^{j+2}}} = \frac{i+j}{(j+1)\, 2^{i+1}}.$$

On pourrait vérifier ici la formule de Bayes / probabilités totales.

EXERCICE 9. *On considère deux v.a. indépendantes* X, Y *à valeurs dans* $\mathbb{N}$*. La fonction de répartition est par définition* $F_X : t \mapsto \mathcal{P}(X \leqslant t)$.[8] *Montrer que la fonction de répartition du couple est le produit des fonctions de répartition des deux v.a. :*

$$\mathcal{P}(X \leqslant u \textit{ et } Y \leqslant v) = \mathcal{P}(X \leqslant u) \times \mathcal{P}(Y \leqslant v)$$

On peut généraliser les couples à des vecteurs :

7. Cela se ≪ voit à l'œil nu ≫ que $\mathcal{P}(X = i, Y = j)$ n'est pas le terme général d'une série produit.

DÉFINITION 13. *La loi conjointe du* vecteur aléatoire $(X_1, X_2 \dots X_n)$ *est définie par les valeurs*

$$p_{i_1, i_2, \dots i_n} = \mathcal{P}(X_1 = i_1 \dots X_n = i_n) = \mathcal{P}(\bigcap_{k=1}^{n} \{X_k = i_k\})$$

La loi marginale de X_1 *(par exemple) vérifie*

$$\mathcal{P}(X_1 = i_1) = \sum_{i_2, i_3, \dots i_n} p_{\mathbf{i_1}, i_2, \dots i_n}$$

La notion d'indépendance se généralise à une telle famille, et même à une famille infinie (toute sous-famille finie devant être indépendante).

Exemple : Considérons n lancers de pièce à pile ou face, i.e. n v.a. de Bernoulli. Sauf à considérer que les pièces ont une mémoire, ces v.a. sont indépendantes. La loi de leur somme est une loi binômiale.
On en déduit aisément la propriété : si $S_1 \sim \mathcal{B}(n_1, p)$ et $S_2 \sim \mathcal{B}(n_2, p)$ alors la loi de $S_1 + S_2$ est $\mathcal{B}(n_1 + n_2, p)$.

2.3.3 Transfert d'indépendance

Au sens intuitif, si X et Y sont indépendantes, alors toute fonction de X est indépendante de toute fonction de Y. C'est vrai aussi au sens probabiliste :

THÉORÈME (LEMME DES COALITIONS). *Si* X *et* Y *sont deux lois indépendantes sur une même tribu à valeurs dans* F, G *; et si* f, g *sont deux fonctions de* F, G *dans* F', G'*, alors les v.a.* $f(X)$ *et* $g(Y)$ *sont indépendantes.*
Plus généralement, si $X_1 \dots X_n$ *sont des v.a. mutuellement indépendantes et si* $1 < k < n$ *alors* $f(X_1, \dots X_k)$ *et* $g(X_{k+1} \dots X_n)$ *sont indépendantes.*

Démonstration. La forme générale vient de ce que $X = (X_1, \dots X_k)$ est indépendant de $Y = (X_{k+1} \dots X_n)$ ce qui se vérifie par la définition. Prouvons donc la forme simple. Il s'agit de montrer que

$$\mathcal{P}((f(X), g(Y)) = (a, b)) = \mathcal{P}(f(X) = a) \times \mathcal{P}(g(Y) = b).$$

Notons $f^{-1}(a) = A, g^{-1}(b) = B$ (ce sont des éléments de la tribu par hypothèse). Il vient

$$(f(X), g(Y)) = (a, b) \iff X \in A, Y \in B.$$

Et par indépendance,

$$\mathcal{P}(X \in A \text{ et } Y \in B) = \mathcal{P}(X \in A) \times \mathcal{P}(Y \in B) = \mathcal{P}(f(X) = a) \times \mathcal{P}(g(Y) = b).$$

Je crois qu'on y est. ♦

On peut retrouver facilement avec ce théorème le résultat de l'exercice 9 (prendre Max(X)).

3 Espérance, variance

3.1 E, V

Ici on a (enfin ?. . .) une différence notable avec le programme de première année :
Contrairement au cas fini, il n'y a pas toujours une espérance (ou une variance) pour une v.a., car un problème de **convergence** se pose.

DÉFINITION 14. *Soit* X *une v.a. réelle discrète prenant les valeurs* $(x_n)_n$*, telle que la famille des* $(x_n \mathcal{P}(X = x_n))_{n \in \mathbb{N}}$ *soit* **sommable***, l'espérance* $E(X)$ *est la somme de la série (absolument convergente) :* $E(X) = \sum_n x_n.\mathcal{P}(X = x_n)$.

Une loi telle que $\mathcal{P}(X=n)=\dfrac{6}{\pi^2 n^2}$ pour $n\in\mathbb{N}^*$, par exemple, n'a pas d'espérance car $\sum n\dfrac{6}{\pi^2n^2}$ diverge. Bien sûr on retrouve la définition connue de l'espérance dans le cas où Ω est fini. La signification heuristique est toujours la même, $E(X)$ est le résultat *moyen* de X – on précisera de manière plus rigoureuse cet énoncé.

PROPOSITION.

- *Si $Y\geqslant 0$ a une espérance et $|X|\leqslant Y$ alors X a une espérance.*
- *L'espérance est positive : si X prend des valeurs positives alors $E(X)\geqslant 0$.*
- $|E(X)|\leqslant E(|X|)$.

Propriétés immédiates mais très importantes en pratique.

THÉORÈME DE TRANSFERT.
Si X est une v.a. réelle discrète, si $\varphi:\mathbb{R}\to\mathbb{R}$ est telle que $Y=\varphi(X)$ ait une espérance, alors $E(\varphi(X))=\sum y\mathcal{P}(Y=y)$ vaut $\sum\limits_n \varphi(x_n)\mathcal{P}(X=x_n)=\sum\limits_{x\in X(\Omega)}\varphi(x)\mathcal{P}(X=x)$.

Par exemple l'espérance d'une fonction constante est égale à la valeur de la constante :
$E(1)=\sum\limits_n 1\mathcal{P}(X=x_n)=1$. Ce théorème est capital puisqu'il permet de calculer l'espérance !

Démonstration. Il n'est pas évident que cela nécessite une preuve, et pourtant... Elle est délicate en plus ! Même dans le cas où φ est une bijection, on a déjà besoin du **théorème de convergence commutative** d'une série ACV (entre nous c'est pour ça qu'il est au programme).
Déjà notons que Y ne prend qu'un nombre dénombrable de valeurs (image d'un ensemble dénombrable). Soit $\{y_j\}_{j\in\mathbb{N}}$ cet ensemble. Ensuite on ramène Ω à $\mathbb{N}$, en considérant les ensembles d'indices

$$A_j=(\varphi\circ X)^{-1}(y_j)=\{i\in\mathbb{N}\mid \varphi(x_i)=y_j\}:$$

ces A_j forment une partition de $\mathbb{N}$ et avec $\mathcal{P}(A_j)=\sum\limits_{x_i\in A_j}\mathcal{P}(X=x_i)$ on a donc

$$E(Y)=\sum y_j\mathcal{P}(A_j)=\sum_j\sum_{i\in A_j}\varphi(x_i)\mathcal{P}(X=x_i)$$

On conclut en constatant qu'on a permuté les termes $\varphi(x_i)\mathcal{P}(X=x_i)$, ce qui ne change pas leur somme en vertu de la CV absolue de la série (thm de convergence commutative). ♦

On démontre similairement (on peut aussi le déduire du Thm. précédent) la linéarité de l'espérance E :

THÉORÈME 6. *L'espérance est linéaire : si X et Y ont une espérance alors pour tous $\lambda,\mu\in\mathbb{R}$ la v.a. $\lambda X+\mu Y$ a une espérance $E(\lambda X+\mu Y)=\lambda E(X)+\mu E(Y)$.*

Démonstration. Soient X,Y deux v.a. ayant une espérance et posons $Z=X+Y$ ainsi que $A_k=\{(i,j)\mid x_i+y_j=z_k$ où les z_k désignent les valeurs de Z (elles sont en nombre dénombrable car image d'un produit d'ensemble dénombrables).
Observons que

$$\sum_j\mathcal{P}(X=x_i,Y=y_j)=\mathcal{P}(X=x_i)$$

car l'évènement $(X=x_i)$ est partitionné en les $(X=x_i,Y=y_j)$ et l'ordre de sommation est arbitraire (par convergence absolue...). Donc, partant de deux sommes de séries ACV par hypothèse, i.e. de familles sommables,

$$
\begin{aligned}
E(X)+E(Y) &= \sum_i x_i \mathcal{P}(X=x_i) + \sum_j y_j \mathcal{P}(Y=y_j) \\
&= \sum_i x_i \sum_j \mathcal{P}(X=x_i, Y=y_j) + \sum_j y_j \sum_i \mathcal{P}(X=x_i, Y=y_j) \\
&= \sum_{i,j} (x_i+y_j)\mathcal{P}(X=x_i, Y=y_j) = \sum_k \sum_{A_k} z_k \mathcal{P}(A_k) = E(Z)
\end{aligned}
$$

en utilisant l'interversion pour les série doubles sommables et la sommation par paquets. ♦

Ceci munit l'espace des v.a. définies sur une tribu et ayant une espérance d'une structure d'evn ($\mathcal{L}^1_{\mathcal{A},\Omega}$).
Dans le cas particulier de $\varphi(x)=x^2$, on a un résultat agréable :

THÉORÈME 7. *Si X^2 a une espérance, alors X aussi.*

Démonstration. Posons pour simplifier $\mathcal{P}(X=x_n)=p_n$. On utilise une petite ruse calculatoire. Par hypothèse les deux séries $\sum p_n$ et $\sum x_n^2 p_n$ convergent (absolument). Or $|2x_n p_n| \leqslant p_n + x_n^2 p_n$, donc la série $\sum x_n p_n$ converge absolument. ♦

DÉFINITION 15. *Si X^2 est d'espérance finie, la variance de X est définie par*

$$V(X) = E(X^2) - E(X)^2.$$

THÉORÈME 8. *On a $V(X) = E\big((X-E(X))^2\big)$.*

Elle est donc toujours $\geqslant 0$, et on définit **l'écart-type** par $\sigma(X) = \sqrt{V(X)}$.

Démonstration. La même que dans le cas fini, on vérifie seulement que les quantités invoquées ont un sens, ce qui est le cas puisqu'on suppose que $E(X^2)$ existe [9]

$$E\big((X-E(X))^2\big) = E(X^2 - 2E(X)X + E(X)^2) = E(X^2) - 2E(X)E(X) + E(X)^2 = V(X)$$

en utilisant la linéarité et l'espérance d'une constante. ♦

On voit avec la deuxième formule que $V(X)$ mesure la *dispersion* autour de la valeur moyenne, $E(X)$.
Dans le cas de deux (voire plusieurs) v.a. on a

PROPOSITION. *Si X, Y sont deux v.a. discrètes réelles* **indépendantes** *et ont une espérance, alors $X \times Y$ a une espérance et $E(XY) = E(X)E(Y)$.*

Démonstration. On écrit la définition et on trouve un produit de séries absolument convergentes :

$$E(XY) = \sum xy\mathcal{P}(XY=xy) = \sum x\mathcal{P}(X=x) \sum y\mathcal{P}(Y=y) = E(X)E(Y) \quad \text{car } \sum |x|\mathcal{P}(X=x) \sum |y|\mathcal{P}(Y=y) \text{ CV.}$$

En fait c'est une bonne explication du théorème sur les séries produits et leur sommation par tranches ! ♦

9. Les traders utilisent des modèles des fluctuations boursières fondées sur l'hypothèse qu'elles ont une variance, appellée la *volatilité* dans ce contexte. L'ennui est qu'il a été prouvé que cette hypothèse est non fondée (wikier 'Fat-tail distribution'), ce qui peut expliquer bien des choses. . .

EXERCICE 10. *Redémontrer l'identité ci-dessus "à la main", en écrivant*

$$E(X) = \sum p_n x_n \qquad E(X^2) = \sum p_n x_n^2 \qquad \textit{etc...}$$

en explicitant les arguments de sommabilité.

3.2 Espérance et variance des lois usuelles

PROPOSITION. *Pour la loi de Poisson* $X \sim \mathcal{P}(m)$*, on a* $E(X) = m$ *et* $V(X) = m$.
Pour la loi géométrique $Y \sim \mathcal{G}(p)$*, on a* $E(Y) = \frac{1}{p}$ *et* $V(Y) = \frac{1-p}{p^2}$.

Démonstration. On utilise des propriétés de séries entières connues.
La série $\sum n e^{-m} \frac{m^n}{n!} = \sum e^{-m} \frac{m^n}{(n-1)!}$ converge absolument et sa somme vaut m en réindexant. De même

$$E(X^2) = \sum n^2 e^{-m} \frac{m^n}{n!} = \sum (n(n-1)+n) e^{-m} \frac{m^n}{n!} = m^2 \sum e^{-m} \frac{m^{n-2}}{(n-2)!} + m \sum e^{-m} \frac{m^{n-1}}{(n-1)!} = m^2 + m$$

d'où $V(X) = E(X^2) - E(X)^2 = m^2 + m - m^2 = m$.

Pour la loi géométrique on écrit

$$\sum_{n \geqslant 1} np(1-p)^{n-1} = p \sum_{n \geqslant 1} n(1-p)^{n-1} = -p \frac{d}{dp} \sum (1-p)^n = -p \frac{d}{dp} \frac{1}{1-(1-p)} = \frac{-p}{-p^2} = \frac{1}{p}$$

et similairement on a

$$\sum_{n \geqslant 1} n(n+1)(1-p)^{n-1} = \frac{d^2}{dp^2} \sum (1-p)^{n+1} = \frac{d^2}{dp^2} \frac{1}{p} = \frac{2}{p^3}$$

d'où

$$E(Y^2) = \sum_{n \geqslant 1} n^2 p(1-p)^{n-1} = \sum_{n \geqslant 1} (n(n+1) - n) p(1-p)^{n-1} = p\frac{2}{p^3} - \frac{p}{p^2} = \frac{2-p}{p^2}$$

d'où la variance

$$V(Y) = E(Y^2) - E(Y)^2 = \frac{1-p}{p^2}$$

♦

L'espérance de la loi géométrique a un sens intuitif : si vous avez une chance sur 100 de gagner, alors il faut en moyenne essayer 100 fois pour y arriver [la première fois]. La variance l'est moins : pour la même valeur on trouve $\sigma = \sqrt{\frac{99/100}{1/100^2}} = \sqrt{99 \times 100} \approx 100$, une dispersion qui remonte presque jusqu'à 0 !

EXERCICE 11. *Faites-vous un tableau qui contienne aussi les espérances et variances des lois étudiées en Sup (Bernoulli, binomiale...)*

EXERCICE 12. *On a vu que la loi binômiale* $\mathcal{B}(n, p)$ *a pour espérance* np *et qu'elle tend vers la loi* $\mathcal{P}(\lambda)$ *si* $np \to \lambda$ *quand* $n \to +\infty$*. Ici l'espérance de la limite est bien la limite de l'espérance. En est-il de même pour la variance ?*

Observons que, si l'espérance est linéaire, en revanche

$$V(aX + b) = E(aX + b - E(aX + b))^2 = E(a(X - E(X))^2 = a^2 V(X)$$

Ceci justifie

DÉFINITION 16. *Si* X *admet une variance, alors* $\dfrac{X - E(X)}{\sigma(X)}$ *est une* **v.a. centrée réduite**, *i.e. une v.a. d'espérance nulle et d'écart-type unité.*

3.3 Estimations

On a deux indicateurs précieux pour apprécier une v.a. : sa valeur moyenne et sa dispersion (quand elles existent). On peut raffiner une peu plus notre connaissance en étudiant la probabilité des valeurs lointaines. On établit d'abord qu'elle est inversement proportionnelle à l'éloignement (voire moins que ça) :

PROPOSITION (INÉGALITÉ DE MARKOV).
On considère une v.a. X *réelle et même (presque sûrement) positive ou nulle, qui admet une espérance. Alors*

$$\mathcal{P}(X \geqslant a) \leqslant \frac{E(X)}{a}.$$

Démonstration.

$$E(X) = \sum x_n p(X = x_n) \geqslant \sum_{(x_n \geqslant a)} x_n \mathcal{P}(X = x_n) \geqslant \sum_{(x_n \geqslant a)} a\mathcal{P}(X = x_n) = aP(X \geqslant a);$$

♦

COROLLAIRE (INÉGALITÉ DE BIENAYMÉ-TCHÉBYCHEF).
Soit X *une v.a. discrète réelle admettant une variance. Alors pour tout* $a > 0$

$$\mathcal{P}(|X - E(X)| \geqslant a) \leqslant \frac{V(X)}{a^2} = \frac{\sigma(X)^2}{a^2}.$$

Démonstration. On pose $Y = (X - E(X))^2$ et on applique à Y l'inégalité précédente. On obtient

$$\mathcal{P}(Y \geqslant a^2) \leqslant \frac{E(Y)}{a^2};$$

Compte tenu de ce que $E(Y) = V(X)$, c'est le résultat annoncé. ♦

Ces résultats sont identiques au cas fini. [10] Notez que la deuxième inégalité a une hypothèse plus forte que la première.

Exemple *:* Le nombre d'erreurs de calcul qui dépriment le correcteur d' une copie de MP suit une loi de Poisson de paramètre 10. La probabilité de faire 20 erreurs ou plus est majorée par 1/2 (Markov), ou mieux $\dfrac{10}{(20-10)^2}$ soit 1/10 par B-T.

EXERCICE 13. (idem avec la loi géométrique)
Il vous faut 5 essais en moyenne pour réussir un carreau à la pétanque. Que disent les inégalités précédentes quant à la probabilité qu'il en faille au moins 10 ? Que vaut exactement cette probabilité ?

Remarque : ne vous prenez pas la tête quant à l'orthographe de « Tchébychef », qui relève de l'alphabet cyrillique : Чебышёв (l'initiale est un "Tch"). Au fait, Bienaymé n'est pas son prénom (lequel est encore plus incongru, cherchez-le sur Internet).

10. Au pire des cas (divergence) on a une majoration par $+\infty$... ce qui n'est jamais faux !

3.4 Covariance

Considérons deux v.a. X et Y. Il n'est pas évident de calculer la variance de $X+Y$, car un terme en $X \times Y$ intervient. Étudions cela de plus près :

DÉFINITION 17. *La* **covariance** *de* (X,Y) *est* $\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$.

PROPOSITION. *(Par linéarité)* $\mathrm{Cov}(X,Y) = E((X-E(X))(Y-E(Y))$.

Cela mesure donc de combien X et Y s'écartent *simultanément* de leur moyenne. En particulier la Covariance est négative quand X et Y varient en sens contraire.

PROPOSITION. *Si* X, Y *sont indépendantes alors* $\mathrm{Cov}(X,Y) = 0$.

Réciproque fausse. Osons le dire, cela peut arriver... par hasard ! La covariance pouvant être positive, ou négative, peut aussi être nulle, accidentellement et sans rien signifier. Ce phénomène est source de nombreuses interprétations erronées.

PROPOSITION.

- $\mathrm{Cov}(X,X) = V(X)$.
- Cov *est bilinéaire symétrique, en particulier* $\mathrm{Cov}(Y,X) = \mathrm{Cov}(X,Y)$ *et* $\mathrm{Cov}(aX,Y) = a\,\mathrm{Cov}(X,Y)$.
- $\mathrm{Cov}(X+\tau, Y) = \mathrm{Cov}(X,Y)$ *(invariance par translation)*

On peut généraliser :

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

PROPOSITION. $V(X_1 + \dots X_n) = V(X_1) + \dots V(X_n) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j) = \sum V(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j)$.

COROLLAIRE 2. *Si* X, Y *sont indépendantes alors* $V(X+Y) = V(X) + V(Y)$.
Plus généralement, si les v.a. $(X_1, \dots X_n)$ *sont* **2 à 2 indépendantes** *alors*

$$V(X_1 + \dots X_n) = V(X_1) + \dots V(X_n)$$

À noter qu'on n'a pas besoin de l'indépendance *mutuelle*. Attention ! la réciproque est fausse.

EXERCICE 14. *Trouver un contre-exemple pour la réciproque.*

EXERCICE 15. *Classer les propriétés :*

- *Les v.a.* $(X_1, \dots X_n)$ *sont 2 à 2 indépendantes.*
- *Les v.a.* $(X_1, \dots X_n)$ *sont mutuellement indépendantes.*
- *Les* $\mathrm{Cov}(X_i, X_j)$ *sont nulles pour* $i \neq j$.
- $V(X_1 + \dots X_n) = V(X_1) + \dots V(X_n)$

LEMME 1. *Pour toute famille finie* $I \subset \Omega$,

$$\Big(\sum_{n \in I} p_n x_n y_n\Big)^2 \leqslant \Big(\sum_{n \in I} p_n x_n^2\Big)\Big(\sum_{n \in I} p_n y_n^2\Big)$$

C'est tout simplement l'inégalité de Cauchy-Schwarz pour le produit scalaire de poids p_k. En passant à la limite (licite car toutes les séries sont absolument CV) :

COROLLAIRE 3. *Si* X *et* Y *ont une variance alors* $X \times Y$ *a une espérance, et*

$$(E(XY))^2 \leqslant E(X^2)E(Y^2).$$

Cela signifie que l'on peut alors calculer

DÉFINITION 18. *Le* **coefficient de corrélation [linéaire]** *de* (X, Y) *est défini*[11] *par*

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

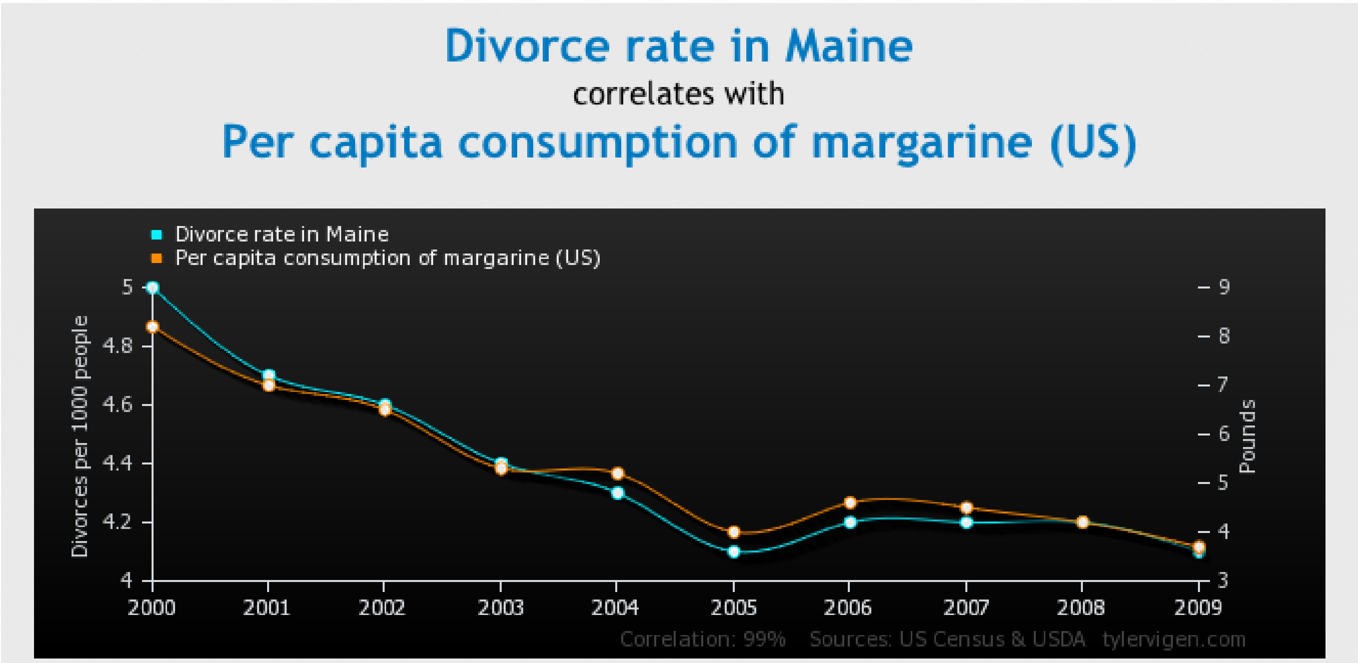
PROPOSITION. ρ *est toujours compris entre -1 et 1.*

Démonstration. Appliquer Cauchy-Schwarz au couple $(X - E(X), Y - E(Y))$. ♦

Intuitivement, si $\rho \approx 1$ c'est que X et Y agissent main dans la main. Si $\rho \approx -1$ c'est que quand X dit blanc, Y dit noir ! Enfin

PROPOSITION. X, Y *sont indépendantes* $\Rightarrow\ \mathrm{Cov}(X,Y) = 0 \iff \rho = 0.$

Attention, décorrélées n'est pas *équivalent* à indépendantes ! Et “corrélées” ne signifie pas qu'il y a une causalité, c'est l'erreur d'interprétation la plus fréquente des probabilités... Cf. entre autres exemples hilarants (de la Baltique) :



	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Divorce rate in Maine *Divorces per 1000 people (US Census)*	5	4.7	4.6	4.4	4.3	4.1	4.2	4.2	4.2	4.1
Per capita consumption of margarine (US) *Pounds (USDA)*	8.2	7	6.5	5.3	5.2	4	4.6	4.5	4.2	3.7
Correlation: 0.992558										

Moins hilarant, les annonces fracassantes de décès consécutifs à des vaccinations anti-Covid. Qu'en pensez-vous en tant que probabilistes aguerri(e)s ? Indication : dans la campagne de test de Pfizer (80,000 tests), il y a eu 6 mots. Dont 4 avaient reçu le placebo...

Exemple : On reprend la loi conjointe donnée par $\mathcal{P}((X,Y) = (i,j)) = \dfrac{i+j}{2^{i+j+3}}$.

On a trouvé que $\mathcal{P}(X=i)=\dfrac{i+1}{2^{i+2}}$ et similairement pour Y. À l'aide des séries entières déjà utilisées – et de quelques autres – on calcule

$$E(X)=E(Y)=\sum_{i\geqslant 0}\frac{i(i+1)}{2^{i+2}}=2\quad E(X^2)=\sum_{i\geqslant 0}\frac{i^2(i+1)}{2^{i+2}}=8,\quad V(X)=V(Y)=8-2^2=4$$

$$E(XY)=\sum_{i,j\geqslant 0}\frac{ij(i+j)}{2^{i+j+3}}=3\quad \mathrm{Cov}(X,Y)=\frac{E(XY)-E(X)E(Y)}{V(X)}=\frac{3-2^2}{4}=-\frac{1}{4}$$

3.5 Loi faible des grands nombres

Intuitivement, la probabilité de réussite d'une expérience est donnée par la moyenne d'un grand nombre d'expériences...

THÉORÈME 9. *Soit* $(X_n)_n$ *une suite de v.a. deux à deux indépendantes, de même loi, ayant une variance* σ^2*. Alors leur moyenne* $M_n=\dfrac{X_1+\dots X_n}{n}$ *tend vers* $m=E(X_1)=\dots E(X_n)$ *au sens où*

$$\forall\varepsilon>0\ \mathcal{P}(|M_n-m|\geqslant\varepsilon)\to 0 \textit{ quand } n\to+\infty$$

Démonstration. On utilise l'inégalité de BIENAYMÉ-TCHÉBYCHEV. L'espérance de M_n est m par linéarité, l'indépendance des X_i donne que

$$V(X_1+\dots X_n)=\sum V(X_i)=nV(X_1)=n\sigma^2,$$

d'où $V(M_n)=\dfrac{\sigma^2}{n}$ et donc

$$\mathcal{P}(|M_n-m|\geqslant\varepsilon)\leqslant\frac{\sigma^2}{n\varepsilon^2}\to 0$$

♦

Cela signifie que la v.a. M_n se concentre autour de son espérance m. Un exemple célèbre est celui où les X_n sont des Bernoulli : la probabilité que le nombre de pile soit très différent du nombre de face après un grand nombre de lancers tend vers 0, ce qu'on voit fort bien sur le graphe d'une loi binomiale.

Remarque : le calcul suggère que $\dfrac{\sqrt{n}}{\sigma}(M_n-m)$ est une v.a. de moyenne nulle et de variance unité. La loi forte des grands nombres dit qu'en un sens à définir, cette suite de v.a. tend vers une loi normale (Gaussienne).

4 Fonction génératrice

4.1 Définition

Nous introduisons un outil remarquablement commode pour les probabilités dénombrables, qui utilise la flexibilité des séries entières.

DÉFINITION 19. *Soit* X *une v.a. discrète* positive, *sa fonction génératrice est* $G_X : t\mapsto E(t^X)$. *Quand* X *prend des valeurs entières on écrit* $G_X(t)=\sum_{n\geqslant 0}p(X=n)t^n$ *:* G_X *est DSE.*

***Exemple** :*

— $\mathcal{P}(m) : G(t)=\sum e^{-m}\dfrac{m^n}{n!}t^n=e^{m(t-1)}$.

— $\mathcal{G}(p) : G(t)=\sum_{n\geqslant 1}p(1-p)^{n-1}t^n=\dfrac{pt}{1-(1-p)t}$.

— $\mathcal{B}(n,p)$: même si Ω est fini on peut écrire $G(t) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} t^k = (1-p+pt)^n$.

Vu que la série converge toujours en $t = 1$ (vers 1) et même converge absolument, on peut affirmer que

PROPOSITION. *G est toujours définie au moins pour $t \in [-1, 1]$ (le rayon de convergence de la série entière est au moins 1).*

EXERCICE 16. *Préciser les rayons de CV pour les exemples donnés ci-dessus.*

EXERCICE 17. *Exprimer G_{2X} en fonction de G_X (non, ce n'est ni son double, ni son carré. . .)*

EXERCICE 18. *Si $np \to \lambda$, calculer la limite de la suite de fonctions $(1-p+pt)^n$ quand $n \to +\infty$. Que retrouve-t-on ?*

PROPOSITION (LE RETOUR DU PISE).
Deux v.a. discrètes ayant même fonction génératrice ont nécessairement même loi.

4.2 Moments

On constate que $E(X) = \sum k\mathcal{P}(X = k)1^{k-1}$ n'est autre que la valeur de la série dérivée de G_X en 1. On a donc (vu que la CV absolue en 1 entraîne la CVN, ce qui permet d'invoquer le théorème de dérivation terme à terme) :

PROPOSITION. *Si X admet une espérance alors G_X est dérivable en 1, et $E(X) = G'_X(1)$.*
X admet une variance dès que G_X est deux fois dérivable en 1, et $E(X^2) = G''_X(1) + G'_X(1)$.

On en déduit une expression de la variance : $\boxed{V(X) = G''_X(1) + G'_X(1) - G'_X(1)^2}$ (savoir la retrouver).

EXERCICE 19. *Vérifier les valeurs d'espérance et variance pour les lois citées en exemple, en les calculant par la fonction génératrice.*

EXERCICE 20. ** Démontrer la réciproque : si G_X est dérivable en 1, alors l'espérance existe (c'et vrai aussi aux ordres supérieurs).*

Plus généralement, on a en itérant le TDTàT

DÉFINITION 20. *Le* **moment d'ordre n** *d'une v.a. X à valeurs dans $\mathbb{N}$ est $E(X^n)$ (quand la série $\sum k^n \mathcal{P}(X = k)$ converge absolument). Si il existe, alors G_X est n fois dérivable et le moment peut s'exprimer en fonction des dérivées en 1 de G_X jusqu'à l'ordre n.*

REMARQUE 7. *On peut montrer que réciproquement, l'existence d'une dérivée d'ordre k de G_X (à gauche) en 1 implique l'existence des moments d'ordre $\leqslant k$.*

EXERCICE 21. *Expression du moment d'ordre 3 en fonction des dérivées de G_X ?*

4.3 Propriétés des fonctions génératrices

Elles montrent leur puissance pour confronter plusieurs v.a.

THÉORÈME 10. *Soient X, Y deux v.a. indépendantes à valeurs dans $\mathbb{N}$ [pour simplifier]. Alors*

$$G_{X+Y} = G_X \times G_Y$$

— La réciproque est fausse.
— Ceci se généralise immédiatement à un nombre fini de v.a. mutuellement indépendantes.
— Quand les lois sont toutes identiques, on trouve le résultat plaisant $G_{X_1+\dots X_n} = (G_X)^n$.

Démonstration. C'est essentiellement le théorème de sommation d'une série produit :

$$\mathcal{P}(X+Y=n) = \sum_{k=0}^{n} \mathcal{P}(X=k)\mathcal{P}(Y=n-k) \quad \text{(par indépendance) d'où}$$

$$G_{X+Y}(t) = \sum_{n} \sum_{k=0}^{n} \mathcal{P}(X=k)\mathcal{P}(Y=n-k)t^n = \sum_{k=0}^{\infty} \mathcal{P}(X=k)t^k \sum_{\ell=0}^{\infty} \mathcal{P}(Y=\ell)t^\ell = G_X(t)G_Y(t)\ \forall t \in [-1,1]$$

car les deux séries sont absolument convergentes pour $|t| \leqslant 1$ (pour $|t| < 1$ il suffit de dire qu'on a des séries entières !) ♦

COROLLAIRE 4. *La somme de deux (ou plus) v.a. indépendantes suivant une loi de Poisson est une loi de Poisson, de paramètre la somme de leurs paramètres.*

EXERCICE 22. *Retrouver par produit des fonctions génératrices la loi de la somme de deux v.a. indépendantes de lois $\mathcal{B}(n,p)$ et $\mathcal{B}(m,p)$ (avec le même p).*