

## VI. VARIABLES ALÉATOIRES.

### VI.1. Le concept de variable aléatoire.

Dans les livres sur le Calcul des probabilités, on définit usuellement une variable aléatoire en disant que c'est une application définie sur l'espace  $\Omega$  des épreuves, et à valeurs réelles.

Ceci est une définition mathématique; l'idée sous-jacente est celle d'une grandeur qui prend des valeurs "au hasard". Voyons cela sur des exemples. Imaginons une expérience de physique avec un détecteur, par exemple on veut détecter les impacts d'électrons sur un écran fluorescent. La position de l'électron sur l'écran est imprévisible (peu importe que ce soit pour des raisons quantiques ou thermiques ou autres). On peut repérer cette position par deux coordonnées  $x$  et  $y$ . Mais ni  $x$ , ni  $y$ , n'ont une valeur déterminée. Elles peuvent prendre n'importe quelle valeur (entre les limites de l'écran), et d'un impact à l'autre ces valeurs varient au hasard: elles sont variables et aléatoires. C'est pourquoi on dit que  $x$  et  $y$  sont des *variables aléatoires*. Il s'agit maintenant de comprendre le rapport entre cette idée et la définition mathématique.

Au chapitre *I* nous avons vu que la structure de l'espace des épreuves  $\Omega$  contenait implicitement toute l'information sur la manière dont intervient le hasard pur dans le problème considéré. On parvient en quelque sorte au *hasard pur* lorsqu'on a identifié le niveau auquel se situent les choix équiprobables (nous avons insisté alors sur l'analogie entre le hasard *pur* et les repères galiléens). Ainsi, dans l'exemple des trois boules qui remplissent deux cases (figure 1), on a vu qu'il y a huit distributions différentes, dont l'équiprobabilité reflète les symétries spatio-temporelles: chaque boule a une chance sur deux exactement de choisir l'une ou l'autre case (symétrie spatiale) et l'expérience avec une boule est rigoureusement reproductible (invariance temporelle), de sorte que la deuxième ou la troisième boule n'est pas influencée par ce qui est arrivé aux précédentes. Si les deux cases ont la même probabilité  $\frac{1}{2}$  d'être choisies, on considère que c'est le hasard pur qui a présidé (si l'on peut dire) au choix; si l'une des deux cases avait plus de chances que l'autre, on penserait qu'une cause, peut-être inconnue, modifie les choix purs du hasard. Comme nous l'avons vu à la fin du chapitre **I**, quand il y a équiprobabilité quelque part, la quête d'une cause s'arrête, car on a alors atteint en quelque sorte le niveau de

connaissance maximale sur la partie compréhensible du phénomène. Cela est tout particulièrement convainquant dans le cas des quantons<sup>(1)</sup> : la propriété de non-séparabilité quantique fait que pour des bosons par exemple il n'y a pas huit distributions, mais quatre modes d'occupation équiprobables. Une conséquence de cela est que (si on lance les particules l'une après l'autre) le deuxième boson a une probabilité  $\frac{2}{3}$  d'aller dans l'état déjà occupé par le premier et une probabilité  $\frac{1}{3}$  d'aller dans l'état encore vide. On dit alors qu'il y a une cause qui influence le second boson afin qu'il préfère aller dans l'état déjà occupé et on donne des noms à cette cause : non-séparabilité, condensation de Bose, etc. Mais cela signifie simplement que le hasard pur n'intervient pas au niveau du choix, séparé pour chaque quanton, de l'état qu'il va occuper. Si on veut chercher à quel niveau intervient le hasard pur, il faut chercher ce qui est équiprobable et on trouve que ce sont les modes d'occupation (figure 1, colonne de droite).

Dans l'expérience avec un détecteur, nous mesurons les coordonnées  $x$  et  $y$  des impacts d'électrons sur un écran ; si nous constatons une certaine répartition, qui favorise certains points et en défavorise d'autres (par exemple si les impacts s'accumulent le long de certaines lignes et forment ainsi une figure d'interférence), nous dirons que le hasard pur n'agit pas directement au niveau du choix du point sur l'écran, et qu'il y a une cause qui favorise les crêtes des franges d'interférence. L'espace  $\Omega$  devra être recherché "en amont". Par contre si les impacts se répartissent uniformément sur l'écran, on pourra prendre pour  $\Omega$  l'ensemble des pixels de l'écran. Dans le cas où  $\Omega$  devra être recherché en amont, chaque élément de  $\Omega$  devra évidemment correspondre à un résultat possible ; autrement dit, à *chaque* épreuve  $\omega$ , élément de  $\Omega$ , doit correspondre un impact *déterminé* sur l'écran. Ici, attention ! Le mot *déterminé* ne doit pas être compris comme signifiant qu'il y a du déterminisme. Pour chaque électron, il y a un choix *non déterministe*, en ce sens que s'il y a eu un déterminisme quelconque dans ce choix, il a été effacé par un brouillage ; mais le choix "par le pur hasard" ne se situe pas au niveau de la position sur l'écran, puisqu'il n'y a pas équiprobabilité à ce niveau, et que donc l'action du hasard est mélangée avec celle d'une cause qui favorise les crêtes de franges. L'action du pur hasard se situe "en amont", à un niveau où il y a une équiprobabilité des choix, un niveau qu'il s'agit de découvrir et qui sera celui de l'espace des épreuves  $\Omega$ . Alors, pour chaque choix non-déterministe fait par le hasard à *ce niveau*, il correspondra un impact *déterminé* sur l'écran, de coordonnées  $x$  et  $y$ . Ainsi,  $x$  et  $y$  sont des fonctions de ce choix.

---

<sup>(1)</sup> particules soumises à la mécanique quantique ; cf. l'ouvrage de J. M. LEVY-LEBLOND et F. BALIBAR, *Quantique, rudiments*.

D'où la définition mathématique “une variable aléatoire est une fonction définie sur l'espace des épreuves, à valeurs réelles”. Cela signifie tout simplement qu'à chaque choix équiprobable  $\omega$  (une épreuve, un élément de  $\Omega$ ), la fonction fait correspondre un nombre (par exemple la coordonnée de l'impact sur l'écran. Le hasard pur choisit aveuglément, au niveau où il intervient, une épreuve parmi toutes les épreuves équiprobables possibles, sans en favoriser aucune, et on observe une valeur numérique  $x$ , qui est un effet de ce choix, mais qui n'a aucune raison d'être répartie uniformément sur l'ensemble des valeurs possibles. C'est cette valeur numérique qui constitue la variable aléatoire.

## VI.2. La loi d'une variable aléatoire. Espérance, variance.

### Fonction génératrice. Fonction caractéristique.

On peut donc résumer ce qui précède de la façon suivante: une variable aléatoire prend une valeur numérique déterminée pour chaque épreuve  $\omega$  (en langage mathématique: c'est une fonction de  $\Omega$  dans  $\mathbb{R}$ ). Ce n'est pas la valeur numérique prise par la variable aléatoire qui est “choisie au hasard”, mais l'épreuve  $\omega$ ; le choix de  $\omega$  parmi toutes les épreuves équiprobables détermine alors la valeur que prendra la variable aléatoire. Il est clair que les valeurs numériques prises par la variable aléatoire ne sont pas, sauf exception, équiprobables. Par conséquent chacune des valeurs numériques possibles a une certaine probabilité. On appelle alors la donnée de ces probabilités la **loi de la variable aléatoire**. En termes plus mathématiques:

Soit  $X$  une variable aléatoire. L'ensemble  $X(\Omega)$  de toutes les valeurs prises par  $X$  est nécessairement fini (et de cardinal au plus égal à celui de  $\Omega$ ), appelons  $x_1, x_2, x_3, \dots, x_N$  ses éléments. Pour chaque  $x_k$  soit  $A_k$  l'ensemble des épreuves  $\omega$  telles que  $X(\omega) = x_k$  ( $A_k = \{\omega \mid X(\omega) = x_k\}$ ).  $A_k$  est donc un événement qu'on peut désigner dans le langage imagé par “la variable aléatoire  $X$  prend la valeur  $x_k$ ”. Chaque  $A_k$  a une probabilité  $p_k = \#A_k/\#\Omega$ . La loi de  $X$  est la donnée des  $x_k$  et des  $p_k$ : on la notera  $\{x_k, (p_k)\}$ .

**Exemple a.** Des cordes étant distribuées au hasard sur un cercle de rayon  $R$  (selon l'un des trois procédés que nous avons vu au chapitre I), la *longueur* d'une corde est une variable aléatoire. Appelons-la  $\ell$ ; l'événement  $A$  dont nous avons calculé la probabilité au chapitre I était “ $\ell > R\sqrt{3}$ ”. Une autre variable aléatoire est  $d$ , la distance au centre du cercle. L'événement  $A$  est aussi “ $d < \frac{1}{2}R$ ”. Pour fixer les idées, mettons que les cordes sont distribuées selon le cas 2 (cf. figure 2), et que leur distribution est discrétisée en 360 degrés d'angle pour repérer (à 1 degré près, donc) la direction de la corde, et que le rayon  $R = 1m.$  est divisé en 100 *cm* (ainsi  $\#\Omega = 36000$ ). Alors la variable aléatoire  $d$  prend les valeurs numériques (en

cm.):  $x_0 = 0, x_1 = 1, x_2 = 2, x_3 = 3, \dots, x_{100} = 100$ , avec les probabilités  $p_0 = 0.01, p_1 = 0.01, p_2 = 0.01, p_3 = 0.01, \dots, p_{99} = 0.01$  (elles sont toutes égales, on peut dire que la loi est *uniforme*).

Si on s'intéresse à la variable  $\ell$ , on discrétisera les valeurs prises par la longueur. Celle-ci variant entre 0 cm et 200 cm, il est naturel de discrétiser en cm. On trouve alors que la probabilité pour que  $\ell = k$  cm. (ou de façon plus précise  $k \text{ cm.} \leq \ell < k + 1 \text{ cm.}$ ) est  $k / 2\sqrt{40000 - k^2}$  (cette fois la loi n'est plus uniforme).

**Exemple b.** On considère les marches aléatoires à  $2n$  pas. L'abscisse  $X$  atteinte par la marche au  $2n^{\text{e}}$  pas est une variable aléatoire qui prend les valeurs  $-2n, -2n+2, -2n+4, \dots, +2n-4, +2n-2, +2n$  avec les probabilités suivantes (cf. chapitre III) :

$$\begin{aligned} \mathcal{P}(X = -2n) &= 2^{-2n} \\ \mathcal{P}(X = -2n + 2) &= 2^{-2n} \binom{2n}{1} \\ \mathcal{P}(X = -2n + 4) &= 2^{-2n} \binom{2n}{2} \\ &\dots \\ \mathcal{P}(X = -2n + 2k) &= 2^{-2n} \binom{2n}{k} \\ &\dots \\ \mathcal{P}(X = 2n - 4) &= 2^{-2n} \binom{2n}{2} \\ \mathcal{P}(X = 2n - 2) &= 2^{-2n} \binom{2n}{1} \\ \mathcal{P}(X = 2n) &= 2^{-2n} \end{aligned}$$

Nous avons vu au chapitre II que les coefficients binômiaux  $\binom{2n}{k}$  atteignaient leur maximum pour  $k = n$ , que ce maximum, pour  $n$  grand, pouvait être approché par  $2^{2n} / \sqrt{\pi n}$ , et que au voisinage du maximum on avait (cf. II.6.)

$$\binom{2n}{n+j} \simeq \binom{2n}{n} \cdot e^{-\frac{j^2}{n}} \simeq \frac{2^{2n}}{\sqrt{\pi n}} \cdot e^{-\frac{j^2}{n}}$$

On en déduit que *si  $n$  est grand*, ces probabilités (c'est-à-dire la loi de  $X$ ) sont

$$\mathcal{P}(X = 2j) \simeq \frac{1}{\sqrt{\pi n}} \cdot e^{-\frac{j^2}{n}}$$

Pour  $j^2 \gg n$  ces probabilités sont infinitésimales; elles ne sont notables que lorsque  $j$  est du même ordre de grandeur que  $\sqrt{n}$  ou plus petit. Pour se faire une idée plus concrète, soit  $n = 1000$ ; alors  $\sqrt{n} \simeq 32$ ; si  $j > 100$ ,  $e^{-j^2/n} < e^{-10} \simeq 0.0000454$ , si  $j > 150$ ,  $e^{-j^2/n} < e^{-22.5} \simeq 1.7 \cdot 10^{-10}$ . On voit qu'il est pratiquement impossible que  $X$  prenne des valeurs supérieures à 300 ou inférieures à  $-300$ . La probabilité pour que  $X = -2000$  ou  $X = +2000$  est  $2^{-2000} \simeq 10^{-602}$ , alors que la probabilité pour que  $X = 0$  est  $2^{-2000} \cdot \binom{2000}{1000} \simeq 1/\sqrt{1000\pi} \simeq 1/56$ . Ainsi, alors que toutes les marches possibles sont équiprobables, on observe une énorme disparité pour les probabilités des valeurs prises par la variable  $X$ . Sur l'ensemble  $\Omega$  de tous les parcours possibles, le hasard n'en favorise aucun; mais précisément de ce fait il en résulte, *uniquement par l'intermédiaire des mécanismes de combinaisons*, que le hasard favorise énormément les valeurs de  $X$  comprises entre  $-300$  et  $+300$ , au détriment des valeurs extrêmes.

Si on avait pris  $n = 1\,000\,000$ , ce phénomène serait encore plus marqué; en effet, on aurait alors  $\mathcal{P}(X = \pm 2\,000\,000) = 2^{-2\,000\,000} \simeq 10^{-602\,060}$ , et  $\mathcal{P}(X = 0) \simeq 1/1772$ . La probabilité pour que  $X$  soit compris entre  $-5000$  et  $+5000$  est alors supérieure à 0.9996. Si on compte en valeur relative, c'est-à-dire à l'échelle de  $n$ , en posant  $x = X/n$ , cela signifie que la probabilité pour que  $x$  soit compris entre  $-0.005$  et  $+0.005$  est supérieure à 0.9996. Pour  $n = 1\,000\,000\,000$ ,  $x$  serait compris entre  $-0.0002$  et  $+0.0002$  avec une probabilité supérieure à 0.9999. Autrement dit, *par le seul mécanisme des combinaisons*, un choix aléatoire entre des parcours équiprobables se traduit, lorsque le nombre de pas  $n$  est très grand, par un quasi-déterminisme au niveau de l'ordonnée atteinte: on est pratiquement certain que celle-ci est proche de 0. En général, la loi est la suivante:

La probabilité pour que  $X/\sqrt{n}$  s'écarte de 0 de plus de 5 est inférieure à 0.0005; ou encore: la probabilité pour que  $x = X/n$  s'écarte de 0 de plus de  $5/\sqrt{n}$  est inférieure à 0.0005.

On est donc pratiquement certain que  $x$  sera dans un intervalle de largeur  $10/\sqrt{n}$  autour de 0. La largeur de cet intervalle diminue proportionnellement à  $1/\sqrt{n}$  quand  $n$  augmente, elle devient donc minuscule quand  $n$  est très grand. Le facteur 10 au numérateur de la fourchette  $10/\sqrt{n}$  est lié au degré de certitude (déterminé ici par la probabilité 0.0005); si on prend une fourchette moins large, le degré de certitude baisse: ainsi la probabilité pour que  $x$  s'écarte de 0 de plus de  $4/\sqrt{n}$  est 0.0046; pour  $3/\sqrt{n}$  elle est de 0.034, pour  $2/\sqrt{n}$  elle est de 0.157, et pour  $1/\sqrt{n}$  elle est de 0.48. On voit donc que si la fourchette est cinq fois plus étroite, il n'y a pratiquement plus de certitude. Tout cela résulte de la variation en  $e^{-t^2}$ : cette fonction garde une valeur notablement différente de zéro tant que  $t$  est inférieur à 2

environ ; pour  $t > 3$ , elle devient très vite extrêmement petite.

La propriété qui se manifeste ainsi est une propriété générale du hasard, que nous voyons ici à l'oeuvre dans un cas particulier. Cette propriété est **la loi des grands nombres** ; elle sera étudiée sous sa forme générale au chapitre suivant. La loi des grands nombres a pour effet de transformer le hasard en déterminisme. L'exemple ci-dessus de la variable aléatoire  $x = X/n$  le montre. Pour de petites valeurs de  $n$ , par exemple  $n = 2$  ou  $n = 5$ , la variable  $x$  est réellement *aléatoire*, c'est-à-dire que les différentes valeurs qu'elle peut prendre ( $-2, -1, 0, +1, +2$  si  $n = 2$  et  $-2, -1.6, -1.2, -0.8, -0.4, 0, +0.4, +0.8, +1.2, +1.6, +2$  pour  $n = 5$ ) sont imprévisibles et ont seulement une certaine probabilité. Mais si  $n$  est très grand, et qu'on mesure les valeurs prises par la variable avec une précision de l'ordre de  $\frac{5}{\sqrt{n}}$ , alors la variable  $x$  prend la valeur 0 avec certitude. Dans un tel processus, le hasard n'a pas cessé d'agir : c'est bien toujours "le hasard pur" qui effectue pour la marche aléatoire des choix équiprobables *sur l'ensemble de tous les parcours possibles* ; mais à un autre niveau d'observation, qui est celui des valeurs prises par la variable  $x$ , cela se traduit par du déterminisme.

Cette loi des grands nombres par laquelle le hasard se transforme en déterminisme est en quelque sorte l'inverse du chaos, qui transforme le déterminisme en hasard. Ce phénomène n'est pas différent de celui que nous avons plusieurs fois déjà souligné, à savoir que des choix équiprobables du hasard à un certain niveau se traduisent par des résultats non équiprobables à un autre niveau (que par exemple un choix équiprobable des deux extrémités d'une corde sur un cercle se traduisait par une répartition non équiprobable de la distance au centre) ; la loi des grands nombres est un effet de ce type, mais amplifié au point que la non-équiprobabilité devient extrême.

Nous avons étudié à la fin du chapitre **II** un exemple typique où intervient cette loi des grands nombres : la loi de Planck. En termes de variables aléatoires, on peut reformuler le problème étudié alors (cf. **II.6.**) en disant que les *nombre d'occupation*  $n_i$  de chaque intervalle d'énergie sont des variables aléatoires. La loi de Planck résultait de la loi des grands nombres, car la probabilité pour que les  $n_i$  soient proches de  $m_i/(\exp(\hbar\omega_i/kT) - 1)$  est pratiquement égale à 1 pour la même raison que ci-dessus pour la variable aléatoire  $x = X/n$ . Lorsqu'on constate expérimentalement la distribution selon les fréquences du rayonnement émis par un corps chauffé, la loi de Planck apparaît comme une loi déterministe. En réalité elle est un effet du hasard.

La Physique a montré que le monde macroscopique est entièrement gou-

verné par la loi des grands nombres<sup>(2)</sup>. Le monde des atomes et des particules est gouverné par le hasard, qui effectue des choix équiprobables sur l'ensemble de tous les états quantiques possibles (par exemple sur l'ensemble des modes d'occupations en statistique de Bose-Einstein). Mais au niveau des observations macroscopiques, cela se traduit par le déterminisme des lois de la thermodynamique ou de la mécanique classique. Il n'y a pas d'autre secret dans cette transformation du hasard en déterminisme, que la loi des grands nombres.

**Exemple c** On considère toujours les marches aléatoires, et on appelle  $T$  l'instant du premier retour à 0. Nous avons vu au chapitre III que

$$\mathcal{P}(T = k) = \begin{cases} 2^{-2\ell} \binom{2\ell}{\ell} \frac{1}{2^{\ell-1}} & \text{si } k = 2\ell \text{ et } \ell \neq 0 \\ 0 & \text{si } k \text{ est impair ou nul} \end{cases}$$

Ainsi  $T$  est une variable aléatoire qui prend les valeurs entières paires, avec les probabilités indiquées ci-dessus. Mais si on considère les marches aléatoires à  $2n$  pas, il peut arriver aussi que le retour à zéro ne se produise jamais ; la variable  $T$  peut donc prendre aussi la valeur “jamais”, qui n'est évidemment pas numérique si on prend la définition au pied de la lettre. On peut montrer que la probabilité de cette valeur “jamais” est

$$\sum_{j=n+1}^{\infty} 2^{-2j} \binom{2j}{j} \frac{1}{2^{j-1}} \simeq \frac{1}{\sqrt{\pi n}}$$

Ainsi la variable  $T$  prend les valeurs  $2, 4, 6, \dots, 2n-2, 2n$ , et “jamais” avec les probabilités respectives données ci-dessus.

**Remarque :** En calcul des probabilités on dit qu'une variable aléatoire qui peut ainsi n'avoir aucune valeur définie pour certaines épreuves est *incomplète*. Toutefois on voit que si  $n$  tend vers l'infini, la probabilité de la valeur “jamais” tend vers 0. Ceci donne envie de considérer parfois des espaces  $\Omega$  infinis. Par exemple dans le cas présent, on aurait pour  $\Omega$  l'ensemble de toutes les marches de durée infinie (donc sans limitation du nombre de pas). C'est la raison pour laquelle beaucoup (et même la plupart) des livres sur le Calcul des probabilités présentent une théorie pour des espaces  $\Omega$  infinis. Le prix à payer pour cela est toutefois exorbitant : la plupart des propriétés que nous avons rencontrées dans ce cours sont extrêmement faciles à établir tant que  $\Omega$  est un ensemble fini. Elles exigent des mathématiques très sophistiquées dès lors que  $\Omega$  est infini (intégrale de Lebesgue, analyse fonctionnelle, etc. En outre, elles ne permettent pas de

---

<sup>(2)</sup> Pour en savoir plus sur le rôle joué par la loi des grands nombres dans le comportement des corps macroscopiques, on pourra consulter : *Cours de Physique de Berkeley*, vol. 5, *Physique statistique* (Armand Colin, éd., coll. U), chapitre 1 : propriétés caractéristiques des systèmes macroscopiques.

calculer davantage de choses : comme cela a été montré dans un ouvrage récent<sup>(3)</sup>, tout ce qui peut être effectivement calculé dans le cadre de ces théories sophistiquées peut aussi être calculé en traitant d'abord le problème avec un espace  $\Omega$  fini, et en faisant tendre *ensuite*, dans les résultats obtenus, le cardinal de  $\Omega$  vers l'infini.

On peut faire la convention que “jamais” soit représentée par la valeur numérique 0. Cette valeur ne risque pas, en effet, d'être déjà prise par la variable  $T$ , puisque l'instant zéro est celui du départ et ne peut donc être celui d'un *retour*. Cela élimine l'inélégance d'une valeur non numérique sans recourir pour autant à l'infini. On peut pratiquement toujours se débrouiller avec de tels expédients.

Étant donnée une variable aléatoire  $X$  qui prend les valeurs  $x_1, x_2, x_3, \dots, x_{N-1}, x_N$  avec les probabilités respectives  $p_1, p_2, p_3, \dots, p_{N-1}, p_N$ , on appelle *moyenne* ou *espérance mathématique* de  $X$  la grandeur

$$\mathbf{E}(X) = \sum_{k=1}^N x_k p_k \tag{VI.1.}$$

On appelle *variance* de  $X$  la grandeur

$$\mathbf{Var}(X) = \sum_{k=1}^N (x_k - m)^2 p_k \tag{VI.2.}$$

avec  $m = \mathbf{E}(X)$ . En développant le carré  $(x_k - m)^2 = x_k^2 - 2mx_k + m^2$  on peut écrire aussi :

$$\begin{aligned} \mathbf{Var}(X) &= \sum_{k=1}^N x_k^2 p_k - 2m \sum_{k=1}^N x_k p_k + m^2 \sum_{k=1}^N p_k \\ &= \mathbf{E}(X^2) - 2m^2 + m^2 = \mathbf{E}(X^2) - m^2 \end{aligned} \tag{VI.2 a.)}$$

La moyenne ne nécessite aucune explication, elle relève de l'intuition première. La variance mérite par contre un commentaire. Comme on le voit dans la définition VI.2, c'est une somme de termes positifs (des carrés) ; la variance est donc d'autant plus grande que les nombres  $x_k - m$  sont plus grands, mais la somme étant pondérée par les probabilités  $p_k$ , une grande valeur pour  $x_k - m$  va peser d'autant plus que la probabilité correspondante,  $p_k$ , est plus grande.  $x_k - m$  est en fait l'écart de  $x_k$  par rapport à la moyenne  $m$ . La variance est la moyenne du carré de l'écart. Elle est d'autant plus grande que la variable  $X$  s'écarte davantage de sa moyenne, et d'autant plus grande aussi que ces écarts se produisent avec une forte probabilité.

---

<sup>(3)</sup> Edward NELSON *Radically Elementary Probability Theory*, Princeton University Press, 1989.

Autrement dit, la variance est une estimation de la tendance de la variable aléatoire à s'écarter de la moyenne, à se disperser. On aurait pu aussi mesurer cette tendance à la dispersion par la moyenne de la valeur absolue  $\sum p_k |x_k - m|$  (il faut évidemment prendre la valeur absolue, car si on calcule la moyenne de  $x_k - m$  en valeur algébrique, on trouve 0), ou par la moyenne de la puissance quatrième  $\sum p_k (x_k - m)^4$ . Comparé à la valeur absolue, le carré fait payer plus cher les grands écarts. La puissance quatrième les ferait payer encore plus cher. Il n'y a pas d'estimation absolue ou universelle de la dispersion des valeurs d'une variable aléatoire. Le choix du carré est purement conventionnel et s'il sert de référence c'est simplement parce qu'on n'a pas trop à s'en plaindre; entre la valeur absolue qui sous-estime les grands écarts et la puissance quatrième qui les surestime, c'est peut-être un bon compromis; il se peut aussi que l'analogie avec le moment d'inertie ait contribué à en faire un *standard mondial*. Pour voir les choses plus concrètement, on peut examiner un exemple très simple.

Soient les variables aléatoires :

$X$ , prenant les valeurs  $-2, -1, 0, +1, +2$  avec les probabilités respectives  $\frac{1}{20}, \frac{1}{10}, \frac{7}{10}, \frac{1}{10}, \frac{1}{20}$

$Y$ , prenant les valeurs  $-2, -1, 0, +1, +2$  avec les probabilités respectives  $\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}$

$Z$ , prenant les valeurs  $-2, -1, 0, +1, +2$  avec les probabilités respectives  $\frac{4}{10}, \frac{1}{10}, 0, \frac{1}{10}, \frac{4}{10}$ .

Pour les trois variables, la dispersion des valeurs est la même, mais la probabilité d'une grande dispersion est élevée pour  $Z$  et faible pour  $X$ . Leurs variances sont :

$$\mathbf{Var}(X) = 0.6 \quad \mathbf{Var}(Y) = 2.0 \quad \mathbf{var}(Z) = 3.4$$

ce qui reflète bien les différents types de dispersion (la variable  $Z$  à forte dispersion a une variance nettement plus élevée que la variable  $X$  à faible dispersion). Si on avait pris la moyenne des valeurs absolues pour mesurer la dispersion on aurait obtenu les valeurs :

$$\text{pour } X : 0.4 \quad \text{pour } Y : 1.2 \quad \text{pour } Z : 1.8$$

Comme prévu on voit que les trois cas sont moins nettement départagés (les grands écarts ne sont pas renforcés). Enfin, si on prend la moyenne des quatrièmes puissances :

$$\text{pour } X : 0.9 \quad \text{pour } Y : 6.8 \quad \text{pour } Z : 13.0$$

On voit l'effet de la surestimation des grands écarts. La variance est une convention très répandue, mais peut aussi devenir un pont-aux-ânes ; dans tel ou tel cas particulier on peut avoir de bonnes raisons de préférer estimer la dispersion par la quatrième puissance que par la variance, et il faut donner la priorité aux bonnes raisons sur le conformisme.

Un défaut de la variance est de ne pas mesurer l'écart dans les mêmes unités que la variable : si les  $x_k$  sont des *cm*, la variance sera en  $cm^2$ . C'est pourquoi on introduit une autre estimation de l'écart qui est la racine carrée de la variance ou *écart-type*. Pour les variables  $X, Y, Z$  les écarts-types sont

$$\text{pour } X : 0.77 \quad \text{pour } Y : 1.41 \quad \text{pour } Z : 1.8$$

Pour l'estimation par les puissances quatrièmes, on pourra également prendre la racine quatrième du résultat, et on obtient ainsi un écart-type d'ordre 4 :

$$\text{pour } X : 0.97 \quad \text{pour } Y : 1.61 \quad \text{pour } Z : 1.90$$

On voit que la surestimation des grands écarts est réduite. En règle générale, on peut dire ceci : avec un écart-type d'ordre  $n$  (racine  $n^e$  de la moyenne des valeurs absolues des puissances  $n^e$ ), plus  $n$  est élevé, plus on pénalise les grands écarts, même s'ils ont une probabilité faible.

Nous avons ainsi vu qu'on pouvait analyser la loi d'une variable aléatoire à l'aide de paramètres quantitatifs tels que la variance, l'écart-type, etc. Nous avons introduit la moyenne des puissances quatrièmes des écarts, qui comme les carrés, ou comme toute autre puissance paire, sont toujours positives. Si on voulait prendre des puissances impaires, il faudrait bien sûr prendre la moyenne *de leurs valeurs absolues* pour avoir une estimation significative de la dispersion. Toutefois, la moyenne d'une puissance impaire, sans valeur absolue, contient aussi de l'information sur la loi de la variable aléatoire ; quoique elle n'exprime pas la dispersion, une telle grandeur joue un rôle dans l'analyse de Fourier des lois de probabilité. On appelle **moments** ce type de grandeurs. De façon précise :

On appelle **moment d'ordre  $n$**  d'une variable aléatoire  $X$  la moyenne  $\sum p_k x_k^n$  des  $n^e$  puissances des valeurs de  $X$ . La moyenne de  $X$  est ainsi le moment d'ordre 1 de  $X$  et la variance de  $X$  est le moment d'ordre 2 de  $X - m$ .

Les paramètres quantitatifs tels que les moments ou la moyenne des valeurs absolues, les écarts-types, etc. contiennent une information sur la loi de la variable aléatoire, mais ne contiennent pas toute l'information. On ne peut pas retrouver la loi d'une variable aléatoire si on n'en connaît que la moyenne et l'écart-type. Sous certaines conditions assez compliquées mais pas trop restrictives, la donnée de tous les moments détermine la loi de la variable, mais nous laissons cette question de côté. Par contre il y a deux manières très efficaces de condenser la totalité de l'information sur la

loi d'une variable aléatoire, c'est d'introduire la *fonction génératrice* et la *fonction caractéristique*.

Étant donnée une variable aléatoire de loi  $\{x_k, (p_k)\}$ , on appelle *fonction caractéristique* de  $X$  la fonction

$$\Phi_X(t) = \sum_k p_k e^{itx_k} \quad (VI.3.)$$

En fait c'est simplement la transformée de Fourier de la loi. La fonction génératrice est

$$G_X(z) = \sum_k p_k z^{x_k} \quad (VI.4.)$$

La fonction génératrice est plutôt recommandée lorsque les  $x_k$  sont des nombres entiers; dans ce cas c'est un polynôme, ou, si on fait tendre le nombre des  $x_k$  vers l'infini, une fonction analytique de  $z$ . On bénéficie alors de tous les outils mathématiques liés aux polynômes ou aux fonctions d'une variable complexe. Rien n'interdit dans le principe de la prendre en considération même lorsque les  $x_k$  sont non entiers, mais dans un tel cas elle est beaucoup moins commode et on lui préférera alors la fonction caractéristique.

De toute façon, fonction caractéristique et fonction génératrice sont liées par la relation

$$\Phi_X(t) = G_X(e^{it}) \quad (VI.5.)$$

Pour une variable aléatoire  $X$  à valeurs entières, on obtient les probabilités de chaque valeur (c'est-à-dire la loi de la variable aléatoire  $X$ ) en développant la fonction génératrice  $G_X(z)$  en série entière, ou en série de Laurent s'il y a des valeurs négatives.

On peut aussi déduire directement de la fonction génératrice d'autres grandeurs liées à la variable aléatoire telles que la moyenne et la variance. Ainsi, la moyenne n'est autre que  $G'_X(1)$  (la dérivée de  $G_X(z)$  au point  $z = 1$ ). En effet

$$G'_X(z) = \sum_n n p_n z^{n-1}$$

donc pour  $z = 1$  cela donne

$$G'_X(1) = \sum_n n p_n = \mathbf{E}(X) \quad (VI.6.)$$

La dérivée seconde fournira la variance :

$$G''_X(z) = \sum_n n(n-1) p_n z^{n-2}$$

ce qui pour  $z = 1$  donne  $G_X''(1) = \sum_n n(n-1)p_n = \sum_n n^2 p_n - \sum_n n p_n = \mathbf{E}(X^2) - \mathbf{E}(X)$ . En utilisant VI.2 a on peut avoir la variance sous la forme :

$$\mathbf{Var}(X) = G_X''(1) + G_X'(1) - G_X'(1)^2 \quad (VI.7.)$$

Des expressions analogues pour la moyenne ou la variance peuvent être obtenues à partir des fonctions caractéristiques. Ces expressions seront utilisées lorsque la variable aléatoire n'est pas à valeurs entières. On obtient en effet par dérivation :

$$\begin{aligned} \Phi_X'(t) &= \sum_k i x_k p_k e^{i x_k t} \\ \Phi_X''(t) &= \sum_k -x_k^2 p_k e^{i x_k t} \end{aligned}$$

donc pour  $t = 0$

$$\begin{aligned} \Phi_X'(0) &= i \sum_k x_k p_k = i \mathbf{E}(X) \\ \Phi_X''(0) &= - \sum_k x_k^2 p_k = - \mathbf{E}(X^2) \end{aligned}$$

Cela conduit (compte tenu aussi de VI.2 a) aux expressions suivantes de la moyenne et de la variance :

$$\begin{aligned} \mathbf{E}(X) &= -i \Phi_X'(0) \\ \mathbf{Var}(X) &= \Phi_X'(0)^2 - \Phi_X''(0) \end{aligned} \quad (VI.8.)$$

Comme nous connaissons déjà quelques variables aléatoires, nous pouvons à titre d'exemple illustratif calculer leurs fonctions génératrices ou leurs fonctions caractéristiques. Reprenons les exemples *b* et *c* ci-dessus : il s'agissait de (*a*) l'abscisse atteinte par une marche aléatoire au bout de  $2n$  pas ; (*b*) l'instant du premier retour en zéro. Les fonctions génératrices correspondantes sont

$$\begin{aligned} G_a(z) &= \sum_{j=-n}^{+n} 2^{-2n} \binom{2n}{n+j} z^{2j} = \left[ \frac{1}{2} \left( z + \frac{1}{z} \right) \right]^{2n} \\ G_b(z) &= \sum_{j=1}^n 2^{-2n} \binom{2j}{j} \frac{1}{2j-1} z^{2j} + \text{reste} \end{aligned}$$

Le reste n'est pas vraiment défini puisqu'il correspond à la valeur "*jamais*" du premier retour mais il tend vers zéro lorsque  $n$  tend vers l'infini, et on peut alors montrer que la limite est  $G_b(z) = 1 - \sqrt{1-z^2}$ . Si on convient,

comme nous en avons évoqué la possibilité plus haut, de réserver la valeur 0 pour “jamais”, le reste sera tout simplement le terme constant

$$\sum_{j=n+1}^{\infty} 2^{-2j} \binom{2j}{j} \frac{1}{2j-1}$$

On voit que l’expédient qui consiste à prendre la valeur zéro pour “jamais” n’apporte aucune simplification dans les calculs : les fonctions génératrices sont en général un puissant outil de calcul. Dans ce cas particulier, elles le restent quand  $n$  est grand, car  $G_b(z)$  est alors une fonction dont l’expression analytique est simple.

Quant aux fonctions caractéristiques, elles s’obtiennent en remplaçant  $z$  par  $e^{it}$  dans les expressions de la fonction génératrice, soit :

$$\begin{aligned} \Phi_a(t) &= \left[ \frac{1}{2} (e^{it} + e^{-it}) \right]^{2n} = (\cos(t))^{2n} \\ \Phi_b(t) &\simeq 1 - \sqrt{1 - e^{2it}} \end{aligned}$$

(ici  $\simeq$  signifie que l’égalité est approchée pour  $n$  grand).

### IV.3. L’indépendance stochastique de deux (ou plus) variables aléatoires.

Au chapitre IV nous avons étudié l’indépendance stochastique de deux événements ou de deux familles d’événements. Nous avons vu que l’indépendance stochastique était toujours liée à une factorisation de l’espace  $\Omega$  : deux événements  $A$  et  $B$  étaient alors stochastiquement indépendants si on pouvait factoriser l’ensemble  $\Omega$  sous la forme d’un tableau (une matrice) rectangulaire, dans lequel l’événement  $A$  était un ensemble de lignes (plus exactement une réunion de lignes complètes) et l’événement  $B$  un ensemble de colonnes. De même une famille d’événements  $\{A_i\}_{i=1,2,\dots,m}$  est stochastiquement indépendante d’une famille d’événements  $\{B_j\}_{j=1,2,\dots,n}$ , si on peut factoriser  $\Omega$  de telle manière que les  $A_i$  soient des ensembles de lignes et les  $B_j$  des ensembles de colonnes. On a alors la propriété  $\mathcal{P}(A_i B_j) = \mathcal{P}(A_i) \times \mathcal{P}(B_j)$ .

On dit que deux variables aléatoires  $X$  et  $Y$  de lois respectives  $\{x_i(p_i)\}$  et  $\{y_j(q_j)\}$  sont stochastiquement indépendantes si la famille d’événements  $A_i : X = x_i$  est stochastiquement indépendante de la famille  $B_j : Y = y_j$ . Il y a alors une factorisation de  $\Omega$  dans laquelle les événements  $A_i$  sont formés de lignes complètes et les  $B_j$  de colonnes complètes ; cela signifie que la variable aléatoire  $X$  est constante le long des lignes et  $Y$  est constante le long des colonnes. Une épreuve  $\omega$  dans le tableau est définie par la donnée

de son numéro de ligne  $\ell(\omega)$  et son numéro de colonne  $col(\omega)$ . On peut alors exprimer l'indépendance stochastique de  $X$  et  $Y$  en disant que  $X(\omega)$  ne dépend que de  $\ell(\omega)$ , et  $Y(\omega)$  ne dépend que de  $col(\omega)$ . En tous cas, on a la propriété de factorisation  $\mathcal{P}(X = x_i \text{ et } Y = y_j) = \mathcal{P}(X = x_i) \times \mathcal{P}(Y = y_j) = p_i q_j$ , qui évidemment n'aurait pas lieu si  $X$  et  $Y$  n'étaient pas stochastiquement indépendantes.

Lorsque deux variables aléatoires sont stochastiquement indépendantes, il y a des règles simples pour calculer leur somme. En effet, l'événement  $X + Y = z_k$  est la réunion des événements disjoints  $X = x_i$  et  $Y = y_j$  pour lesquels  $x_i + y_j = z_k$ ; or d'après l'indépendance stochastique on peut dire que  $\mathcal{P}(X = x_i \text{ et } Y = y_j) = p_i q_j$ , donc

$$\mathcal{P}(X + Y = z_k) = \sum_{x_i + y_j = z_k} p_i q_j$$

(la sommation sur  $x_i + y_j = z_k$  signifie qu'on ne retient pour la somme que les indices  $i$  et  $j$  pour lesquels  $x_i + y_j = z_k$ ). Si on applique cela à la fonction caractéristique de la variable aléatoire  $X + Y$ , on obtient

$$\begin{aligned} \Phi_{X+Y}(t) &= \sum_k \mathcal{P}(X + Y = z_k) e^{itz_k} \\ &= \sum_k \left[ \sum_{x_i + y_j = z_k} p_i q_j \right] e^{itz_k} \\ &= \sum_k \left[ \sum_{x_i + y_j = z_k} p_i q_j \right] e^{it(x_i + y_j)} \\ &= \sum_k \left[ \sum_{x_i + y_j = z_k} p_i e^{itx_i} q_j e^{ity_j} \right] \\ &= \sum_{i,j} p_i e^{itx_i} q_j e^{ity_j} \end{aligned}$$

qui n'est autre que le produit  $\Phi_X(t) \times \Phi_Y(t)$  des fonctions caractéristiques de  $X$  et de  $Y$ .

On peut donc énoncer :

**Théorème :** Si  $X$  et  $Y$  sont des variables aléatoires stochastiquement indépendantes, alors la fonction caractéristique de leur somme est le produit des fonctions caractéristiques de chacune :

$$\Phi_{X+Y}(t) = \Phi_X(t) \times \Phi_Y(t) \tag{VI.9.}$$

Bien entendu, puisque  $\Phi_X(t) = G_X(e^{it})$ , il est clair qu'on a la même propriété pour les fonctions génératrices :

$$G_{X+Y}(z) = G_X(z) \times G_Y(z) \tag{VI.10.}$$

mais comme on le verra dans les exemples et exercices, cette relation pour les fonctions génératrices, quoique toujours vraie, n'est d'un emploi commode que si les  $x_i$  et  $y_j$  sont entiers.

On peut interpréter l'abscisse atteinte au bout de  $2n$  pas par une marche aléatoire comme la somme de  $2n$  variables aléatoires  $X_j$  :

$$X_j = \begin{cases} +1 & \text{(avec prob. } 1/2) \text{ si le } j^{\text{e}} \text{ pas est un pas en avant} \\ -1 & \text{(avec prob. } 1/2) \text{ si le } j^{\text{e}} \text{ pas est un pas en arrière} \end{cases}$$

l'abscisse atteinte après  $2n$  pas est alors égale à  $S = X_1 + X_2 + X_3 + \dots + X_{2n}$ . Ces variables aléatoires sont stochastiquement indépendantes, puisque chaque pas d'une marche aléatoire est stochastiquement indépendant des autres. La fonction génératrice de chacune des  $X_j$  est  $G_j(z) = \frac{1}{2}(z + 1/z)$  et d'après le théorème ci-dessus on doit avoir  $G_S(z) = [\frac{1}{2}(z + 1/z)]^{2n}$ , ce qui effectivement recoupe le calcul direct précédent. On pouvait donc utiliser la multiplication des fonctions génératrices pour calculer la loi de  $S$ , au cas où on ne l'aurait pas déjà calculée par un autre procédé.

#### VI.4. La loi conjointe de deux (ou plus) variables aléatoires.

Lorsque deux variables aléatoires sont indépendantes, on peut calculer la probabilité d'événements tels que  $\{X = x_j ; Y = y_k\}$  à partir de la seule connaissance des lois respectives de  $X$  et  $Y$ . Il suffit pour cela d'appliquer la règle du produit. En effet, l'événement  $\{X = x_j ; Y = y_k\}$  (" $X = x_j$  et  $Y = y_k$ ") est l'intersection  $\{X = x_j\} \cap \{Y = y_k\}$ ; donc sa probabilité est, d'après la règle du produit, égale à  $\mathcal{P}(X = x_j) \cdot \mathcal{P}(Y = y_k)$ .

Par contre, si  $X$  et  $Y$  ne sont pas stochastiquement indépendantes, on ne peut pas déduire les probabilités  $\mathcal{P}(X = x_j ; Y = y_k)$  des probabilités  $\mathcal{P}(X = x_j)$  et  $\mathcal{P}(Y = y_k)$ . Il faut disposer d'une information supplémentaire, qui caractérise la dépendance entre  $X$  et  $Y$ .

Supposons que  $Y$  soit une fonction de  $X$ , par exemple  $Y = X^2$ . Dans ce cas on peut calculer les probabilités  $\mathcal{P}(X = x_j ; Y = y_k)$  à partir des probabilités  $\mathcal{P}(X = x_j)$  et  $\mathcal{P}(Y = y_k)$ .

On peut même faire plus fort : les calculer à partir des seules  $\mathcal{P}(X = x_j)$ . En effet, les  $y_k$  sont alors nécessairement les carrés des  $x_j$ ; si deux  $x_j$  distincts, disons  $x_{j_1}$  et  $x_{j_2}$ , ont le même carré  $y_k$  (cela équivaut à dire que  $x_{j_2} = -x_{j_1}$ ), alors  $\mathcal{P}(Y = y_k) = \mathcal{P}(X = x_{j_1}) + \mathcal{P}(X = x_{j_2})$ ; si un seul des  $x_j$ , disons  $x_{j_0}$ , a pour carré  $y_k$ , alors  $\mathcal{P}(Y = y_k) = \mathcal{P}(X = x_{j_0})$ . On voit que dans ce cas la loi de  $Y$  se déduit directement de celle de  $X$ . Quant aux probabilités  $\mathcal{P}(X = x_j ; Y = y_k)$ , elles sont évidemment nulles si  $y_k$  n'est le carré d'aucun  $x_j$ , et égales à  $\mathcal{P}(X = x_j)$  sinon. On peut dire que non

seulement il n'y a pas plus d'information dans la donnée des probabilités des événements faisant intervenir les deux variables aléatoires  $X$  et  $Y$  que dans la donnée des probabilités des événements faisant intervenir séparément chacune des deux variables aléatoires  $X$  et  $Y$ , mais même qu'il n'y en a pas plus que dans la donnée des probabilités des événements faisant intervenir la seule variable aléatoire  $X$ .

L'exemple que nous venons de voir est en fait un cas extrême de dépendance. Il s'agit du cas où  $X$  et  $Y$  sont fonction l'une de l'autre.  $X$  est bien aléatoire, c'est-à-dire que les valeurs prises par  $X$  dépendent d'un choix (fait sur  $\Omega$ ) du hasard; mais la valeur prise par  $X$  détermine alors univoquement la valeur prise par  $Y$ , de sorte qu'il n'y a aucune intervention *supplémentaire* du hasard pour déterminer  $Y$ . Entre les deux extrêmes, il peut y avoir des formes de dépendance moins radicales qu'une telle relation fonctionnelle: que par exemple, la valeur de  $X$  étant fixée, plusieurs valeurs puissent être prises par  $Y$ , chacune avec une certaine probabilité, mais que ces probabilités dépendent de la valeur de  $X$ . L'indépendance stochastique signifie que les probabilités pour que  $Y$  prenne une valeur  $y_k$  ne sont pas influencées par la valeur prise par  $X$ . Si ces probabilités dépendent de la valeur de  $X$ , il y a une dépendance *stochastique* de  $Y$  par rapport à  $X$ , mais  $Y$  n'est pas pour autant une *fonction* de  $X$ . Pour rendre cela plus concret, prenons les trois exemples que voici.

$X$  est une variable aléatoire qui prend les valeurs  $-2, -1, 0, +1, +2$ .  $Y$  est une autre variable aléatoire qui prend les valeurs  $0, +1, +4$ . Chacun des trois tableaux ci-contre représente une répartition différente des probabilités  $\mathcal{P}(X = x_j ; Y = y_k)$ . Dans les trois tableaux, les lois de probabilités de  $X$  et de  $Y$  sont les mêmes, c'est-à-dire que les probabilités

$$p_j = \mathcal{P}(X = j) = \sum_{k=0,1,4} \mathcal{P}(X = j ; Y = k)$$

pour  $j = -2, -1, 0, +1, +2$ , ainsi que les probabilités

$$q_k = \mathcal{P}(Y = k) = \sum_{j=0,\pm 1,\pm 2} \mathcal{P}(X = j ; Y = k)$$

pour  $k = 0, 1, 4$ , sont les mêmes dans les trois tableaux (on obtient les  $p_j$  en faisant la somme des éléments de la colonne  $j$  et les  $q_k$  en faisant la somme des éléments de la ligne  $j$ ). Par contre ces trois exemples diffèrent par la répartition bidimensionnelle des probabilités, qui reflète la dépendance entre les deux variables.

Dans le premier tableau la répartition est faite de façon à assurer l'indépendance stochastique entre  $X$  et  $Y$ . En effet, si on calcule les

	$X = -2$	$X = -1$	$X = 0$	$X = +1$	$X = +2$
$Y = 0$	1/25	1/25	1/25	1/25	1/25
$Y = +1$	2/25	2/25	2/25	2/25	2/25
$Y = +4$	2/25	2/25	2/25	2/25	2/25

**tableau 1**

Ce tableau représente la loi de probabilité conjointe de deux variables aléatoires  $X$  et  $Y$  *indépendantes*. On trouve la probabilité pour que  $X = j$  et  $Y = k$  (avec  $j = -2, -1, 0, +1, +2$  et  $k = 0, 1, 4$ ) à l'intersection de la ligne  $Y = k$  et de la colonne  $X = j$ . On peut constater que la probabilité conditionnelle pour que  $Y$  prenne l'une de ses trois valeurs est la même quelle que soit la valeur de  $X$  (les rapports des nombres figurant dans une même ligne à la somme des éléments de la ligne sont les mêmes pour chaque ligne ; de même pour les colonnes).

	$X = -2$	$X = -1$	$X = 0$	$X = +1$	$X = +2$
$Y = 0$	0	0	1/5	0	0
$Y = +1$	0	1/5	0	1/5	0
$Y = +4$	1/5	0	0	0	1/5

**tableau 2**

Ce tableau représente la loi de probabilité conjointe de deux variables aléatoires  $X$  et  $Y$  *liées* par la relation  $Y = X^2$ . On voit que les probabilités sont nulles lorsque  $Y \neq X^2$ . Mais la loi de  $X$  (somme des probabilités figurant sur une même colonne) et la loi de  $Y$  (somme des probabilités figurant sur une même ligne) sont les mêmes que dans le tableau 1.

	$X = -2$	$X = -1$	$X = 0$	$X = +1$	$X = +2$
$Y = 0$	1/60	1/30	1/10	1/30	1/60
$Y = +1$	1/15	1/10	1/15	1/10	1/15
$Y = +4$	7/60	1/15	1/30	1/15	7/60

**tableau 3**

Ce tableau représente la loi de probabilité conjointe de deux variables aléatoires  $X$  et  $Y$  qui ne sont ni indépendantes, ni fonction l'une de l'autre. La probabilité conditionnelle pour que  $Y = k$  sachant que  $X = j$  dépend de  $j$ , mais par exemple pour  $X = 0$  il y a comme on voit trois valeurs possibles pour  $Y$ , avec probabilités non nulles 1/10, 1/15, 1/30. La probabilité conditionnelle pour que  $Y = 0$  sachant que  $X = 0$  est  $\frac{1}{2}$ ; mais la probabilité conditionnelle pour que  $Y = 0$  sachant que  $X = 1$  est  $\frac{1}{6}$  et la probabilité conditionnelle pour que  $Y = 0$  sachant que  $X = 2$  est  $\frac{1}{12}$ : elle dépend fortement de la valeur de  $X$ , mais n'est pas égale à 0 ou 1 comme c'était le cas pour le tableau 2.

probabilités conditionnelles

$$\mathcal{P}(X = j | Y = k) = \frac{\mathcal{P}(X = j ; Y = k)}{\mathcal{P}(Y = k)} = \frac{\mathcal{P}(X = j ; Y = k)}{\sum_j \mathcal{P}(X = j ; Y = k)}$$

qui sont donc pour chaque ligne le quotient des éléments de la ligne par la somme des éléments de la ligne, on constate que ces probabilités conditionnelles ne dépendent pas de  $k$ , ce qui est bien la marque de l'indépendance stochastique; au demeurant  $\mathcal{P}(X = j | Y = k) = \mathcal{P}(X = j)$ . De même les probabilités conditionnelles

$$\mathcal{P}(Y = k | X = j) = \frac{\mathcal{P}(X = j ; Y = k)}{\mathcal{P}(X = j)} = \frac{\mathcal{P}(X = j ; Y = k)}{\sum_k \mathcal{P}(X = j ; Y = k)}$$

qui sont donc pour chaque colonne le quotient des éléments de la colonne par la somme des éléments de la colonne, ne dépendent pas de  $j$ . (Vérifiez cela sur le tableau avec un crayon).

Le second tableau représente la répartition des probabilités lorsque  $Y = X^2$ . Cette dépendance absolue fait que  $Y$  ne peut pas prendre une valeur qui ne soit pas le carré de  $X$ . Si  $X = 2$ ,  $Y$  ne peut pas être égal à 0 ou à 1, ce qui se traduit sur le tableau par le fait que la probabilité pour que  $X = 2$  et  $Y = 0$ , ou pour que  $X = 2$  et  $Y = 1$ , est nulle.

Enfin, dans le troisième tableau, on a un exemple de dépendance *stochastique*, c'est-à-dire que la variable  $Y$  n'est pas une fonction de  $X$ , mais n'est pas non plus stochastiquement indépendante de  $X$  : lorsque  $X$  prend la valeur  $+2$ ,  $Y$  ne prend pas *forcément* la valeur  $+4$ , et peut prendre n'importe laquelle des trois valeurs  $0$ ,  $1$ , ou  $4$ . Mais cette fois, contrairement à ce qui se passe sur le tableau 1, les probabilités conditionnelles  $\mathcal{P}(Y = k \mid X = j)$  dépendent de  $j$ , et aussi les probabilités conditionnelles  $\mathcal{P}(X = j \mid Y = k)$  dépendent de  $k$ . Par exemple, la probabilité (conditionnelle) pour que  $Y = 0$  sachant que  $X = 0$  est  $\frac{1}{2}$ , mais la probabilité conditionnelle pour que  $Y = 0$  sachant que  $X = +1$  est  $\frac{1}{6}$  et la probabilité conditionnelle pour que  $Y = 0$  sachant que  $X = +2$  est  $\frac{1}{12}$ . Dans le tableau 2, si  $X = 0$ , il est *certain* que  $Y = 0$  ; dans le tableau 1, la probabilité conditionnelle pour que  $Y = 0$  sachant que  $X = j$  est *toujours*  $\frac{1}{5}$  quelle que soit la valeur  $j$  de  $X$ .

Tout cela montre que, si deux variables aléatoires  $X$  et  $Y$  ne sont ni stochastiquement indépendantes, ni liées par une relation fonctionnelle, on ne peut déduire à *partir de leurs deux lois* les probabilités de toutes les combinaisons de valeurs qu'elles peuvent prendre conjointement : il faut disposer d'un tableau qui donne directement l'information complète. On appelle *loi conjointe* des deux variables aléatoires  $X$  et  $Y$  la donnée de ce tableau, c'est-à-dire la donnée de

- a) toutes les valeurs  $x_j$  que peut prendre  $X$  et toutes les valeurs  $y_k$  que peut prendre  $Y$  ;
- b) toutes les probabilités  $r_{j,k} = \mathcal{P}(X = x_j ; Y = y_k)$

Pour désigner commodément une loi conjointe on écrira  $\{x_j, y_k ; (r_{j,k})\}$ . Cela signifiera que la probabilité pour que  $X = x_j$  et  $Y = y_k$  est  $r_{j,k}$ . Cette loi conjointe contient donc une information qui *n'est pas* déjà contenue dans les deux lois  $\{x_j ; (p_j)\}$  et  $\{y_k ; (q_k)\}$ , sauf, comme nous l'avons déjà dit, si  $X$  et  $Y$  sont stochastiquement indépendantes (alors  $r_{j,k} = p_j q_k$ ) ou si  $Y = f(X)$ ,  $f$  étant une fonction non aléatoire (alors  $r_{j,k} = 0$  si  $y_k \neq f(x_j)$  et  $r_{j,k} = p_j$  si  $y_k = f(x_j)$ , comme sur le tableau 2).

Les deux lois  $\{x_j ; (p_j)\}$  et  $\{y_k ; (q_k)\}$ , c'est-à-dire les lois de  $X$  et de  $Y$  sont appelées *lois marginales*, par opposition à la loi conjointe du couple  $X, Y$ . Lorsque la loi conjointe de deux variables est représentée sur un tableau à double entrée comme les tableaux 1, 2, et 3, on obtient les lois marginales en effectuant les sommes sur les lignes et sur les colonnes.

Cela peut se généraliser à un nombre quelconque de variables aléatoires : au lieu d'un tableau à deux indices (ou matrice), on aurait un tableau à  $n$  indices ou dimensions.

Tout cela n'est bien sûr qu'une question de langage; si dès le départ (et certains auteurs procèdent ainsi) on avait annoncé que les variables aléatoires sont des grandeurs vectorielles, que ce sont bien des fonctions définies sur l'espace  $\Omega$ , mais à valeurs dans  $\mathbb{R}^n$  et non dans  $\mathbb{R}$ , on n'aurait pas eu à distinguer le cas de plusieurs variables aléatoires du cas d'une seule; la loi d'une variable aléatoire vectorielle  $X$  serait la donnée des valeurs vectorielles  $x_j$  et de leurs probabilités, qui serait évidemment la même chose que la loi conjointe des  $n$  composantes de  $X$ . Dans ce langage vectoriel on ne parlerait pas de l'indépendance stochastique entre  $n$  variables aléatoires, mais entre les  $n$  composantes d'une variable aléatoire.

Ce qui en tous cas reste certain, c'est que (dit dans le langage non vectoriel) si deux variables aléatoires ne sont ni stochastiquement indépendantes, ni liées par une relation fonctionnelle, on ne peut calculer les probabilités qui leur sont relatives que si on connaît leur loi conjointe.

Précédemment (§VI.3) nous avons associé à la loi d'une variable aléatoire des grandeurs comme les moments (la variance) ou des fonctions comme les fonctions génératrices ou caractéristiques. On peut en faire autant avec les lois conjointes. Ainsi on appelle *covariance* des deux variables aléatoires  $X$  et  $Y$  (de loi conjointe  $\{x_j, y_k ; (r_{j,k})\}$ ) la grandeur

$$\mathbf{Cov}(X, Y) = \sum_{j,k} r_{j,k} [x_j - \mathbf{E}(X)][y_k - \mathbf{E}(Y)] \quad (VI.11.)$$

De même on appellera *fonction génératrice conjointe* de  $X$  et  $Y$  la fonction de *deux* variables complexes

$$G_{X,Y}(z, w) = \sum_{j,k} r_{j,k} z^{x_j} w^{y_k} \quad (VI.12.)$$

(peu pratique si les nombres  $x_j$  et  $y_k$  ne sont pas entiers) et *fonction caractéristique conjointe* de  $X$  et  $Y$  la fonction de *deux* variables réelles

$$\Phi_{X,Y}(s, t) = \sum_{j,k} r_{j,k} e^{isx_j} e^{ity_k} \quad (VI.13.)$$

Bien entendu si  $X$  et  $Y$  sont stochastiquement indépendantes on aura

$$G_{X,Y}(z, w) = G_X(z) \cdot G_Y(w) \quad (VI.14.)$$

et

$$\Phi_{X,Y}(s, t) = \Phi_X(s) \cdot \Phi_Y(t) \quad (VI.15.)$$

Si  $X$  et  $Y$  sont stochastiquement indépendantes, leur covariance est nécessairement nulle: en effet, dans ce cas on aura  $r_{j,k} = p_j q_k$ , et par

conséquent l'expression (VI.11.) de la covariance se factorise

$$\begin{aligned} \mathbf{Cov}(X, Y) &= \sum_{j,k} r_{j,k} [x_j - \mathbf{E}(X)][y_k - \mathbf{E}(Y)] \\ &= \sum_j p_j [x_j - \mathbf{E}(X)] \times \sum_k q_k [y_k - \mathbf{E}(Y)] \\ &= 0 \end{aligned}$$

Une erreur tentante chez les débutants est de croire que la réciproque a lieu : que si la covariance de deux variables aléatoires est nulle, alors elles sont indépendantes. Cela est faux et on en a deux exemples avec les tableaux 2 et 3 : si on calcule la covariance de  $X$  et  $Y$  avec la loi conjointe du tableau 2 ou du tableau 3, on trouve qu'elle est nulle, alors que  $X$  et  $Y$  ne sont indépendantes dans aucun de ces deux cas.

Nous avons vu au §VI.2. que si deux variables aléatoires  $X$  et  $Y$  sont indépendantes, on peut facilement calculer la loi de leur somme en faisant le produit des fonctions génératrices ou des fonctions caractéristiques. Lorsque deux variables ne sont pas indépendantes, mais qu'on connaît leur loi conjointe, et par conséquent leur fonction génératrice ou caractéristique conjointe, on obtient la fonction génératrice ou caractéristique de leur somme par les formules suivantes :

$$\begin{aligned} \Phi_{X+Y}(t) &= \Phi_{X,Y}(t, t) \\ G_{X+Y}(z) &= G_{X,Y}(z, z) \end{aligned} \tag{VI.16.}$$

Si on combine cela avec (VI.15.) ou (VI.14.) on retrouve (VI.9.) ou (VI.10.)

Cela s'étend comme on le devine au cas de 2, 3, 4... variables aléatoires.