8

Probabilités discrètes

Le hasard est un élément qui intervient dans beaucoup de nos tentatives d'appréhender le monde dans lequel nous vivons. La *théorie des probabilités* permet de calculer la probabilité de réalisation d'événements complexes en supposant qu'ils obéissent à certains axiomes. Cette théorie, qui trouve des applications dans toutes les branches scientifiques, est fortement liée aux notions que nous avons vues dans les chapitres qui précèdent.

Les probabilités sont dites "discrètes" si on peut calculer la probabilité de tout événement en effectuant une sommation (et non une intégration). Comme nous commençons à être plutôt à l'aise avec les sommes, rien d'étonnant à ce que nous soyons tout à fait prêts à appliquer nos connaissances sur des calculs de probabilités et de moyennes.

8.1 DÉFINITIONS

(Les lecteurs qui ne sont pas familiers avec les probabilités pourront lire avec profit l'ouvrage bien connu de Feller, qui présente une très bonne introduction au sujet [120]).

La base des probabilités est la notion d'*espace de probabilité*. Un espace de probabilité est constitué d'un ensemble Ω de toutes les choses qui peuvent arriver dans un problème donné et d'une règle ou d'un ensemble de règles qui associent une probabilité $\Pr(\omega)$ à tout événement élémentaire $\omega \in \Omega$. La probabilité $\Pr(\omega)$ est toujours un nombre réel positif ou nul, et tout espace de probabilités discret doit satisfaire la condition

$$\sum_{\omega \in \Omega} \Pr(\omega) = 1 . \tag{8.1}$$

Ainsi, toute valeur $\Pr(\omega)$ est comprise dans l'intervalle $[0\,..\,1]$. On dit que Pr est une *loi de probabilité*, ou une *distribution de probabilité*, car son rôle est de partager une probabilité totale de 1 entre tous les événements ω.

Par exemple, l'ensemble Ω des événements élémentaires lorsqu'on jette une paire de dés est $D^2 = \{$ ⚀⚀, ⚀⚁, ..., ⚅⚅ $\}$, où

$D = \{$ ⚀, ⚁, ⚂, ⚃, ⚄, ⚅ $\}$

désigne l'ensemble des six faces d'un dé. Nous considérerons que deux résultats comme ⚀⚁ et ⚁⚀ sont distincts (en prenant par exemple deux dés de couleurs différentes). Ainsi, l'espace de probabilité contient $6^2 = 36$ éléments.

On suppose généralement que les dés sont "justes", c'est-à-dire que chacune des faces d'un dé a une probabilité $\frac{1}{6}$ d'apparaître, et donc que chacun des événements de Ω a la probabilité $\frac{1}{36}$. On peut aussi considérer que les dés sont pipés, donc qu'ils donnent lieu à une distribution de probabilité différente. Soient par exemple

Attention, ça va se gâter.

$$\begin{aligned} \mathrm{Pr}_1(\boxed{\cdot}) &= \mathrm{Pr}_1(\boxed{\vdots}) = \tfrac{1}{4}\,; \\ \mathrm{Pr}_1(\boxed{\cdot\cdot}) &= \mathrm{Pr}_1(\boxed{\cdot\cdot}) = \mathrm{Pr}_1(\boxed{\cdot\cdot}) = \mathrm{Pr}_1(\boxed{\cdot\cdot}) = \tfrac{1}{8}\,. \end{aligned}$$

On a $\sum_{d\in D} \mathrm{Pr}_1(d) = 1$, donc Pr_1 est bien une distribution de probabilité sur l'ensemble D, et on peut associer des probabilités aux éléments de $\Omega = D^2$ en appliquant la règle

$$\mathrm{Pr}_{11}(d\,d') = \mathrm{Pr}_1(d)\,\mathrm{Pr}_1(d')\,. \tag{8.2}$$

Par exemple, $\mathrm{Pr}_{11}(\boxed{\vdots}\boxed{\cdot\cdot}) = \frac{1}{4}\cdot\frac{1}{8} = \frac{1}{32}$. Cette distribution est correcte car

$$\begin{aligned} \sum_{\omega\in\Omega} \mathrm{Pr}_{11}(\omega) &= \sum_{d\,d'\in D^2} \mathrm{Pr}_{11}(d\,d') = \sum_{d,d'\in D} \mathrm{Pr}_1(d)\,\mathrm{Pr}_1(d') \\ &= \sum_{d\in D} \mathrm{Pr}_1(d) \sum_{d'\in D} \mathrm{Pr}_1(d') = 1\cdot 1 = 1\,. \end{aligned}$$

On peut aussi considérer le cas où l'un des dés est juste tandis que l'autre est pipé,

$$\mathrm{Pr}_{01}(d\,d') = \mathrm{Pr}_0(d)\,\mathrm{Pr}_1(d')\,, \qquad \text{où } \mathrm{Pr}_0(d) = \tfrac{1}{6}, \tag{8.3}$$

auquel cas $\mathrm{Pr}_{01}(\boxed{\vdots}\boxed{\cdot\cdot}) = \frac{1}{6}\cdot\frac{1}{8} = \frac{1}{48}$. Dans le "monde réel", les dés ne sont jamais tout à fait justes car il est impossible d'obtenir une symétrie parfaite ; toutefois, la valeur $\frac{1}{6}$ se trouve généralement très proche de la réalité.

Si toutes les faces d'un cube étaient identiques, comment pourrions-nous savoir quelle face se trouve au-dessus ?

Un *événement* est un sous-ensemble de Ω. Lorsqu'on joue aux dés, par exemple, l'ensemble

$$\{\boxed{\cdot}\boxed{\cdot},\ \boxed{\cdot\cdot}\boxed{\cdot\cdot},\ \boxed{\cdot\cdot}\boxed{\cdot\cdot},\ \boxed{\cdot\cdot}\boxed{\cdot\cdot},\ \boxed{\cdot\cdot}\boxed{\cdot\cdot},\ \boxed{\vdots}\boxed{\vdots}\}$$

constitue l'événement "on a tiré un double". Les éléments ω de Ω sont appelés *événements élémentaires* car ils ne peuvent pas décomposés ; l'élément ω peut être considéré comme l'événement singleton $\{\omega\}$.

La probabilité d'un événement A est définie par la formule

$$\Pr(\omega \in A) \;=\; \sum_{\omega \in A} \Pr(\omega)\,. \tag{8.4}$$

Si $R(\omega)$ est une proposition concernant ω, on désigne par "$\Pr\big(R(\omega)\big)$" la somme de toutes les probabilités $\Pr(\omega)$ telles que $R(\omega)$ est vraie. Ainsi, par exemple, la probabilité de tirer un double avec des dés justes est égale à $\frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{1}{36} = \frac{1}{6}$; en revanche, si les deux dés sont lancés avec la distribution de probabilité $\Pr_1$, alors on trouve $\frac{1}{16}+\frac{1}{64}+\frac{1}{64}+\frac{1}{64}+\frac{1}{64}+\frac{1}{16} = \frac{3}{16} > \frac{1}{6}$. En pipant les dés de cette manière, on rend l'événement "tirer un double" plus probable.

Vous avez peut-être remarqué que nous utilisons ici la notation $\sum$ dans un sens plus général que celui que nous avons défini au chapitre 2 : les sommes de (8.1) et (8.4) portent sur tous les éléments ω d'un ensemble qui n'est pas nécessairement un ensemble de nombres entiers. Malgré tout, il n'y a pas de raison de s'inquiéter : il suffit, comme nous l'avons fait, d'utiliser une notation particulière sous le signe $\sum$ chaque fois qu'on somme sur autre chose que des entiers pour qu'il n'y ait pas de confusion possible. Les autres définitions du chapitre 2 sont toujours correctes dans ce nouveau contexte. En particulier, les sommes infinies, telles qu'elles ont été définies au chapitre 2, correspondent bien à l'interprétation des sommes qui nous intéressent actuellement lorsque l'ensemble Ω est infini. Comme toutes les probabilités sont positives ou nulles et comme leur somme est majorée par la constante 1, la probabilité de l'événement A de (8.4) est bien définie pour tout $A \subseteq \Omega$.

Une *variable aléatoire* est une fonction définie sur les événements élémentaires ω d'un espace de probabilité. Par exemple, si $\Omega = D^2$, on peut définir $S(\omega)$ comme étant la somme des valeurs des deux dés du tirage ω, de sorte que $S(\boxed{\vdots}\boxed{\cdot\cdot})$ $= 6+3 = 9$. La probabilité de tirer un sept est égale à la probabilité de l'événement $S(\omega) = 7$, soit

$$\begin{aligned}&\Pr(\boxed{\cdot}\boxed{\vdots}) + \Pr(\boxed{\cdot}\boxed{\cdot\cdot}) + \Pr(\boxed{\cdot\cdot}\boxed{\cdot\cdot})\\ &\qquad + \Pr(\boxed{\cdot\cdot}\boxed{\cdot\cdot}) + \Pr(\boxed{\cdot\cdot}\boxed{\cdot}) + \Pr(\boxed{\vdots}\boxed{\cdot})\,.\end{aligned}$$

Si on jette des dés justes ($\Pr = \Pr_{00}$), cela arrive avec une probabilité $\frac{1}{6}$; si les dés sont pipés ($\Pr = \Pr_{11}$), la probabilité devient $\frac{1}{16}+\frac{1}{64}+\frac{1}{64}+\frac{1}{64}+\frac{1}{64}+\frac{1}{16} = \frac{3}{16}$, la même que pour les doubles.

On peut généralement se permettre de ne pas écrire "(ω)" lorsqu'on manipule des variables aléatoires car, dans un problème donné, on n'a affaire dans la plupart des cas qu'à un seul espace de probabilité. Ainsi, nous pouvons simplement écrire "$S = 7$" pour désigner l'événement "on a tiré 7", et "$S = 4$" pour désigner l'événement $\{\boxed{\cdot}\boxed{\cdot\cdot}, \boxed{\cdot}\boxed{\cdot}, \boxed{\cdot\cdot}\boxed{\cdot}\}$.

Toute variable aléatoire peut être caractérisée par la distribution de probabilité de ses valeurs. Ainsi, par exemple, on peut écrire dans un tableau les probabilités des onze valeurs $\{2, 3, \ldots, 12\}$ que peut prendre S :

s	2	3	4	5	6	7	8	9	10	11	12
$\Pr_{00}(S=s)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$\Pr_{11}(S=s)$	$\frac{4}{64}$	$\frac{4}{64}$	$\frac{5}{64}$	$\frac{6}{64}$	$\frac{7}{64}$	$\frac{12}{64}$	$\frac{7}{64}$	$\frac{6}{64}$	$\frac{5}{64}$	$\frac{4}{64}$	$\frac{4}{64}$

Si on travaille sur un problème concernant la variable aléatoire S, mais qui ne préjuge d'aucune autre propriété des dés, les seules probabilités de la table ci-dessus suffisent pour trouver la réponse, sans qu'on ait à se préoccuper de la structure de l'ensemble $\Omega = D^2$. En fait, on pourrait considérer que l'espace de probabilité est l'ensemble $\Omega = \{2, 3, \ldots, 12\}$, associé à la distribution de probabilité $\Pr(s)$ désirée. Dans ce cas, "$S = 4$" constituerait un événement élémentaire. Retenons donc qu'on peut très souvent se permettre d'ignorer l'espace de probabilité sous-jacent Ω et travailler directement sur les variables aléatoires et leurs distributions.

Si deux variables aléatoires X et Y sont définies sur le même espace de probabilité Ω, on peut déterminer leur comportement sans rien savoir de Ω, à condition de connaître la "distribution conjointe"

$$\Pr(X=x \text{ et } Y=y)$$

pour toute valeur x que peut prendre X et toute valeur y que peut prendre Y. On dit que X et Y sont des variables aléatoires *indépendantes* si

$$\Pr(X=x \text{ et } Y=y) \;=\; \Pr(X=x)\cdot\Pr(Y=y) \tag{8.5}$$

pour tout x et tout y. Intuitivement, cela signifie que la valeur de X n'a aucun effet sur celle de Y, et réciproquement.

Par exemple, si Ω est l'ensemble D^2 des lancers de deux dés, soit S_1 le chiffre tiré sur le premier dé et S_2 le chiffre tiré sur le second. Alors les variables aléatoires S_1 et S_2 sont indépendantes, pour chacune des trois distributions de probabilité $\Pr_{00}$, $\Pr_{11}$ et $\Pr_{01}$. En effet, nous avons défini la probabilité de chaque événement élémentaire dd' comme le produit de la probabilité de $S_1 = d$ et de la probabilité de $S_2 = d'$. Nous aurions pu décider d'une définition différente, pour que, par exemple,

$$\Pr(\boxed{\cdot}\,\boxed{\cdot\cdot\cdot}) \,/\, \Pr(\boxed{\cdot}\,\boxed{\cdot\cdot\cdot}) \;\neq\; \Pr(\boxed{\cdot}\,\boxed{\cdot\cdot\cdot}) \,/\, \Pr(\boxed{\cdot}\,\boxed{\cdot\cdot\cdot});$$

nous ne l'avons pas fait car deux dés différents ne sont pas sensés s'influencer mutuellement. Avec les définitions dont nous avons convenu, ces rapports valent tous deux $\Pr(S_2=5)/\Pr(S_2=6)$.

La variable aléatoire S est la somme $S_1 + S_2$ des valeurs des dés. Considérons maintenant une autre variable aléatoire P, définie par le produit

S_1S_2. Peut-on dire que S et P sont indépendantes ? Intuitivement non, car si $S = 2$, alors P est forcément égal à 1. Formellement non plus, car, à l'évidence, la condition d'indépendance (8.5) n'est pas respectée (au moins dans le cas où les dés sont justes) : pour chaque valeur possible de s et p, on a $0 < \Pr_{00}(S=s)\cdot\Pr_{00}(P=p) \leqslant \frac{1}{6}\cdot\frac{1}{9}$; cela ne peut pas être égal à $\Pr_{00}(S=s \text{ et } P=p)$ qui est un multiple de $\frac{1}{36}$.

Pour avoir une idée générale du comportement d'une variable aléatoire donnée, on cherche souvent sa moyenne. La *moyenne* d'une suite de nombres est égale à la somme de ses valeurs divisée par le nombre de valeurs. Il existe deux autres notions proches de celle-ci : la *médiane*, qui est la valeur du "milieu" de la suite, c'est-à-dire le nombre de la suite tel qu'il existe dans la suite autant de nombres qui lui sont supérieurs ou égaux que de nombres qui lui sont inférieurs ou égaux (si la suite contient un nombre pair d'éléments, la médiane est égale à la moyenne des deux nombres "centraux") ; et le *mode*, qui est la valeur la plus fréquente dans la suite. Par exemple, la moyenne de la suite $(3,1,4,1,5)$ est $\frac{3+1+4+1+5}{5} = 2,8$; sa médiane vaut 3, et son mode est 1.

Ces trois notions peuvent être définies aussi sur des variables aléatoires plutôt que sur des suites. La *moyenne* d'une variable aléatoire X à valeurs réelles sur un espace de probabilité Ω est égale, par définition, à

$$\sum_{x\in X(\Omega)} x\cdot\Pr(X=x) \tag{8.6}$$

à condition que cette somme éventuellement infinie existe (ici, $X(\Omega)$ désigne l'ensemble des valeurs que peut prendre X). La *médiane* de X est l'ensemble de tous les x tels que

$$\Pr(X\leqslant x) \geqslant \frac{1}{2} \quad \text{et} \quad \Pr(X\geqslant x) \geqslant \frac{1}{2}. \tag{8.7}$$

Enfin, le *mode* de X est l'ensemble de tous les x tels que

$$\Pr(X=x) \geqslant \Pr(X=x') \qquad \text{pour tout } x' \in X(\Omega). \tag{8.8}$$

Remarquez que ces trois définitions, concernant des variables aléatoires, s'accordent bien avec celles concernant les suites de nombres. En effet, supposons qu'on répète une même expérience (par exemple jeter deux dés) un grand nombre de fois, de sorte que chaque valeur de X (ou S ou P, dans le cas des dés) finisse par apparaître avec une fréquence à peu près proportionnelle à sa probabilité. Alors les moyenne, médiane et mode de la suite de nombres obtenue sont proches des moyenne, médiane et mode de la variable aléatoire X.

Dans notre exemple du jeter de dés, la moyenne de S vaut $2\cdot\frac{1}{36}+3\cdot\frac{2}{36}+\cdots+12\cdot\frac{1}{36} = 7$ dans la distribution $\Pr_{00}$; elle est aussi égale à 7 dans $\Pr_{11}$.

La médiane et le mode valent $\{7\}$ aussi dans les deux distributions. La moyenne de P dans la distribution $\Pr_{00}$ est égale à $\frac{49}{4} = 12,25$, sa médiane vaut $\{10\}$ et son mode $\{6,12\}$. Dans la distribution de $\Pr_{11}$, sa moyenne reste la même mais sa médiane tombe à $\{8\}$ tandis que son mode devient le singleton $\{6\}$.

Les théoriciens des probabilités ont un autre nom pour la moyenne d'une variable aléatoire : ils l'appellent l'*espérance* et l'écrivent

$$\mathrm{E}X = \sum_{\omega\in\Omega} X(\omega)\Pr(\omega)\,. \tag{8.9}$$

Dans notre exemple, cette somme contient 36 termes (un pour chaque élément de Ω), alors que (8.6) est une somme de onze termes seulement. Toutefois, elles sont toutes deux égales à

$$\sum_{\substack{\omega\in\Omega\\ x\in X(\Omega)}} x\Pr(\omega)\big[x = X(\omega)\big]\,.$$

Parmi les trois notions que nous venons d'introduire, la plus utile est sans conteste la moyenne. C'est pourquoi, à partir de maintenant, nous allons complètement oublier tout ce qui concerne la médiane et le mode. Au cours du chapitre, nous utiliserons indifféremment les termes "moyenne" et "espérance".

J'ai compris, du moins je l'espère : en moyenne, "moyenne" signifie "espérance".

Si X et Y sont deux variables aléatoires définies sur un même espace de probabilité, alors $X + Y$ est une variable aléatoire sur cet espace. D'après la formule (8.9), la moyenne de leur somme est égale à la somme de leurs moyennes :

$$\begin{aligned}\mathrm{E}(X+Y) &= \sum_{\omega\in\Omega}\big(X(\omega)+Y(\omega)\big)\Pr(\omega)\\ &= \mathrm{E}X + \mathrm{E}Y\,.\end{aligned} \tag{8.10}$$

De même, pour toute constante α on a

$$\mathrm{E}(\alpha X) = \alpha\mathrm{E}X\,. \tag{8.11}$$

La règle correspondante pour la multiplication de deux variables aléatoires est en général plus compliquée, car l'espérance est définie comme une somme sur des événements élémentaires et la plupart des sommes de produits sont difficiles à simplifier. Néanmoins, dans le cas particulier où les deux variables aléatoires sont indépendantes, on a une jolie formule :

$$\mathrm{E}(XY) = (\mathrm{E}X)(\mathrm{E}Y), \qquad \text{si } X \text{ et } Y \text{ sont indépendantes.} \tag{8.12}$$

Ceci peut se prouver à l'aide de la règle de distributivité :

$$\begin{aligned}
E(XY) &= \sum_{\omega\in\Omega} X(\omega)Y(\omega)\cdot \Pr(\omega) \\
&= \sum_{\substack{x\in X(\Omega)\\ y\in Y(\Omega)}} xy\cdot \Pr(X=x \text{ et } Y=y) \\
&= \sum_{\substack{x\in X(\Omega)\\ y\in Y(\Omega)}} xy\cdot \Pr(X=x)\Pr(Y=y) \\
&= \sum_{x\in X(\Omega)} x\Pr(X=x) \cdot \sum_{y\in Y(\Omega)} y\Pr(Y=y) = (EX)(EY).
\end{aligned}$$

Par exemple, nous savons que $S = S_1 + S_2$ et $P = S_1S_2$, où S_1 et S_2 sont les chiffres respectifs donnés par le premier et le second dé après un jet. Comme $ES_1 = ES_2 = \frac{7}{2}$, on a $ES = 7$; de plus, du fait que S_1 et S_2 sont indépendantes, EP vaut bien $\frac{7}{2}\cdot\frac{7}{2} = \frac{49}{4}$, comme nous le prétendions il y a peu. D'autre part, $E(S+P) = ES + EP = 7 + \frac{49}{4}$. Cependant, comme S et P ne sont pas indépendantes, on n'a pas le droit d'affirmer que $E(SP) = 7\cdot\frac{49}{4} = \frac{343}{4}$. En réalité, l'espérance de SP vaut $\frac{637}{6}$ dans la distribution $\Pr_{00}$ et 112 (exactement) dans la distribution $\Pr_{11}$.

8.2 MOYENNE ET VARIANCE

La deuxième caractéristique importante d'une variable aléatoire, après son espérance, est sa *variance*

$$VX = E\big((X-EX)^2\big). \tag{8.13}$$

Si on pose $\mu = EX$, la variance VX est égale à la moyenne de $(X-\mu)^2$. Elle permet de se faire une idée de la "répartition" de la distribution de X.

Voici un exemple de calcul de variance. Supposons qu'on nous donne de quoi acheter deux tickets de loterie, et que c'est une offre que nous ne pouvons pas refuser. Les organisateurs de cette loterie impriment 100 tickets par semaine. L'un de ces tickets, dont le numéro est tiré uniformément au hasard (cela signifie que tous les tickets ont même probabilité), permet à son possesseur de gagner la coquette somme de cent millions de francs ; les 99 autres ne gagnent rien du tout.

(C'est un tout petit peu subtil : il y a deux espaces de probabilité différents selon la stratégie que nous choisissons, mais EX_1 et EX_2 sont égales).

Nous avons le choix entre deux possibilités : soit prendre nos deux tickets pour le même tirage (la même semaine), soit en prendre un pour un tirage donné et l'autre pour un tirage différent (donc ne pas les acheter la même semaine). Quelle est la meilleure stratégie ? Soient X_1 et X_2 les variables aléatoires qui représentent la somme que nous gagnons avec le

premier et le second ticket respectivement. L'espérance de X_1 est

$$EX_1 = \tfrac{99}{100} \cdot 0 + \tfrac{1}{100} \cdot 100 = 1,$$

exprimée en millions de francs ; on trouve exactement la même chose pour l'espérance de X_2. Par conséquent, d'après (8.10), notre gain moyen sera

$$E(X_1 + X_2) = EX_1 + EX_2 = 2 \text{ millions},$$

quelle que soit la stratégie adoptée.

Pourtant, ces deux stratégies ont l'air différentes. Pour nous en assurer, allons plus loin en étudiant l'exacte distribution de probabilité de $X_1 + X_2$:

Au-delà de nos espérances...

	gains (en millions)		
	0	100	200
même tirage	0,9800	0,0200	
tirages différents	0,9801	0,0198	0,0001

Si nous achetons les deux tickets la même semaine, nous avons 98% de chances de ne rien gagner et 2% de chances de gagner 100 millions. Si nous les achetons pendant deux semaines différentes, nous avons 98,01% de chances de ne rien gagner ce qui fait un tout petit peu plus que précédemment, 0,01% de chances de gagner 200 millions ce qui est aussi un tout petit peu supérieur à ce que nous avions avant, tandis que nos chances de gagner 100 millions diminuent pour atteindre et 1,98%. Dans ce deuxième cas, la distribution de $X_1 + X_2$ est donc plus "étalée" ; la valeur du milieu, 100 millions, est un peu moins probable mais les valeurs extrêmes sont un peu plus probables.

La variance permet de quantifier cette notion de répartition plus ou moins étalée d'une variable aléatoire. Nous la mesurons en considérant le carré de la déviation de la variable aléatoire par rapport à sa moyenne. Dans le premier cas, la variance vaut

$$0,98(0M - 2M)^2 + 0,02(100M - 2M)^2 = 196M^2$$

alors que dans le deuxième cas elle est égale à

$$0,9801(0M - 2M)^2 + 0,0198(100M - 2M)^2 + 0,0001(200M - 2M)^2 = 198M^2.$$

Nous trouvons bien une variance plus grande dans ce dernier cas, du fait que la distribution est légèrement plus étalée que dans le premier cas.

Voilà qui est intéressant : la variance d'une somme en franc s'exprime en francs carrés.

Lorsqu'on travaille avec des variances, les nombres peuvent devenir très grands car tout est élevé au carré (le facteur M^2 vaut un billion, ce qui est

plutôt impressionnant, même pour des joueurs invétérés). Pour revenir à une échelle plus raisonnable, on considère habituellement la racine carrée de la variance. On obtient ainsi ce qu'on appelle l'*écart-type*, que l'on désigne généralement par la lettre grecque σ :

$$\sigma = \sqrt{VX}. \tag{8.14}$$

Les écarts-types des variables aléatoires $X_1 + X_2$ de nos deux stratégies valent respectivement $\sqrt{196M^2} = 14,00M$ et $\sqrt{198M^2} \approx 14,071247M$. Dans un certain sens, la seconde stratégie est plus risquée, et le risque supplémentaire "vaut" en quelque sorte 71 247 francs.

Pour réduire les risques, il y a un bon moyen : soudoyer les organisateurs. On obtient alors des probabilités indiscrètes.

(N.B. : la direction décline toute responsabilité quant aux opinions exprimées dans ces marges).

Comment la variance peut-elle nous aider à choisir une stratégie ? Il n'y a pas de réponse catégorique à cette question. Bien sûr, la stratégie qui admet la plus grande variance est un peu plus risquée ; mais vaut-il mieux jouer "pépère" ou prendre un peu plus de risques ? Supposez que nous ayons la possibilité d'acheter 100 tickets au lieu de deux seulement. Dans ce cas, le gain est garanti si nous les achetons tous au même tirage (et la variance est nulle). D'un autre côté, si nous participons à cent tirages différents, nous ne gagnerons rien avec une probabilité $0,99^{100} \approx 0,366$, mais nous aurons une probabilité non nulle de gagner jusqu'à 10 milliards de francs. Notre but n'est pas de décider quelle est la meilleure alternative ; contentons-nous simplement de faire les calculs.

Il existe en fait une manière plus simple de calculer la variance que celle donnée dans la définition (8.13). Comme (EX) est une constante, on a

$$\begin{aligned} E\big((X - EX)^2\big) &= E\big(X^2 - 2X(EX) + (EX)^2\big) \\ &= E(X^2) - 2(EX)(EX) + (EX)^2 . \end{aligned}$$

Par conséquent,

$$VX = E(X^2) - (EX)^2 . \tag{8.15}$$

"La variance est égale à la moyenne des carrés moins le carré de la moyenne".

Il n'y a donc rien d'étonnant à ce que les variances de notre loterie aient été exactement des nombres entiers de billions de francs. Par exemple, la moyenne de $(X_1 + X_2)^2$ vaut $0,98(0M)^2 + 0,02(100M)^2 = 200M^2$ ou $0,9801(0M)^2 + 0,0198(100M)^2 + 0,0001(200M)^2 = 202M^2$, selon le cas. En soustrayant $4M^2$ (le carré de la moyenne) nous retombons sur les résultats précédemment calculés.

Voyons maintenant un résultat particulièrement utile pour calculer $V(X+Y)$ si X et Y sont indépendantes : comme $E(XY) = (EX)(EY)$ dans ce cas, on a

$$E\big((X+Y)^2\big) = E(X^2 + 2XY + Y^2) = E(X^2) + 2(EX)(EY) + E(Y^2) .$$

Par conséquent,

$$\begin{aligned} V(X+Y) &= E\big((X+Y)^2\big) - (EX+EY)^2 \\ &= E(X^2) + 2(EX)(EY) + E(Y^2) \\ &\quad - (EX)^2 - 2(EX)(EY) - (EY)^2 \\ &= E(X^2) - (EX)^2 + E(Y^2) - (EY)^2 \\ &= VX + VY. \end{aligned} \tag{8.16}$$

"La variance d'une somme de variables aléatoires indépendantes est égale à la somme de leurs variances". Par exemple, la variance de la somme que nous pouvons gagner avec un unique ticket vaut

$$E(X_1^2) - (EX_1)^2 \;=\; 0,99(0M)^2 + 0,01(100M)^2 - (1M)^2 \;=\; 99M^2\,.$$

Ainsi, la variance du gain total de deux tickets dans deux tirages différents (donc indépendants) est égale à $2 \times 99M^2 = 198M^2$. Plus généralement, la variance pour n tickets dans n tirages indépendants vaut $n \times 99M^2$.

Revenons un instant à nos jets de dés. La variance de la somme des valeurs des dés S se calcule avec la formule ci-dessus, car $S = S_1 + S_2$ est la somme de deux variables aléatoires indépendantes. Comme

$$VS_1 \;=\; \frac{1}{6}(1^2+2^2+3^2+4^2+5^2+6^2) - \left(\frac{7}{2}\right)^2 \;=\; \frac{35}{12}$$

si les dés sont justes, on a $VS = \frac{35}{12} + \frac{35}{12} = \frac{35}{6}$. Si on lance un dé pipé, alors

$$VS_1 \;=\; \frac{1}{8}(2\cdot 1^2+2^2+3^2+4^2+5^2+2\cdot 6^2) - \left(\frac{7}{2}\right)^2 \;=\; \frac{45}{12}\,;$$

par conséquent, $VS = \frac{45}{6} = 7,5$ si les deux dés sont pipés. Remarquez que la variance de S est plus élevée si les dés sont pipés, mais que tirer la valeur 7, moyenne dans les deux cas, est plus probable si les dés sont pipés que s'ils sont justes. Si le but du jeu est de tirer le plus possible de 7, la variance n'est absolument pas le meilleur critère de réussite.

Bon. Nous savons maintenant calculer des variances. Par contre, nous ne savons pas vraiment pourquoi nous calculons ce paramètre. Par quelle vertu est-il naturel que d'autres que nous pourrions imaginer ? La réponse réside dans l'*inégalité de Tchebychev* ([29] et [57]), qui met en évidence une propriété importante de la variance :

$$\Pr\big((X-EX)^2 \geqslant \alpha\big) \;\leqslant\; VX/\alpha, \qquad \text{pour tout } \alpha > 0 \tag{8.17}$$

(à ne pas confondre avec les inégalités monotones de Tchebychev que nous avons vues au chapitre 2).

En gros, la formule (8.17) signifie que si la variance VX d'une variable aléatoire X est petite, sa valeur sera rarement loin de sa moyenne EX. La preuve de cette propriété est étonnamment simple :

$$\begin{aligned} VX &= \sum_{\omega\in\Omega} \big(X(\omega)-EX\big)^2 \Pr(\omega) \\ &\geqslant \sum_{\substack{\omega\in\Omega \\ (X(\omega)-EX)^2\geqslant\alpha}} \big(X(\omega)-EX\big)^2 \Pr(\omega) \\ &\geqslant \sum_{\substack{\omega\in\Omega \\ (X(\omega)-EX)^2\geqslant\alpha}} \alpha \Pr(\omega) \quad = \quad \alpha\cdot\Pr\big((X-EX)^2\geqslant\alpha\big)\,; \end{aligned}$$

il suffit de diviser par α pour conclure.

Si on note μ la moyenne et σ l'écart-type et si on remplace α par c^2VX dans (8.17), la condition $(X-EX)^2 \geqslant c^2VX$ devient $(X-\mu)^2 \geqslant (c\sigma)^2$; donc, (8.17) entraîne que

$$\Pr\big(|X-\mu|\geqslant c\sigma\big) \;\leqslant\; 1/c^2\,. \tag{8.18}$$

Ainsi, la valeur de X se situe presque toujours à moins de c fois l'écart-type de sa valeur moyenne ; elle n'est en dehors de l'intervalle ainsi défini qu'avec une probabilité $1/c^2$. Toute variable aléatoire a au moins 75% de chances de se trouver à moins de 2σ de μ ; elle a au moins 99% de chances de se situer entre $\mu-10\sigma$ et $\mu+10\sigma$. Ces résultats constituent les cas particuliers $\alpha = 4VX$ et $\alpha = 100VX$ de l'inégalité de Tchebychev.

Si on jette deux dés n fois de suite, où n est un grand nombre, la somme des valeurs des jets sera presque toujours proche de $7n$. En effet, la variance de n jets indépendants est égale à $\frac{35}{6}n$, donc l'écart-type vaut

$$\sqrt{\tfrac{35}{6}n}\,.$$

L'inégalité de Tchebychev entraîne que la somme finale sera entre

$$7n - 10\sqrt{\tfrac{35}{6}n} \quad \text{et} \quad 7n + 10\sqrt{\tfrac{35}{6}n}$$

dans au moins 99% des cas si on utilise des dés justes. Par exemple, si on jette un million de fois les dés, il y a 99% de chances que le total soit compris entre 6,975 millions et 7,025 millions.

Voyons quelque chose de plus général. Soit X une variable aléatoire *quelconque* sur un espace de probabilité Ω, admettant une moyenne μ et un écart-type σ finis. Considérons alors l'espace de probabilité Ω^n, dont les événements élémentaires sont des n-uplets $(\omega_1, \omega_2, \ldots, \omega_n)$ tels que

$\omega_k \in \Omega$ pour tout k, et dont les probabilités sont définies par

$$\Pr(\omega_1, \omega_2, \ldots, \omega_n) = \Pr(\omega_1)\Pr(\omega_2)\ldots\Pr(\omega_n).$$

Si on définit un ensemble de variables aléatoires X_k par la formule

$$X_k(\omega_1, \omega_2, \ldots, \omega_n) = X(\omega_k),$$

alors la quantité

$$X_1 + X_2 + \cdots + X_n$$

désigne une somme de de n variables aléatoires indépendantes. Cela revient à prendre n "échantillons" indépendants de X sur Ω et les additionner. Comme la moyenne de $X_1 + X_2 + \cdots + X_n$ est égale à $n\mu$ et son écart-type est $\sqrt{n}\,\sigma$, la moyenne des n échantillons,

$$\frac{1}{n}(X_1 + X_2 + \cdots + X_n),$$

se trouvera entre $\mu - 10\sigma/\sqrt{n}$ et $\mu + 10\sigma/\sqrt{n}$ pendant au moins 99% du temps. En d'autres termes, si n est choisi assez grand, alors la moyenne de n échantillons indépendants sera presque toujours extrêmement proche de l'espérance EX. (Il existe même un résultat plus fort, la "loi forte des grands nombres", qui est démontré dans tous les livres de théorie des probabilités ; toutefois, pour notre propos, le résultat simple que nous venons de déduire de l'inégalité de Tchebychev suffit largement).

(Cela signifie que, pour tout n fixé, la moyenne d'une ensemble de n échantillons indépendants sera entre les limites données dans au moins 99% des cas. Ne croyez surtout pas qu'il s'agisse de la moyenne d'une suite infinie X_1, X_2, X_3, ... lorsque n varie).

Il arrive parfois qu'on ait besoin d'estimer la moyenne d'une variable aléatoire X sans rien connaître des caractéristiques de son espace de probabilité (par exemple, si on veut connaître la température moyenne à midi en janvier à San Francisco, ou la durée de vie moyenne des agents d'assurances). Pour cela, on observe des échantillons de X : si on a récolté des valeurs empiriques indépendantes $X_1, X_2, \ldots, X_n$, on peut raisonnablement considérer que la véritable moyenne de X est à peu près égale à

$$\widehat{E}X = \frac{X_1 + X_2 + \cdots + X_n}{n}. \tag{8.19}$$

On peut aussi estimer la variance avec la formule

$$\widehat{V}X = \frac{X_1^2 + X_2^2 + \cdots + X_n^2}{n-1} - \frac{(X_1 + X_2 + \cdots + X_n)^2}{n(n-1)}. \tag{8.20}$$

Notez bien les facteurs $(n-1)$ dans les dénominateurs. A priori, on aurait envie de les remplacer par n, comme dans (8.19), parce que la véritable variance VX se calcule, dans (8.15), en fonction de l'espérance. Pourtant,

on obtient une meilleure estimation avec $n-1$ qu'avec n, car la définition (8.20) implique que

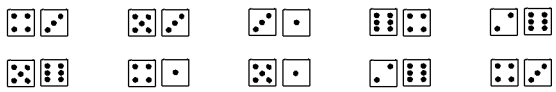
$$E(\widehat{V}X) = VX. \tag{8.21}$$

Voici pourquoi :

$$\begin{aligned}
E(\widehat{V}X) &= \frac{1}{n-1}E\Big(\sum_{k=1}^{n} X_k^2 - \frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n} X_jX_k\Big) \\
&= \frac{1}{n-1}\Big(\sum_{k=1}^{n} E(X_k^2) - \frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n} E(X_jX_k)\Big) \\
&= \frac{1}{n-1}\Big(\sum_{k=1}^{n} E(X^2) - \frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\big(E(X)^2[j\neq k] + E(X^2)[j=k]\big)\Big) \\
&= \frac{1}{n-1}\Big(nE(X^2) - \frac{1}{n}\big(nE(X^2) + n(n-1)E(X)^2\big)\Big) \\
&= E(X^2) - E(X)^2 = VX.
\end{aligned}$$

Remarquez que, dans ce calcul, on fait usage de l'hypothèse d'indépendance des observations lorsqu'on remplace $E(X_jX_k)$ par $(EX)^2[j\neq k]+E(X^2)[j=k]$.

En pratique, les résultats expérimentaux sur une variable aléatoire X sont généralement obtenus en calculant une moyenne $\hat{\mu} = \widehat{E}X$ et un écart-type $\hat{\sigma} = \sqrt{\widehat{V}X}$ empiriques, et en présentant le résultat sous la forme "$\hat{\mu} \pm \hat{\sigma}/\sqrt{n}$". Voici par exemple dix jets de deux dés supposés justes :



La moyenne empirique de la somme S vaut

$$\hat{\mu} = (7+11+8+5+4+6+10+8+8+7)/10 = 7{,}4$$

et la variance empirique est égale à

$$(7^2+11^2+8^2+5^2+4^2+6^2+10^2+8^2+8^2+7^2-10\hat{\mu}^2)/9 \approx 2{,}1^2.$$

Sur la base de ces résultats expérimentaux, nous obtenons une valeur estimée de $7{,}4 \pm 2{,}1/\sqrt{10} \approx 7{,}4 \pm 0.7$, pour la moyenne de la somme S.

Voyons encore un exemple de calcul, théorique cette fois, de la moyenne et de la variance d'une variable aléatoire. Nous avons considéré au chapitre 5 le problème de la victoire au football ; il s'agissait de n chapeaux lancés

en l'air et d'une permutation aléatoire de ces chapeaux. Nous avons montré dans l'équation (5.51) que la probabilité qu'aucun des supporters ne retrouve son chapeau vaut $n_{\underline{1}}/n! \approx 1/e$. Nous avons aussi calculé la probabilité que k personnes exactement retrouvent leurs propres chapeaux :

$$P(n,k) = \frac{1}{n!}\binom{n}{k}(n-k)_{\underline{1}} = \frac{1}{k!}\frac{(n-k)_{\underline{1}}}{(n-k)!}. \tag{8.22}$$

Présentons ces résultats avec le formalisme que nous venons de découvrir. Soit Π_n l'espace de probabilité contenant la totalité des $n!$ permutations π de $\{1,2,\ldots,n\}$, avec $\Pr(\pi) = 1/n!$ pour tout $\pi \in \Pi_n$. La variable aléatoire

$$F_n(\pi) = \text{nombre de "points fixes" de } \pi, \qquad \text{pour } \pi \in \Pi_n,$$

A ne pas confondre avec un nombre de Fibonacci.

compte le nombre d'appariements corrects (chapeau,propriétaire) dans le problème de la victoire au football. Bien que nous sachions pertinemment que l'équation (8.22) implique que $\Pr(F_n = k)$, faisons comme si nous ne connaissions pas cette formule et calculons la valeur moyenne et l'écart-type de F_n.

En fait, le calcul de la moyenne est beaucoup plus facile à effectuer maintenant qu'au chapitre 5. Remarquons simplement que

$$F_n(\pi) = F_{n,1}(\pi) + F_{n,2}(\pi) + \cdots + F_{n,n}(\pi),$$
$$\text{où } F_{n,k}(\pi) = [\text{la kième position de } \pi \text{ est un point fixe}], \text{ pour } \pi \in \Pi_n.$$

Par conséquent, $EF_n = EF_{n,1} + EF_{n,2} + \cdots + EF_{n,n}$. Or, l'espérance de $F_{n,k}$ est tout bonnement égale à la probabilité que $F_{n,k} = 1$, qui vaut $1/n$ car, parmi les $n!$ permutations $\pi = \pi_1\pi_2\ldots\pi_n \in \Pi_n$, il y en a exactement $(n-1)!$ qui satisfont $\pi_k = k$. Par conséquent,

$$EF_n = n/n = 1, \qquad \text{pour } n > 0. \tag{8.23}$$

En moyenne, il y aura un chapeau qui se retrouvera sur sa tête attitrée. "Le nombre moyen de points fixes d'une permutation aléatoire est égal à 1".

La question de l'écart type est plus délicate, car les $F_{n,k}$ ne sont pas indépendants les uns par rapport aux autres. Nous pouvons néanmoins calculer la variance en analysant leurs dépendances mutuelles :

$$\begin{aligned} E(F_n^2) &= E\Big(\Big(\sum_{k=1}^{n} F_{n,k}\Big)^2\Big) = E\Big(\sum_{j=1}^{n}\sum_{k=1}^{n} F_{n,j}F_{n,k}\Big) \\ &= \sum_{j=1}^{n}\sum_{k=1}^{n} E(F_{n,j}F_{n,k}) = \sum_{1\le k\le n} E(F_{n,k}^2) + 2\sum_{1\le j<k\le n} E(F_{n,j}F_{n,k}) \end{aligned}$$

(c'est une astuce du même genre que nous avons utilisée lors du calcul de (2.33) au chapitre 2). Comme $F_{n,k}$ vaut toujours 0 ou 1, on a $F_{n,k}^2 = F_{n,k}$, donc $E(F_{n,k}^2) = EF_{n,k} = 1/n$ comme précédemment. De plus, si $j < k$, alors $E(F_{n,j}\,F_{n,k}) = \Pr(j$ et k sont tous deux des points fixes de $\pi) = (n-2)!/n! = 1/n(n-1)$. Par conséquent,

$$E(F_n^2) \;=\; \frac{n}{n} + \binom{n}{2}\frac{2}{n(n-1)} \;=\; 2\,, \qquad \text{pour } n \geqslant 2. \tag{8.24}$$

Par exemple, pour $n = 3$, on a $\frac{2}{6}0^2 + \frac{3}{6}1^2 + \frac{0}{6}2^2 + \frac{1}{6}3^2 = 2$. La variance est $E(F_n^2) - (EF_n)^2 = 1$, donc l'écart-type vaut aussi 1 (tout comme la moyenne). "Une permutation aléatoire de $n \geqslant 2$ éléments a 1 ± 1 points fixes en moyenne."

8.3 FG DE PROBABILITE

Si X est une variable aléatoire à valeurs entières positives ou nulles, les techniques présentées au chapitre 7 offrent un moyen efficace de calculer sa distribution de probabilité. La *fonction génératrice de probabilité* (ou fgp) de X est définie par

$$G_X(z) \;=\; \sum_{k\geqslant 0} \Pr(X=k)\, z^k\,. \tag{8.25}$$

Cette série entière en z contient toutes les informations concernant la variable aléatoire X. Elle peut aussi s'exprimer de deux autres façons :

$$G_X(z) \;=\; \sum_{\omega\in\Omega} \Pr(\omega)\, z^{X(\omega)} \;=\; E(z^X)\,. \tag{8.26}$$

Les coefficients de $G_X(z)$ sont positifs ou nuls et leur somme vaut 1 ; cette dernière condition peut s'écrire

$$G_X(1) \;=\; 1\,. \tag{8.27}$$

Réciproquement, toute série entière $G(z)$ à coefficients positifs ou nuls et telle que $G(1) = 1$ est la fgp d'une certaine variable aléatoire.

Généralement, l'usage des fgp simplifie le calcul de moyennes et de variances. La moyenne, notamment, s'exprime sans difficulté :

$$\begin{aligned} EX \;&=\; \sum_{k\geqslant 0} k\cdot \Pr(X=k) \\ &=\; \sum_{k\geqslant 0} \Pr(X=k)\cdot kz^{k-1}\big|_{z=1} \\ &=\; G_X'(1)\,. \end{aligned} \tag{8.28}$$

Il suffit de dériver la fgp par rapport à z, puis de poser $z = 1$.

La variance est un tout petit peu plus compliquée :

$$\begin{aligned} E(X^2) &= \sum_{k\geqslant 0} k^2 \cdot \Pr(X=k) \\ &= \sum_{k\geqslant 0} \Pr(X=k) \cdot \big(k(k-1)z^{k-2} + kz^{k-1}\big)\big|_{z=1} \\ &= G''_X(1) + G'_X(1)\,. \end{aligned}$$

Par conséquent,

$$VX = G''_X(1) + G'_X(1) - G'_X(1)^2\,. \tag{8.29}$$

Donc, selon les équations (8.28) et (8.29), il suffit, pour pouvoir calculer la moyenne et la variance, de savoir calculer les valeurs $G'_X(1)$ et $G''_X(1)$ des deux dérivées. Nous n'avons pas besoin de connaître une forme close des probabilités, ni même une forme close de $G_X(z)$.

Pour simplifier les calculs ultérieurs, convenons d'écrire, pour toute fonction G,

$$\mathrm{Moy}(G) = G'(1)\,, \tag{8.30}$$

$$\mathrm{Var}(G) = G''(1) + G'(1) - G'(1)^2\,,. \tag{8.31}$$

Les fgp sont appréciables aussi parce que, dans la plupart des cas, elles se trouvent être des fonctions relativement simples de z. Observons par exemple la *distribution uniforme* d'ordre n, dans laquelle la variable aléatoire prend chacune des valeurs $\{0, 1, \ldots, n-1\}$ avec la probabilité $1/n$. La fgp correspondante

$$U_n(z) = \frac{1}{n}(1 + z + \cdots + z^{n-1}) = \frac{1}{n}\frac{1-z^n}{1-z}\,, \qquad \text{pour } n \geqslant 1. \tag{8.32}$$

admet une forme close en tant que série géométrique.

Pourtant, cette forme close s'avère quelque peu embarrassante : si on pose $z = 1$, on obtient le rapport indéfini $0/0$, bien que $U_n(z)$ soit un polynôme parfaitement défini pour toute valeur de z. A partir de la forme non close $(1 + z + \cdots + z^{n-1})/n$, la valeur $U_n(1) = 1$ est immédiate ; si on veut la déduire de la forme close, il semble qu'il faille recourir à la règle de L'Hospital pour calculer $\lim_{z\to 1} U_n(z)$. L'évaluation de $U'_n(1)$ par cette même règle sera plus délicate, car il y aura un facteur $(z-1)^2$ au dénominateur ; $U''_n(1)$ sera plus difficile encore.

Fort heureusement, il y a un moyen d'éviter ce dilemme. Si une série $G(z) = \sum_{n\geqslant 0} g_n z^n$ converge au moins pour une valeur de z telle que $|z| > 1$, la série $G'(z) = \sum_{n\geqslant 0} n g_n z^{n-1}$ a la même propriété, ainsi que $G''(z)$,

$G'''(z)$, etc. Par conséquent, on peut appliquer la formule de Taylor pour écrire

$$G(1+t) = G(1) + \frac{G'(1)}{1!}t + \frac{G''(1)}{2!}t^2 + \frac{G'''(1)}{3!}t^3 + \cdots. \tag{8.33}$$

Lorsqu'on développe $G(1+t)$ en série, toutes les dérivées de $G(z)$ en $z = 1$ apparaissent dans les coefficients.

Par exemple, on peut facilement trouver, par ce biais, les dérivées de la fgp uniforme $U_n(z)$:

$$\begin{aligned} U_n(1+t) &= \frac{1}{n}\frac{(1+t)^n - 1}{t} \\ &= \frac{1}{n}\binom{n}{1} + \frac{1}{n}\binom{n}{2}t + \frac{1}{n}\binom{n}{3}t^2 + \cdots + \frac{1}{n}\binom{n}{n}t^{n-1}. \end{aligned}$$

En comparant ceci à (8.33), on obtient

$$U_n(1) = 1; \quad U_n'(1) = \frac{n-1}{2}; \quad U_n''(1) = \frac{(n-1)(n-2)}{3}. \tag{8.34}$$

Plus généralement, $U_n^{(m)}(1) = (n-1)^{\underline{m}}/(m+1)$ (mais les cas $m = 1$ et $m = 2$ nous suffisent pour calculer la moyenne et la variance). La moyenne de la distribution uniforme est donc égale à

$$U_n'(1) = \frac{n-1}{2}, \tag{8.35}$$

et voici sa variance :

$$\begin{aligned} U_n''(1) + U_n'(1) - U_n'(1)^2 &= 4\frac{(n-1)(n-2)}{12} + 6\frac{(n-1)}{12} - 3\frac{(n-1)^2}{12} \\ &= \frac{n^2-1}{12}. \end{aligned} \tag{8.36}$$

Voici maintenant une autre raison d'apprécier les fgp. Nous avons vu aux chapitres 5 et 7 que le produit de deux fonctions génératrices correspond à la convolution des suites correspondantes. Voici un fait plus important encore, qui concerne les fgp : si deux variables aléatoires sont indépendantes, le produit de leurs fgp correspond à leur somme. En effet, si X et Y sont des variables aléatoire qui ne prennent que des valeurs entières, la probabilité que $X + Y = n$ est

$$\Pr(X+Y=n) = \sum_k \Pr(X=k \text{ et } Y=n-k).$$

Si X et Y sont indépendantes, on en déduit que

$$\Pr(X+Y=n) = \sum_k \Pr(X=k)\,\Pr(Y=n-k)\,,$$

ce qui constitue une convolution. Voici donc le résultat que nous obtenons :

$$G_{X+Y}(z) = G_X(z)\,G_Y(z)\,, \qquad \text{si } X \text{ et } Y \text{ sont indépendantes.} \tag{8.37}$$

Un peu plus haut dans ce chapitre, nous avons établi que $V(X+Y) = VX+VY$ lorsque X et Y sont indépendantes. Soient $F(z)$ et $G(z)$ les fgp de X et Y et $H(z)$ la fgp de $X+Y$. Alors

$$H(z) = F(z)G(z)\,,$$

et on déduit des formules (8.28) à (8.31) que

$$\mathrm{Moy}(H) = \mathrm{Moy}(F) + \mathrm{Moy}(G)\,; \tag{8.38}$$
$$\mathrm{Var}(H) = \mathrm{Var}(F) + \mathrm{Var}(G)\,. \tag{8.39}$$

Ces deux propriétés de $\mathrm{Moy}(H) = H'(1)$ et $\mathrm{Var}(H) = H''(1) + H'(1) - H'(1)^2$ ne sont pas vraies pour n'importe quel produit de fonctions $H(z) = F(z)G(z)$. Dans le cas général, on a

$$\begin{aligned} H'(z) &= F'(z)G(z) + F(z)G'(z)\,, \\ H''(z) &= F''(z)G(z) + 2F'(z)G'(z) + F(z)G''(z)\,. \end{aligned}$$

Toutefois, si on pose $z = 1$, on voit qu'il suffit que

$$F(1) = G(1) = 1 \tag{8.40}$$

et que les dérivées existent pour que (8.38) et (8.39) soient vraies. Les "probabilités" n'ont même pas besoin d'être dans l'intervalle $[0\,..\,1]$. Si $F(1)$ et $G(1)$ ne valent pas 1, il suffit de normaliser les fonctions $F(z)$ et $G(z)$ en les divisant respectivement par $F(1)$ et $G(1)$ pour que la condition adéquate soit respectée, pourvu, bien sûr, que les valeurs de $F(1)$ et de $G(1)$ soient non nulles.

La moyenne et la variance font partie d'une série infinie de paramètres, appelés *cumulants*, qui ont été introduits par l'astronome danois Thorvald Nicolai Thiele [351] en 1903. Le deux premiers cumulants, κ_1 et κ_2, d'une variable aléatoire sont exactement les paramètres que nous avons appelé moyenne et variance. Les cumulants d'ordre plus élevé permettent d'exprimer des propriétés plus subtiles d'une distribution. Si $G(z)$ est la fgp d'une variable aléatoire, la série de ses cumulants est définie par la formule

$$\ln G(e^t) = \frac{\kappa_1}{1!}t + \frac{\kappa_2}{2!}t^2 + \frac{\kappa_3}{3!}t^3 + \frac{\kappa_4}{4!}t^4 + \cdots. \tag{8.41}$$

Regardons cela de plus près. Si $G(z)$ est la fgp de X, on a

$$\begin{aligned} G(e^t) &= \sum_{k\geqslant 0} \Pr(X=k)e^{kt} = \sum_{k,m\geqslant 0} \Pr(X=k)\frac{k^m t^m}{m!} \\ &= 1 + \frac{\mu_1}{1!}t + \frac{\mu_2}{2!}t^2 + \frac{\mu_3}{3!}t^3 + \cdots , \end{aligned} \tag{8.42}$$

où μ_m, défini par

$$\mu_m = \sum_{k\geqslant 0} k^m \Pr(X=k) = E(X^m), \tag{8.43}$$

est appelé le "moment d'ordre m" de X. Si on prend l'exponentielle de chacun des deux membres de (8.41), on obtient une nouvelle formule pour $G(e^t)$:

$$\begin{aligned} G(e^t) &= 1 + \frac{(\kappa_1 t + \frac{1}{2}\kappa_2 t^2 + \cdots)}{1!} + \frac{(\kappa_1 t + \frac{1}{2}\kappa_2 t^2 + \cdots)^2}{2!} + \cdots \\ &= 1 + \kappa_1 t + \tfrac{1}{2}(\kappa_2 + \kappa_1^2)t^2 + \cdots . \end{aligned}$$

Il ne reste plus qu'à mettre en équations les coefficients des puissances de t pour en déduire une suite de formules

$$\kappa_1 = \mu_1 , \tag{8.44}$$
$$\kappa_2 = \mu_2 - \mu_1^2 , \tag{8.45}$$
$$\kappa_3 = \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3 , \tag{8.46}$$
$$\kappa_4 = \mu_4 - 4\mu_1\mu_3 + 12\mu_1^2\mu_2 - 3\mu_2^2 - 6\mu_1^4 , \tag{8.47}$$
$$\begin{aligned} \kappa_5 = \mu_5 &- 5\mu_1\mu_4 + 20\mu_1^2\mu_3 - 10\mu_2\mu_3 \\ &+ 30\mu_1\mu_2^2 - 60\mu_1^3\mu_2 + 24\mu_1^5 , \end{aligned} \tag{8.48}$$
$$\vdots$$

qui exprime les cumulants en fonction des moments. Notez que κ_2 est bien égal à la variance $E(X^2) - (EX)^2$.

"For these higher half-invariants we shall propose no special names."
—T. N. Thiele [351]

On peut immédiatement déduire de l'équation (8.41) que les cumulants définis par le produit $F(z)G(z)$ de deux fgp sont égaux aux sommes des cumulants correspondants de $F(z)$ et $G(z)$, tout simplement parce que le logarithme d'un produit est une somme. Par conséquent, tout cumulant d'une somme de variables aléatoires indépendantes est additif, comme le sont la moyenne et la variable. En raison de cette propriété, les cumulants se trouvent être des paramètres plus importants que les moments.

Changeons maintenant de point de vue et écrivons

$$G(1+t) = 1 + \frac{\alpha_1}{1!}t + \frac{\alpha_2}{2!}t^2 + \frac{\alpha_3}{3!}t^3 + \cdots .$$

Les α_m sont des "moments factoriels" car, selon (8.33),

$$\begin{aligned}
\alpha_m &= G^{(m)}(1) \\
&= \sum_{k\geqslant 0} \Pr(X=k)k^{\underline{m}} z^{k-m}\Big|_{z=1} \\
&= \sum_{k\geqslant 0} k^{\underline{m}} \Pr(X=k) \\
&= E(X^{\underline{m}})\,. \qquad (8.49)
\end{aligned}$$

Il s'ensuit que

$$\begin{aligned}
G(e^t) &= 1 + \frac{\alpha_1}{1!}(e^t-1) + \frac{\alpha_2}{2!}(e^t-1)^2 + \cdots \\
&= 1 + \frac{\alpha_1}{1!}(t + \tfrac{1}{2}t^2 + \cdots) + \frac{\alpha_2}{2!}(t^2 + t^3 + \cdots) + \cdots \\
&= 1 + \alpha_1 t + \tfrac{1}{2}(\alpha_2 + \alpha_1)t^2 + \cdots\,,
\end{aligned}$$

et on peut exprimer les cumulants en fonction des dérivées $G^{(m)}(1)$:

$$\kappa_1 = \alpha_1\,, \qquad (8.50)$$

$$\kappa_2 = \alpha_2 + \alpha_1 - \alpha_1^2\,, \qquad (8.51)$$

$$\kappa_3 = \alpha_3 + 3\alpha_2 + \alpha_1 - 3\alpha_2\alpha_1 - 3\alpha_1^2 + 2\alpha_1^3\,, \qquad (8.52)$$

$$\vdots$$

Ce formules constituent une généralisation de (8.38) et (8.39) à tous les cumulants.

Revenons sur terre et appliquons ces notions nouvelles sur des exemples simples. La variable aléatoire la plus simple possible est la "constante aléatoire" : X a une valeur fixée x avec probabilité 1. Dans ce cas, $G_X(z) = z^x$ et $\ln G_X(e^t) = xt$, donc la moyenne est égale à x et tous les autres cumulants sont nuls. Il s'ensuit que l'action de multiplier une fgp par z^x a pour effet d'augmenter la moyenne de x en laissant inchangés la variance et tous les autres cumulants.

Voyons maintenant comment appliquer les fonction génératrices de probabilité à notre jeu de dés. La fgp de la distribution des valeurs d'un dé est

$$G(z) = \frac{z + z^2 + z^3 + z^4 + z^5 + z^6}{6} = zU_6(z)\,,$$

où U_6 est la fgp de la distribution uniforme d'ordre 6. Comme le facteur "z" ajoute 1 à la moyenne, celle-ci est égale à $3,5$ au lieu de la valeur $\frac{n-1}{2} = 2,5$ calculée selon (8.35). Par contre, ce facteur "z" ne modifie pas le calcul de la variance (8.36), donc celle-ci est égale à $\frac{35}{12}$.

La fgp de la somme des valeurs de deux dés indépendants est égale au carré de la fgp d'un unique dé :

$$\begin{aligned} G_S(z) &= \frac{z^2+2z^3+3z^4+4z^5+5z^6+6z^7+5z^8+4z^9+3z^{10}+2z^{11}+z^{12}}{36} \\ &= z^2 U_6(z)^2 . \end{aligned}$$

Si on jette deux dés n fois de suite, la probabilité d'obtenir une somme totale égale à k vaut donc

$$\begin{aligned} [z^k]\, G_S(z)^n &= [z^k]\, z^{2n} U_6(z)^{2n} \\ &= [z^{k-2n}]\, U_6(z)^{2n} . \end{aligned}$$

La distribution de chapeaux est une sorte de distribution d'uniformes.

Dans le problème de la victoire au football, ou problème d'énumération des points fixes d'une permutation aléatoire, on sait d'après (5.49) que la fgp est

$$F_n(z) = \sum_{0 \leqslant k \leqslant n} \frac{(n-k)_{\downarrow}}{(n-k)!} \frac{z^k}{k!} , \qquad \text{pour } n \geqslant 0. \tag{8.53}$$

Par conséquent,

$$\begin{aligned} F_n'(z) &= \sum_{1 \leqslant k \leqslant n} \frac{(n-k)_{\downarrow}}{(n-k)!} \frac{z^{k-1}}{(k-1)!} \\ &= \sum_{0 \leqslant k \leqslant n-1} \frac{(n-1-k)_{\downarrow}}{(n-1-k)!} \frac{z^k}{k!} \\ &= F_{n-1}(z) . \end{aligned}$$

Inutile de calculer les coefficients pour conclure de cette récurrence $F_n'(z) = F_{n-1}(z)$ que $F_n^{(m)}(z) = F_{n-m}(z)$, et donc que

$$F_n^{(m)}(1) = F_{n-m}(1) = [n \geqslant m] . \tag{8.54}$$

Avec cette formule, il nous est facile de calculer la moyenne et la variance. Nous trouvons, comme précédemment (mais plus rapidement), qu'elles sont toutes deux égales à 1 pour tout $n \geqslant 2$.

En fait, nous pouvons même montrer que le mième cumulant κ_m de cette variable aléatoire est égal à 1 pour tout $n \geqslant m$. En effet, le mième cumulant ne dépend que de $F_n'(1)$, $F_n''(1)$, ..., $F_n^{(m)}(1)$, qui sont tous égaux à 1 ; on obtient le même résultat pour le mième cumulant en remplaçant $F_n(z)$ par la fgp limite

$$F_\infty(z) = e^{z-1} , \tag{8.55}$$

qui satisfait $F_\infty^{(m)}(1) = 1$ pour tout m. Les cumulants de F_∞ sont tous égaux à 1 du fait que

$$\ln F_\infty(e^t) = \ln e^{e^t-1} = e^t - 1 = \frac{t}{1!} + \frac{t^2}{2!} + \frac{t^3}{3!} + \cdots .$$

8.4 PILE OU FACE

Nous allons maintenant nous intéresser à des processus qui n'ont que deux résultats possibles. Si on lance une pièce de monnaie en l'air, on a une probabilité p d'obtenir pile et une probabilité q d'obtenir face, avec $p + q = 1$. (nous supposons que la pièce ne tombe pas sur la tranche, ni dans un trou, etc). Dans cette section, la somme des nombres p et q sera toujours égale à 1. Si la pièce est équilibrée, on a $p = q = \frac{1}{2}$; autrement, nous dirons que la pièce est biaisée.

Les escrocs patentés savent bien que $p \approx 0,9$ si on joue à pile ou face avec un penny américain tout neuf sur une surface lisse : à cause de la position du centre de gravité, la tête d'Abraham Lincoln se trouve presque toujours dessous.

Voici la fonction génératrice de probabilité pour le nombre de piles après un unique lancer :

$$H(z) = q + pz. \tag{8.56}$$

Si on lance la pièce n fois de suite, en supposant que deux lancers différents sont indépendants, la fonction génératrice de probabilité du nombre de piles est, selon la formule du binôme,

$$H(z)^n = (q + pz)^n = \sum_{k\geqslant 0} \binom{n}{k} p^k q^{n-k} z^k . \tag{8.57}$$

Ainsi, la probabilité d'obtenir exactement k fois pile dans n lancers est égale à $\binom{n}{k}p^k q^{n-k}$. Cette suite de probabilités est appelée la *loi binomiale*

Supposons maintenant que nous lancions une pièce jusqu'à ce que nous obtenions pile. Quelle est la probabilité que k lancers exactement soient nécessaires ? Il est évident que $k = 1$ avec probabilité p ; et la probabilité que $k = 2$ vaut qp (c'est la probabilité de trouver face d'abord, puis pile). De façon générale, la probabilité d'effectuer k lancers est égale à $q^{k-1}p$. La fonction génératrice correspondante est donc

$$pz + qpz^2 + q^2pz^3 + \cdots = \frac{pz}{1 - qz} . \tag{8.58}$$

Si on répète le processus jusqu'à obtenir n fois pile, on trouve la fgp

$$\begin{aligned}\left(\frac{pz}{1-qz}\right)^n &= p^n z^n \sum_k \binom{n+k-1}{k} (qz)^k \\ &= \sum_k \binom{k-1}{k-n} p^n q^{k-n} z^k . \end{aligned} \tag{8.59}$$

Remarquons en passant qu'elle est égale à z^n multiplié par

$$\left(\frac{p}{1-qz}\right)^n = \sum_k \binom{n+k-1}{k} p^n q^k z^k, \tag{8.60}$$

la fonction génératrice de la *loi binomiale négative*.

L'espace de probabilité de l'exemple (8.59) est tout à fait différent de ceux que nous avons rencontrés jusqu'à présent dans ce chapitre : il est infini, alors que les précédents étaient finis. Chacun de ses éléments est une suite de finie de piles et/ou de faces, contenant exactement n piles, et finissant par pile. La probabilité d'une telle suite est égale à $p^n q^{k-n}$, où $k-n$ est le nombre de faces. Ainsi, par exemple, si $n = 3$ et si on note P pour pile et F pour face, la suite FPFFFPP appartient à l'espace de probabilité, et sa probabilité est égale à $qpqqqpp = p^3q^4$.

Pile je gagne, face tu perds.
Non ? D'accord ; face tu perds, pile je gagne.
Non ? Bon, alors pile tu perds, face je gagne.

Soient X une variable aléatoire de loi binomiale (8.57) et Y une variable aléatoire de loi binomiale négative (8.60). Ces distributions dépendent de n et de p. La moyenne de X est $nH'(1) = np$, car sa fgp est $H(z)^n$; sa variance est

$$n\big(H''(1) + H'(1) - H'(1)^2\big) = n(0 + p - p^2) = npq. \tag{8.61}$$

L'écart-type est donc égal à $\sqrt{npq}$: si on lance n fois une pièce de monnaie, on doit donc obtenir pile à peu près $np \pm \sqrt{npq}$ fois. La moyenne et la variance de Y se calculent de manière similaire : si on pose

$$G(z) = \frac{p}{1-qz},$$

alors on a

$$G'(z) = \frac{pq}{(1-qz)^2},$$
$$G''(z) = \frac{2pq^2}{(1-qz)^3};$$

par conséquent, $G'(1) = pq/p^2 = q/p$ et $G''(1) = 2pq^2/p^3 = 2q^2/p^2$. Il s'ensuit que la moyenne et la variance de Y sont respectivement égales à nq/p et nq/p^2.

Il y a une façon plus simple de trouver ce résultat, en utilisant la fonction génératrice inverse

$$F(z) = \frac{1-qz}{p} = \frac{1}{p} - \frac{q}{p}z, \tag{8.62}$$

et en écrivant

$$G(z)^n = F(z)^{-n}. \tag{8.63}$$

Ce polynôme $F(z)$ n'est pas une fonction génératrice de probabilité car il a des coefficients négatifs. Cependant, il satisfait la condition fondamentale $F(1) = 1$. Formellement, $F(z)$ correspond donc à une pièce avec laquelle la "probabilité" d'obtenir pile est égale à $-q/p$, et $G(z)$ équivaut à lancer cette pièce -1 fois(!). Ainsi, la loi binomiale négative de paramètres (n, p) peut être considérée comme une loi binomiale ordinaire de paramètres $(n', p') = (-n, -q/p)$. Formellement, la moyenne doit donc être $n'p' = (-n)(-q/p) = nq/p$, et la variance $n'p'q' = (-n)(-q/p)(1 + q/p) = nq/p^2$. Ce calcul strictement formel, où apparaissent des probabilités négatives, est tout à fait correct. En effet, les résultats que nous y appliquons sont basés sur des identités de séries formelles pour lesquelles il n'est pas nécessaire que $0 \leqslant p \leqslant 1$.

La probabilité que je rajeunisse est négative.

Ah ? Alors la probabilité que tu vieillisses ou que tu ne changes pas est > 1.

Passons à un autre exemple : Combien de fois faut-il lancer une pièce pour obtenir pile deux fois de suite ? Maintenant, l'espace de probabilité est constitué de toutes les suites de P et de F qui finissent par PP mais qui ne contiennent pas d'autre occurrence de deux P consécutifs :

$$\Omega = \{\mathrm{PP}, \mathrm{FPP}, \mathrm{FFPP}, \mathrm{PFPP}, \mathrm{FFFPP}, \mathrm{FPFPP}, \mathrm{PFFPP}, \ldots\}.$$

La probabilité d'une séquence donnée s'obtient en remplaçant P par p et F par q. Par exemple, la suite FPFPP apparaît avec la probabilité

$$\Pr(\mathrm{FPFPP}) = qpqpp = p^3q^2.$$

Soit S la somme infinie

$$S = \mathrm{PP} + \mathrm{FPP} + \mathrm{FFPP} + \mathrm{PFPP} + \mathrm{FFFPP} + \mathrm{FPFPP} + \mathrm{PFFPP} + \cdots$$

de tous les éléments de Ω. Si on remplace chaque P par pz et chaque F par qz, on obtient la fonction génératrice de probabilité du nombre de lancers nécessaires pour tomber deux fois de suite sur pile.

Il existe une curieuse relation entre S et la somme des pavages par des dominos

$$T = \mathrm{I} + \square + \square\square + \square\square\square + \square\square\square\square + \square\square\square\square\square + \cdots$$

de l'équation (7.1) : S peut être obtenue à partir de T en remplaçant chaque $\square$ par F et chaque $\square\square\square$ par PF, puis en ajoutant PP à la fin. Il est facile de démontrer cette correspondance, car tout élément de Ω est de la forme $(\mathrm{F} + \mathrm{PF})^n\mathrm{PP}$, pour un certain $n \geqslant 0$, et tout terme de T est de la forme $(\square + \square\square\square)^n$. Par conséquent, d'après (7.4),

$$S = (1 - \mathrm{F} - \mathrm{PF})^{-1}\mathrm{PP},$$

et la fonction génératrice de probabilité que nous cherchons est donc

$$\begin{aligned} G(z) &= \big(1 - qz - (pz)(qz)\big)^{-1}(pz)^2 \\ &= \frac{p^2z^2}{1 - qz - pqz^2}\,. \end{aligned} \tag{8.64}$$

Ce que nous avons fait récemment avec la loi binomiale négative nous incite à calculer la moyenne et la variance de (8.64) en écrivant

$$G(z) = \frac{z^2}{F(z)}\,,$$

où

$$F(z) = \frac{1 - qz - pqz^2}{p^2}\,,$$

et en calculant la "moyenne" et la "variance" de cette pseudo-fgp $F(z)$ (qui satisfait $F(1) = 1$). On a

$$\begin{aligned} F'(1) &= (-q - 2pq)/p^2 = 2 - p^{-1} - p^{-2}\,; \\ F''(1) &= -2pq/p^2 = 2 - 2p^{-1}\,. \end{aligned}$$

Par conséquent, comme $z^2 = F(z)G(z)$, $\mathrm{Moy}(z^2) = 2$, et $\mathrm{Var}(z^2) = 0$, voici la moyenne et la variance de la distribution $G(z)$:

$$\mathrm{Moy}(G) = 2 - \mathrm{Moy}(F) = p^{-2} + p^{-1}\,; \tag{8.65}$$

$$\mathrm{Var}(G) = -\mathrm{Var}(F) = p^{-4} + 2p^{-3} - 2p^{-2} - p^{-1}\,. \tag{8.66}$$

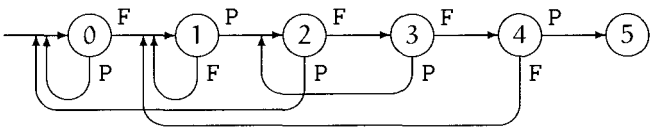
Lorsque $p = \frac{1}{2}$, la moyenne et la variance valent 6 et 22 respectivement. (L'exercice 4 est consacré aux calculs de moyenne et de variance par soustraction).

Tentons maintenant une expérience plus compliquée : nous allons lancer une pièce en l'air jusqu'à obtenir la configuration FPFFP. La somme des suites gagnantes est

$$\begin{aligned} S = {} & \mathrm{FPFFP} + \mathrm{PFPFFP} + \mathrm{FFPFFP} \\ & + \mathrm{PPFPFFP} + \mathrm{PFFPFFP} + \mathrm{FPFPFFP} + \mathrm{FFFPFFP} + \cdots\,; \end{aligned}$$

" "Vous être un automate, une machine à calculer", m'écriai-je. "Il y a parfois en vous quelque chose de positivement inhumain". "
—J. H. Watson [83]

elle est plus difficile à décrire que la précédente. Pour résoudre le problème, nous pouvons considérer S comme un "langage" défini par l'"automate fini" suivant :



Les événements élémentaires de l'espace de probabilité sont exactement les suites de P et de F qui mènent de l'état 0 à l'état 5. Supposons par exemple que nous venions de faire FPF ; alors nous nous trouvons dans l'état 3. A ce moment, si nous tirons face, nous allons dans l'état 4 ; et si nous tirons pile, toujours depuis l'état 3, nous nous retrouvons en 2 (et non en 0, car il suffit que le FP que nous venons de faire soit suivi par FFP pour que nous gagnions.

Soit S_k la somme de toutes les suites de P et F qui conduisent de l'état 0 à l'état k. Alors

$$\begin{aligned}
S_0 &= 1 + S_0 P + S_2 P\,,\\
S_1 &= S_0 F + S_1 F + S_4 F\,,\\
S_2 &= S_1 P + S_3 P\,,\\
S_3 &= S_2 F\,,\\
S_4 &= S_3 F\,,\\
S_5 &= S_4 P\,.
\end{aligned}$$

Notre somme S est égale à S_5. Pour l'obtenir, il suffit de résoudre ce système de six équations à six inconnues $S_0, S_1, \ldots, S_5$. Si on remplace P par pz et F par qz dans l'un des F_k ainsi trouvés, on trouve une fonction génératrice dont le coefficient de z^n représente la probabilité d'être dans l'état k après n tirages.

Ce procédé peut s'appliquer à tout diagramme de transitions entre états, où la transition entre l'état j et l'état k est associée à une probabilité $p_{j,k}$. On obtient ainsi un système d'équations linéaires dont les solutions sont les fonctions génératrices des probabilités des états après n transitions. Les systèmes de ce type sont appelés des *processus de Markov* et l'étude de leur comportement est très liée à la théorie des équations linéaires.

Il existe aussi une manière bien plus simple de résoudre notre problème de pile ou face, qui évite le passage par l'automate fini. Les six équations à six inconnues $S_0, S_1, \ldots, S_5$, peuvent être avantageusement remplacées par seulement deux équations à deux inconnues. Voici l'astuce : considérons la somme $N = S_0 + S_1 + S_2 + S_3 + S_4$ de toutes les suites qui ne contiennent aucune occurrence de FPFFP,

$$N = 1 + P + F + PP + \cdots + FPFPF + FPFFF + \cdots\,.$$

Alors

$$1 + N(P + F) = N + S\,, \tag{8.67}$$

car le membre gauche contient les termes qui finissent par FPFFP (et donc appartiennent à S) et les autres (qui appartiennent à N) ; inversement, tout

terme de droite non vide appartient soit à N P soit à N F. De plus, l'équation

$$\mathrm{N\,FPFFP} \;=\; \mathrm{S} + \mathrm{S\,FFP} \tag{8.68}$$

est vérifiée, car, si un terme du membre gauche n'est pas élément de S, alors il le devient si on supprime ses trois dernières lettres ; et tout terme du membre droit se retrouve dans le membre gauche.

Il est facile de résoudre ce système de deux équations : d'après (8.67), $\mathrm{N} = (1-\mathrm{S})(1-\mathrm{P}-\mathrm{F})^{-1}$, donc

$$(1-\mathrm{S})(1-\mathrm{F}-\mathrm{P})^{-1}\,\mathrm{FPFFP} \;=\; \mathrm{S}(1+\mathrm{FFP})\,.$$

Il suffit, comme précédemment, de remplacer P par pz et F par qz pour obtenir la fonction génératrice de probabilité $G(z)$ du nombre de lancers. Après quelques simplifications (parce que $p+q=1$), on trouve

$$\frac{\big(1-G(z)\big)\,p^2q^3z^5}{1-z} \;=\; G(z)(1+pq^2z^3)\,;$$

par conséquent, la solution est

$$G(z) \;=\; \frac{p^2q^3z^5}{p^2q^3z^5+(1+pq^2z^3)(1-z)}\,. \tag{8.69}$$

Remarquez que $G(1)=1$ si $pq \neq 0$. Nous sommes donc sûrs d'obtenir la configuration FPFFP avec probabilité 1, sauf si la pièce est truquée au point de toujours tomber du même côté.

Pour trouver la moyenne et la variance de la distribution (8.69), inversons $G(z)$ comme nous l'avons fait dans le problème précédent en écrivant $G(z) = z^5/F(z)$, où F est un polynôme :

$$F(z) \;=\; \frac{p^2q^3z^5+(1+pq^2z^3)(1-z)}{p^2q^3}\,. \tag{8.70}$$

Nous en déduisons les dérivées

$$\begin{aligned} F'(1) &= 5-(1+pq^2)/p^2q^3\,,\\ F''(1) &= 20-6pq^2/p^2q^3\,, \end{aligned}$$

et, si X désigne le nombre de lancers,

$$\mathrm{EX} \;=\; \mathrm{Moy}(G) \;=\; 5-\mathrm{Moy}(F) \;=\; p^{-2}q^{-3}+p^{-1}q^{-1}\,; \tag{8.71}$$

$$\begin{aligned} \mathrm{VX} \;=\; \mathrm{Var}(G) &= -\mathrm{Var}(F)\\ &= -25+p^{-2}q^{-3}+7p^{-1}q^{-1}+\mathrm{Moy}(F)^2\\ &= (\mathrm{EX})^2-9p^{-2}q^{-3}-3p^{-1}q^{-1}\,. \end{aligned} \tag{8.72}$$

Pour $p=\frac{1}{2}$, la moyenne et la variance valent respectivement 36 et 996.

Nous allons maintenant voir les choses de façon bien plus générale. Le problème que nous venons de résoudre nous donne de bonnes pistes pour nous attaquer au cas où nous lançons la pièce jusqu'à obtenir une configuration donnée *quelconque* A de P et de F. Soient donc S la somme de toutes les suites gagnantes et N la somme des suites qui n'ont pas encore abouti au motif A recherché. Alors l'équation (8.67) ne change pas, tandis que l'équation (8.68) devient

$$\begin{aligned} NA &= S\big(1 + A^{(1)}\,[A^{(m-1)} = A_{(m-1)}] + A^{(2)}\,[A^{(m-2)} = A_{(m-2)}] \\ &\qquad + \cdots + A^{(m-1)}\,[A^{(1)} = A_{(1)}]\big)\,, \end{aligned} \tag{8.73}$$

où m est la longueur de A, et $A^{(k)}$ et $A_{(k)}$ désignent respectivement les k dernieres lettres et les k premières lettres de A. Par exemple, si A est le motif FPFFP que nous venons d'étudier, alors

$$\begin{aligned} A^{(1)} &= P\,, & A^{(2)} &= FP\,, & A^{(3)} &= FFP\,, & A^{(4)} &= PFFP\,; \\ A_{(1)} &= F\,, & A_{(2)} &= FP\,, & A_{(3)} &= FPF\,, & A_{(4)} &= FPFF\,. \end{aligned}$$

Comme la seule égalité vérifiée est $A^{(2)} = A_{(2)}$, l'équation (8.73) se réduit à (8.68).

Soit $\widetilde{A}$ le résultat obtenu en remplaçant P par p^{-1} et F par q^{-1} dans A. Alors il n'est pas difficile de généraliser nos calculs de (8.71) et (8.72) pour conclure (voir l'exercice 20) que la moyenne vaut

$$EX = \sum_{k=1}^{m} \widetilde{A}_{(k)}\,[A^{(k)} = A_{(k)}]\,; \tag{8.74}$$

et que la variance est donnée par

$$VX = (EX)^2 - \sum_{k=1}^{m} (2k-1)\widetilde{A}_{(k)}\,[A^{(k)} = A_{(k)}]\,. \tag{8.75}$$

Si $p = \frac{1}{2}$, il existe une interprétation particulièrement simple de ces formules. Etant donné un motif A de longueur m, soit

$$A{:}A = \sum_{k=1}^{m} 2^{k-1}\,[A^{(k)} = A_{(k)}]\,. \tag{8.76}$$

Nous allons voir maintenant comment trouver la représentation en base deux de ce nombre : commencez par superposer deux occurrences de A, puis déplaçez la seconde occurrence vers la droite, lettre par lettre ; écrivez "1" sous la première lettre de cette seconde occurrence chaque fois que les

lettres des deux occurrences coïncident exactement, écrivez "0" dans le cas contraire. La figure suivante présente une illustration de ce procédé :

$$
\begin{array}{rl}
A = & \mathrm{PFPFPPFPFP} \\
A{:}A = & (1000010101)_2 = 512 + 16 + 4 + 1 = 533 \\
& \mathrm{PFPFPPFPFP} \quad \surd \\
& \phantom{\mathrm{P}}\mathrm{PFPFPPFPFP} \\
& \phantom{\mathrm{PP}}\mathrm{PFPFPPFPFP} \\
& \phantom{\mathrm{PPP}}\mathrm{PFPFPPFPFP} \\
& \phantom{\mathrm{PPPP}}\mathrm{PFPFPPFPFP} \\
& \phantom{\mathrm{PPPPP}}\mathrm{PFPFPPFPFP} \quad \surd \\
& \phantom{\mathrm{PPPPPP}}\mathrm{PFPFPPFPFP} \\
& \phantom{\mathrm{PPPPPPP}}\mathrm{PFPFPPFPFP} \quad \surd \\
& \phantom{\mathrm{PPPPPPPP}}\mathrm{PFPFPPFPFP} \\
& \phantom{\mathrm{PPPPPPPPP}}\mathrm{PFPFPPFPFP} \quad \surd
\end{array}
$$

"Chem bol'she periodov u nashego slova, tem pozzhe ono poîavlîetsîâ." — A. D. Solov'ev

D'après (8.74), le nombre moyen de lancers nécessaires pour que le motif A apparaisse est exactement égal à $2(A{:}A)$, car $\widetilde{A}_{(k)} = 2^k$ lorsque $p = q = \frac{1}{2}$. Ce résultat, découvert par le mathématicien soviétique A. D. Solov'ev en 1966 [331], peut sembler paradoxal à première vue : les motifs qui ne coïncident pas lorsqu'on les superpose surviennent plus tôt que les autres ! Il faut presque deux fois plus de temps pour rencontrer PPPPP que pour rencontrer PPPPF ou FPPPP.

Voyons maintenant un jeu inventé par Walter Penney [289] en 1969. Alice et Bill jouent à pile ou face jusqu'a ce qu'ils obtiennent PPF ou PFF. Si c'est PPF, c'est Alice qui gagne, sinon c'est Bill. Ce jeu semble bien être équitable si la pièce n'est pas truquée, car les deux motifs PPF et PFF présentent les mêmes caractéristiques si on le considère séparément : la fonction génératrice de probabilité du nombre de lancers nécessaires pour obtenir PPF est

$$G(z) = \frac{z^3}{z^3 - 8(z-1)},$$

Bien sûr que non ! Sur qui auraient-ils l'avantage ?

et on trouve exactement la même pour PFF. Par conséquent, ni Alice ni Bill ne prend l'avantage s'ils jouent en solitaire.

En revanche, lorsqu'on considère les deux motifs simultanément, leur interaction donne quelque chose d'intéressant. Soient A et B les sommes des configurations gagnantes d'Alice et de Bill respectivement :

$$
\begin{aligned}
S_A &= \mathrm{PPF} + \mathrm{PPPF} + \mathrm{FPPF} + \mathrm{PPPPF} + \mathrm{PFPPF} + \mathrm{FPPPF} + \cdots; \\
S_B &= \mathrm{PFF} + \mathrm{FPFF} + \mathrm{PFPFF} + \mathrm{FFPFF} + \mathrm{FPFPFF} + \mathrm{FFFPFF} + \cdots.
\end{aligned}
$$

Soit aussi N la somme de toutes les suites pour lesquelles aucun des deux joueurs n'a encore gagné :

$$N = 1 + P + F + PP + PF + FP + FF + PPP + PFP + FPP + \cdots . \tag{8.77}$$

On peut alors facilement vérifier les équations suivantes :

$$\begin{aligned} 1 + N(P+F) &= N + S_A + S_B ; \\ N\,PPF &= S_A ; \\ N\,PFF &= S_A F + S_B . \end{aligned} \tag{8.78}$$

Si on pose $P = F = \frac{1}{2}$, la valeur de S_A est égale à la probabilité qu'Alice gagne, et la valeur de S_B est égale à la probabilité que Bill gagne. Les trois équations se réduisent à

$$1 + N = N + S_A + S_B ; \qquad \tfrac{1}{8}N = S_A ; \qquad \tfrac{1}{8}N = \tfrac{1}{2}S_A + S_B .$$

Nous en déduisons que $S_A = \frac{2}{3}$ et $S_B = \frac{1}{3}$. Alice gagnera en moyenne deux fois plus souvent que Bill !

On peut généraliser ce jeu en permettant à Alice et Bill de choisir leurs motifs A et B au lieu de les leur imposer. Les deux motifs ne sont pas nécessairement de même longueur, mais il ne faut pas que A soit contenu dans B ou inversement, sinon le jeu dégénérerait (par exemple, si $A = PF$ et $B = FPFP$, le pauvre Bill ne peut jamais gagner, et si $A = PFP$ et $B = FP$, les deux joueurs peuvent prétendre à la victoire en même temps). On peut alors écrire trois équations analogues à (8.73) et (8.78) :

$$\begin{aligned} 1 + N(P+F) &= N + S_A + S_B ; \\ NA &= S_A \sum_{k=1}^{l} A^{(l-k)} [A^{(k)} = A_{(k)}] + S_B \sum_{k=1}^{\min(l,m)} A^{(l-k)} [B^{(k)} = A_{(k)}] ; \\ NB &= S_A \sum_{k=1}^{\min(l,m)} B^{(m-k)} [A^{(k)} = B_{(k)}] + S_B \sum_{k=1}^{m} B^{(m-k)} [B^{(k)} = B_{(k)}] . \end{aligned} \tag{8.79}$$

Ici, l et m désignent respectivement les longueurs de A et B. Par exemple, si $A = PFFPFPFP$ et $B = FPFPFFP$, les deux dernières équations se réduisent à

$$\begin{aligned} N\,PFFPFPFP &= S_A\,FFPFPFP + S_A + S_B\,FFPFPFP + S_B\,FPFP ; \\ N\,FPFPFFP &= S_A\,FPFFP + S_A\,FFP + S_B\,FPFFP + S_B . \end{aligned}$$

Si on suppose que la pièce n'est pas truquée, on obtient les probabilités de

victoire en posant $P = F = \frac{1}{2}$. Nos deux équations deviennent alors

$$\begin{aligned} N &= S_A \sum_{k=1}^{l} 2^k [A^{(k)} = A_{(k)}] + S_B \sum_{k=1}^{\min(l,m)} 2^k [B^{(k)} = A_{(k)}] ; \\ N &= S_A \sum_{k=1}^{\min(l,m)} 2^k [A^{(k)} = B_{(k)}] + S_B \sum_{k=1}^{m} 2^k [B^{(k)} = B_{(k)}] . \end{aligned} \tag{8.80}$$

Nous verrons mieux ce qui se passe si nous généralisons l'opération A:A de (8.76) à une fonction de deux motifs indépendants A et B :

$$A{:}B = \sum_{k=1}^{\min(l,m)} 2^{k-1} [A^{(k)} = B_{(k)}] . \tag{8.81}$$

Les équations (8.80) se réécrivent alors simplement en

$$S_A(A{:}A) + S_B(B{:}A) = S_A(A{:}B) + S_B(B{:}B) ,$$

et l'avantage d'Alice vaut

$$\frac{S_A}{S_B} = \frac{B{:}B - B{:}A}{A{:}A - A{:}B} . \tag{8.82}$$

Cette belle formule est due à John Horton Conway [137].

Par exemple, si A = PFFPFPFP et B = FPFPFFP comme précédemment, alors $A{:}A = (10000001)_2 = 129$, $A{:}B = (0001010)_2 = 10$, $B{:}A = (0001001)_2 = 9$ et $B{:}B = (1000010)_2 = 66$. Donc, le rapport S_A/S_B vaut $(66 - 9)/(129 - 10) = 57/119$. Alice ne gagnera ce jeu que 57 fois sur 176 en moyenne.

Il peut se passer des choses étranges dans le jeu de Penney. Par exemple, le motif PPFP gagne contre le motif PFPP à 3 contre 2, et PFPP gagne contre FPPP à 7 contre 5. Par conséquent, PPFP devrait être bien meilleur que FPPP. Eh bien non, car FPPP gagne contre PPFP, à 7 contre 5 ! La relation entre les motifs n'est pas transitive. L'exercice 57 montre que si Alice choisit un quelconque motif $\tau_1\tau_2 \ldots \tau_l$ de longueur $l \geqslant 3$, Bill est sûr d'avoir plus de chances de gagner s'il choisit le motif $\overline{\tau}_2\tau_1\tau_2 \ldots \tau_{l-1}$, où $\overline{\tau}_2$ est égal à F si τ_2 est égal à P et inversement.

Vous avez dit bizarre ? Comme c'est bizarre !

8.5 HACHAGE

Pour conclure ce chapitre, nous allons appliquer la théorie des probabilités à l'informatique. Un bon nombre d'algorithmes de stockage et de recherche d'informations dans un ordinateur sont basés sur une technique que l'on appelle le "hachage". Voici le problème qu'elle permet de