

IV. L'INDÉPENDANCE STOCHASTIQUE.

IV.1. La signification de l'indépendance stochastique.

L'idée d'indépendance stochastique est assez naturelle; reprenons par exemple le problème des boules qu'on jette successivement dans des cases. Au chapitre *I* nous avons vu que l'équiprobabilité des huit distributions dans la colonne de gauche de la figure 1 résultait des deux faits suivants :

- a) les deux cases sont symétriques de par leurs positions dans l'espace (aucune des deux n'est privilégiée).
- b) chaque boule *ignore* ce qui est arrivé aux précédentes.

La condition b) est bien sûr aussi nécessaire que la première, car si elle n'était pas vérifiée, les deux cases cesseraient d'être symétriques après que la première boule ait été placée (ce qui justement est le cas pour des particules de Bose). La condition b) signifie donc qu'il y a *indépendance* entre les lancers successifs.

C'est cette sorte d'indépendance qui correspond au concept d'*indépendance stochastique*.

Lorsque le Calcul des probabilités fut axiomatisé, on chercha à donner une définition purement mathématique de l'indépendance stochastique. Une définition purement mathématique, cela signifie une définition qui ne recourt pas à des considérations physiques (telles que la séparabilité spatio-temporelle), ni à des considérations philosophiques (telles que la "nature" de l'ignorance mutuelle des boules). Une définition purement mathématique doit porter uniquement sur ce qui se passe à l'intérieur du modèle mathématique, c'est-à-dire, dans le cas du Calcul de probabilités, à l'intérieur de l'espace Ω . Mais Ω n'a aucune structure mathématique en dehors de la structure nue d'ensemble ⁽¹⁾. Il faut donc fournir une définition qui porte sur cette seule structure d'ensemble, c'est-à-dire sur les *éléments* (épreuves), les *parties* (événements), les *ensembles de parties*, etc.

Commençons par observer la première colonne de la figure 1. Dans ce cas Ω est un ensemble à huit éléments, il y a donc exactement $2^8 = 256$

⁽¹⁾ Bien sûr, comme nous l'avons vu au chapitre *I*, la *structure* de cet ensemble doit contenir sous forme implicite toute l'information qu'on possède sur le problème concret sous-jacent, mais la construction ou le choix de cette structure concerne la phase de *modélisation*, et non la partie purement mathématique du Calcul des probabilités qui, elle, présuppose simplement un ensemble Ω déjà donné.

parties (en incluant la partie vide et le tout). Il est bien clair qu'une étude exhaustive, comme celle qui va suivre, serait difficile avec un Ω à plusieurs milliards d'éléments ! Pour faciliter la nomenclature, numérotons les épreuves de 1 à 8, en allant du haut vers le bas sur la figure 1 : ainsi l'épreuve 5 est la cinquième en partant du haut sur la colonne de gauche de la figure 1 (voir cette figure) et correspond au cas : "boule 1 dans la case de gauche ou case $\mathcal{A}$, boules 2 et 3 dans la case de droite ou case $\mathcal{B}$ ", etc. Parmi les 256 sous-ensembles possibles, on peut remarquer les six suivants, qui expriment chacun un événement concernant l'une des trois boules :

$$A_1 : \text{"la boule 1 est dans la case } \mathcal{A} \text{"} = \{1, 2, 4, 5\}$$

$$A_2 : \text{"la boule 2 est dans la case } \mathcal{A} \text{"} = \{1, 2, 3, 6\}$$

$$A_3 : \text{"la boule 3 est dans la case } \mathcal{A} \text{"} = \{1, 3, 4, 7\}$$

$$B_1 : \text{"la boule 1 est dans la case } \mathcal{B} \text{"} = \{3, 6, 7, 8\}$$

$$B_2 : \text{"la boule 2 est dans la case } \mathcal{B} \text{"} = \{4, 5, 7, 8\}$$

$$B_3 : \text{"la boule 3 est dans la case } \mathcal{B} \text{"} = \{2, 5, 6, 8\}$$

Par exemple, A_1 est constitué des éléments N^{os} 1, 2, 4, et 5 c'est-à-dire des épreuves que voici :



En regardant la figure 1 on voit bien que ces quatre épreuves sont la première, la deuxième, la quatrième, et la cinquième en partant du haut.

De même, A_3 est constitué des éléments N^{os} 1, 3, 4, et 7 c'est-à-dire :



Les intersections deux par deux de ces six événements sont :

$$A_1 A_2 = \{1, 2\} \quad A_1 A_3 = \{1, 4\} \quad A_2 A_3 = \{1, 3\}$$

$$A_1 B_2 = \{4, 5\} \quad A_1 B_3 = \{2, 5\} \quad A_2 B_3 = \{2, 6\}$$

$$A_2 B_1 = \{3, 6\} \quad A_3 B_1 = \{3, 7\} \quad A_3 B_2 = \{4, 7\}$$

$$B_1 B_2 = \{7, 8\} \quad B_1 B_3 = \{6, 8\} \quad B_2 B_3 = \{5, 8\}$$

(on n'a pas compté $A_1 B_1$, $A_2 B_2$, $A_3 B_3$, qui sont évidemment vides puisqu'une boule ne peut pas être à la fois dans la case $\mathcal{A}$ et dans la case $\mathcal{B}$)

Les intersections par trois peuvent également être notées ; les seules qui sont non vides sont les suivantes :

$$A_1 A_2 A_3 = \{1\} \quad A_1 A_2 B_3 = \{2\} \quad A_1 B_2 A_3 = \{4\} \quad B_1 A_2 A_3 = \{3\}$$

$$B_1 A_2 B_3 = \{6\} \quad A_1 B_2 B_3 = \{5\} \quad B_1 B_2 A_3 = \{7\} \quad B_1 B_2 B_3 = \{8\}$$

On voit que toutes les intersections par deux sont composées de deux éléments, et que toutes les intersections par trois sont composées d'un seul élément (si elles ne sont pas vides); en termes de probabilités cela se traduit par le fait que les probabilités des événements A_j ou B_j valent $\frac{1}{2}$, celles des intersections par deux valent $\frac{1}{4}$ (qui est le carré de $\frac{1}{2}$), et celles des intersections par trois valent $\frac{1}{8}$ (qui est le cube de $\frac{1}{2}$); autrement dit les égalités $\mathcal{P}(E \cap F) = \mathcal{P}(E) \times \mathcal{P}(F)$ et $\mathcal{P}(E \cap F \cap G) = \mathcal{P}(E) \times \mathcal{P}(F) \times \mathcal{P}(G)$ s'appliquent pour les intersections retenues ci-dessus. En revanche elles ne s'appliquent pas pour les intersections vides, telles que $A_1 B_1$, $A_2 B_2$, $A_1 A_2 B_1$, $A_1 B_2 A_2$, etc, ni pour les intersections trop pleines telles que $A_1 A_1 = A_1$, $A_1 B_2 B_2$, etc. Un examen plus attentif montre que les intersections qui *ne vérifient pas* la règle "probabilité de l'intersection égale produit des probabilités" sont celles où un indice est répété, tandis que les intersections qui *vérifient* la règle sont celles où les indices sont distincts. Or l'indice est le numéro de la boule; on peut donc dire que la règle $\mathcal{P}(E \cap F) = \mathcal{P}(E) \times \mathcal{P}(F)$ s'applique si les événements E et F concernent⁽¹⁾ des boules différentes, et ne s'applique plus lorsque les événements E et F concernent une seule et même boule. C'est ainsi que la règle du produit exprime mathématiquement le fait que par exemple la boule N°2 "ignore" ce qui est arrivé à la boule N°1.

Un examen encore plus approfondi de ce phénomène montre cependant que le modèle mathématique (c'est-à-dire l'ensemble Ω) est plus riche en événements indépendants que la *véritable* causalité physique ne l'impose. Considérons par exemple les événements suivants :

$$C_1 = \{1, 2, 3, 4\} \qquad C_2 = \{3, 4, 5, 6\} \qquad C_3 = \{2, 4, 6, 8\}$$

On observera que la règle du produit s'applique aussi à ces trois événements : en effet

$$C_1 C_2 = \{3, 4\} \quad C_1 C_3 = \{2, 4\} \quad C_2 C_3 = \{4, 6\}$$

ainsi que

$$C_1 C_2 C_3 = \{4\}$$

de sorte qu'on a également $\mathcal{P}(C_1 C_2) = \mathcal{P}(C_2 C_3) = \mathcal{P}(C_1 C_3) = \frac{1}{4}$ et $\mathcal{P}(C_1 C_2 C_3) = \frac{1}{8}$, mais on ne voit pas très bien quelles propriétés se trouvent ainsi *séparées* quant à la causalité. En réalité le cas des événements C_1 , C_2 ,

⁽¹⁾ Lorsqu'on dit par exemple que l'événement E *concerne* la boule N°1, il faut comprendre par là qu'en tant que sous-ensemble de Ω il regroupe toutes les épreuves où la boule N°1 joue un rôle particulier : par exemple l'événement A_1 est l'ensemble de toutes les épreuves où la boule N°1 est dans la case $\mathcal{A}$.

C_3 , n'est qu'un artéfact du modèle mathématique. Cela apparaît encore plus clairement si au lieu de se limiter à trois boules et deux cases on considère le cas général de m boules et N cases : les événements du type A_j , B_j , etc. conservent leur sens ainsi que la règle du produit ; on trouvera pour tout m et tout N des événements tels que les C_j , qui vérifient aussi la règle du produit, mais sans qu'on puisse leur trouver un sens qui soit conservé lorsque le nombre de boules ou de cases change.

On peut donc introduire la définition mathématique suivante :

Deux événements E et F seront dits *stochastiquement indépendants* si $\mathcal{P}(E \cap F) = \mathcal{P}(E) \times \mathcal{P}(F)$; plus généralement, si $E_1, E_2, E_3, \dots, E_m$ et $F_1, F_2, F_3, \dots, F_n$ sont deux familles d'événements sur un espace d'épreuves Ω , on dira que la famille $\{E_i\}$ est *stochastiquement indépendante* de la famille $\{F_j\}$ si pour tout couple (i, j) on a

$$\mathcal{P}(E_i \cap F_j) = \mathcal{P}(E_i) \times \mathcal{P}(F_j)$$

Dans ce cas n'importe quelle réunion de plusieurs d'entre les E_i sera stochastiquement indépendante de n'importe quelle réunion de plusieurs d'entre les F_j .

Mais il ne faut pas oublier qu'une telle définition formelle, par nature, introduit toujours des artéfacts.

On comprendra mieux la signification de l'indépendance stochastique formelle en la reliant à l'arithmétique : elle n'est rien d'autre que le reflet de propriétés purement arithmétiques liées à la divisibilité du nombre entier $\#\Omega$. On pouvait déjà soupçonner cela d'après la propriété du produit : en effet dire que $\mathcal{P}(AB) = \mathcal{P}(A) \times \mathcal{P}(B)$ revient à dire que $\#(AB) \times \#\Omega = \#(A) \times \#(B)$, ce qui montre bien que $\#\Omega$ et $\#(A)$ ou $\#(B)$ doivent avoir des diviseurs communs. Tout cela devient plus clair si on représente l'espace Ω non comme une liste séquentielle, mais comme un tableau qui reflète la divisibilité de $\#\Omega$ par exemple si $\#\Omega = 35$ on peut présenter la liste complète de ses éléments comme un tableau à cinq lignes et sept colonnes (ou vice-versa). La colonne de gauche de la figure 1, dans laquelle $\#\Omega = 8 = 4 \times 2$ peut ainsi être représentée comme sur la figure 8.

La colonne de gauche de la figure 8 est l'événement que nous avons désigné ci-dessus par A_1 et la colonne de droite l'événement B_1 . L'événement A_2 est formé des deux premières *lignes*, l'événement B_2 des deux dernières lignes, l'événement B_3 des lignes N^{os} 2 et 3, l'événement A_3 des lignes N^{os} 1 et 4. La propriété du produit est donc simplement l'expression du fait bien connu que le nombre d'éléments d'un tableau est égal au nombre de lignes multiplié par le nombre de colonnes.

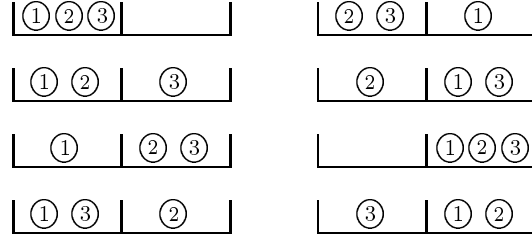


figure 8

Bien sûr dans la figure 8 les configurations ont été rangées en tenant compte du *sens* de l'expérience, c'est-à-dire du rôle joué par les trois boules, de même que le choix des événements $A_1, A_2, A_3, B_1, B_2, B_3$, avait été dicté par ce sens. Mais on conserve cette propriété de multiplication en permutant n'importe comment les éléments de la deuxième colonne : dans ce cas les colonnes sont globalement inchangées (elles consistent toujours dans les événements A_1 et B_1) mais les lignes sont différentes : par exemple sur la figure 8 les lignes $N^{\text{os}} 2$ et 3 avaient en commun que la boule $N^{\circ} 3$ était dans la case $\mathcal{B}$ (événement B_3) ; après permutation de la deuxième colonne il ne reste plus rien qui ait un rapport avec la boule $N^{\circ} 3$: le sens est perdu, mais la propriété de multiplication subsiste.

En fait $\#\Omega = 8$ peut se décomposer davantage ; au lieu de 2×4 , c'est $2 \times 2 \times 2$, donc au lieu d'un tableau à quatre lignes et deux colonnes, c'est un cube, ou un tableau à trois entrées de deux éléments chacune : la première entrée correspond aux événements A_1 et B_1 , la deuxième entrée aux événements A_2 et B_2 , et la troisième entrée aux événements A_3 et B_3 (voir figure 9).

On retrouve ainsi ce qui en réalité avait présidé dès le départ à la construction du modèle mathématique Ω . L'espace des épreuves a une structure qui reflète directement les invariances physiques ou la causalité qui sont à l'origine de l'équiprobabilité. Mais en construisant un tel modèle, on introduit ipso facto des régularités supplémentaires (qui dérivent de celles qu'on a postulé a priori par des permutations) qui n'ont plus aucune signification par rapport au problème qu'on voulait modéliser.

Ainsi, on peut voir sur la figure 9 que l'événement A_1 ("la boule $N^{\circ}1$ est dans la case $\mathcal{A}$ ") correspond à la face gauche du cube ; l'événement A_2 à la face avant ; l'événement A_3 à la face de dessous ; l'événement B_1 à la face de droite ; l'événement B_2 à la face arrière ; l'événement B_3 à la face de dessus. Les intersections A_1A_2, A_1A_3 , etc. sont donc des intersections de deux faces, c'est-à-dire des arêtes (ou rien si les faces sont opposées comme A_1B_1). Les

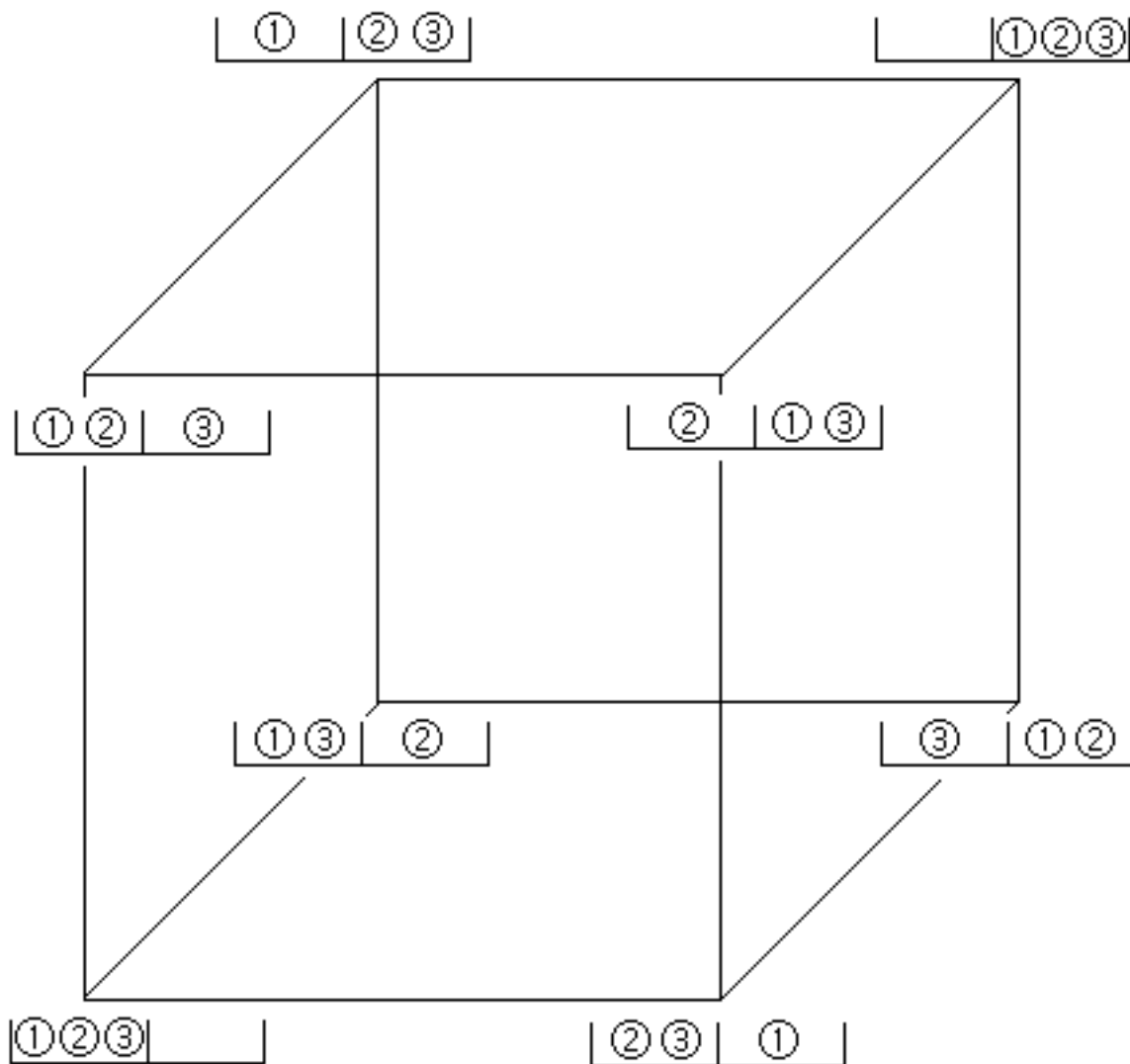


figure 9

intersections par trois $A_1A_2A_3$, $A_1B_2A_3$, etc sont des sommets. Tout ce qui avait été dit plus haut s'interprète donc selon la structure de *produit cartésien* de l'espace Ω . Le fait que chacune des trois boules ignore ce que font les deux autres (ce qui constitue la propriété causale caractéristique du problème à modéliser) est traduit dans le modèle Ω par un produit cartésien d'ensembles : pour chaque boule il y a deux possibilités $\mathcal{A}$ ou $\mathcal{B}$, et il y a trois boules qui s'ignorent mutuellement *donc* l'ensemble Ω sera de la forme $\{\mathcal{A}, \mathcal{B}\}^3$. L'espace Ω de toutes les épreuves possibles pour n cases et r boules serait un hypercube à r dimensions, avec des arêtes à n éléments ; en toute logique, son cardinal est n^r , conformément à (II.1.) Si maintenant on permute n'importe comment chaque arête, on ne change pas la structure

de produit, mais on fait disparaître le rapport avec les boules.

L'indépendance stochastique formelle d'une famille d'événements par rapport à une autre, c'est-à-dire le simple fait d'avoir $\mathcal{P}(AB) = \mathcal{P}(A) \times \mathcal{P}(B)$, est toujours liée à une décomposition de Ω en produit d'ensembles plus petits (et donc aussi à la décomposition de l'entier $\#\Omega$ en produit de facteurs premiers). Tout problème de probabilité qui comporte une indépendance *causale* se modélisera sous forme d'un produit cartésien qui reflète directement et par construction cette indépendance causale; mais cela n'entraîne de loin pas que réciproquement, toute autre décomposition de Ω en produit implique une signification causale.

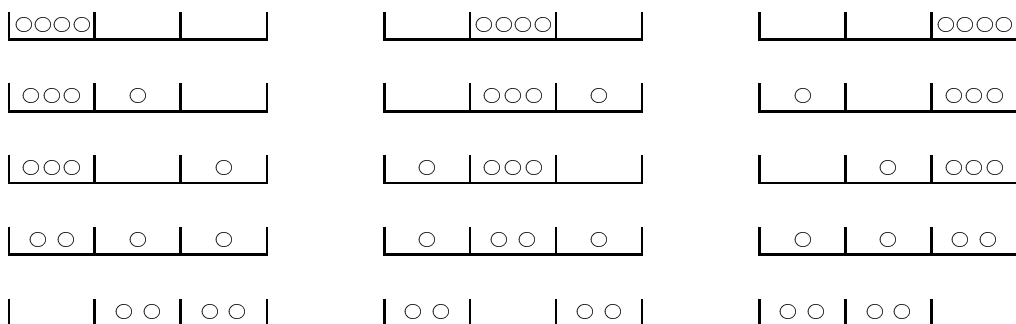


figure 10

C'est ce qu'on peut constater par exemple dans le cas de la statistique de Bose-Einstein. Le nombre de configurations pour r particules et N états est $(r + N - 1)! / r! (N - 1)!$. Ce nombre n'est certes jamais premier et on peut donc toujours décomposer Ω d'une façon ou d'une autre en produit. Mais contrairement à ce qui était le cas pour les boules, où le nombre de facteurs du produit était toujours égal au nombre de boules, l'espace Ω pour m particules de Bose et N états quantiques est bien décomposable en produit, mais de différentes manières, qui varient selon les valeurs de m et de N , sans qu'on puisse observer une règle permanente; ceci reflète le fait que lorsque des particules de Bose viennent occuper des états quantiques, elles ne s'y rangent pas indépendamment les unes des autres (cette propriété est appelée la *non-séparabilité quantique*). Il n'y a pas d'indépendance causale entre les particules, mais cela n'empêche pas de trouver dans chaque cas particulier (c'est-à-dire pour chaque valeur particulière de m ou N) des factorisations possibles pour l'espace Ω .

On a par exemple représenté sur la figure 10 l'espace Ω pour une statistique de Bose à 4 particules et 3 états; il y a $15 = 5 \times 3$ modes d'occupation possibles, donc une factorisation en tableau à deux dimensions.

La même structure que celle représentée sur la figure 10 peut être décrite par le tableau de chiffres que voici :

(4, 0, 0)	(0, 4, 0)	(0, 0, 4)
(3, 1, 0)	(0, 3, 1)	(1, 0, 3)
(3, 0, 1)	(1, 3, 0)	(0, 1, 3)
(2, 1, 1)	(1, 2, 1)	(1, 1, 2)
(0, 2, 2)	(2, 0, 2)	(2, 2, 0)

On peut s'amuser à constater que par exemple la famille d'événements $\{A_j\}$ ($j = 1, 2, 3$) correspondant aux colonnes du tableau peut recevoir le sens physique suivant : A_j est l'événement "l'état $N^\circ j$ possède le nombre d'occupation non ex-aequo le plus élevé". La famille $\{B_i\}$ correspondant à des lignes recevrait le sens suivant :

- B_1 "les nombres d'occupation sont 4, 0, et 0" (1^{re} ligne);
- B_2 "les nombres d'occupation sont 3, 1, et 0" (2^e et 3^e lignes);
- B_3 "les nombres d'occupation sont 2, 1, et 1" (4^e ligne);
- B_4 "les nombres d'occupation sont 2, 2, et 0" (5^e ligne);

Les événements de la famille A sont stochastiquement indépendants de ceux de la famille B . Ceci a-t-il une signification quant à la causalité physique ? Par exemple cela aurait-il la signification que les particules, une fois que leur répartition en groupes d'effectifs fixés est faite, vont placer le groupe le plus nombreux *indépendamment* de la répartition ? Pour que cela corresponde à une propriété physique, il faudrait au moins que cela reste vrai quel que soit le nombre d'états et le nombre de particules ; or si on prend trois particules occupant trois états, il y a $10 = 2 \times 5$ configurations, données par le tableau suivant :

(3, 0, 0)	(0, 2, 1)
(2, 0, 1)	(0, 1, 2)
(1, 2, 0)	(0, 3, 0)
(1, 0, 2)	(0, 0, 3)
(1, 1, 1)	(2, 1, 0)

On peut toujours fabriquer des paires d'événements indépendants en regroupant des lignes et des colonnes, mais la propriété physique qui avait semblé apparaître dans le cas précédent a complètement disparu : il n'y a aucune chance de finir par la faire apparaître en permutant correctement la deuxième colonne par rapport à la première, puisque tout simplement le nombre de colonnes n'est plus égal au nombre d'états, et ne *peut pas* l'être, parce que 3 n'est pas un diviseur de 10.

En fait tout est lié à la divisibilité de l'entier $\#\Omega$: plus il possède de diviseurs, plus nombreuses sont les familles d'événements stochastiquement

indépendantes. Mais il est rarissime que cette indépendance stochastique formelle soit le reflet d'une indépendance causale. Lorsque c'est le cas, c'est que l'indépendance causale a été postulée a priori et a *présidé* à la fabrication du modèle : comme dans le cas des trois boules, le modèle est alors fabriqué délibérément sous forme de produit cartésien, précisément pour refléter une indépendance causale sous la forme de l'indépendance stochastique. Par permutations il s'y ajoute encore une foule d'autres familles stochastiquement indépendantes d'événements, mais qui n'ont aucune signification causale. Dans le cas des particules de Bose, le modèle n'est pas fabriqué au départ sous forme de produit cartésien, et s'il se trouve qu'il est factorisable, cela ne reflète rien de physique.

Il est bien sûr impossible d'affirmer avec certitude qu'aucune de ces factorisations “parasites” ne peut être le reflet d'une causalité encore inconnue ; mais ces factorisations non voulues à l'avance proviennent de propriétés arithmétiques des nombres entiers, et disparaîtraient sous leur forme exacte si par exemple l'équiprobabilité postulée a priori n'était qu'approximative au lieu d'être mathématiquement exacte. Or on ne peut donner un sens physique à une propriété qui disparaît dès lors que les paramètres physiques subissent une variation arbitrairement petite. C'est pourquoi une propriété du type $\mathcal{P}(AB) \simeq \mathcal{P}(A) \times \mathcal{P}(B)$ (avec $\simeq$ au lieu de $=$), mais qui ne changerait pas complètement pour chaque nouvelle valeur de m ou N , aurait bien plus de chances d'avoir un sens physique que des factorisations exactes.

IV.2. L'effacement de la causalité.

Afin de bien comprendre ce que signifie réellement l'indépendance stochastique, il faut encore la relier à la *nature* du hasard. Tout au long de ce chapitre, nous avons beaucoup insisté sur l'indépendance stochastique en tant que reflet d'une indépendance causale sous-jacente. Ainsi l'exemple des trois boules “qui s'ignorent mutuellement”. Dans ce cas, l'ignorance mutuelle s'expliquait par la Mécanique classique. On pourrait étudier plus en détail le rôle joué par la Mécanique en reprenant le modèle simplifié de roulette présenté au chapitre I. Les trois boules seraient une bille de roulette qu'on aurait lancée trois fois. L'événement que la boule tombe dans la boîte $\mathcal{A}$ correspondrait à un résultat tel que “rouge”, “impair”, ou “manque”. Par exemple on peut supposer que le disque est divisé en trente six secteurs égaux (d'ouverture 10 degrés chacun), numérotés de 1 à 36 ; on dira que le résultat est “impair” si la bille s'arrête dans un secteur portant un numéro impair. Lorsqu'on lance la bille avec un certain angle θ et un coefficient de frottement k , le point du disque où la bille s'arrêtera est, comme nous avons vu au chapitre I, parfaitement déterminé, mais il y a un phénomène

de chaos déterministe qui crée du hasard. Le résultat, par exemple “impair”, est déterminé par les conditions initiales de lancement, c’est-à-dire par les valeurs – rigoureusement exactes – de θ et k . L’indépendance causale entre les trois lancers signifie que les valeurs de θ et k au moment où on lance une deuxième fois la bille sont indépendantes du point où celle-ci était arrivée la première fois, et de même, que ces valeurs au moment où on lance une troisième fois la bille sont indépendantes des points où celle-ci était arrivée la première et la deuxième fois. Lorsqu’on voit les choses ainsi, il est clair qu’il n’y a eu aucune intervention du hasard entre le moment où la bille a été lâchée sur le disque et la manifestation du résultat : le hasard serait intervenu au moment du “choix” des valeurs de θ et k et plus jamais ensuite. Or toute la discussion présentée en **I.4.** sur le hasard montrait au contraire, que non seulement *du* hasard est créé par le mouvement lui-même, mais que c’est même *le* hasard qui par nature est toujours créé comme cela. Même si un *autre* hasard est intervenu au moment du choix des valeurs de θ et k , celui-ci n’est peut-être pas d’une autre nature, il peut lui aussi avoir été produit par un effet de chaos (par exemple les valeurs de θ et k peuvent être choisies par une fonction **random**, ou bien la bille peut avoir été lancée par un croupier alcoolique dont la main tremble), mais en renvoyant tout à ce hasard premier nous n’expliquons rien, nous ne faisons que déplacer le problème.

C’est pourquoi nous allons maintenant voir comment le mouvement déterministe de la bille crée l’indépendance stochastique là où au départ il n’y a pas d’indépendance causale, en effaçant en quelque sorte la causalité. On complètera ainsi la discussion du chapitre **I**, car le mécanisme par lequel le chaos transforme le déterminisme en hasard consiste précisément à rendre *stochastiquement* indépendants des phénomènes qui sont *causalement* dépendants.

Imaginons que la bille soit bien lancée trois fois, mais que les choix initiaux de θ et k , au lieu d’être causalement indépendants entre eux, soient étroitement dépendants : pour fixer les idées, disons que, si θ_1 et k_1 sont les valeurs initiales du premier lancer et sont, elles, “choisies au hasard”, alors le second lancer sera effectué avec $\theta_2 = \alpha \theta_1$ et $k_2 = k_1$, et le troisième avec $\theta_3 = \theta_1$ et $k_3 = \beta k_1$, α et β étant des paramètres fixés à l’avance. On suppose bien entendu comme au chapitre **I** que k_1 est petit, de l’ordre de 10^{-4} , de sorte que la bille ne s’arrêtera qu’après plusieurs milliers de tours. Dans ces conditions les trois lancers ne sont absolument pas indépendants, puisque θ_2 , θ_3 , k_2 , et k_3 sont fonctions de θ_1 et k_1 . De ce fait les choix de θ_1 , θ_2 , et θ_3 (ou de k_1 , k_2 , et k_3) ne sont pas stochastiquement indépendants, ce qu’on peut vérifier aisément : si on divise l’intervalle des valeurs possibles de θ , soit $]0, \pi[$ en deux demi-intervalles, par exemple en $]0, \frac{\pi}{2}[$ et $[\frac{\pi}{2}, \pi[$,

alors la probabilité pour que θ_1 , qui est choisi au hasard, soit dans l'un ou l'autre de ces deux demi-intervalles, par exemple $]0, \frac{\pi}{2}[$, est bien sûr $\frac{1}{2}$; mais (compte tenu que les valeurs de θ_1 , θ_2 , et θ_3 sont fonctions les unes des autres) la probabilité pour que θ_1 , θ_2 , et θ_3 soient *tous les trois* dans ce même demi-intervalle n'est pas $\frac{1}{8}$; un rapide calcul montre qu'elle est égale à $\frac{1}{2\alpha}$ si $\alpha > 1$ et à $\frac{1}{2}$ si $\alpha \leq 1$; en effet, puisque $\theta_2 = \alpha \theta_1$ et $\theta_3 = \theta_1$, on voit bien que si $\alpha > 1$, la condition nécessaire et suffisante pour que θ_1 , θ_2 , et θ_3 soient tous les trois dans l'intervalle $]0, \frac{\pi}{2}[$ est que θ_1 soit dans $]0, \frac{\pi}{2\alpha}[$, événement dont la probabilité est bien $\frac{1}{2\alpha}$; si par contre $\alpha < 1$, cette condition nécessaire et suffisante est que θ_1 soit dans $]0, \frac{\pi}{2}[$ (il est vrai que si $\alpha = 4$, la probabilité est quand même $\frac{1}{8}$, mais ceci est un cas particulier sans signification). Cela exprime la forte dépendance entre les trois valeurs. S'il y avait indépendance stochastique entre les trois choix de θ_1 , θ_2 , et θ_3 , alors, puisque chacun des trois choix aurait une probabilité $\frac{1}{2}$ de tomber dans le demi-intervalle $]0, \frac{\pi}{2}[$, la probabilité pour que tous les trois y tombent à la fois serait le produit $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$. Mais comme $\theta_3 = \theta_1$, si θ_1 y tombe, θ_3 y sera automatiquement aussi et θ_2 y sera également si θ_1 est dans $]0, \frac{\pi}{2\alpha}[$. Un constat analogue vaut pour k_1 , k_2 , et k_3 , ou pour les neuf couples (θ_i, k_j) .

Puisque les conditions initiales des trois lancers ne sont pas indépendantes, et que les positions finales atteintes par la bille sont rigoureusement déterminées par les conditions initiales, il ne peut pas y avoir non plus indépendance causale rigoureuse entre les trois positions finales. Et pourtant, si on fait l'expérience, on obtiendra malgré cela des résultats stochastiquement indépendants. C'est-à-dire que la dépendance causale des choix de θ_1 , θ_2 , et θ_3 se traduit bien par leur dépendance stochastique (la probabilité que les trois soient dans le demi-intervalle $]0, \frac{\pi}{2}[$ est $\frac{1}{2\alpha}$ et non $\frac{1}{8}$), mais les trois positions *en fin de trajectoire*, bien qu'étant toujours, en toute rigueur, causalement liées, seront, elles, stochastiquement indépendantes.

Le phénomène de chaos déterministe produit par le très grand nombre de réflexions sur le bord, que nous avons analysé au chapitre I, aura eu pour effet d'*effacer* la dépendance entre les trois lancers.

Faire l'expérience signifie ici recommencer les trois lancers un très grand nombre de fois, par exemple un million de fois, et mesurer les probabilités par la fréquence de chaque chiffre. Cette expérience peut fort bien n'être qu'une simulation numérique. La probabilité d'avoir "impair" étant $\frac{1}{2}$ avec le premier des trois lancers, on aura environ 500 000 fois "impair" et 500 000 fois "pair" pour le premier des trois résultats. Mais, bien que les trois lancers ne soient pas du tout indépendants, et que par exemple sur le million d'essais il arrivera environ $500\,000/\alpha$ fois que les trois valeurs θ_1 , θ_2 , et θ_3 soient toutes trois inférieures à $\frac{\pi}{2}$, il n'arrivera que 125 000 fois environ

que les trois résultats à la fois soient impairs (révélant ainsi une probabilité de $\frac{1}{8}$), c'est-à-dire que tout se passera comme si les trois résultats étaient stochastiquement indépendants. C'est ainsi que le chaos déterministe *efface* la causalité. Comment cela est-il possible ?

Dans cette question nous laissons volontairement inexpliqué le hasard qui choisit la valeur de θ_1 ; nous ne nous intéressons qu'au hasard relatif qui est fabriqué par le chaos pour les positions finales aux second et troisième lancer.

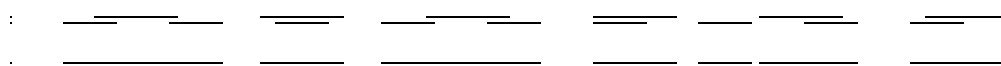
Si on réalisait effectivement cette expérience par simulation numérique, on *constaterait* qu'au premier lancer la bille s'arrête une fois sur deux sur un chiffre impair, et que une fois sur huit toutes les trois s'arrêtent sur un chiffre impair. On *constaterait*, mais on ne *comprendrait* pas pour autant. Une explication qualitative est dans de tels cas plus instructive. La clé du problème est bien sûr que de petits changements sur θ entraînent de grands changements sur le résultat final. Ainsi, si on prend $\theta_2 = \theta_1 + \varepsilon$, avec ε très petit, nous savons d'après ce qui a été vu en **I.4.** que le point où la bille aboutira au second lancer se situera à une distance de l'ordre de $NR\varepsilon$ (N étant le nombre de cordes parcourues et R le rayon du disque) du point atteint au premier lancer (en supposant que k n'a pas changé). Cette distance est en général nettement supérieure à la largeur d'un secteur de 10 degrés. Si $\theta_2 = \alpha \theta_1$, α n'étant pas spécialement proche de 1, la disparité des positions finales sera encore bien plus considérable. La probabilité pour que θ_1 et θ_2 soient tous deux dans le même demi-intervalle $]0, \frac{\pi}{2}[$ est $\frac{1}{2\alpha}$ et non $\frac{1}{4}$, parce que les deux valeurs θ_1 et θ_2 sont liées. Supposons maintenant que, au lieu d'avoir deux demi-intervalles de longueur 90 degrés, on ait découpé l'intervalle de 0 à 180 degrés en 10800 petits intervalles de longueur une minute d'angle, alternativement pairs et impairs. On peut représenter cela graphiquement en noircissant les intervalles impairs et en laissant en blanc les intervalles pairs :

. — — — — — —

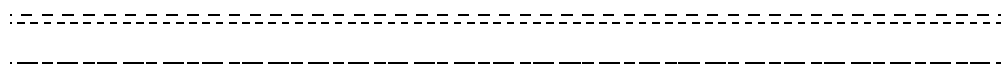
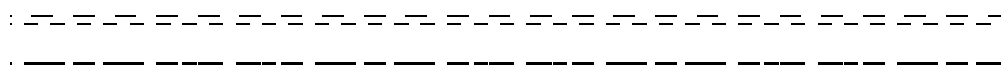
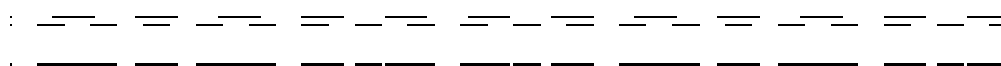
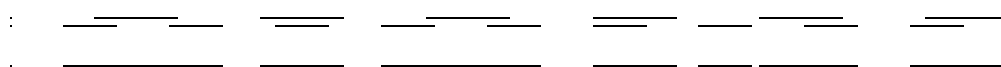
alors, pour que θ_1 et θ_2 soient tous deux dans un intervalle pair, il faut que non seulement θ_1 soit tombé par hasard dans un intervalle pair, mais que $\alpha \theta_1$ soit lui aussi dans un intervalle pair : cela revient donc à découper l'intervalle de 0 à 180 degrés en $10800 \times \alpha$ petits intervalles de longueur $1/\alpha$ minute d'angle, alternativement pairs et impairs, et à exiger que θ_1 soit dans un intervalle pair de cette nouvelle série, qu'on peut représenter graphiquement de la même façon :

. — — — — — — — — —

Ainsi la condition que θ_1 et θ_2 soient tous deux dans un intervalle pair de la première série, équivaut à dire que θ_1 soit à la fois dans un intervalle pair de la première série et dans un intervalle pair de la deuxième série; autrement dit, θ_1 doit être dans une zone qui est blanche à la fois sur la première et la deuxième série, ce qu'on peut visualiser en superposant les deux séries :



Sur la ligne du bas, les deux séries sont complètement superposées, alors que sur les deux lignes rapprochées du haut on les voit encore séparées. La condition pour que θ_1 et θ_2 soient tous deux dans un intervalle pair de la première série se traduit par le fait que θ_1 se trouve dans une région restée blanche après la superposition. Puisque θ_1 est choisi au hasard, la probabilité pour qu'il en soit ainsi est simplement le rapport entre la somme des longueurs des régions restées blanches et la longueur totale de 180 degrés. Ce rapport n'est pas facile à calculer car il varie avec la finesse de la subdivision; mais il tend vers une limite quand la subdivision devient de plus en plus fine (des intervalles d'une minute d'angle sont en pratique assez fins pour être assimilés au cas limite). Voici par exemple ce que donnent graphiquement des subdivisions de plus en plus fines :



On voit que les régions restées blanches sont de plus en plus petites, mais en même temps de plus en plus nombreuses, de sorte que la proportion de blanc sur la longueur totale (qui est la probabilité pour que θ_1 et θ_2 soient tous deux dans un intervalle pair) tend vers une limite.

On peut calculer numériquement cette limite (il n'y a pas de formule analytique simple pour l'exprimer, mais on peut faire un petit programme simple qui la calcule numériquement); le tableau suivant en donne les valeurs x pour différentes valeurs de α . À titre de comparaison on a donné entre

parenthèses la valeur de $1/2\alpha$, qui était la probabilité pour que θ_1 et θ_2 soient tous deux dans le demi-intervalle $]0, \frac{\pi}{2}[$.

α	x	$(\frac{1}{2\alpha})$	α	x	$(\frac{1}{2\alpha})$
3/2	0.25	(0.333)	4/3	0.25	(0.375)
5/3	0.2667	(0.300)	5/4	0.25	(0.400)
6/5	0.25	(0.417)	7/5	0.2571	(0.357)
7/6	0.25	(0.428)	8/7	0.25	(0.4375)
9/7	0.254	(0.3889)	9/8	0.25	(0.444)
10/7	0.25	(0.350)	10/9	0.25	(0.450)
11/8	0.25	(0.3636)	11/9	0.2525	(0.409)
11/10	0.25	(0.454)	12/11	0.25	(0.458)
13/10	0.25	(0.3846)	13/11	0.2517	(0.423)
13/12	0.25	(0.4615)	14/11	0.25	(0.3928)
14/13	0.25	(0.4643)	15/11	0.2515	(0.3667)
15/13	0.2513	(0.4333)	15/14	0.25	(0.4667)
16/13	0.25	(0.406)	16/15	0.25	(0.469)
17/13	0.2511	(0.382)	17/14	0.25	(0.412)
17/15	0.2510	(0.441)	17/16	0.25	(0.471)
18/13	0.25	(0.361)	18/17	0.25	(0.472)
19/15	0.2509	(0.395)	19/16	0.25	(0.421)
19/17	0.2508	(0.447)	19/18	0.25	(0.4737)
129/113	0.25	(0.438)	79/73	0.25	(0.462)

On voit que la valeur de x est toujours proche de $\frac{1}{4}$; le plus souvent elle est même exactement $\frac{1}{4}$ (sur le tableau les valeurs marquées 0.25 correspondent au cas où x est *exactement* égal à $\frac{1}{4}$; on aurait marqué 0.2500 si c'était une valeur approchée à 10^{-4} près). Le cas où x est différent de $\frac{1}{4}$ provient de contraintes liées à la divisibilité des nombres entiers qui interviennent ici (α est fractionnaire); on pourrait montrer mathématiquement que x vaut toujours $\frac{1}{4}$ quand α est irrationnel, mais cela est sans intérêt pour la question qui nous préoccupe ici; il suffit de voir que x est presque toujours égal à $\frac{1}{4}$, et, même quand il n'est pas exactement égal à $\frac{1}{4}$, en est toujours proche.

Cela montre que l'indépendance stochastique n'est pas liée à l'existence d'une indépendance effective entre θ_1 et θ_2 , mais seulement à la finesse de la subdivision: avec la même relation liant θ_1 et θ_2 , on constate que la propriété du produit n'est pas vérifiée lorsque l'intervalle de 0 à 180 degrés est divisé en deux moitiés, et qu'il suffit de subdiviser en un grand nombre d'intervalles pour qu'elle le devienne. Toutefois, comme le montre le tableau précédent, il y a des exceptions: pour certaines valeurs de α , la propriété du produit n'est pas exactement vérifiée, ce qui reflète qu'une certaine dépendance stochastique subsiste; cela signifie simplement que le *brouillage*

n'a pas été suffisant pour créer un hasard parfait. Le mécanisme exact de ce résidu de dépendance est une affaire d'arithmétique et de divisibilité due à la périodicité de la subdivision. Il suffirait de subdiviser en intervalles de longueur variable (mais comportant chacun une moitié paire et une moitié impaire) pour supprimer cette périodicité; et en effet, avec des subdivisions à pas variable, on supprimerait les contraintes de divisibilité et la valeur limite x serait alors toujours exactement $\frac{1}{4}$. Le brouillage serait alors devenu suffisant.

Or, pour le point d'arrivée de la bille, le brouillage est encore bien plus considérable: les 36 secteurs du disque ont une largeur de 10 degrés, ce qui est trop large pour que ce qui vient d'être mis en évidence se produise avec l'angle θ , mais largement suffisant pour que cela se produise avec le point d'arrivée de la bille, grâce à l'amplification caractéristique du chaos déterministe: si $\theta_2 = \alpha\theta_1$, le second point d'arrivée de la bille sera une fonction parfaitement déterminée du premier point d'arrivée, mais avec une variabilité telle que cette fois le brouillage sera parfait: toutes les contraintes de dépendance auront été *effacées* par le très grand nombre de rebondissements de la bille sur le bord du disque. La division du cercle en 36 secteurs de 10 degrés chacun équivaut à une subdivision infiniment fine à cause de l'amplification; la transformation du déterminisme en hasard est donc due à la finesse effective de la subdivision. On peut ainsi deviner que le brouillage par le chaos se ramène toujours d'une manière ou d'une autre à raffiner des subdivisions.

Ce n'est bien sûr pas par hasard que la roulette, même sous la forme d'un modèle très simplifié, illustre si bien tous les mécanismes du hasard. Toutes nos discussions montrent que la roulette est conçue de manière à accumuler la plus grande variété possible de *brouillages* dans un processus qui par nature est déterministe. On peut dire que la roulette est une véritable machine à effacer la causalité. Un dé est aussi un système destiné à effacer la causalité, mais moins sophistiqué; le roulement du dé, par exemple, a pour fonction de changer un grand nombre de fois la face dirigée vers le haut avant l'arrêt, ce qui joue le même rôle que les nombreuses réflexions de la bille de la roulette, à savoir créer l'amplification de la sensibilité aux conditions initiales. L'effacement des causes est d'autant plus radical que les subdivisions sont plus fines comparées à la variabilité des paramètres; cette finesse relative peut être renforcée soit en augmentant le nombre des subdivisions (c'est pourquoi la roulette comporte 36 chiffres plutôt que 5, 7, ou 10; dans le cas du dé cela reviendrait à faire des faces plus nombreuses et plus petites), soit en augmentant l'amplification de la sensibilité aux conditions initiales (c'est pourquoi la roulette agence des mouvements très chaotiques). Dans le cas d'une bille de roulette, la sensibilité aux conditions

initiales du mouvement est telle que pour pouvoir prédire le point d'arrivée à partir des lois de la Mécanique classique, il faudrait contrôler la position et la vitesse initiales de la bille ainsi que la forme exacte du plateau avec une précision bien plus grande que la dimension des atomes, c'est-à-dire avec une précision telle que la Mécanique classique ne pourrait même plus être tenue pour valide.

L'indépendance stochastique n'est donc pas l'expression exacte d'une indépendance causale réelle et absolue; elle résulte d'un *effacement* de la causalité par des procédés de brouillage adéquats (subdivisions fines et amplification). Ces procédés de brouillage peuvent être aussi bien artificiels (roulette, dé, fonctions **random**, publication d'une information sur le Net, etc.) que naturels (agitation thermique des molécules, combinaison des gènes, dispersion des semences par le vent).

IV.3. Probabilités conditionnelles.

Lorsque deux événements A et B ne sont pas stochastiquement indépendants, c'est-à-dire lorsque $\mathcal{P}(AB) \neq \mathcal{P}(A) \times \mathcal{P}(B)$, on peut introduire le rapport

$$\mathcal{P}(A | B) = \frac{\mathcal{P}(AB)}{\mathcal{P}(B)} \quad (IV.1.)$$

Si A et B étaient stochastiquement indépendants, on aurait bien sûr $\mathcal{P}(A | B) = \mathcal{P}(A)$. On dit que ce rapport est la **probabilité conditionnelle** de A par rapport à B ; dans le langage imagé, on dit aussi que c'est la probabilité conditionnelle de A "sachant que B s'est produit". Le concept mathématique ainsi introduit est évidemment censé refléter l'idée suggérée par le langage imagé, à savoir que la probabilité conditionnelle est la probabilité sur un ensemble restreint d'épreuves (celles qui vérifient une condition).

On peut aussi écrire la relation (IV.1.) sous la forme

$$\mathcal{P}(A | B) = \frac{\#(AB)}{\#B}$$

Si on la rapproche de (I.2.) cela signifie que cette fois on considère l'événement AB dans l'espace des épreuves B . En interprétant, cela revient à dire qu'on traite le problème suivant : parmi l'ensemble des épreuves équiprobables qui constituent l'événement B , quelle est la proportion de celles qui satisfont en outre à l'événement A ? Autrement dit, il s'agit de la probabilité pour que A se produise, mais sur l'espace restreint des épreuves qui correspondent à la réalisation de B . Cela exprime bien l'idée de la probabilité conditionnelle de A , *sachant que B s'est produit*. L'indépendance

stochastique entre A et une famille B_i peut donc se traduire en disant que la probabilité conditionnelle de A sachant que B_i s'est produit, est la même pour tous les B_i .

Du point de vue purement formel, une probabilité conditionnelle est donc simplement une probabilité sur un espace d'épreuves restreint. Dans n'importe quel problème de probabilités, on doit construire un espace d'épreuves Ω adapté à la résolution du problème, et nous voyons ici qu'on peut toujours considérer Ω comme une partie d'un espace plus gros, les probabilités définies sur Ω étant alors des probabilités conditionnelles sur le plus grand espace. Le meilleur choix de Ω est toujours le plus petit, mais il est peu commode de modifier Ω en cours de problème. Le meilleur Ω est donc le plus petit possible, parmi ceux qui contiennent tous les événements étudiés.

Pour illustrer cela, prenons une urne qui contient p bulletins OUI et q bulletins NON, donc $n = p + q$ bulletins en tout. On tire au hasard un bulletin ; il est intuitivement évident que la probabilité pour que ce bulletin porte OUI est p/n , et la probabilité pour qu'il porte NON est q/n ; un calcul rigoureusement dérivé des principes aboutirait, heureusement, au même résultat. Si on fait deux tirages successifs (sans remplacement), il est plus difficile de répondre immédiatement avec la seule perception intuitive, et le calcul rigoureusement dérivé des principes peut être préférable ; il passe par le dénombrement systématique : pour un seul tirage on pouvait prendre pour espace Ω l'ensemble des bulletins de l'urne, pour deux tirages il faut prendre l'ensemble des couples de bulletins (mais sans répétition). D'après II.2. son cardinal est $n(n - 1)$. Le nombre de couples formés de deux bulletins OUI est $p(p - 1)$, le nombre de couples formés de deux bulletins NON est $q(q - 1)$, et le nombre de couples formés d'un bulletin OUI et d'un bulletin NON est pq (quel que soit l'ordre). Donc la probabilité d'avoir deux bulletins OUI en deux tirages est $p(p - 1)/n(n - 1)$, celle d'avoir deux bulletins NON est $q(q - 1)/n(n - 1)$, et celle d'avoir un OUI puis un NON (ou l'inverse) est $pq/n(n - 1)$.

Il n'y a pas indépendance stochastique entre les tirages (il y en aurait si on effectuait des tirages *avec* remplacement). Cela peut se constater en introduisant les événements relatifs à chaque tirage, tout comme on avait introduit les événements relatifs à chaque boule dans l'exemple discuté en IV.1.

On posera donc :

A_1 : "le premier bulletin est un OUI"

B_1 : "le premier bulletin est un NON"

A_2 : "le deuxième bulletin est un OUI"

B_2 : “le deuxième bulletin est un NON”

L'événement “les deux bulletins sont des OUI sera alors A_1A_2 , l'événement “les deux bulletins sont des NON sera B_1B_2 , l'événement “le premier bulletin est un OUI et le deuxième est un NON” sera A_1B_2 , et l'événement “le premier bulletin est un NON et le deuxième est un OUI” sera A_2B_1 .

On a ainsi, comme cela a été dit plus haut :

$$\mathcal{P}(A_1) = \frac{p}{n} \quad \mathcal{P}(B_1) = \frac{q}{n}$$

$$\mathcal{P}(A_1A_2) = \frac{p(p-1)}{n(n-1)}$$

$$\mathcal{P}(A_1B_2) = \frac{pq}{n(n-1)}$$

$$\mathcal{P}(B_1A_2) = \frac{pq}{n(n-1)}$$

$$\mathcal{P}(B_1B_2) = \frac{q(q-1)}{n(n-1)}$$

Les probabilités $\mathcal{P}(A_2)$ et $\mathcal{P}(B_2)$ n'avaient pas encore été évoquées avant ; mais on les obtient sans difficulté : $A_2 = A_2A_1 \cup A_2B_1$ et A_1, B_1 sont disjoints, donc

$$\mathcal{P}(A_2) = \mathcal{P}(A_2A_1) + \mathcal{P}(A_2B_1) = \frac{p(p-1)}{n(n-1)} + \frac{pq}{n(n-1)} = \frac{p}{n}$$

$$\mathcal{P}(B_2) = \mathcal{P}(B_2A_1) + \mathcal{P}(B_2B_1) = \frac{pq}{n(n-1)} + \frac{q(q-1)}{n(n-1)} = \frac{q}{n}$$

On peut voir que la règle du produit n'est pas satisfaite :

$$\mathcal{P}(A_1) \times \mathcal{P}(A_2) = \frac{p}{n} \times \frac{p}{n} \neq \frac{p(p-1)}{n(n-1)}$$

$$\mathcal{P}(A_1) \times \mathcal{P}(B_2) = \frac{p}{n} \times \frac{q}{n} \neq \frac{pq}{n(n-1)}$$

$$\mathcal{P}(B_1) \times \mathcal{P}(A_2) = \frac{q}{n} \times \frac{p}{n} \neq \frac{pq}{n(n-1)}$$

$$\mathcal{P}(B_1) \times \mathcal{P}(B_2) = \frac{q}{n} \times \frac{q}{n} \neq \frac{q(q-1)}{n(n-1)}$$

Calculons alors les probabilités conditionnelles $\mathcal{P}(A_2 \mid A_1)$, $\mathcal{P}(B_2 \mid A_1)$,

$\mathcal{P}(A_2 \mid B_1)$, et $\mathcal{P}(B_2 \mid B_1)$. On trouve

$$\mathcal{P}(A_2 \mid A_1) = \frac{\mathcal{P}(A_2 A_1)}{\mathcal{P}(A_1)} = \frac{p(p-1)/n(n-1)}{p/n} = \frac{p-1}{n-1}$$

$$\mathcal{P}(A_2 \mid B_1) = \frac{\mathcal{P}(A_2 B_1)}{\mathcal{P}(B_1)} = \frac{pq/n(n-1)}{q/n} = \frac{p}{n-1}$$

$$\mathcal{P}(B_2 \mid A_1) = \frac{\mathcal{P}(B_2 A_1)}{\mathcal{P}(A_1)} = \frac{pq/n(n-1)}{q/n} = \frac{q}{n-1}$$

$$\mathcal{P}(B_2 \mid B_1) = \frac{\mathcal{P}(B_2 B_1)}{\mathcal{P}(B_1)} = \frac{q(q-1)/n(n-1)}{q/n} = \frac{q-1}{n-1}$$

Cette suite de calculs était fastidieuse, quoique facile, mais maintenant nous pouvons *interpréter*.

Par exemple $\mathcal{P}(A_2 \mid A_1)$, la probabilité conditionnelle pour que A_2 se produise sachant que A_1 s'est produit, est dans le langage imagé "la probabilité pour que le second bulletin tiré soit un OUI, sachant que le premier était un OUI". Si on raisonne selon l'intuition première, on dira que si le premier bulletin tiré a été un OUI il reste dans l'urne $p-1$ OUI et q NON; donc la probabilité de tirer un bulletin OUI la deuxième fois est toujours le rapport du nombre de bulletins OUI au nombre total de bulletins, soit $(p-1)/(n-1)$. De même, si le premier bulletin tiré a été un NON, il reste dans l'urne $n-1$ bulletins, dont p OUI et $q-1$ NON, de sorte que la probabilité de tirer un OUI est cette fois $p/(n-1)$.

L'intuition première que tout le monde peut avoir a priori ne permet pas de voir immédiatement et de manière évidente quelle doit être la valeur de probabilités telles que $\mathcal{P}(A_1 A_2)$, $\mathcal{P}(A_1 B_2)$, $\mathcal{P}(B_1 A_2)$, et $\mathcal{P}(B_1 B_2)$; pour les trouver, il a fallu dénombrer les couples de bulletins en utilisant II.2. Par contre il était immédiatement évident que les probabilités $\mathcal{P}(A_1)$ et $\mathcal{P}(B_1)$ valent respectivement p/n et q/n . Il y a dans tous les domaines des choses qu'on peut percevoir par intuition immédiate et d'autres qu'on ne peut que trouver par un raisonnement complexe, c'est-à-dire composé de plusieurs déductions dont une seule à la fois peut être reconnue comme immédiatement évidente, et il en est également ainsi en Calcul des probabilités. Or il est aussi immédiatement évident que la probabilité pour que le second bulletin tiré soit un OUI *sachant que le premier était un OUI* est $(p-1)/(n-1)$, puisque cela résulte directement, tout comme le calcul de $\mathcal{P}(A_1)$, du fait que le nombre de bulletins OUI présents dans l'urne est $p-1$. Pour dénombrer tous les couples possibles de bulletins, il faut déjà faire appel à des notions mathématiques telles que le produit cartésien de deux ensembles, qui sont abstraites et ne relèvent pas de l'évidence immédiate.

Il faut en effet *déduire* le nombre de couples possibles à partir du nombre donné de bulletins, alors que relier la probabilité et le quotient de p par n (ou de $p - 1$ par $n - 1$) relève de l'intuition première.

Ci-dessus, nous avons présenté le calcul de $\mathcal{P}(A_2 | A_1)$ de telle façon qu'il se déduise, par la définition mathématique (IV.1.), du calcul préalable, et non évident, de $\mathcal{P}(A_2 A_1)$. En procédant ainsi, nous retrouvons un résultat qui aurait pu être obtenu directement par l'intuition première. Cette façon de procéder est didactiquement intéressante, car elle montre que la théorie marche. On fait souvent cela dans l'enseignement : retrouver par une voie indirecte et compliquée un résultat qui était évident a priori. C'est très mauvais pour calculer, mais très bon pour comprendre. En effet, lorsqu'on expose une théorie élaborée, sous forme axiomatique, les étudiants ont le sentiment d'un savoir qui tombe du ciel. Mais les mathématiciens du passé qui ont créé de toutes pièces ces concepts mathématiques (espaces d'épreuves, indépendance stochastique, probabilités conditionnelles, etc.) sont partis de l'intuition première et ont construit peu à peu la théorie par tâtonnements. Le critère essentiel de validité qui guidait les créateurs était que la bonne théorie serait celle qui, dans tous les cas où l'intuition première permet de calculer un résultat, concorderait avec cette intuition première. En Physique, il faut que la théorie soit confirmée par l'expérience : chaque fois que la théorie permet de calculer un résultat et l'expérience de le mesurer, les deux résultats doivent concorder. Il en va de même pour le Calcul des probabilités, sauf que ce qui joue le rôle de l'expérience est l'intuition première. En effet, le Calcul des probabilités n'est pas une science expérimentale ; on ne *mesure* jamais les probabilités : c'est, comme pour la géométrie, à partir de symétries fondamentales qu'on postule l'équiprobabilité d'un ensemble d'épreuves (voir chapitre I). Si dans beaucoup d'applications du Calcul des probabilités on recoupe les faits expérimentaux, c'est parce que les symétries fondamentales (de l'espace-temps ou des particules quantiques, par exemple) sont d'origine expérimentale. Mais à l'intérieur du Calcul des probabilités ces symétries fondamentales sont postulées et on en déduit des *probabilités a priori*. C'est pourquoi le rôle joué par l'intuition première est absolument essentiel et constitue le véritable fondement du Calcul des probabilités. Si la définition (IV.1.) permet de retrouver, par l'application rigoureuse et abstraite des principes de base, ce que l'intuition tient pour évident, en l'occurrence que $\mathcal{P}(A_2 | A_1) = (p - 1) / (n - 1)$, c'est un signe que la théorie est bonne. Mais s'il en est ainsi, c'est bien entendu parce que tout avait été fabriqué pour qu'il en soit ainsi : le choix du modèle Ω et l'équiprobabilité des épreuves que ce choix sous-entend contient *en germe* le résultat. La théorie a été faite exprès pour qu'il en soit ainsi.

Cela dit, si retrouver par la théorie un résultat qui était évident a priori

est conceptuellement instructif, on peut aussi utiliser le chemin inverse dans un but strictement pragmatique. Supposons que nous veuillions calculer $\mathcal{P}(A_2 A_1)$ (dont la valeur ne se voit pas immédiatement à partir de l'intuition première). Au lieu de procéder par dénombrement comme nous l'avons fait, on pourrait aussi suivre le chemin inverse: il est *évident* que $\mathcal{P}(A_1) = p/n$; il est *évident aussi* que $\mathcal{P}(A_2 | A_1) = (p-1)/(n-1)$. En inversant la définition (IV.1.) on obtient

$$\mathcal{P}(A_2 A_1) = \mathcal{P}(A_2 | A_1) \times \mathcal{P}(A_1) \quad (IV.2.)$$

et on en déduit alors que $\mathcal{P}(A_2 A_1) = \frac{p}{n} \cdot \frac{p-1}{n-1}$, c'est-à-dire la même chose que ce qu'on avait obtenu par dénombrement direct.

Cet exemple montre l'intérêt pratique des probabilités conditionnelles. En pratique (IV.2.) est plus intéressant et plus fréquemment utilisé que (IV.1.), car il arrive souvent, du moins dans les problèmes de niveau élémentaire, que les probabilités conditionnelles soient plus "évidentes" ou en tous cas plus faciles à calculer que celles des intersections, et les probabilités des intersections d'événements s'obtiennent alors par produits successifs.

On remarquera aussi comment nous avons calculé $\mathcal{P}(A_2)$ et $\mathcal{P}(B_2)$. Contrairement aux probabilités $\mathcal{P}(A_1)$ et $\mathcal{P}(B_1)$, il n'y a aucune évidence immédiate, car on ignore ce qui s'est passé au premier tirage et par conséquent on ne connaît pas le nombre de bulletins OUI présents dans l'urne au moment où on effectue le second tirage. Lorsqu'on avait à évaluer $\mathcal{P}(A_2 | A_1)$ ou $\mathcal{P}(A_2 | B_1)$ on pouvait connaître le nombre de bulletins OUI présents dans l'urne car on se plaçait dans l'hypothèse où le premier tirage était donné; mais si on veut connaître la probabilité d'avoir OUI au second tirage sans savoir ce qui s'est produit au premier, il n'y a plus d'argument immédiatement évident. L'intuition première peut donner le sentiment vague que la probabilité d'avoir OUI au deuxième tirage doit être une sorte de moyenne sur les deux possibilités qui peuvent se produire au premier tirage. Mais pour donner un contenu quantitatif clair et solide à ce sentiment vague, il faut un raisonnement mathématique plus sophistiqué, comme celui que nous avons effectivement suivi: ce raisonnement consiste à décomposer l'événement A_2 en deux parties disjointes $A_2 A_1$ et $A_2 B_1$ dont on peut calculer séparément les probabilités; alors $\#A_2 = \#(A_2 A_1) + \#(A_2 B_1)$, ou $\mathcal{P}(A_2) = \mathcal{P}(A_2 A_1) + \mathcal{P}(A_2 B_1)$. Mais si on utilise (IV.2.) on peut écrire cela sous la forme

$$\mathcal{P}(A_2) = \mathcal{P}(A_2 | A_1) \cdot \mathcal{P}(A_1) + \mathcal{P}(A_2 | B_1) \cdot \mathcal{P}(B_1) \quad (IV.3.)$$

formule qui exprime effectivement la probabilité de A_2 comme une moyenne

des probabilités conditionnelles sur les deux possibilités qui peuvent se produire au premier tirage; les coefficients qui interviennent dans ce calcul de moyenne sont les probabilités respectives de chacune de ces deux possibilités.

L'intérêt de cette formule est qu'elle permet de calculer une probabilité non évidente a priori, par composition de probabilités évidentes. Mais elle n'est rien d'autre qu'une manière un peu sophistiquée de répéter que la probabilité de la réunion de deux événements disjoints est la somme des probabilités de chacun.

Cette formule (IV.3.) peut d'ailleurs être généralisée comme suit. On découpe Ω en r morceaux disjoints $E_1, E_2, E_3, \dots, E_r$ (une telle famille d'événements disjoints et dont la réunion est Ω tout entier est appelée une *famille exhaustive d'événements*). Soit A un événement quelconque. Alors on peut dire que les événements $AE_1, AE_2, AE_3, \dots, AE_r$ sont également disjoints et que leur réunion est A . On en déduit que $\mathcal{P}(A) = \mathcal{P}(AE_1) + \mathcal{P}(AE_2) + \mathcal{P}(AE_3) + \dots + \mathcal{P}(AE_r)$, ou encore

$$\begin{aligned} \mathcal{P}(A) = \mathcal{P}(A | E_1) \cdot \mathcal{P}(E_1) + \mathcal{P}(A | E_2) \cdot \mathcal{P}(E_2) + \\ + \dots + \mathcal{P}(A | E_r) \cdot \mathcal{P}(E_r) \end{aligned} \quad (IV.4.)$$

Ceci est vrai quel que soit l'événement A et quelle que soit la famille exhaustive $E_1, E_2, E_3, \dots, E_r$, mais pour en tirer un résultat intéressant dans un problème donné il faut bien sûr choisir astucieusement les E_j .

Nous verrons une application importante de cette formule au chapitre **VIII** (Processus en cascade) : voir problème N°2. Nous en discuterons une autre application (à la génétique) dans la section suivante (IV.4). On peut pratiquer le Calcul des probabilités sans jamais y recourir, en décomposant l'événement A en la réunion disjointe des AE_j . Mais la discussion ci-dessus montre que les probabilités conditionnelles sont souvent plus commodes ou plus évidentes que celles des intersections: c'est ce qui rend cette formule très pratique dans beaucoup de cas.

Plus haut nous avons justifié l'introduction des probabilités conditionnelles par un argument de commodité: dans l'exemple des bulletins qu'on sort de l'urne, la valeur de $\mathcal{P}(A_2)$ n'était pas "évidente a priori", tandis que $\mathcal{P}(A_1)$ et $\mathcal{P}(A_2 | A_1)$ l'étaient.

Pour *vraiment* comprendre les probabilités conditionnelles il est utile de revenir sur les raisons de cette évidence. Dès le début de cet ouvrage, nous avons appris que pour calculer une probabilité il fallait opérer de la manière suivante:

- a) trouver l'ensemble Ω de toutes les épreuves équiprobables possibles;
- b) traduire l'événement A dont on cherche la probabilité en terme de sous-ensemble de Ω .

c) dénombrer Ω et A .

Les probabilités évidentes sont celles où le quotient $\#A / \#\Omega$ est tout de suite évident, sans qu'on ait besoin de réfléchir pour effectuer a), b), c). Il en est ainsi pour la probabilité de tirer un bulletin OUI dans une urne qui en contient p , sur n bulletins en tout, car on voit tout de suite que $\#\Omega = n$ et $\#A = p$. Les cardinaux sont évidents avant même d'avoir réfléchi à la nature des éléments. Lorsque les épreuves sont des mots de n lettres, des permutations d'un ensemble, ou des choses encore plus compliquées, il faut d'abord un raisonnement abstrait (les étapes a et b ci-dessus) pour comprendre que Ω est bien l'ensemble des mots de n lettres. Ensuite il faut appliquer une des formules de dénombrement du chapitre **II**. Ces formules ne peuvent pas être devinées en vitesse, elles font partie des connaissances accumulées par les mathématiciens au cours des siècles. En outre elles ne suffisent pas à elles seules et leur application ne constitue que l'étape c): auparavant il faut avoir fait le travail de modélisation, qui constitue les étapes a) et b). Même dans le cas très simple de A_1A_2 (tirer deux fois de suite un OUI), il faut réfléchir un peu, et abstraitement, pour comprendre que Ω est l'ensemble de tous les tirages de deux bulletins distincts, et qu'il relève du deuxième cas de dénombrement (**II.2.** tirages sans remise) d'où $\#\Omega = n(n-1)$; il faut encore recourir une seconde fois au même type de raisonnement abstrait pour $\#(A_1A_2) = p(p-1)$. C'est pourquoi une personne n'ayant jamais étudié le Calcul des probabilités et à peine fait de mathématiques dans sa vie pourra deviner la valeur de $\mathcal{P}(A_1)$, mais pas celle de $\mathcal{P}(A_1A_2)$; et encore moins celle de $\mathcal{P}(A_2)$, car pour calculer cette dernière il faut en outre la décomposition ensembliste $A_2 = A_1A_2 \cup B_1A_2$.

Or la même personne n'ayant jamais étudié le calcul des probabilités comprendra aisément que si le premier bulletin tiré était un OUI, la probabilité de tirer un OUI une seconde fois est $(p-1)/(n-1)$. Les étapes a) et b) sont ici aussi inutiles, car il saute aux yeux que pour ce problème l'ensemble des épreuves est l'ensemble des bulletins restants, et l'événement l'ensemble des bulletins OUI restants. Si on considère ce sous-problème: "on a déjà tiré un bulletin OUI au premier tirage, quelle est la probabilité d'en tirer un au second tirage" et qu'on veut malgré tout passer par a) et b), on prendra pour Ω l'ensemble des $n-1$ bulletins restant après le premier tirage, et pour A l'ensemble des $p-1$ bulletins OUI restant après le premier tirage: c'est en tous cas la démarche implicite dans le raisonnement naïf. Toutefois, si on veut coûte que coûte procéder selon les règles formelles a), b), c), il vaut mieux prendre en compte pour ce premier tirage tous les cas possibles, ce qui revient à prendre pour Ω l'ensemble de tous les tirages de deux bulletins dont le premier est OUI, et pour A l'ensemble de tous les tirages de deux bulletins qui sont tous deux des OUI; autrement dit, on prend pour espace

des épreuves A_1 et pour événement $A = A_1A_2$. Comparés (respectivement) à l'espace de tous les tirages de deux bulletins, et à l'événement "le second est un OUI", ces deux ensembles s'obtiennent en ne retenant que les épreuves qui satisfont la condition "le premier tirage est OUI".

Dans ce problème "réduit", où l'espace des épreuves est A_1 et l'événement A_1A_2 , la probabilité s'obtient toujours selon la formule *cardinal de l'événement sur cardinal de l'espace*, mais cela devient alors exactement ce que dit IV.1.

Le but de ces remarques est le suivant: il n'y a absolument aucune différence de nature entre une probabilité conditionnelle et une probabilité tout court. On ne parle de probabilité conditionnelle que si on se situe dans un problème donné, et qu'on y distingue un sous-problème. N'importe quel problème de probabilité peut toujours être considéré comme un sous-problème d'un problème plus vaste, de sorte qu'on puisse le traiter en termes de probabilités conditionnelles. Par exemple, dans le problème des bulletins on peut toujours supposer qu'avant le premier tirage avait eu lieu un tirage N^o . La probabilité p/n de tirer un bulletin OUI au premier tirage est identique à la probabilité conditionnelle de tirer ce bulletin OUI sachant qu'au tirage N^o on avait tiré un NON, et qu'il y avait alors $n + 1$ bulletins en tout dont p OUI, ou à la probabilité conditionnelle de tirer ce bulletin OUI sachant qu'au tirage N^o on avait tiré un OUI, et qu'il y avait alors $n + 1$ bulletins en tout dont $p + 1$ OUI.

IV. 4. Relativité du hasard.

Ce constat concernant la nature essentiellement relative des probabilités conditionnelles a des conséquences pratiques. Jusqu'ici, nous avons toujours rencontré des problèmes où la première chose à faire était de chercher les invariances afin de déterminer a priori ce qui est équiprobable. (en langage imagé "le niveau où intervient le hasard pur"). Dans le problème longuement discuté ci-dessus, où on tirait un bulletin au hasard dans une urne, il était aisé de postuler que tous les bulletins sont équiprobables, parce que nous savions que nous tirions des bulletins individuels. Transformons un peu l'expérience: imaginons qu'à l'intérieur de l'urne se trouvent, non des bulletins, mais des dés: une partie des dés est blanche, les autres sont rouges; disons p dés blancs et q dés rouges, le nombre total de dés étant $n = p + q$. Les dés rouges ont une face gagnante: si un dé rouge est tiré au sort, et qu'en le faisant rouler il marque "six", je gagne. Mais les dés blancs n'ont aucune face gagnante. Il est aisé de déterminer l'espace des épreuves Ω correspondant, c'est l'ensemble de toutes les faces de dés possibles, dont le nombre est $6n$: $\#\Omega = 6n$. L'événement A "je gagne" est l'ensemble des

faces gagnantes, c'est-à-dire l'ensemble des faces “six” de dés rouges, soit q faces en tout. La probabilité de gagner est donc $\#A / \#\Omega = q/6n$.

Nous avons ainsi raisonné une fois de plus par équiprobabilité a priori, comme nous le faisons depuis le début. Ici, le “niveau où intervenait le hasard pur” était le choix d'une face de dé parmi $6n$ faces équiprobables.

Mais supposons que ce niveau nous reste caché; par exemple, supposons que nous ignorions ce qui se trouve à l'intérieur de l'urne, que le tirage d'un dé se fasse à l'intérieur de l'urne et à l'abri des regards, soit par un mécanisme caché, soit par un génie qui habite l'urne. Chaque fois que nous frottons l'urne, le génie tire au hasard un dé: si ce dé est rouge, le génie le jette hors de l'urne puis nous (qui sommes à l'extérieur) le lançons; si le dé est blanc, le génie ne se manifeste pas et rien ne sort de l'urne.

Dans cette seconde version de l'expérience il nous serait difficile de deviner comment intervient le hasard pur. En effet, rien ne nous dit que la décision du génie de lancer un dé hors de l'urne, provient d'un tirage entre des dés équiprobables; nous ignorons qu'il y a aussi des dés blancs; le constat que les dés qui sortent sont toujours rouges ne permet aucune déduction. Certes, nous voyons bien comment intervient le hasard dans la partie visible du processus: lorsque nous lançons le dé que le génie nous a jeté, nous sommes assurés que ses six faces sont équiprobables; mais la décision du génie de lancer un dé ou non, peut être attribuée à n'importe quelle forme de hasard. Et nous n'avons aucun moyen de savoir a priori quelle va être la probabilité pour qu'il lance un dé hors de l'urne; nous ne pouvons que la mesurer statistiquement: en frottant un très grand nombre de fois l'urne, on peut par exemple constater empiriquement que dans 30% des cas, un dé est jeté hors de l'urne.

Le point essentiel est alors le suivant: malgré notre ignorance de la nature du phénomène, nous pouvons calculer la probabilité de gagner en utilisant la formule IV.4, à condition d'y introduire les probabilités empiriques mesurées. Soit en effet S l'événement “un dé sort de l'urne”, R l'événement “rien ne sort”, et G l'événement “je gagne”. En utilisant IV.4, on peut dire que

$$\mathcal{P}(G) = \mathcal{P}(G | S) \cdot \mathcal{P}(S) + \mathcal{P}(G | R) \cdot \mathcal{P}(R) = \frac{1}{6} \cdot 0.3 + 0 \cdot 0.7 = 0.05$$

Le raisonnement a priori aurait donné $q/6n$, donc (pour $q/n = 0.3$), le même résultat. Du moment que les probabilités $\mathcal{P}(S)$ et $\mathcal{P}(R)$ sont connues (peu importe que ce soit a priori ou empiriquement), la probabilité $\mathcal{P}(G)$ est déterminée *indépendamment de la manière dont le hasard est intervenu dans l'urne*. Si par exemple le choix aléatoire effectué à notre insu par le

génie dans l'urne était d'une grande complexité, de sorte que le "véritable" espace Ω aurait un cardinal énorme, que le quotient p/n serait une fraction compliquée (c'est-à-dire que p et n seraient tous deux de très grands entiers), mais proche de 0.3, le calcul ci-dessus, quoique approché, serait tout aussi correct. Mathématiquement, la formule IV.4 se *démontre* à partir de l'existence d'un espace des épreuves. Mais, une fois fixées les probabilités $\mathcal{P}(E_j)$, la probabilité $\mathcal{P}(A)$ sera déterminée indépendamment de la nature de l'espace Ω . On peut même oublier qu'il existe quelque part un niveau où intervient le pur hasard. Si toutefois ce niveau n'*existe réellement pas*, alors la formule IV.4 peut perdre toute validité. C'est ce qui se produit dans la Mécanique quantique, lorsque des amplitudes interfèrent (voir une discussion sur ce point à la section **XIII.3**).

Ainsi la formule IV.4 est utilisable dans les cas où les invariances ne sont pas intégralement connues, de sorte qu'une partie seulement des probabilités peuvent être calculées a priori et que les autres, d'origine empirique, ne peuvent être déterminées qu'a posteriori par des mesures statistiques.

Autrement dit, la formule IV.4 permet de remplacer une partie des probabilités a priori par des probabilités empiriques. Cet expédient est inévitable lorsqu'on ne peut pas accéder au niveau où intervient une invariance (cf. la discussion de la section **I.2**).

Nous allons immédiatement appliquer cela à un problème de génétique, celui des mariages consanguins. Imaginons que dans une population on isole un couple de parents, de sorte que les enfants de ces parents se reproduisent ensuite entre eux. Nous nous intéressons à la manière dont un gène particulier, G , se combinera à ses compagnons allèles du même locus (voir **III.5** et **III.6**). Nous supposons d'abord que la combinaison homozygote GG n'est pas invalidante et que les individus GG accèdent autant que les autres à la procréation.

Plusieurs cas de figure peuvent se présenter. Chacun des deux parents peut appartenir à l'une des quatre catégories XX , XG , GX , GG . La distinction entre XG et GX ne correspond à aucune différence biologique entre les individus porteurs, mais représente le choix qui s'opère entre les deux chromosomes au moment de la fécondation : ainsi, dire que le père est XG signifie que c'est le chromosome porteur de X qui sera retenu pour constituer une nouvelle paire : la première des deux lettres correspond au chromosome retenu. Ainsi il y aura seize catégories pour le couple. Si les parents sont tous deux GX , les enfants seront tous GG , et par conséquent si ces enfants se croisent ensuite entre eux les petits enfants le seront aussi ; si les parents sont XG et GX (ou inversement), aucun enfant ne sera GG mais un enfant sur deux sera GX (les autres seront alors XG). Il y aura alors

une chance sur quatre que les petits-enfants soient GG . Si les parents sont XX et GX (ou XX et XG), les enfants seront aussi pour moitié XG et pour l'autre moitié XG , et donc les petits-enfants auront aussi une chance sur quatre d'être GG . En regroupant systématiquement les différents cas qui se présentent on peut distinguer les trois événements suivants, selon la catégorie des parents :

$$\begin{aligned} E_1: & \quad GG + GG, GG + GX, GX + GG, \text{ ou } GX + GX \\ E_2: & \quad GG + XX, GG + XG, XG + GG, XX + GG, \\ & \quad GX + XX, GX + XG, XG + GX, \text{ ou } XX + GX \\ E_3: & \quad XG + XG, XG + XX, XX + XG, \text{ ou } XX + XX \end{aligned}$$

Dans chacun de ces trois événements on aura une probabilité spécifique pour que les petits-enfants soient GG :

parents	enfants	petits-enfants
E_1	100% GG	100% GG
E_2	50% GX et 50% XG	25% GG
E_3	100% XX	0% GG

Si on appelle E l'événement "un petit-enfant est GG ", on peut interpréter le nombre qui figure dans la troisième colonne de ce tableau comme la probabilité conditionnelle de E sachant que E_1 , E_2 , ou E_3 s'est produit. Ainsi

$$\mathcal{P}(E \mid E_1) = 1 \quad ; \quad \mathcal{P}(E \mid E_2) = 0.25 \quad ; \quad \mathcal{P}(E \mid E_3) = 0 \quad ;$$

Pour connaître la probabilité de E lorsque les parents sont pris au hasard dans la population (donc on ignore s'ils sont dans E_1 , dans E_2 , ou dans E_3), on utilise IV.4, ce qui donne :

$$\mathcal{P}(E) = \mathcal{P}(E \mid E_1) \cdot \mathcal{P}(E_1) + \mathcal{P}(E \mid E_2) \cdot \mathcal{P}(E_2) + \mathcal{P}(E \mid E_3) \cdot \mathcal{P}(E_3)$$

Tout comme pour l'expérience avec le génie caché dans l'urne, on ne peut pas, dans ce problème, trouver $\mathcal{P}(E_1)$, $\mathcal{P}(E_2)$ et $\mathcal{P}(E_3)$ en cherchant des épreuves équiprobables a priori car les mécanismes d'apparition des gènes G et X nous sont inconnus et certainement si complexes que nous serions bien incapables d'y trouver des invariances. Conformément aux remarques faites au début de cette section, on peut cependant ignorer ces invariances, il n'est même pas nécessaire qu'il existe un moyen de les mettre en évidence. Il suffit de remplacer les coefficients $\mathcal{P}(E_j)$ par leurs valeurs observées. Ainsi, l'espace Ω reste ici inconnu, mais nous pouvons nous en passer en utilisant des données empiriques. Celles-ci nous fournissent la valeur du nombre ε (la proportion d'individus GG dans la population, d'où nous déduisons par la loi de Hardy-Weinberg que si on choisit au hasard un individu dans la

population, la probabilité pour qu'il soit GG sera ε , la probabilité pour qu'il soit hétérozygote GX/XG sera $2(\sqrt{\varepsilon} - \varepsilon)$, et la probabilité pour qu'il soit XX sera $\eta = 1 - 2\sqrt{\varepsilon} + \varepsilon = (1 - \sqrt{\varepsilon})^2$. Si les couples de parents sont pris au hasard, la probabilité pour que par exemple les deux parents soient GG sera ε^2 . L'événement E_1 du tableau ci-dessus aura donc la probabilité

$$\begin{aligned} \mathcal{P}(GG + GG) + \mathcal{P}(GG + GX) + \mathcal{P}(GX + GG) + \mathcal{P}(GX + GX) \\ = \varepsilon^2 + \varepsilon \cdot \frac{x}{2} + \frac{x}{2} \cdot \varepsilon + \frac{x}{2} \cdot \frac{x}{2} \\ = \varepsilon^2 + x\varepsilon + \frac{x^2}{4} = \varepsilon \end{aligned}$$

L'événement E_2 aura la probabilité

$$\begin{aligned} \mathcal{P}(GG + XG) + \mathcal{P}(GG + XX) + \mathcal{P}(GX + XG) + \mathcal{P}(GX + XX) + \\ + \mathcal{P}(XG + GG) + \mathcal{P}(XG + GX) + \mathcal{P}(XX + GG) + \mathcal{P}(XX + GX) \\ = \varepsilon \cdot \frac{x}{2} + \varepsilon \cdot \eta + \frac{x}{2} \cdot \frac{x}{2} + \frac{x}{2} \cdot \eta + \frac{x}{2} \cdot \varepsilon + \frac{x}{2} \cdot \frac{x}{2} + \eta \cdot \varepsilon + \eta \cdot \frac{x}{2} \\ = 2(\sqrt{\varepsilon} - \varepsilon) = x \end{aligned}$$

Enfin, l'événement E_3 aura la probabilité

$$\begin{aligned} \mathcal{P}(XG + XG) + \mathcal{P}(XG + XX) + \mathcal{P}(XX + XG) + \mathcal{P}(XX + XX) \\ = \frac{x}{2} \cdot \frac{x}{2} + \frac{x}{2} \cdot \eta + \eta \cdot \frac{x}{2} + \eta^2 \\ = (1 - \sqrt{\varepsilon})^2 = \eta \end{aligned}$$

Ces trois résultats auraient pu être obtenus sans calcul: l'événement E_1 regroupe en effet les couples de parents qui procréeront des GG , l'événement E_2 les couples de parents qui procréeront des hétérozygotes, et l'événement E_3 les couples de parents qui procréeront des XX ; il est donc logique (puisque l'on suppose que tous les couples ont la même fécondité) que les poids relatifs de ces événements soient égaux à celui de la progéniture correspondante.

Appliquons alors la règle IV.4 des probabilités conditionnelles: $\mathcal{P}(E) = \varepsilon + 0.25x + 0\eta = \frac{1}{2}(\sqrt{\varepsilon} + \varepsilon)$.

On peut donc énoncer le résultat suivant: dans une population où la proportion d'individus GG est ε , un couple de parents pris au hasard a (pour chaque enfantement) la probabilité ε de mettre au monde un individu GG . Mais un couple formé de frère et soeur a la probabilité $\frac{1}{2}(\sqrt{\varepsilon} + \varepsilon)$ de mettre au monde un individu GG . Si par exemple il y a un GG sur 10000 dans la population, la procréation entre frère et soeur a une chance sur deux cent de produire un GG .

La loi des accouplements consanguins peut être généralisée mais nous ne donnons ici que les résultats⁽¹⁾. Par le même procédé que celui utilisé plus haut (application de la règle *IV.4* des probabilités conditionnelles) on peut analyser l'effet de croisements moins fortement consanguins; dans le cas de croisements entre frères et soeurs, nous avons divisé les choix possibles de parents en catégories; pour des cousins germains, qui n'ont pas les mêmes parents, mais les mêmes grands-parents, il faudra remonter à ces derniers. Les lois obtenues sont simples: pour les exprimer on introduit un nombre appelé *facteur de consanguinité*. Ce nombre vaut $\frac{1}{2}$ pour des frères et soeurs; il est égal à $\frac{1}{8}$ pour un couple formé d'un demi-frère et sa demi-soeur ou de doubles cousins germains, à $\frac{1}{16}$ pour des cousins germains, etc. Si f est le facteur de consanguinité d'un couple, la probabilité pour qu'il engendre un enfant GG est $f\sqrt{\varepsilon} + (1 - f)\varepsilon$ (dans l'hypothèse où les individus GG ont autant de chances de procréer que les autres). Pour les maladies génétiques très rares, ε est négligeable devant $\sqrt{\varepsilon}$; la probabilité de transmettre la maladie est donc beaucoup plus élevée en cas de mariage consanguin, car le facteur f est alors beaucoup plus grand que $\sqrt{\varepsilon}$. Pour la mucoviscidose, qui n'est pas rarissime, $\varepsilon = 1/2500$ et $\sqrt{\varepsilon} = 1/50$, donc un mariage entre cousins germains donnera une probabilité $1/615$, quatre fois plus grande que pour une procréation exogamique.

Il ne faut pas oublier qu'il s'agit là des probabilités correspondant à un choix au hasard de cousins germains parmi la population. Il est bien clair que si les cousins savent que l'un d'entre eux est homozygote XX , la probabilité d'engendrer un GG sera nulle, et s'ils savent être tous deux hétérozygotes, la probabilité sera $\frac{1}{4}$.

⁽¹⁾ On trouvera un exposé assez complet (mais rapide) des lois probabilistes de l'hérédité dans le petit ouvrage de Gustave Malécot *Les mathématiques de l'hérédité* (Masson, Paris, 1948). Le lecteur qui ne craint pas d'effectuer tous les calculs lui-même pourra trouver le thème de la consanguinité traité en exercices dans l'ouvrage de William Feller (chapitre V, problèmes N^{os} 33 à 40, pages 144 – 145).