4

INTRODUCTION À LA THÉORIE PROBABILISTE

STATISTIQUES APPLIQUÉES

La NASA*

Washsington, D.C.

La NASA (National Aeronautics and Space Administration) est l'agence gouvernementale américaine en charge du programme spatial civil américain et de la recherche en aéronautique et aérospatiale. Plus connue pour l'exploration de l'espace en vol habité, la mission de la NASA est d'être à la pointe des avancées dans le domaine de l'exploration spatiale, des découvertes scientifiques et de la recherche en aéronautique. La NASA, avec ses 18 800 employés, travaille actuellement à la conception d'un nouveau système de lancement qui emmènera les astronautes plus loin dans l'espace et sera la pierre angulaire des explorations futures de l'espace par l'homme.

Alors que la mission première de la NASA est l'exploration de l'espace, son expertise a été mise au service de nombreux pays et organisations à travers le monde. Par exemple, la NASA est intervenue lors de l'effondrement de la mine de cuivre et d'or San José à Copiapo au Chili, piégeant 33 mineurs à plus de 2 000 pieds sous terre. Pour ramener ces hommes à la surface aussi vite que possible, il était impératif que les équipes de secours soient correctement guidées pour sauver autant de mineurs que possible. Le gouvernement chilien a demandé à la NASA de l'assister pour concevoir un plan de secours. En réponse, la NASA a dépêché sur place quatre personnes, un ingénieur, deux physiciens et un psychologue ayant une expertise en matière de conception des véhicules et de situations de confinement longue durée.

Les probabilités de succès et d'échec de plusieurs plans de secours occupaient tous les esprits. En l'absence de données historiques face à cette situation inédite, les scientifiques de la NASA ont développé des probabilités subjectives de succès et d'échec des différents plans de secours en se basant sur des circonstances similaires auxquelles des astronautes ont fait face lors de leur retour de missions plus ou moins longues dans l'espace. Les probabilités fournies par la NASA ont guidé les choix des responsables chiliens en fournissant des indications sur la façon dont les mineurs pouvaient survivre à l'ascension dans une cage de secours.

Le plan de secours conçu par les autorités chiliennes en coordination avec l'équipe de la NASA a conduit à la construction d'une cage de secours de 13 pieds de long et pesant 924 livres dans le but de remonter les mineurs à la surface un par un. Tous les mineurs ont été sauvés, le dernier remontant à la surface 68 jours après l'effondrement de la mine.

Dans ce chapitre, vous serez initié au calcul et à l'interprétation des probabilités dans de nombreuses situations. En plus de la définition de probabilités subjectives, vous apprendrez à assigner des probabilités en utilisant les méthodes classiques et la méthode des fréquences relatives. Les relations probabilistes de base, les probabilités conditionnelles et le théorème de Bayes seront également abordés.

* Les auteurs remercient les docteurs Michael Duncan et Clinton Cragg, de la NASA, de leur avoir fourni ce Statistiques appliquées.

Les responsables fondent souvent leurs décisions sur une analyse d'éléments incertains, comme par exemple :

1. Quelles sont les chances que les ventes baissent si on augmente les prix ?

2. Quelle est la probabilité qu'une nouvelle méthode d'assemblage augmente la productivité ?

3. Quelle est la probabilité que le projet soit fini à temps ?

4. Quelles sont les chances qu'un nouvel investissement soit rentable ?

La **probabilité** est une mesure numérique de la vraisemblance d'occurrence d'un événement. Ainsi, les probabilités peuvent être utilisées pour mesurer le degré d'incertitude associé aux quatre événements cités ci-dessus. Si les probabilités étaient connues, nous pourrions déterminer la vraisemblance que chaque événement survienne.

> Une série de lettres entre Pierre de Fermat et Blaise Pascal, dans les années 1650, est à l'origine des travaux sur les probabilités.

La valeur d'une probabilité est toujours comprise entre 0 et 1. Une probabilité proche de zéro signifie qu'un événement a peu de chance de se produire ; une probabilité proche de 1 signifie qu'un événement se produira très certainement. Les probabilités comprises entre 0 et 1 représentent les degrés de vraisemblance qu'un événement se réalise. Par exemple, si nous considérons l'événement « il pleut demain », nous comprenons que lorsque le bulletin météo indique « une probabilité proche de zéro qu'il pleuve », cela signifie qu'il n'y a presque aucune chance qu'il pleuve. Cependant, si la probabilité qu'il pleuve est de 0,90, nous savons qu'il est très vraisemblable qu'il pleuve. Une probabilité de 0,50 indique qu'il y a une chance sur deux qu'il pleuve. La figure 4.1 illustre la présentation de la probabilité comme une mesure numérique de la vraisemblance d'un événement.

4.1 EXPÉRIENCE, RÈGLES DE COMPTAGE ET ATTRIBUTION DE PROBABILITÉS

En termes probabilistes, une **expérience** est un processus qui génère un ensemble de résultats prédéfinis. Lorsque l'expérience n'est pas répétée, un seul des résultats possibles de l'expérience se produit. Plusieurs exemples d'expériences, et leurs résultats possibles sont présentés ci-dessous.

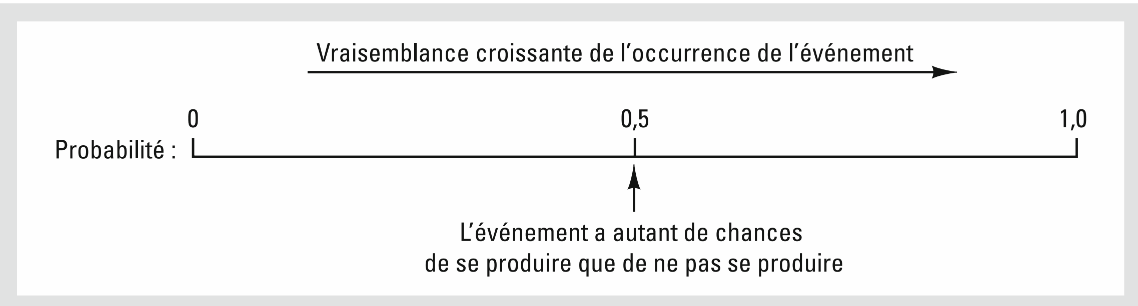

Figure 4.1 ***Probabilité, mesure numérique de la vraisemblable de l'occurrence d'un événement***

Expérience	Résultats de l'expérience
Lancer une pièce de monnaie	Pile, Face
Sélectionner une pièce pour l'inspecter	Défectueuse, non défectueuse
Faire une offre de vente	Achat, pas d'achat
Lancer un dé	1, 2, 3, 4, 5, 6
Jouer au foot	Gagner, perdre, match nul

L'ensemble des résultats possibles d'une expérience est également appelé « **espace-échantillon** ».

► **Espace-échantillon**
L'espace-échantillon d'une expérience correspond à l'ensemble des résultats possibles.

Un résultat possible de l'expérience est également appelé « **élément de l'échantillon** », pour souligner le fait qu'il s'agit d'un élément de l'espace-échantillon.

Les résultats possibles de l'expérience sont également appelés « éléments de l'échantillon ».

Considérons la première expérience inscrite dans le tableau précédent, lancer une pièce de monnaie. Les résultats de l'expérience (les éléments de l'échantillon) correspondent à la face visible de la pièce – pile ou face. Si l'on note l'espace-échantillon S, on peut le décrire de la manière suivante :

$$S = \{\text{Pile, Face}\}$$

L'espace-échantillon de la seconde expérience inscrite dans le tableau – sélectionner une pièce pour l'inspecter – est décrit par :

$$S = \{\text{Défectueuse, Non défectueuse}\}$$

Les deux expériences décrites ci-dessus ont deux résultats possibles (l'échantillon est composé de deux éléments). Considérons la quatrième expérience inscrite dans le tableau, lancer un dé. Les résultats possibles de l'expérience, définis comme le nombre de points apparaissant sur la face supérieure du dé, sont les six éléments de l'espace-échantillon de cette expérience :

$$S = \{1, 2, 3, 4, 5, 6\}$$

4.1.1 Règles de comptage, combinaisons et permutations

Être capable d'identifier et de dénombrer les résultats possibles de l'expérience est une étape nécessaire dans la détermination des probabilités. Nous discutons maintenant de trois règles de comptage, très utiles.

Expériences à plusieurs étapes. La première règle de comptage considérée est appropriée pour les expériences à étapes multiples. Considérons l'expérience consistant à lancer deux pièces de monnaie. Les résultats de l'expérience correspondent au côté visible des deux pièces (pile ou face). Combien de résultats sont possibles pour cette expérience ? Elle peut être considérée comme une expérience à deux étapes, dans laquelle l'étape 1 correspond au lancer de la première pièce et l'étape 2 au lancer de la seconde pièce. Si l'on note l'apparition du côté pile par P et l'apparition du côté face par F, le résultat (F, F) indique que le côté face est apparu lors des deux lancers. En utilisant cette notation, l'espace-échantillon (S) de cette expérience de lancer de pièces est :

$$S = \{(F, F), (P, F), (F, P), (P, P)\}$$

Ainsi, quatre résultats sont possibles. Dans ce cas, il n'est pas difficile d'énumérer tous les résultats possibles.

La règle de comptage des expériences à plusieurs étapes permet de dénombrer les résultats possibles sans les énumérer.

► **Règle de comptage des expériences à plusieurs étapes**
Si une expérience peut être décrite par une séquence de k étapes, avec n_1 résultats possibles à la première étape, n_2 résultats possibles à la seconde étape et ainsi de suite, alors le nombre total de résultats possibles de l'expérience est égal à $(n_1)(n_2)\ldots(n_k)$.

En considérant l'expérience du lancer de deux pièces comme la séquence d'un premier lancer ($n_1 = 2$) puis d'un second lancer ($n_2 = 2$), d'après la règle de comptage, l'expérience a quatre résultats possibles différents ($2 \times 2 = 4$). Comme nous l'avons montré, $S = \{(F, F), (P, F), (F, P), (P, P)\}$. Le nombre de résultats possibles dans une expérience impliquant six lancers de pièces est égal à 64 ($2 \times 2 \times 2 \times 2 \times 2 \times 2 = 64$).

Un **diagramme arborescent** est une représentation graphique utile pour visualiser une expérience à plusieurs étapes. La figure 4.2 représente un diagramme arborescent pour le lancer de deux pièces. Les étapes successives sont représentées de gauche à droite sur le graphique. L'étape 1 correspond au lancer de la première pièce et l'étape 2 au lancer de la seconde pièce. À chaque étape, les deux résultats possibles sont pile ou face. Notez que pour chaque résultat possible de l'étape 1, deux branches représentent les deux résultats possibles de l'étape 2. Finalement, chacun des points qui terminent le graphique correspond à un résultat possible de l'expérience. Chaque chemin à travers les branches de l'arbre, depuis le nœud le plus à gauche jusqu'à un des nœuds à droite de l'arbre, correspond à une séquence unique de résultat.

Sans diagramme arborescent, on peut penser qu'il y a seulement trois résultats à l'expérience consistant aux deux lancers d'une pièce : 0 face, 1 face et 2 faces.

Voyons, à présent, comment utiliser la règle de comptage pour des expériences à plusieurs étapes dans l'analyse du projet d'expansion de la capacité de production de la société Kentucky Power & Light (KP&L). KP&L étudie un projet d'expansion de

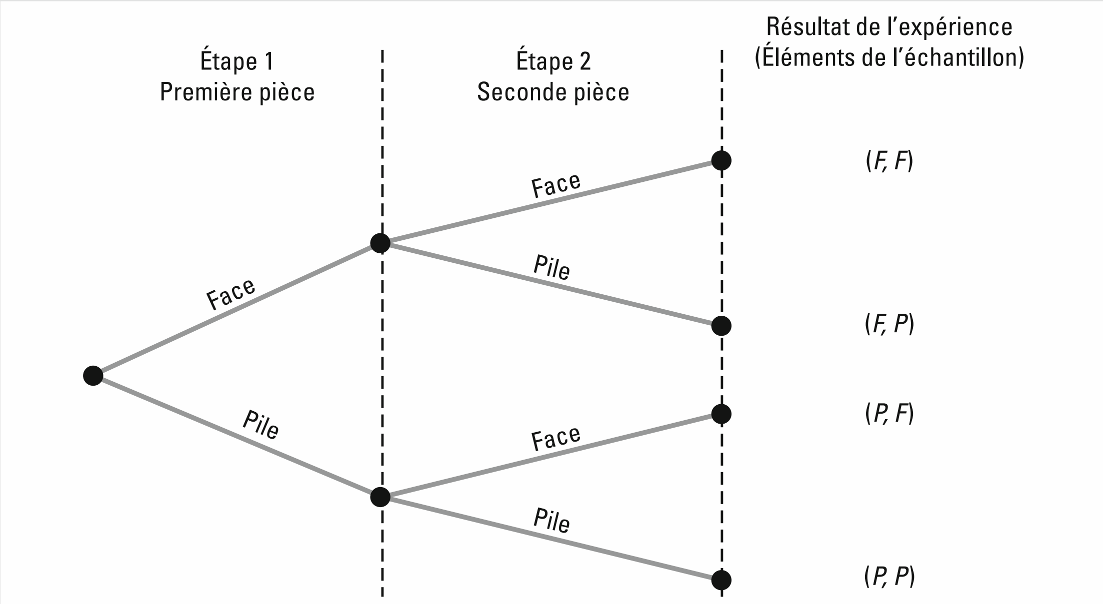

Figure 4.2 *Diagramme arborescent du lancer de deux pièces*

la capacité de production de l'une de ses usines dans le Nord du Kentucky. Le projet comporte deux phases successives : phase 1, conception ; phase 2, construction. Bien que chaque phase soit programmée et contrôlée autant que possible, la direction ne peut pas prédire à l'avance le temps exact nécessaire à la réalisation de chacune des phases du projet. Une analyse des projets de construction similaires a révélé que la phase de conception pouvait durer 2, 3 ou 4 mois et la phase de construction 6, 7 ou 8 mois. De plus, à cause de la nécessité impérative de modifier l'installation électrique, la direction a fixé à 10 mois maximum la durée de réalisation du projet entier.

Tableau 4.1 *Liste des résultats possibles de l'expérience (éléments de l'échantillon) pour le problème de la société KP&L*

Temps de réalisation (en mois)			
Phase 1 (Conception)	**Phase 2 (Construction)**	**Notation des résultats possibles**	**Temps de la réalisation du projet entier (en mois)**
2	6	(2, 6)	8
2	7	(2, 7)	9
2	8	(2, 8)	10
3	6	(3, 6)	9
3	7	(3, 7)	10
3	8	(3, 8)	11
4	6	(4, 6)	10
4	7	(4, 7)	11
4	8	(4, 8)	12

Puisque trois durées différentes sont possibles pour chaque phase, en appliquant la règle de comptage pour des expériences à plusieurs étapes, on obtient un total de 9 résultats possibles de l'expérience ($3 \times 3 = 9$). Pour décrire ces résultats, on utilise une notation à deux chiffres ; par exemple, (2, 6) indique que la phase de conception est achevée en 2 mois et la phase de construction en 6 mois. Avec ce résultat, le projet entier est réalisé en 8 mois ($2 + 6 = 8$). Le tableau 4.1 résume les neuf résultats possibles du problème KP&L. La figure 4.3 représente le diagramme arborescent de l'expérience.

La règle de comptage et l'arbre permettent au responsable du projet d'identifier les résultats possibles et de déterminer les temps de réalisation envisageables. À partir des informations contenues dans la figure 4.3, on peut conclure que la durée d'achèvement du projet varie entre 8 et 12 mois, six des neuf résultats possibles de l'expérience fournissant le temps de réalisation souhaité, d'au plus 10 mois. Bien qu'il soit utile d'identifier les résultats de l'expérience, il est nécessaire de déterminer les probabilités de chaque résultat possible avant d'estimer la probabilité que le projet soit achevé en 10 mois au plus.

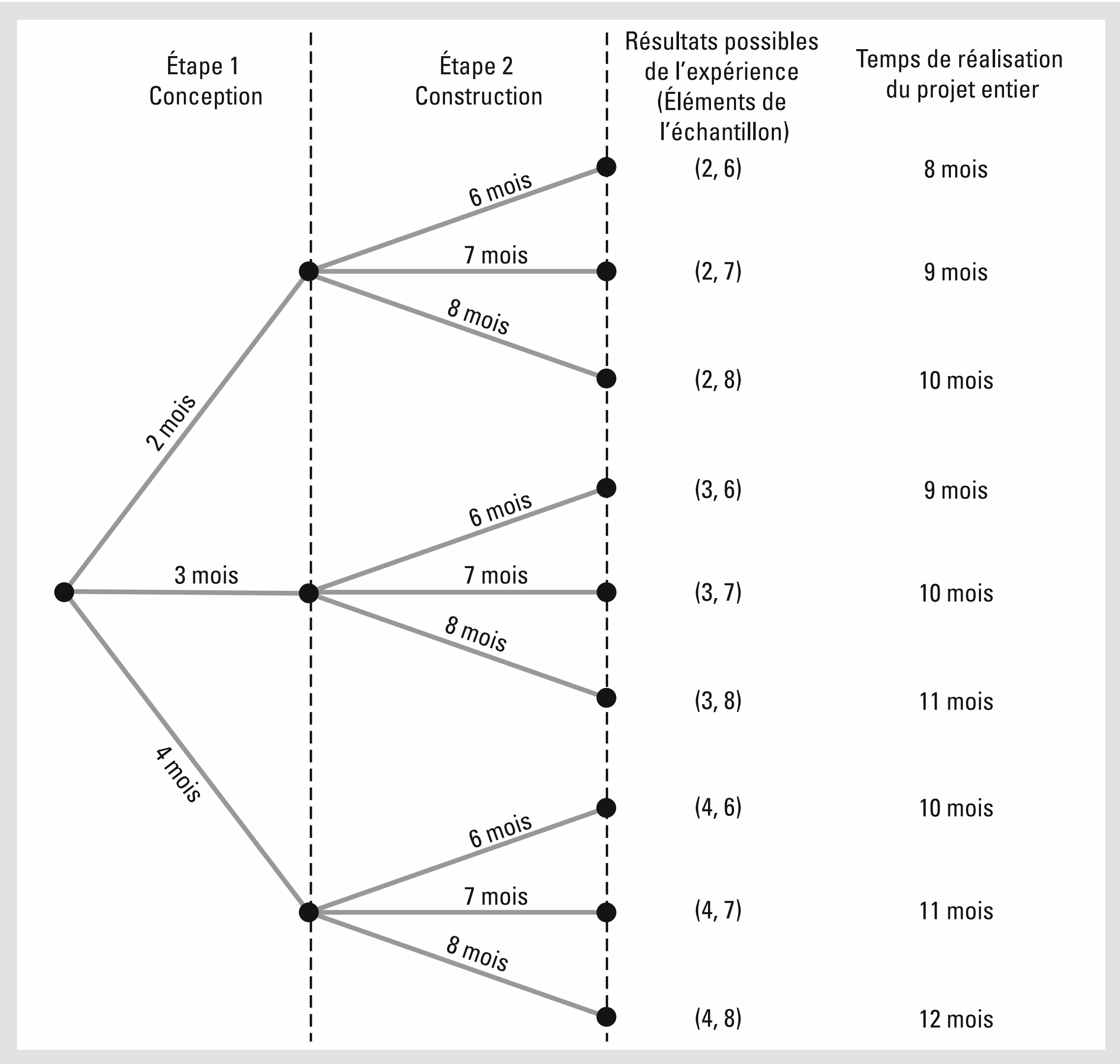

Figure 4.3 ***Diagramme arborescent du projet de la société KP&L***

Combinaisons. Une seconde règle de comptage qui est souvent utile, permet de compter le nombre de résultats obtenus en sélectionnant *n* objets parmi un ensemble (généralement plus large) de *N* objets. Il s'agit de la règle de comptage par combinaisons.

► **Règle de comptage par combinaisons**

Le nombre de combinaisons obtenues avec *n* objets sélectionnés parmi *N* est :

$$C_n^N = \binom{N}{n} = \frac{N!}{n!(N-n)!} \tag{4.1}$$

où $N! = N(N-1)(N-2)\ldots(2)(1)$

$n! = n(n-1)(n-2)\ldots(2)(1)$

et par définition $0! = 1$

La notation ! signifie *factorielle* ; par exemple, factorielle 5 est égale à $5! = (5)(4)(3)(2)(1) = 120$.

Pour illustrer la règle de comptage par combinaisons, considérons une procédure de contrôle de la qualité, dans laquelle un inspecteur sélectionne aléatoirement deux pièces sur cinq pour tester leur qualité. Dans un groupe de cinq pièces, combien de combinaisons de deux pièces peuvent être sélectionnées ? La règle de comptage définie par l'équation (4.1) montre qu'avec $N = 5$ et $n = 2$, nous avons

$$C_2^5 = \binom{5}{2} = \frac{5!}{2!(5-2)!} = \frac{(5)(4)(3)(2)(1)}{(2)(1)(3)(2)(1)} = \frac{120}{12} = 10$$

Dans un échantillon issu d'une population de taille finie *N*, la règle de comptage par combinaisons permet de déterminer le nombre d'échantillons différents de taille *n* qui peuvent être sélectionnés.

Ainsi, dix résultats sont possibles pour l'expérience de sélection aléatoire de deux pièces parmi cinq. Si on nomme les cinq pièces A, B, C, D et E, les dix combinaisons ou résultats possibles de l'expérience sont AB, AC, AD, AE, BC, BD, BE, CD, CE et DE.

Considérons un autre exemple : le système de loterie de Floride utilise une sélection aléatoire de six numéros parmi 53 pour déterminer le gagnant chaque semaine. La règle de comptage par combinaisons définie par l'équation (4.1) permet de déterminer le nombre de façon de sélectionner 6 nombres entiers parmi 53.

$$\binom{53}{6} = \frac{53!}{6!(53-6)!} = \frac{(53)(52)(51)(50)(49)(48)}{(6)(5)(4)(3)(2)(1)} = 22\,957\,480$$

Selon la règle de comptage par combinaisons, près de 23 millions de combinaisons sont possibles à la loterie. Un individu qui achète un billet de loterie a 1 chance sur 22 957 480 de gagner.

La règle de comptage par combinaisons prouve que les chances de gagner à la loterie sont très minces.

Permutations. Une troisième règle de comptage, parfois utile, est la règle de comptage par permutations. Elle nous permet de calculer le nombre de résultats possibles lorsque n objets sont sélectionnés parmi N, en tenant compte de l'ordre de tirage. Les mêmes n objets tirés dans un ordre différent constituent un autre résultat de l'expérience.

► **Règle de comptage par permutations**

Le nombre de permutations de n objets sélectionnés parmi N est égal à

$$P_n^N = n!\binom{N}{n} = \frac{N!}{(N-n)!} \tag{4.2}$$

La règle de comptage par permutations est proche de celle par combinaisons ; cependant, une expérience aura toujours plus de permutations que de combinaisons pour un même nombre d'objets sélectionnés. Ceci tient au fait que pour chaque tirage de n objets, il y a $n!$ façons différentes de les ordonner.

Considérons de nouveau l'exemple du processus de contrôle de la qualité, dans lequel un inspecteur sélectionne deux pièces parmi cinq. Combien de permutations peuvent être effectuées ? La règle de comptage fournie par l'équation (4.2) montre qu'avec $N = 5$ et $n = 2$,

$$P_2^5 = \frac{5!}{(5-2)!} = \frac{5!}{3!} = \frac{(5)(4)(3)(2)(1)}{(3)(2)(1)} = (5)(4) = 20$$

Ainsi, 20 résultats sont possibles pour cette expérience consistant à sélectionner aléatoirement deux pièces parmi cinq, lorsque l'ordre de tirage est pris en compte. Si on nomme les pièces A, B, C, D et E, les 20 permutations sont AB, BA, AC, CA, AD, DA, AE, EA, BC, CB, BD, DB, BE, EB, CD, DC, CE, EC, DE et ED.

4.1.2 Détermination des probabilités

Voyons maintenant comment déterminer les probabilités des résultats possibles de l'expérience. Les trois approches les plus fréquemment utilisées sont la méthode classique, la méthode de la fréquence relative et la méthode subjective. Quelle que soit la méthode utilisée, les probabilités doivent satisfaire deux **conditions de base**.

► **Conditions de base pour déterminer des probabilités**

1. La probabilité associée à chaque résultat possible de l'expérience doit être comprise entre 0 et 1. Si l'on note E_i le i^e résultat possible de l'expérience et $P(E_i)$ sa probabilité, on a

$$0 \leq P(E_i) \leq 1 \text{ pour tout } i \tag{4.3}$$

2. La somme des probabilités de tous les résultats possibles de l'expérience doit être égale à 1. Pour n résultats possibles, on a

$$P(E_1) + P(E_2) + \ldots + P(E_n) = 1 \tag{4.4}$$

La **méthode classique** de détermination des probabilités est appropriée lorsque les résultats possibles de l'expérience sont équiprobables. Si n résultats sont possibles, une probabilité de $1/n$ est associée à chaque résultat. Cette approche respecte automatiquement les deux conditions de base des probabilités.

Par exemple, considérons le lancer d'une pièce de monnaie équilibrée. Les deux résultats possibles de l'expérience – pile ou face – sont équiprobables. Puisque l'un des deux résultats équiprobables est face, la probabilité d'observer face est ½ ou 0,50. De même, la probabilité d'observer pile est également ½ ou 0,50.

Considérons l'exemple du lancer de dé. Il est raisonnable de penser que les six résultats possibles sont équiprobables et donc à chaque résultat est associée une probabilité de 1/6. Si $P(1)$ correspond à la probabilité que le 1 apparaisse, alors $P(1) = 1/6$. De même, $P(2) = 1/6$, $P(3) = 1/6$, $P(4) = 1/6$, $P(5) = 1/6$ et $P(6) = 1/6$. Notez que les conditions (4.3) et (4.4) sont satisfaites puisque chacune des probabilités est supérieure ou égale à zéro et que leur somme est égale à 1.

La **méthode de la fréquence relative** de détermination des probabilités est appropriée lorsque les données disponibles estiment le nombre de fois où le résultat se produira si l'expérience est répétée un grand nombre de fois. Considérons l'exemple d'une étude des temps d'attente dans le service de radiologie d'un hôpital local. Le nombre de patients ayant rendez-vous à 9 heures a été collecté pendant 20 jours consécutifs. Les résultats suivants ont été obtenus :

Nombre de patients	Nombre de jours au cours desquels le résultat se produit
0	2
1	5
2	6
3	4
4	3
	Total 20

Ces données montrent que sur 2 des 20 jours, aucun patient n'avait rendez-vous ; sur 5 des 20 jours, un patient avait rendez-vous, etc. En utilisant la méthode de la fréquence relative, on peut assigner la probabilité de $2/20 = 0{,}10$ au résultat « aucun patient n'a de rendez-vous », de $5/20 = 0{,}25$ au résultat « un patient a un rendez-vous », $6/20 = 0{,}30$ au résultat « deux patients ont un rendez-vous », $4/20 = 0{,}20$ au résultat « trois patients ont un rendez-vous » et $3/20 = 0{,}15$ au résultat « quatre patients ont un rendez-vous ». Comme avec la méthode classique, les deux conditions de base (4.3) et (4.4) sont automatiquement satisfaites lorsque la méthode de la fréquence relative est utilisée.

La **méthode subjective** de détermination des probabilités est appropriée lorsqu'il est irréaliste de supposer que les résultats de l'expérience sont équiprobables et lorsque peu de données sont disponibles. Lorsque la méthode subjective est utilisée pour assigner des probabilités aux résultats d'une expérience, nous devons utiliser toutes les informations disponibles, comme notre expérience ou notre intuition. Après avoir pris en compte toutes les informations disponibles, nous spécifions une probabilité qui traduit notre *degré de croyance* (sur une échelle allant de 0 à 1) quant à la réalisation du résultat. Puisque les probabilités subjectives traduisent les croyances d'une personne, elles sont personnelles. En utilisant la méthode subjective, il est vraisemblable que différentes personnes associent des probabilités différentes à un même résultat de l'expérience.

Lorsqu'on utilise la méthode subjective de détermination des probabilités, une attention particulière doit être apportée au respect des conditions de base (4.3) et (4.4). Quelles que soient les croyances d'une personne, la probabilité associée à chaque résultat de l'expérience doit être comprise entre 0 et 1, et la somme des probabilités de tous les résultats possibles de l'expérience doit être égale à 1.

Considérons l'exemple d'une offre d'achat d'une maison, faite par Tom et Judy Elsbernd. Deux résultats sont possibles :

E_1 = leur offre est acceptée
E_2 = leur offre est refusée

Judy pense que la probabilité que leur offre soit acceptée est égale à 0,8 ; ainsi, pour Judy, $P(E_1) = 0{,}8$ et $P(E_2) = 0{,}2$. Tom, cependant, croit que la probabilité que leur offre soit acceptée est de 0,6 ; ainsi, pour Tom, $P(E_1) = 0{,}6$ et $P(E_2) = 0{,}4$. Notez que les croyances de Tom reflètent le fait qu'il est plus pessimiste que Judy, quant à l'acceptation de leur offre.

À la fois Judy et Tom ont déterminé des probabilités qui satisfont les deux conditions de base. Le fait que leurs croyances soient différentes illustre la nature personnelle de la méthode subjective.

Même dans des situations commerciales, où les méthodes classique et de la fréquence relative peuvent être appliquées, les responsables peuvent vouloir obtenir des estimations subjectives des probabilités. Dans de tels cas, les meilleures estimations des probabilités sont souvent obtenues en combinant méthode classique ou de la fréquence relative et approche subjective.

> Le théorème de Bayes (cf. section 4.5) est un moyen de combiner les probabilités a priori, déterminées subjectivement, avec les probabilités obtenues par d'autres méthodes, de manière à obtenir des probabilités révisées, dites probabilités a posteriori.

4.1.3 Les probabilités pour le projet de la société KP&L

Nous poursuivons l'analyse du projet de la société KP&L en développant les probabilités pour chacun des neuf résultats possibles de l'expérience, énumérés dans le tableau 4.1. En se basant sur son expérience, la direction a conclu que les différents résultats possibles de l'expérience n'étaient pas équiprobables. Par conséquent, la méthode classique de détermination des probabilités ne peut pas être utilisée. La direction a alors décidé de mener une étude sur les temps de réalisation de projets similaires effectués par KP&L, au cours des trois années précédentes. Les résultats de l'étude de 40 projets similaires sont résumés dans le tableau 4.2.

Tableau 4.2 ***Résultats concernant la réalisation de 40 projets de la société KP&L***

Temps de réalisation (en mois)			
Phase 1 Conception	**Phase 2 Construction**	**Éléments de l'échantillon**	**Nombre d'anciens projets ayant ces temps de réalisation**
2	6	(2, 6)	6
2	7	(2, 7)	6
2	8	(2, 8)	2
3	6	(3, 6)	4
3	7	(3, 7)	8
3	8	(3, 8)	2
4	6	(4, 6)	2
4	7	(4, 7)	4
4	8	(4, 8)	6
			Total 40

Après avoir examiné les résultats de cette étude, la direction a décidé d'utiliser la méthode de la fréquence relative pour déterminer les probabilités. La direction aurait pu estimer de façon subjective les probabilités mais elle considère le projet actuel assez semblable aux 40 projets antérieurs. La méthode de la fréquence relative a donc été jugée la plus appropriée.

En utilisant les données du tableau 4.2 pour calculer les probabilités, on note que le résultat (2, 6) – phase 1 achevée en 2 mois et phase 2 achevée en 6 mois – survient 6 fois parmi les 40 projets considérés. Nous utilisons la méthode de la fréquence relative pour associer une probabilité de 6/40 = 0,15 à ce résultat. De même, le résultat (2, 7)

Tableau 4.3 ***Détermination des probabilités pour le problème de la société KP&L basée sur la méthode de la fréquence relative***

Éléments de l'échantillon	Temps de réalisation du projet	Probabilité des éléments de l'échantillon
(2, 6)	8 mois	$P(2, 6) = 6/40 = 0{,}15$
(2, 7)	9 mois	$P(2, 7) = 6/40 = 0{,}15$
(2, 8)	10 mois	$P(2, 8) = 2/40 = 0{,}05$
(3, 6)	9 mois	$P(3, 6) = 4/40 = 0{,}10$
(3, 7)	10 mois	$P(3, 7) = 8/40 = 0{,}20$
(3, 8)	11 mois	$P(3, 8) = 2/40 = 0{,}05$
(4, 6)	10 mois	$P(4, 6) = 2/40 = 0{,}05$
(4, 7)	11 mois	$P(4, 7) = 4/40 = 0{,}10$
(4, 8)	12 mois	$P(4, 8) = 6/40 = 0{,}15$
		Total 1,00

survient 6 fois parmi les 40 projets, soit avec une probabilité de 6/40 = 0,15. En poursuivant ce raisonnement, nous obtenons les probabilités pour tous les points d'échantillon du projet de la société KP&L, regroupées dans le tableau 4.3. Notez que $P(2,6)$ correspond à la probabilité du point d'échantillon (2, 6), $P(2,7)$ correspond à la probabilité du point d'échantillon (2, 7) et ainsi de suite.

REMARQUES

1. En statistiques, la notion d'expérience est quelque peu différente de celle qui prévaut en sciences physiques. En sciences physiques, une expérience est généralement menée dans un laboratoire ou dans un environnement contrôlé, dans le but d'en découvrir les causes et les effets. Les résultats des expériences statistiques sont déterminés par une probabilité. Même si l'expérience est répétée exactement de la même façon, un résultat totalement différent peut survenir. À cause de cette influence des probabilités sur le résultat, les expériences statistiques sont parfois appelées expériences aléatoires.
2. Lors du tirage d'un échantillon aléatoire sans remise à partir d'une population de taille N, la règle de comptage par combinaisons est utilisée pour déterminer le nombre d'échantillons différents de taille n qui peuvent être sélectionnés.

EXERCICES

Méthode

1. Une expérience en trois étapes a trois résultats possibles à la première étape, deux résultats possibles à la seconde étape et quatre résultats possibles à la troisième étape. Combien de résultats possibles existe-il pour l'expérience considérée dans son ensemble ?
2. De combien de façons peut-on sélectionner trois éléments parmi six ? Utiliser les lettres A, B, C, D, E et F pour identifier les éléments et énumérer chaque combinaison possible de trois éléments.
3. Combien de permutations de trois éléments peut-on faire avec six éléments ? Utiliser les lettres A, B, C, D, E et F pour identifier les éléments et énumérer chaque permutation comprenant les éléments B, D et F.
4. Considérer l'expérience qui consiste à lancer trois fois une pièce de monnaie.
   a) Construire le diagramme arborescent de l'expérience.
   b) Énumérer les résultats possibles de l'expérience.
   c) Quelle est la probabilité de chaque résultat possible ?
5. Supposez qu'une expérience a cinq résultats possibles équiprobables : E_1, E_2, E_3, E_4, E_5. Déterminer les probabilités de chaque résultat et montrer que les conditions (4.3) et (4.4) sont vérifiées. Quelle méthode avez-vous utilisée ?

6. Une expérience qui a trois résultats possibles, a été répétée 50 fois : E_1 est apparu 20 fois, E_2 13 fois et E_3 17 fois. Déterminer la probabilité de chacun des résultats. Quelle méthode avez-vous utilisée ?

7. Un responsable a subjectivement attribué les probabilités suivantes aux quatre résultats possibles d'une expérience : $P(E_1) = 0{,}10$, $P(E_2) = 0{,}15$, $P(E_3) = 0{,}40$ et $P(E_4) = 0{,}20$. L'attribution de ces probabilités est-elle correcte ? Expliquer.

Applications

8. Dans la ville de Milford, les propositions pour modifier la répartition des zones sont soumises à un processus en deux étapes : un examen par la commission d'urbanisme et un examen par le conseil municipal qui prend la décision finale. À l'étape 1, la commission d'urbanisme examine la demande de changement de la répartition des zones et émet un avis, positif ou négatif, quant à ce changement. À l'étape 2, le conseil municipal examine l'avis de la commission d'urbanisme puis vote pour approuver ou désapprouver le changement. Supposez que le promoteur d'un complexe immobilier fait une demande de modification des zones. Considérer le processus de décision comme une expérience à deux étapes.

- **a)** Combien y a t-il d'éléments d'échantillon dans cette expérience ? Énumérez-les.
- **b)** Construire un diagramme arborescent pour cette expérience.

9. L'échantillonnage aléatoire simple utilise un échantillon de taille n, issu d'une population de taille N, pour obtenir des données permettant d'inférer sur les caractéristiques de la population. Supposez que nous ayons une population de 50 comptes bancaires et que nous voulions faire de l'inférence sur cette population à partir d'un échantillon de quatre comptes. Combien d'échantillons aléatoires différents peut-on obtenir ?

10. Beaucoup d'étudiants ont contracté des dettes durant leurs études. Le tableau suivant indique le pourcentage d'étudiants endettés et le montant moyen de leur dette parmi les étudiants de quatre universités et de quatre écoles des beaux-arts (*U.S. News and World Report, America's Best Colleges*, 2008).

Université	% d'étudiants endettés	Montant ($)	École	% d'étudiants endettés	Montant ($)
Pace	72	32 980	Wartburg	83	28 758
Iowa State	69	32 130	Morehouse	94	27 000
Massachusetts	55	11 227	Wellesley	55	10 206
SUNY-Albany	64	11 856	Wofford	49	11 012

- **a)** Si nous choisissons aléatoirement un étudiant de Morehouse College, quelle est la probabilité qu'il soit endetté ?
- **b)** Si nous choisissons aléatoirement une de ces huit institutions dans le cadre d'une étude sur les prêts aux étudiants, quelle est la probabilité que l'institution choisie ait plus de 60 % des étudiants endettés ?
- **c)** Si nous choisissons aléatoirement une de ces huit institutions dans le cadre d'une étude sur les prêts aux étudiants, quelle est la probabilité que dans cette institution, les étudiants endettés aient une dette moyenne de plus de 30 000 dollars ?

d) Quelle est la probabilité qu'un étudiant de l'université Pace ne soit pas endetté ?

e) Parmi les étudiants de l'université de Pace endettés, le montant moyen de la dette est de 32 980 dollars. En considérant tous les étudiants de l'université de Pace, quelle est la dette moyenne par étudiant ?

11. L'enquête nationale sur l'utilisation d'équipements de protection (NOPUS) a été menée pour fournir des données probabilistes sur le port du casque par les motards aux États-Unis. L'enquête fut menée en envoyant des observateurs sur des sites routiers sélectionnés aléatoirement où ils collectèrent des données sur le nombre de motards portant un casque, ainsi que sur le nombre de motards portant un casque conforme aux réglementations du Département des Transports (site de l'administration nationale de sécurité routière, 7 janvier 2010). Un échantillon de données représentatif de l'enquête NOPUS est fourni ci-dessous.

	Type de casque	
Région	**Conforme à la réglementation**	**Non-conforme à la réglementation**
Nord-Est	96	62
Centre Ouest	86	43
Sud	92	49
Ouest	76	16
Total	350	170

a) Utiliser les données d'échantillon pour estimer la probabilité qu'un motard porte un casque conforme à la réglementation.

b) La probabilité qu'un motard porte un casque conforme à la réglementation cinq ans plus tôt était de 0,48 et l'année dernière, cette probabilité était de 0,63. Est-ce que la Sécurité Routière peut être satisfaite des résultats de cette dernière enquête ?

c) Quelle est la probabilité que les motards portent des casques conformes à la réglementation par région ? Quelle région a la plus forte probabilité que les motards portent des casques conformes à la réglementation ?

12. La loterie Powerball se déroule deux fois par semaine dans 31 États américains, les îles Vierges et le district de Columbia. Pour participer à la loterie Powerball, un individu doit acheter un ticket à 2 dollars, choisir cinq numéros compris entre 1 et 59 et ensuite le numéro Powerball compris entre 1 et 35. Pour déterminer les numéros gagnants, cinq boules blanches sont tirées au hasard parmi 59 boules blanches numérotées de 1 à 59 et une boule rouge est tirée parmi 35 boules rouges numérotées de 1 à 35. Pour gagner la cagnotte, les numéros d'un participant doivent correspondre aux numéros des cinq boules blanches tirées au hasard, quel que soit l'ordre de tirage, et au numéro de la boule rouge. Les nombres 5-16-22-23-29 et le nombre Powerball 6 ont donné lieu au jackpot historique de 580 millions de dollars (site Internet de Powerball, 29 novembre 2012).

a) Combien de résultats sont possibles ? Astuce : Considérez une expérience en deux étapes : sélectionner les numéros des 5 boules blanches puis le numéro d'une boule rouge.

b) Quelle est la probabilité de gagner à la loterie Powerball ?

13. Une société qui produit du dentifrice, étudie cinq emballages différents. En supposant qu'un emballage particulier a autant de chance d'être choisi par un consommateur qu'un autre, quelle probabilité attribueriez-vous au choix de chaque emballage ? Lors d'une expérience réelle, on a demandé à 100 clients de choisir l'emballage qu'ils préfèrent. Les résultats suivants ont été obtenus. Est-ce que les données confirment l'hypothèse selon laquelle un emballage a autant de chance d'être choisi qu'un autre ? Expliquer.

Emballage	Nombre de fois choisi
1	5
2	15
3	30
4	40
5	10

4.2 ÉVÉNEMENTS ET PROBABILITÉS

Dans l'introduction de ce chapitre, nous avons utilisé le mot « événement » dans le sens courant du terme. Ensuite, dans la section 4.1, nous avons introduit le concept d'expérience et de résultats d'expérience, appelés éléments de l'échantillon. Les éléments de l'échantillon et les événements constituent les bases de l'analyse probabiliste. Nous devons maintenant introduire la définition formelle d'un **événement** lié aux éléments de l'échantillon. Cela constitue la base de la détermination de la probabilité d'un événement.

► **Événement**
Un *événement* est un ensemble d'éléments d'échantillon.

Par exemple, revenons au problème de la société KP&L et supposons que le responsable du projet soit intéressé par l'événement correspondant à la réalisation du projet en 10 mois, maximum. En se référant au tableau 4.3, on s'aperçoit que six points d'échantillon – (2, 6), (2, 7), (2, 8), (3, 6), (3, 7) et (4, 6) – offrent un temps de réalisation inférieur ou égal à 10 mois. Soit C l'événement « le projet est réalisé en, au plus, 10 mois » ; on écrit

$$C = \{(2, 6), (2, 7), (2, 8), (3, 6), (3, 7), (4, 6)\}$$

L'événement C se produit si le résultat de l'expérience correspond à l'un de ces six points d'échantillon.

D'autres événements peuvent intéresser la direction de la société KP&L, comme par exemple :

L = « le projet est réalisé en moins de 10 mois »
M = « le projet est réalisé en plus de 10 mois »

En utilisant les informations contenues dans le tableau 4.3, on s'aperçoit que ces événements sont constitués des points d'échantillon suivants :

$$L = \{(2, 6), (2, 7), (3, 6)\}$$
$$M = \{(3, 8), (4, 7), (4, 8)\}$$

De nombreux autres événements peuvent être définis pour le problème de la société KP&L, mais dans tous les cas, l'événement est identifié par un ensemble de points d'échantillon de l'expérience.

Étant données les probabilités des points d'échantillon (cf. tableau 4.3), on peut utiliser la définition suivante pour calculer la probabilité de n'importe quel événement lié au problème de la société KP&L.

► **Probabilité d'un événement**
La probabilité d'un événement est égale à la somme des probabilités des points d'échantillon qui constituent cet événement.

Selon cette définition, on calcule la probabilité d'un événement particulier en additionnant les probabilités des points d'échantillon (les résultats possibles de l'expérience) qui constituent l'événement. Nous pouvons maintenant calculer la probabilité que le projet soit réalisé en 10 mois, maximum. Puisque cet événement est donné par $C = \{(2, 6), (2, 7), (2, 8), (3, 6), (3, 7), (4, 6)\}$, la probabilité ($P$) de l'événement C est égale à

$$P(C) = P(2,6) + P(2,7) + P(2,8) + P(3,6) + P(3,7) + P(4,6)$$

En se référant aux probabilités des points d'échantillon fournies dans le tableau 4.3, nous avons

$$P(C) = 0,15 + 0,15 + 0,05 + 0,10 + 0,20 + 0,05 = 0,70$$

De même, puisque l'événement « le projet est réalisé en moins de 10 mois » correspond à $L = \{(2, 6), (2, 7), (3, 6)\}$, la probabilité de cet événement est égale à

$$P(L) = P(2,6) + P(2,7) + P(3,6) = 0,15 + 0,15 + 0,10 = 0,40$$

Pour finir, l'événement « le projet est réalisé en plus de 10 mois » est défini par $M = \{(3, 8), (4, 7), (4, 8)\}$ et donc

$$P(M) = P(3,8) + P(4,7) + P(4,8) = 0,05 + 0,10 + 0,15 = 0,30$$

En utilisant ces probabilités, nous sommes maintenant en mesure de dire à la direction de KP&L qu'il y a une probabilité de 0,70 que le projet soit réalisé en, au plus, 10 mois ; une probabilité de 0,40 que le projet soit réalisé en moins de 10 mois et une probabilité de 0,30 que le projet soit réalisé en plus de 10 mois. Cette procédure de calcul de la probabilité d'un événement peut être répétée pour n'importe quel autre événement qui intéresse la direction de KP&L.

Lorsque les éléments d'échantillon d'une expérience sont identifiés, ainsi que leurs probabilités, on peut utiliser la définition précédente pour calculer la probabilité d'un événement. Cependant, dans de nombreuses expériences, le nombre de points d'échantillon est grand, rendant l'identification de ces éléments d'échantillon et de leur probabilité extrêmement difficile, voire impossible. Dans la suite de ce chapitre, nous présenterons quelques relations probabilistes fondamentales qui permettent de calculer la probabilité d'un événement sans connaître la probabilité de chaque élément d'échantillon.

REMARQUES

1. L'espace-échantillon, S, est un événement. Puisqu'il contient tous les résultats possibles de l'expérience, il a une probabilité égale à 1 ; c'est-à-dire, $P(S) = 1$.
2. Lorsque la méthode classique est utilisée pour déterminer les probabilités, on suppose que les résultats possibles de l'expérience sont équiprobables. Dans ce cas, la probabilité d'un événement peut être calculée en comptant le nombre de résultats possibles qui forment cet événement et en divisant ce chiffre par le nombre total de résultats possibles.

EXERCICES

Méthode

14. Une expérience a quatre résultats possibles équiprobables : E_1, E_2, E_3 et E_4.

a) Quelle est la probabilité que E_2 se réalise ?

b) Quelle est la probabilité que deux des résultats possibles se réalisent (par exemple, E_1 ou E_3) ?

c) Quelle est la probabilité que trois des résultats se réalisent (par exemple, E_1ou E_2 ou E_4) ?

15. Considérez l'expérience qui consiste à choisir une carte dans un jeu qui en compte 52. Chaque carte correspond à un élément de l'échantillon avec une probabilité de 1/52.

a) Énumérer les éléments de l'échantillon qui constituent l'événement « un as a été tiré ».

b) Énumérer les éléments de l'échantillon qui constituent l'événement « un trèfle a été tiré ».

c) Énumérer les éléments de l'échantillon qui constituent l'événement « une figure (valet, dame ou roi) a été tirée ».

d) Trouver les probabilités associées à chacun des événements cités dans les questions (a), (b) et (c).

16. Considérez l'expérience qui consiste à lancer une paire de dés. Supposez que nous nous intéressions à la somme de la valeur des deux dés.

a) Combien d'éléments de l'échantillon sont possibles ? (Astuce : Utilisez la règle de comptage pour des expériences à plusieurs étapes).

b) Énumérer les éléments de l'échantillon.

c) Quelle est la probabilité d'obtenir la valeur 7 ?

d) Quelle est la probabilité d'obtenir une valeur supérieure ou égale à 9 ?

e) Puisque chaque lancer a six possibilités de donner une valeur paire (2, 4, 6, 8, 10 et 12) et seulement cinq possibilités de donner une valeur impaire (3, 5, 7, 9 et 11), on devrait obtenir plus souvent une valeur paire qu'une valeur impaire. Êtes-vous d'accord avec ce raisonnement ? Expliquer.

f) Quelle méthode avez-vous utilisée pour déterminer les probabilités demandées ci-dessus ?

Applications

17. Reprendre les éléments de l'échantillon relatif à l'exemple de la société KP&L et leurs probabilités, regroupés dans les tableaux 4.2 et 4.3.

a) Le budget de la phase de conception (étape 1) sera dépassé si quatre mois sont nécessaires à sa réalisation. Énumérer les éléments de l'échantillon qui constituent l'événement « le budget de la phase de conception est dépassé ».

b) Quelle est la probabilité que le budget de la phase de conception soit dépassé ?

c) Le budget de la phase de construction (étape 2) sera dépassé si huit mois sont nécessaires à sa réalisation. Énumérer les éléments de l'échantillon qui constituent l'événement « le budget de la phase de construction est dépassé ».

d) Quelle est la probabilité que le budget de la phase de construction soit dépassé ?

e) Quelle est la probabilité que le budget des deux phases soit dépassé ?

18. Le magazine *Fortune* publie une liste annuelle des 500 plus grandes sociétés américaines. Les sièges sociaux de ces 500 sociétés sont situés dans 38 États différents. Le tableau suivant indique les 8 États dans lesquels on trouve le plus grand nombre de sociétés appartenant au classement *Fortune* 500 (site Internet de *Money*/CNN, 12 mai 2012).

État	Nombre de sociétés	État	Nombre de sociétés
Californie	53	Ohio	28
Illinois	32	Pennsylvanie	23
New Jersey	21	Texas	52
New York	50	Virginie	24

Supposez qu'une des 500 sociétés soit sélectionnée de façon aléatoire dans le cadre d'une enquête de suivi.

a) Quelle est la probabilité que la société sélectionnée ait son siège en Californie ?

b) Quelle est la probabilité que la société sélectionnée ait son siège en Californie, à New York ou au Texas ?

c) Quelle est la probabilité que la société sélectionnée ait son siège dans l'un des huit États listés ci-dessus ?

19. Pensez-vous que le gouvernement protège de façon appropriée les investisseurs ? Cette question faisait partie d'une enquête en ligne sur les investisseurs de moins de 65 ans vivant aux États-Unis et en Grande-Bretagne (sondage *Financial Times*/Harris, 1er octobre 2009). Le nombre d'investisseurs vivant aux États-Unis et en Grande-Bretagne qui ont répondu Oui, Non ou Incertain à cette question, est fourni ci-dessous.

Réponse	États-Unis	Grande-Bretagne
Oui	187	197
Non	334	411
Incertain	256	213

a) Estimer la probabilité qu'un investisseur vivant aux États-Unis pense que le gouvernement ne protège pas correctement les investisseurs.

b) Estimer la probabilité qu'un investisseur vivant en Grande-Bretagne pense que le gouvernement ne protège pas correctement les investisseurs ou n'est pas sûr qu'il le fasse.

c) Pour un investisseur sélectionné aléatoirement dans ces deux pays, estimer la probabilité qu'il pense que le gouvernement ne protège pas correctement les investisseurs.

d) D'après les résultats de l'enquête, y a-t-il une grande différence d'appréciation entre les investisseurs vivant aux États-Unis et ceux vivant en Grande-Bretagne quant à la protection offerte par le gouvernement vis-à-vis des investisseurs ?

20. Junior Achievement USA et la fondation Allstate ont mené une enquête auprès des adolescents âgés de 14 à 18 ans. Il leur a été demandé à quel âge ils pensaient devenir financièrement indépendants (*USA Today*, 30 avril 2012). Les réponses fournies par 944 adolescents qui ont répondu à cette question figurent ci-dessous.

Âge d'indépendance financière	Nombre de réponses
Entre 16 et 20 ans	191
Entre 21 et 24 ans	467
Entre 25 et 27 ans	244
À partir de 28 ans	42

Supposez qu'un adolescent soit sélectionné aléatoirement au sein de la population des adolescents âgés de 14 à 18 ans.

a) Calculer la probabilité d'être financièrement indépendant pour chacune des quatre tranches d'âge.

b) Quelle est la probabilité d'être financièrement indépendant avant 25 ans ?

c) Quelle est la probabilité d'être financièrement indépendant après 24 ans ?

d) Les probabilités suggèrent-elles que les adolescents sont quelque peu irréalistes au regard de leurs attentes en matière d'âge d'indépendance financière ?

21. Des données sur les types d'accident du travail survenant aux États-Unis sont fournies ci-dessous (*The World Almanac*, 2012).

Type d'accident	Nombre d'accidents
Incident de transport	1 795
Agression et acte de violence	837
Contact avec des objets et des équipements	741
Chute	645
Exposition à des substances ou des environnements nocifs	404
Incendie et explosion	113

Supposez qu'un accident soit sélectionné aléatoirement à partir de cette population.

a) Quelle est la probabilité que l'accident soit lié à une chute ?

b) Quelle est la probabilité que l'accident soit lié à un incident de transport ?

c) Quel est le type d'accident le moins probable ? Quelle est la probabilité que ce type d'accident survienne ?

4.3 QUELQUES RELATIONS PROBABILISTES FONDAMENTALES

4.3.1 Complément d'un événement

Étant donné un événement A, le **complément** de A est défini comme l'événement composé de tous les points d'échantillon qui ne constituent pas A. Le complément de A est noté A^c. Le **diagramme de Venn**, présenté à la figure 4.4, illustre le concept de complément. Le rectangle représente l'espace-échantillon d'une expérience et donc contient tous les points d'échantillon possibles. Le cercle représente l'événement A et contient uniquement les points d'échantillon appartenant à A. La région grisée du rectangle contient tous les points d'échantillon qui n'appartiennent pas à l'événement A et donc, par définition, correspond au complément de A.

Dans toute application probabiliste, soit l'événement A, soit son complément doit se produire. Par conséquent,

$$P(A) + P(A^c) = 1$$

En réarrangeant les termes, on obtient le résultat suivant :

► **Calculer une probabilité en se servant de son complément**

$$P(A) = 1 - P(A^c) \tag{4.5}$$

L'équation (4.5) permet de calculer facilement la probabilité d'un événement A, dans la mesure où la probabilité de son complément, $P(A^c)$, est connue.

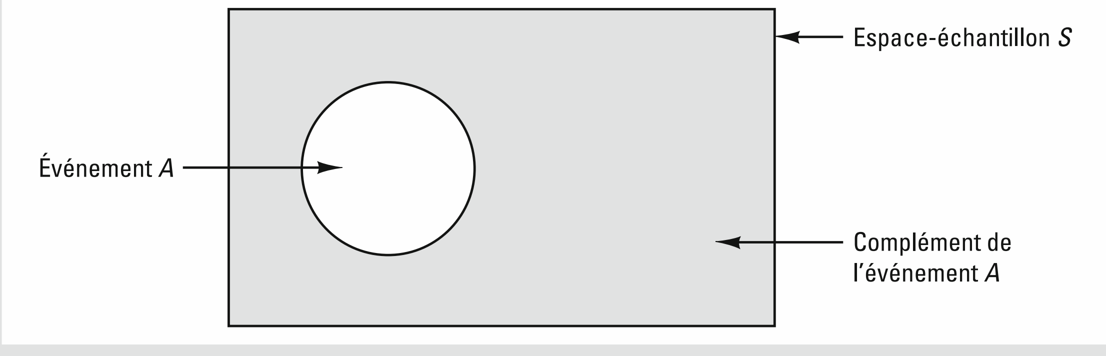

Figure 4.4 ***Complément de l'événement A***

Considérons l'exemple d'un responsable des ventes qui, après avoir examiné les rapports de vente, a constaté que 80 % des contacts établis avec de nouveaux clients ne se concluaient pas par une vente. En notant A l'événement « vente » et A^c l'événement « pas de vente », le responsable a établi que $P(A^c) = 0{,}80$. En utilisant la formule (4.5), on s'aperçoit que

$$P(A) = 1 - P(A^c) = 1 - 0{,}80 = 0{,}20$$

Nous pouvons en conclure qu'un contact établi avec un nouveau client a une probabilité de 0,20 d'aboutir à une vente.

Dans un autre exemple, un responsable des achats déclare qu'il y a une probabilité de 0,90 qu'un fournisseur livre une cargaison sans défaut. En utilisant l'événement complémentaire, on peut conclure qu'il y a une probabilité de 0,10 (1 – 0,90 = 0,10) que la cargaison contienne des pièces défectueuses.

4.3.2 La loi de la somme

La loi de la somme est utile lorsque l'on a deux événements et que l'on s'intéresse à la probabilité qu'au moins un des deux événements se produise. C'est-à-dire, avec les événements A et B, on s'intéresse à la probabilité que l'événement A ou l'événement B ou les deux se produisent.

Avant de présenter la loi de la somme, nous discuterons de deux concepts liés à la combinaison d'événements : l'union d'événements et l'intersection d'événements. Étant donnés les deux événements A et B, l'**union de A et B** est définie par :

► **Union de deux événements**
L'*union* de A et B est l'événement qui contient tous les points d'échantillon appartenant à A ou B ou les deux. L'union est notée $A \cup B$.

Le diagramme de Venn de la figure 4.5 illustre l'union des événements A et B. Notez que les deux cercles contiennent tous les points d'échantillon de l'événement A, ainsi que tous les points d'échantillon de l'événement B. Le fait que les cercles se coupent, indique que certains points d'échantillon sont contenus à la fois dans A et dans B.

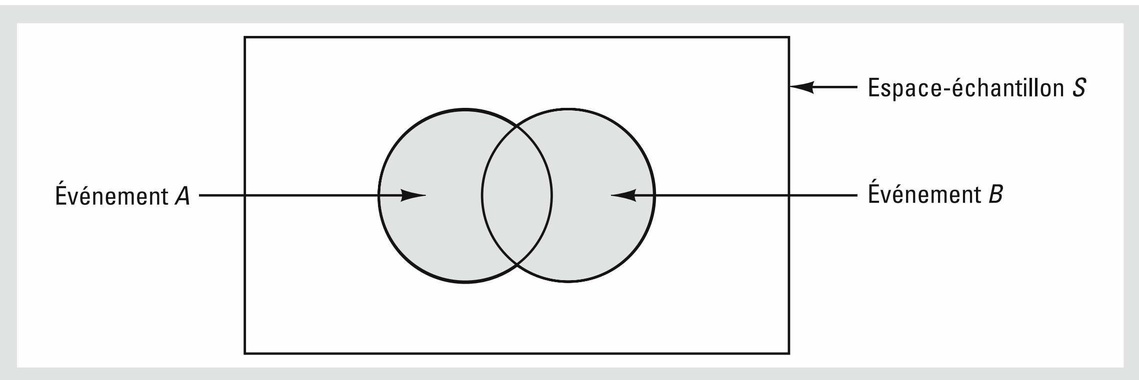

Figure 4.5 ***Union des événements A et B***

La définition de l'**intersection de *A* et *B*** est donnée ci-dessous :

► **Intersection de deux événements**
Étant donnés les événements A et B, l'*intersection* de A et de B correspond à l'événement contenant les points d'échantillon appartenant *à la fois à A et à B*. L'intersection est notée $A \cap B$.

Le diagramme de Venn présenté à la figure 4.6 illustre l'intersection de deux événements. L'intersection correspond à la partie grisée où les deux cercles se coupent ; elle contient les points d'échantillon qui appartiennent à la fois à A et à B.

Discutons maintenant de la loi de la somme. La **loi de la somme** est un moyen de calculer la probabilité de l'événement A ou B ou à la fois A et B. En d'autres termes, la loi de la somme permet de calculer la probabilité de l'union de deux événements, $A \cup B$. Sa formule est donnée ci-dessous :

► **Loi de la somme**

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad (4.6)$$

Pour comprendre de manière intuitive la loi de la somme, notez que les deux premiers termes de la loi de la somme, $P(A) + P(B)$, représentent l'ensemble des points d'échantillon contenus dans $A \cup B$. Cependant, puisque les points d'échantillon contenus dans l'intersection $A \cap B$ sont à la fois dans A et dans B, lorsque l'on calcule $P(A) + P(B)$, on compte deux fois chaque point d'échantillon contenu dans $A \cap B$. On corrige cela en soustrayant $P(A \cap B)$.

Pour illustrer la loi de la somme, considérons une petite usine d'assemblage employant 50 salariés. Chaque salarié est supposé terminer son travail en un temps donné et de façon à ce que le produit assemblé passe avec succès le test d'inspection finale. Parfois, certains travailleurs ne finissent pas leur travail à temps et/ou assemblent des pièces défectueuses. À la fin d'une période d'évaluation des performances, le responsable de la production a trouvé que 5 des 50 salariés n'avaient pas fini leur travail dans les

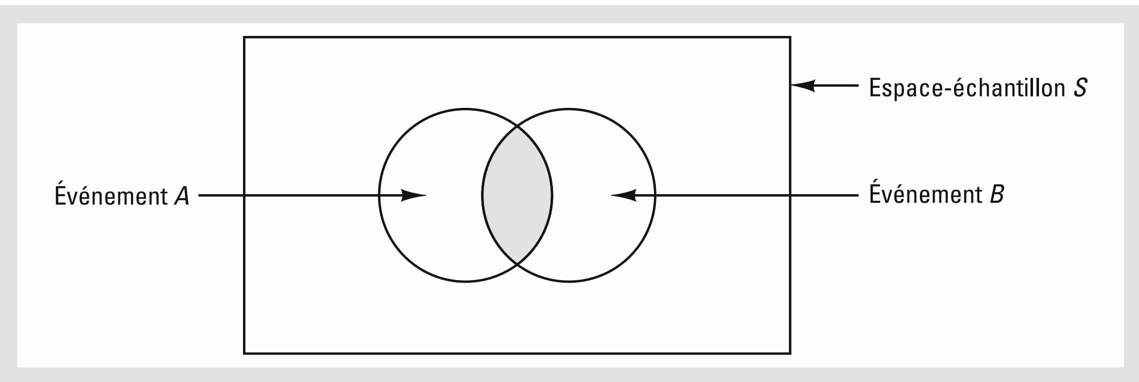

Figure 4.6 ***Intersection des événements A et B***

temps, 6 avaient assemblé des pièces défectueuses et 2 n'avaient pas fini leur travail à temps et avaient assemblé des pièces défectueuses.

Soient les événements

L = « le travail n'est pas fini à temps »
D = « le produit assemblé est défectueux »

Les fréquences relatives permettent d'obtenir les probabilités suivantes :

$$P(L) = \frac{5}{50} = 0{,}10$$

$$P(D) = \frac{6}{50} = 0{,}12$$

$$P(L \cap D) = \frac{2}{50} = 0{,}04$$

Après avoir examiné les données sur les performances, le responsable de la production a décidé d'attribuer une mauvaise évaluation à tout employé dont le travail est soit en retard, soit défectueux ; il s'intéresse donc à l'événement $L \cup D$. Quelle est la probabilité que le responsable de la production attribue une mauvaise évaluation à un employé ?

Notez que la probabilité demandée concerne l'union de deux événements. Nous voulons connaître $P(L \cup D)$. En utilisant la formule (4.6),

$$P(L \cup D) = P(L) + P(D) - P(L \cap D)$$

Connaissant la valeur des trois probabilités apparaissant dans le membre de droite de cette équation, on obtient

$$P(L \cup D) = 0{,}10 + 0{,}12 - 0{,}04 = 0{,}18$$

Ce calcul nous permet de conclure que la probabilité qu'un employé sélectionné aléatoirement reçoive une mauvaise évaluation est égale à 0,18.

Considérons un autre exemple, celui d'une étude récente menée par le responsable du personnel d'une grande société de logiciels. Il est apparu que 30 % des employés qui ont quitté l'entreprise au cours des deux années précédentes, l'ont fait parce qu'ils n'étaient pas satisfaits de leur salaire, 20 % parce qu'ils n'étaient pas satisfaits de leur fonction et 12 % parce qu'ils n'étaient satisfaits ni de leur salaire, ni de leur fonction. Quelle est la probabilité qu'un employé parti au cours des deux années précédentes, l'ait fait parce qu'il n'était pas satisfait de son salaire, de sa fonction ou des deux ?

Soient les événements

S = « l'employé est parti à cause de son salaire »
T = « l'employé est parti à cause de sa fonction »

Nous avons $P(S) = 0{,}30$, $P(T) = 0{,}20$ et $P(S \cap T) = 0{,}12$. En utilisant la loi de la somme, nous avons

$$P(S \cup T) = P(S) + P(T) - P(S \cap T) = 0{,}30 + 0{,}20 - 0{,}12 = 0{,}38.$$

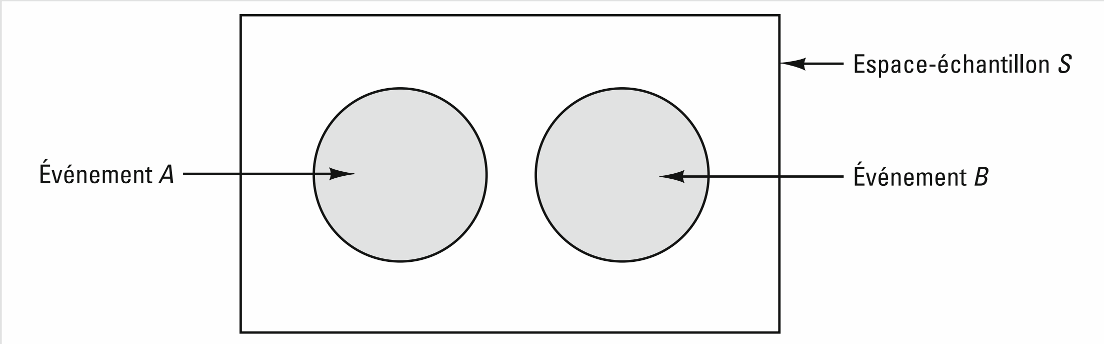

Figure 4.7 ***Événements mutuellement exclusifs***

Il y a donc une probabilité de 0,38 qu'un employé soit parti pour des raisons de salaire ou de fonction.

Avant de clore notre discussion sur la loi de la somme, considérons le cas particulier des **événements mutuellement exclusifs**.

► **Événements mutuellement exclusifs**
Deux événements sont dits mutuellement exclusifs si les événements n'ont aucun point d'échantillon en commun.

Les événements A et B sont mutuellement exclusifs si, lorsqu'un événement se produit, l'autre ne peut pas se produire. Ainsi, une condition pour que A et B soient mutuellement exclusifs est que leur intersection soit vide. Le diagramme de Venn, présenté à la figure 4.7, illustre deux événements A et B mutuellement exclusifs. Dans ce cas, $P(A \cap B) = 0$ et la formule de la loi de la somme se réduit à

► **Loi de la somme pour des événements mutuellement exclusifs**

$$P(A \cup B) = P(A) + P(B)$$

EXERCICES

Méthode

22. Supposez qu'un espace-échantillon soit composé de cinq résultats possibles équiprobables : E_1, E_2, E_3, E_4, E_5. Soient

$$A = \{E_1, E_2\}$$

$$B = \{E_3, E_4\}$$

$$C = \{E_2, E_3, E_5\}$$

a) Calculer $P(A)$, $P(B)$ et $P(C)$.

b) Calculer $P(A \cup B)$. Les événements A et B sont-ils mutuellement exclusifs ?

c) Déterminer A^c, C^c, $P(A^c)$ et $P(C^c)$.

d) Déterminer $A \cup B^c$ et $P(A \cup B^c)$.

e) Calculer $P(B \cup C)$.

23. Supposez qu'un espace-échantillon S soit composé de sept éléments : $S = \{E_1, E_2, E_3, E_4, E_5, E_6, E_7\}$. Les probabilités attribuées à ces éléments de l'échantillon sont les suivantes : $P(E_1) = 0{,}05$, $P(E_2) = 0{,}20$, $P(E_3) = 0{,}20$, $P(E_4) = 0{,}25$, $P(E_5) = 0{,}15$, $P(E_6) = 0{,}10$ et $P(E_7) = 0{,}05$. Soient

$$A = \{E_1, E_4, E_6\}$$
$$B = \{E_2, E_4, E_7\}$$
$$C = \{E_2, E_3, E_5, E_7\}$$

a) Calculer $P(A)$, $P(B)$ et $P(C)$.

b) Déterminer $A \cup B$ et $P(A \cup B)$.

c) Déterminer $A \cap B$ et $P(A \cap B)$.

d) Les événements A et C sont-ils mutuellement exclusifs ?

e) Déterminer B^c et $P(B^c)$.

Applications

24. L'université Clarkson a effectué une enquête d'opinion auprès de ses anciens élèves. En particulier, il était demandé aux anciens élèves d'indiquer si leur passage à Clarkson avait répondu à leurs attentes, les avait surpassées ou ne les avait pas satisfaites. Les résultats de l'enquête ont montré que 4 % des anciens élèves n'ont pas répondu, 26 % considéraient que leurs attentes n'avaient pas été satisfaites et 65 % ont répondu que leur expérience à Clarkson correspondait à leurs attentes.

a) Quelle est la probabilité qu'un ancien élève sélectionné aléatoirement réponde que son expérience a surpassé ses attentes ?

b) Quelle est la probabilité qu'un ancien élève sélectionné aléatoirement réponde que son expérience a répondu ou surpassé ses attentes ?

25. Dans l'enquête Eco Pulse menée par la société de marketing Shelton Group, on demandait aux personnes interrogées d'indiquer les actions qui leur procuraient un sentiment de culpabilité (*Los Angeles Times*, 15 août 2012). Selon les résultats de l'enquête, il y a une probabilité de 0,39 qu'une personne sélectionnée aléatoirement se sente coupable de gaspiller de la nourriture et une probabilité de 0,27 qu'une personne sélectionnée aléatoirement se sente coupable de laisser les lumières allumées alors qu'elle n'est pas dans la pièce. De plus, il y a une probabilité de 0,12 qu'une personne sélectionnée aléatoirement se sente coupable pour ces deux raisons.

a) Quelle est la probabilité qu'une personne sélectionnée aléatoirement se sente coupable soit de gaspiller de la nourriture, soit de laisser les lumières allumées lorsqu'elle n'est pas dans la pièce ?

b) Quelle est la probabilité qu'une personne sélectionnée aléatoirement ne se sente pas coupable pour l'une ou l'autre de ces raisons ?

26. Les informations sur les fonds mutuels fournies par Morningstar Investment Research incluent le type de fonds mutuels (domestique, international ou à revenu fixe) et le classement Morningstar. Le classement est exprimé en nombre d'étoiles de 1 (le plus faible) à 5 (le plus élevé). Un échantillon de 25 fonds mutuels appartenant au classement *Morningstar Funds 500* (2008) est sélectionné. Les informations suivantes ont été collectées :

- Seize fonds mutuels étaient domestiques.
- Treize fonds mutuels avaient au plus 3 étoiles.
- Sept des fonds domestiques avaient 4 étoiles.
- Deux des fonds domestiques avaient 5 étoiles.

Supposez que l'un de ces 25 fonds mutuels soit sélectionné de façon aléatoire afin d'en apprendre davantage sur ce fonds et la stratégie d'investissement.

a) Quelle est la probabilité de sélectionner un fonds domestique ?

b) Quelle est la probabilité de sélectionner un fonds ayant 4 ou 5 étoiles ?

c) Quelle est la probabilité de sélectionner un fond qui soit domestique *et* qui ait 4 ou 5 étoiles ?

d) Quelle est la probabilité de sélectionner un fond qui soit domestique *ou* qui ait 4 ou 5 étoiles ?

27. Quelles rencontres de basket universitaire de la NCAA ont la plus forte probabilité de voir s'affronter une équipe engagée dans le championnat national de basket ? Au cours des 20 dernières années, la rencontre de la côte atlantique (ACC) est arrivée en tête du palmarès en ayant eu à 10 reprises une équipe engagée dans le championnat. La rencontre du Sud-Est (SEC) s'est classée seconde : à 8 reprises, une équipe engagée dans le championnat a joué durant ces rencontres. Cependant, ces deux rencontres n'ont eu, qu'une seule fois, une équipe engagée simultanément dans le championnat, lorsque l'équipe d'Arkansas (SEC) a battu l'équipe de Duke (ACC) 76 à 70 en 1994 (site Internet NCAA, avril 2009). Utiliser ces données pour estimer les probabilités suivantes.

a) Quelle est la probabilité que lors d'une rencontre ACC, une équipe engagée dans le championnat joue ?

b) Quelle est la probabilité que lors d'une rencontre SEC, une équipe engagée dans le championnat joue ?

c) Quelle est la probabilité qu'à la fois lors d'une rencontre ACC et d'une rencontre SEC, une équipe engagée dans le championnat joue ?

d) Quelle est la probabilité qu'au moins une équipe issue de ces deux rencontres soit engagée dans le championnat ? C'est-à-dire, quelle est la probabilité qu'une équipe issue de l'ACC ou du SEC participe au championnat ?

e) Quelle est la probabilité que le championnat se déroule sans équipe issue de ces deux rencontres ?

28. Une étude sur les abonnés d'un magazine a révélé que 45,8 % d'entre eux ont loué une voiture au cours des 12 derniers mois pour des raisons professionnelles, 54 % pour des raisons personnelles et 30 % à la fois pour des raisons professionnelles et personnelles.

a) Quelle est la probabilité qu'un abonné ait loué une voiture au cours des 12 derniers mois pour des raisons professionnelles ou personnelles ?

b) Quelle est la probabilité qu'un abonné n'ait loué aucune voiture au cours des 12 derniers mois que ce soit pour des raisons professionnelles ou personnelles ?

29. Les élèves de terminale les plus brillants candidatent dans les grandes écoles et les universités les plus prestigieuses en plus grand nombre chaque année. Puisque le nombre de places reste relativement stable, certaines écoles rejettent davantage de candidatures. L'université de Pennsylvanie a reçu 2 851 candidatures en première année. Dans ce groupe, 1 033 étudiants ont été acceptés sur dossier, 854 rejetés définitivement et 964 soumis au vote d'une commission d'admission. Par le passé, l'université a admis environ 18 % des candidats passés devant la commission sur un nombre total d'étudiants (candidats admis sur dossier et candidats admis après passage en commission) égal à 2 375. Soient D, R et C les événements « un candidat est admis sur dossier », « un candidat est rejeté » et « un candidat est renvoyé devant la commission d'admission ». Soit A l'événement « le candidat passé devant la commission est admis ».

a) Utiliser les données pour estimer $P(D), P(R)$ et $P(C)$.

b) Les événements D et C sont-ils mutuellement exclusifs ? Calculer $P(D \cap C)$.

c) Sur les 2 375 étudiants admis par le passé à l'université de Pennsylvanie, quelle est la probabilité qu'un étudiant sélectionné aléatoirement ait été accepté sur dossier ?

d) Supposons qu'un étudiant soumette aujourd'hui sa candidature à l'université de Pennsylvanie. Quelle est la probabilité que l'étudiant soit admis sur dossier ou accepté par la commission d'admission ?

4.4 PROBABILITÉ CONDITIONNELLE

Souvent, la probabilité d'un événement est influencée par le fait qu'un événement, lié au premier, se soit produit. Considérons un événement A avec une probabilité $P(A)$. Si nous apprenons qu'un événement B, lié à A, s'est déjà produit, nous pouvons tirer parti de cette information pour calculer une nouvelle probabilité de l'événement A. Cette nouvelle probabilité de l'événement A, appelée **probabilité conditionnelle**, est notée $P(A|B)$. La notation | est utilisée pour souligner le fait que nous considérons la probabilité de l'événement A *sachant que* l'événement B s'est produit. Par conséquent, la notation $P(A|B)$ se lit « probabilité de A sachant B ».

Comme exemple d'application des probabilités conditionnelles, considérons les possibilités de promotion professionnelle des policiers, hommes et femmes, d'une grande métropole à l'Est des États-Unis. Les forces de police de cette ville comptent 1 200 officiers, 960 hommes et 240 femmes. Au cours des deux dernières années, 324 policiers ont été promus. La répartition de ces promotions entre hommes et femmes est détaillée dans le tableau 4.4.

Tableau 4.4 ***Promotion des policiers au cours des deux dernières années***

	Homme	Femme	Totaux
Promu	288	36	324
Non promu	672	204	876
Totaux	960	240	1 200

Après avoir examiné ces chiffres, un comité de femmes policiers a entamé une procédure judiciaire pour discrimination, en se basant sur le fait que 288 hommes policiers avaient été promus contre seulement 36 femmes. L'administration policière a rétorqué que le nombre relativement bas de femmes policiers promues n'était pas dû à un comportement discriminatoire mais au fait que peu de femmes font partie des forces de police. Montrons comment utiliser les probabilités conditionnelles pour analyser l'accusation de discrimination.

Soient les événements

H = « le policier est un homme »
F = « le policier est une femme »
A = « le policier est promu »
A^c = « le policier n'est pas promu »

Diviser les données du tableau 4.4 par le nombre total de policiers (1 200) nous permet de résumer les informations disponibles par les probabilités suivantes :

$P(H \cap A) = 288/1200 = 0{,}24$ = probabilité qu'un policier choisi aléatoirement soit un homme et ait été promu

$P(H \cap A^c) = 672/1200 = 0{,}56$ = probabilité qu'un policier choisi aléatoirement soit un homme et n'ait pas été promu

$P(F \cap A) = 36/1200 = 0{,}03$ = probabilité qu'un policier choisi aléatoirement soit une femme et ait été promu

$P(F \cap A^c) = 204/1200 = 0{,}17$ = probabilité qu'un policier choisi aléatoirement soit une femme et n'ait pas été promu

Puisque ces valeurs correspondent à la probabilité d'intersection de deux événements, ces probabilités sont appelées **probabilités jointes**. Le tableau 4.5, qui fournit un résumé des informations, en termes de probabilités, sur les promotions au sein de la police, est dit *tableau des probabilités jointes*.

Tableau 4.5 ***Tableau des probabilités jointes pour les promotions***

Les probabilités jointes apparaissent à l'intérieur du tableau

	Homme (H)	Femme (F)	Totaux
Promu (A)	0,24	0,03	0,27
Non promu (A^c)	0,56	0,17	0,73
Totaux	0,80	0,20	1,00

Les probabilités marginales apparaissent dans les marges du tableau

Les valeurs inscrites dans les marges du tableau des probabilités jointes fournissent les probabilités de chaque événement séparément. C'est-à-dire, $P(H) = 0{,}80, P(F) = 0{,}20$, $P(A) = 0{,}27$ et $P(A^c) = 0{,}73$. Ces **probabilités** sont dites **marginales**, du fait de leur localisation dans les marges du tableau des probabilités jointes. Les probabilités marginales sont obtenues en additionnant les probabilités jointes, associées à l'événement, dans les lignes ou les colonnes du tableau des probabilités jointes. Par exemple, la probabilité marginale d'être promu est égale à $P(A) = P(H \cap A) + P(F \cap A) = 0{,}24 + 0{,}03 = 0{,}27$. D'après les probabilités marginales, 80 % des policiers sont des hommes, 20 % sont des femmes, 27 % des officiers (hommes et femmes confondus) ont été promus et 73 % ne l'ont pas été.

Commençons l'analyse des probabilités conditionnelles en calculant la probabilité qu'un policier soit promu, sachant qu'il s'agit d'un homme. Nous cherchons donc à déterminer $P(A \mid H)$. Cette notation signifie simplement que nous nous intéressons à la probabilité de l'événement A (promotion) sachant que la condition décrite par l'événement H (le policier est un homme) est satisfaite. Ainsi, nous nous intéressons maintenant seulement aux possibilités de promotion des 960 hommes policiers. Puisque 288 des 960 hommes policiers ont reçu une promotion, la probabilité d'être promu sachant que le policier est un homme est égale à 288/960, soit 0,30. En d'autres termes, sachant que le policier est un homme, ce policier avait 30 % de chances de recevoir une promotion au cours des deux dernières années.

Cette procédure est facile à mettre en œuvre, car le tableau 4.4 fournit le nombre de policiers dans chaque catégorie. Nous allons maintenant montrer comment des probabilités conditionnelles, comme $P(A \mid H)$, peuvent être directement calculées à partir des probabilités des événements, plutôt qu'à partir des fréquences du tableau 4.4.

Nous avons montré que $P(A \mid H) = 288/960 = 0{,}30$. Divisons à la fois le numérateur et le dénominateur de cette fraction par 1 200, le nombre total de policiers.

$$P(A \mid H) = \frac{288}{960} = \frac{288/1200}{960/1200} = \frac{0{,}24}{0{,}80} = 0{,}30$$

Nous voyons maintenant que la probabilité conditionnelle $P(A|H)$ est égale à $0{,}24/0{,}80$. En vous référant au tableau des probabilités jointes (tableau 4.5), notez en particulier que 0,24 est la probabilité jointe de A et de H ; c'est-à-dire, $P(A \cap H) = 0{,}24$. Notez également que 0,80 est la probabilité marginale qu'un policier sélectionné aléatoirement soit un homme ; c'est-à-dire, $P(H) = 0{,}80$. Ainsi, la probabilité conditionnelle $P(A|H)$ est égale au ratio entre la probabilité jointe $P(A \cap H)$ et la probabilité marginale $P(H)$.

$$P(A|H) = \frac{P(A \cap H)}{P(H)} = \frac{0{,}24}{0{,}80} = 0{,}30$$

Le fait que les probabilités conditionnelles correspondent au ratio entre une probabilité jointe et une probabilité marginale, fournit la formule générale pour calculer la probabilité conditionnelle de deux événements A et B :

► **Probabilité conditionnelle**

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{4.7}$$

ou

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \tag{4.8}$$

Le diagramme de Venn, de la figure 4.8, permet de comprendre intuitivement les probabilités conditionnelles. Le cercle de droite correspond à l'événement B qui s'est réalisé ; la partie du cercle commune avec l'événement A correspond à l'événement $(A \cap B)$. Nous savons qu'une fois l'événement B réalisé, la seule façon de pouvoir encore observer l'événement A est que l'événement $(A \cap B)$ se réalise. Ainsi, le ratio $P(A \cap B)/P(B)$ fournit la probabilité conditionnelle que nous observions l'événement A sachant que l'événement B s'est déjà produit.

Revenons à la question d'une éventuelle discrimination envers les femmes policiers. La probabilité marginale de la colonne 1 du tableau 4.5 montre que la probabilité qu'un policier reçoive une promotion est égale à $P(A) = 0{,}27$ (que ce soit un homme ou une femme). Cependant, la question fondamentale dans cette affaire de discrimination implique deux probabilités conditionnelles : $P(A|H)$ et $P(A|F)$. C'est-à-dire, quelle est la probabilité qu'un policier soit promu sachant qu'il s'agit d'un homme ? Quelle est la probabilité qu'un policier soit promu sachant qu'il s'agit d'une femme ? Si ces deux probabilités sont égales, il n'y a aucun fondement à l'accusation de discrimination puisque les chances de promotion sont les mêmes pour les femmes et pour les hommes. Par contre, une différence entre les deux probabilités conditionnelles accréditerait la thèse selon laquelle les policiers sont traités différemment en matière de promotion, en fonction de leur sexe.

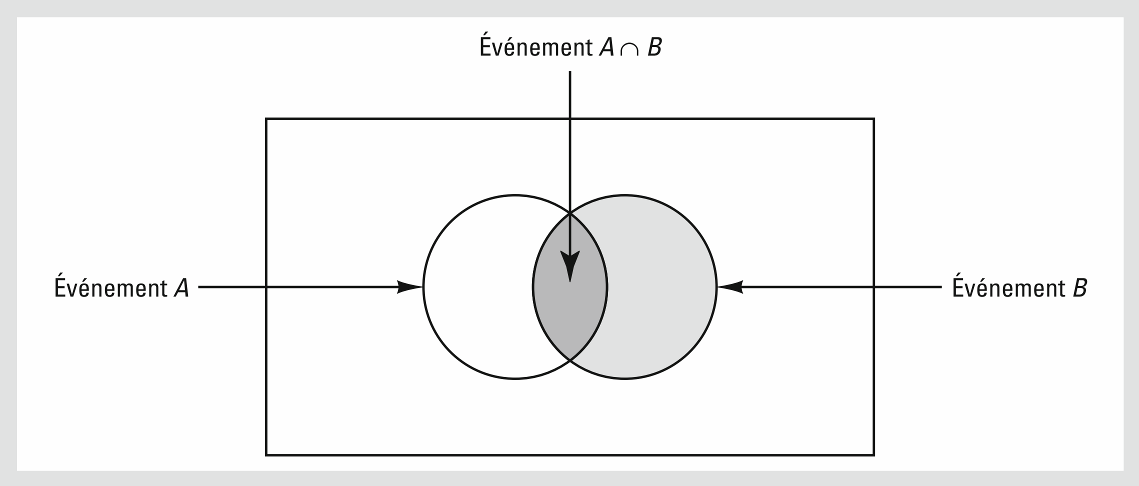

Figure 4.8 ***Probabilité conditionnelle***

Nous avons déjà déterminé que $P(A|H) = 0{,}30$. Utilisons maintenant les probabilités du tableau 4.5 et la relation (4.7) pour calculer la probabilité qu'un policier reçoive une promotion sachant qu'il s'agit d'une femme, c'est-à-dire $P(A|F)$. On obtient :

$$P(A|F) = \frac{P(A \cap F)}{P(F)} = \frac{0{,}03}{0{,}20} = 0{,}15$$

Quelles conclusions pouvez-vous en tirer ? La probabilité de recevoir une promotion est deux fois plus importante pour un homme que pour une femme. Bien que l'utilisation des probabilités conditionnelles ne prouve pas en elle-même l'existence d'une discrimination envers les femmes, les valeurs des probabilités conditionnelles soutiennent l'argument avancé par les femmes policiers.

4.4.1 Événements indépendants

Dans l'exemple précédent, $P(A) = 0{,}27$, $P(A|H) = 0{,}30$ et $P(A|F) = 0{,}15$. Nous avons vu que la probabilité de recevoir une promotion (événement A) était affectée ou influencée par le sexe du policier. En particulier, puisque $P(A|H) \neq P(A)$, les événements A et H sont dépendants. C'est-à-dire que la probabilité de l'événement A (promotion) est affectée ou altérée par le fait que l'événement H (le policier est un homme) se produise avec certitude. De manière similaire, puisque $P(A|F) \neq P(A)$, les événements A et F sont dépendants. Cependant, si la probabilité de l'événement A n'était pas affectée par l'existence de l'événement H – c'est-à-dire, si $P(A|H) = P(A)$ – alors, les événements A et H seraient dits **indépendants**. Ceci conduit à la définition suivante d'indépendance de deux événements :

► **Événements indépendants**

Deux événements A et B sont indépendants si

$$P(A|B) = P(A) \qquad (4.9)$$

ou

$$P(B|A) = P(B) \qquad (4.10)$$

Sinon, les événements sont dépendants.

4.4.2 Loi de la multiplication

Alors que la loi de la somme des probabilités permet de calculer la probabilité de l'union de deux événements, la loi de la multiplication permet de calculer la probabilité de l'intersection de deux événements. La loi de la multiplication est basée sur la définition de la probabilité conditionnelle. En réarrangeant les termes des formules (4.7) et (4.8), on obtient la **loi de la multiplication**.

► **Loi de la multiplication**

$$P(A \cap B) = P(B)P(A|B) \qquad (4.11)$$

ou

$$P(A \cap B) = P(A)P(B|A) \qquad (4.12)$$

Pour illustrer l'utilisation de la loi de la multiplication, considérons le service de diffusion d'un journal, auquel 84 % des ménages d'une région particulière sont abonnés quotidiennement. Si l'on note Q l'événement « un ménage est abonné à l'édition quotidienne », $P(Q) = 0{,}84$. De plus, on sait que la probabilité qu'un ménage déjà abonné à l'édition quotidienne, soit également abonné à l'édition du dimanche (événement D), est égale à 0,75 ; c'est-à-dire, $P(D|Q) = 0{,}75$. Quelle est la probabilité qu'un ménage soit abonné à la fois à l'édition quotidienne et à l'édition du dimanche ? En utilisant la loi de la multiplication, la probabilité désirée, $P(D \cap Q)$, est égale à

$$P(D \cap Q) = P(Q)P(D|Q) = 0{,}84 \times 0{,}75 = 0{,}63$$

Nous savons maintenant que 63 % des ménages sont abonnés aux éditions quotidiennes et du dimanche.

Avant de conclure cette section, considérons le cas spécial de la loi de la multiplication pour des événements indépendants. Rappelons que deux événements sont indépendants si $P(A|B) = P(A)$ ou $P(B|A) = P(B)$. Par conséquent, d'après les formules (4.11) et (4.12), la loi de la multiplication pour des événements indépendants s'écrit :

► **Loi de la multiplication pour événements indépendants**

$$P(A \cap B) = P(A)P(B) \qquad (4.13)$$

Pour calculer la probabilité de l'intersection de deux événements indépendants, on multiplie simplement leurs probabilités respectives. Notez que la loi de la multiplication pour des événements indépendants fournit un autre moyen de déterminer si A et B sont indépendants. En effet, si $P(A \cap B) = P(A)P(B)$, alors A et B sont indépendants ; si $P(A \cap B) \neq P(A)P(B)$, alors A et B sont dépendants.

Pour illustrer la loi de la multiplication appliquée à des événements indépendants, considérons l'exemple du responsable d'une station-service qui sait, de par son expérience, que 80 % des clients payent l'essence par carte de crédit. Quelle est la probabilité que les deux prochains clients utilisent chacun une carte de crédit ? Si l'on note A l'événement « le premier client utilise une carte de crédit » et B l'événement « le second client utilise une carte de crédit », alors l'événement qui nous intéresse est $A \cap B$. Sans autre information, on peut raisonnablement supposer que les deux événements sont indépendants. Ainsi,

$$P(A \cap B) = P(A)P(B) = 0{,}80 \times 0{,}80 = 0{,}64$$

Pour résumer cette section, notez que l'intérêt des probabilités conditionnelles réside dans le fait que les événements sont souvent liés. Dans de tels cas, les événements sont dits dépendants et les formules des probabilités conditionnelles fournies par les équations (4.7) et (4.8) permettent de calculer la probabilité des événements. Si deux événements ne sont pas liés, ils sont indépendants ; dans ce cas, la probabilité d'un événement n'est pas affectée par le fait que l'autre événement se réalise ou non.

REMARQUES

Ne confondez pas la notion d'événements mutuellement exclusifs avec celle d'événements indépendants. Deux événements de probabilité non nulle ne peuvent pas être à la fois mutuellement exclusifs et indépendants. Si un événement mutuellement exclusif est certain de se produire, la probabilité que l'autre événement se produise est nulle. Ils sont donc dépendants.

EXERCICES

Méthode

30. Supposez que nous ayons deux événements, A et B, avec $P(A) = 0{,}50$, $P(B) = 0{,}60$ et $P(A \cap B) = 0{,}40$.

a) Calculer $P(A \mid B)$.

b) Calculer $P(B \mid A)$.

c) Les événements A et B sont-ils indépendants ? Pourquoi ?

31. Supposez que nous ayons deux événements, A et B, mutuellement exclusifs. Supposez de plus que $P(A) = 0,30$ et $P(B) = 0,40$.

a) Calculer $P(A \cap B)$.

b) Calculer $P(A|B)$.

c) Un étudiant en statistiques affirme que les concepts d'événements mutuellement exclusifs et d'événements indépendants sont identiques et que si des événements sont mutuellement exclusifs, ils doivent être indépendants. Êtes-vous d'accord avec lui ? Utiliser les probabilités de cet exemple pour justifier votre réponse.

d) Quelle conclusion générale pouvez-vous tirer de vos résultats concernant des événements mutuellement exclusifs et indépendants ?

Applications

32. L'industrie automobile a vendu 657 000 véhicules aux États-Unis en janvier 2009 (*The Wall Street Journal*, 4 février 2009). Du fait des mauvaises conditions économiques, ce chiffre est en baisse de 37 % par rapport à janvier 2008. Les trois principaux constructeurs automobiles américains – General Motors, Ford et Chrysler – ont vendu 280 500 véhicules, en baisse de 48 % par rapport à janvier 2008. Un résumé des ventes par constructeur automobile et par type de véhicule vendu est fourni dans le tableau ci-dessous. Les données sont exprimées en milliers de véhicules. Les principaux constructeurs non-américains sont Toyota, Honda et Nissan. La catégorie Camion léger comprend les pickups, les mini-vans, les SUV et les crossover.

		Type de véhicule	
		Voiture	**Camion léger**
Constructeur	**Américain**	87,4	193,1
	Non américain	228,5	148,0

a) Construire un tableau des probabilités jointes pour ces données et utiliser ce tableau pour répondre aux questions suivantes.

b) Quelles sont les probabilités marginales ? Que vous apprennent-elles sur les probabilités associées au constructeur et au type de véhicule vendu ?

c) Si un véhicule est fabriqué par un des constructeurs américains, quelle est la probabilité que le véhicule soit une voiture ? Quelle est la probabilité que ce soit un camion léger ?

d) Si un véhicule n'est pas fabriqué par un des constructeurs américains, quelle est la probabilité que le véhicule soit une voiture ? Quelle est la probabilité que ce soit un camion léger ?

e) Si le véhicule est un camion léger, quelle est la probabilité qu'il soit fabriqué par un des constructeurs américains ?

f) Que vous disent les probabilités à propos des ventes ?

33. On a demandé aux étudiants passant le test d'admission au diplôme en management (GMAT) quelle était leur discipline principale l'année précédente et s'ils avaient l'intention de poursuivre leur MBA en tant qu'étudiant à plein temps ou à temps partiel. Un résumé de leurs réponses est fourni ci-dessous.

		Discipline principale			
		Commerce	**Ingénierie**	**Autres**	**Totaux**
Statut d'inscription	**Plein temps**	421	393	76	890
	Temps partiel	400	593	46	1 039
	Totaux	821	986	122	1 929

a) Construire le tableau des probabilités jointes pour ces données.

b) Utiliser les probabilités marginales de la discipline principale (commerce, ingénierie, autre) pour déterminer quelle discipline produit le plus d'étudiants en MBA potentiels.

c) Si un étudiant a l'intention de s'inscrire à plein temps en MBA, quelle est la probabilité que cet étudiant ait suivi principalement des cours d'ingénierie l'année précédente ?

d) Si un étudiant a suivi principalement des cours de commerce, quelle est la probabilité qu'il ait l'intention de suivre le MBA en étant inscrit à temps plein ?

e) Soient F l'événement « un étudiant a l'intention de s'inscrire à plein temps » et B l'événement « l'étudiant a suivi des cours de commerce l'an passé ». Les événements F et B sont-ils indépendants ? Justifier votre réponse.

34. Le département américain des transports rapporte des statistiques sur la ponctualité des vols dans les principaux aéroports américains. Les compagnies JetBlue, United et US Airways se partagent le terminal C de l'aéroport Logan de Boston. Le pourcentage de vols arrivés à l'heure en août 2012 était de 76,8 % pour JetBlue, 71,5 % pour United et 82,2 % pour US Airways (site Internet du département américain des transports, octobre 2012). Supposez que 30 % des vols arrivant au terminal C sont des vols de la compagnie JetBlue, 32 % de la compagnie United et 38 % de la compagnie US Airways.

a) Construire le tableau des probabilités jointes avec trois lignes (les compagnies aériennes) et deux colonnes (arrivées à l'heure versus arrivées en retard).

b) L'annonce de l'arrivée du vol 1 382 en porte 20 du terminal C vient d'être faite. Quelle est la probabilité que ce vol soit à l'heure ?

c) Quelle compagnie a, de façon la plus probable, assuré ce vol ? Quelle est la probabilité que ce vol ait été assuré par cette compagnie ?

d) Supposez qu'une annonce soit faite prévenant du retard du vol 1 382. Quelle la compagnie a, de façon la plus probable, assuré ce vol ? Quelle est la probabilité que ce vol ait été assuré par cette compagnie ?

35. Selon l'étude Ameriprise Financial Money Across Generation, 9 parents sur 10 ayant des enfants adultes, âgés entre 20 et 35 ans, ont aidé financièrement leurs enfants d'une façon ou d'une autre : études, voiture, loyer, factures, couverture d'un découvert et/ou hébergement à titre gracieux (*Money*, janvier 2009). Le tableau suivant issu d'un échantillon de données représentatives de l'étude, indique le nombre de fois où les parents ont fourni une assistance financière à leurs enfants adultes pour acheter une voiture et payer leur loyer.

		Paiement du loyer	
		Oui	Non
Achat d'une voiture	Oui	56	52
	Non	14	78

a) Construire le tableau des probabilités jointes et l'utiliser pour répondre aux questions suivantes.

b) D'après les probabilités marginales d'achat d'une voiture ou de paiement du loyer, les parents sont-ils plus susceptibles d'aider leurs enfants adultes en achetant une voiture ou en payant le loyer ? Quelle est votre interprétation des probabilités marginales ?

c) Si les parents ont fourni une assistance financière pour l'achat d'une voiture, quelle est la probabilité que les parents payent également le loyer ?

d) Si les parents n'ont pas fourni une assistance financière pour l'achat d'une voiture, quelle est la probabilité que les parents payent le loyer ?

e) L'assistance financière pour l'achat d'une voiture est-elle indépendante de l'assistance financière pour payer le loyer ? Utiliser les probabilités pour justifier votre réponse.

f) Quelle est la probabilité que les parents aient fourni une assistance financière à leurs enfants adultes soit pour les aider à acheter une voiture, soit pour payer leur loyer ?

36. Jama Crawford de l'équipe des Trail Blazers de Portland de l'Association nationale de basketball est le meilleur lanceur-franc de l'équipe, réussissant 93 % de ces lancers (site Internet de ESPN, 5 avril 2012). Supposez qu'à la fin d'un match, Jamal Crawford soit bousculé et ait l'occasion de réaliser deux lancers.

a) Quelle est la probabilité qu'il réussisse ses deux lancers ?

b) Quelle est la probabilité qu'il réussisse au moins un lancer ?

c) Quelle est la probabilité qu'il rate ses deux lancers ?

d) Souvent, au cours d'un match, une équipe commet intentionnellement une faute sur un joueur adverse pour stopper le jeu. La stratégie habituelle consiste à commettre intentionnellement une faute sur le plus mauvais lanceur-franc de l'équipe adverse. Supposons que le joueur central des Trail Blazers de Portland réussisse 58 % de ses lancers-francs. Calculer les probabilités évoquées aux questions (a), (b) et (c) dans le cas du joueur central et démontrer que commettre intentionnellement une faute sur le joueur central des Trail Blazers de Portland est une meilleure stratégie que commettre une faute intentionnelle sur Jamal Crawford. Supposez que, comme dans les questions (a), (b) et (c), deux lancers soient autorisés.

37. Une enquête conjointe menée par le magazine Parade et Yahoo a révélé que 59 % des travailleurs américains déclarent que s'ils pouvaient tout recommencer, ils choisiraient une carrière différente (*USA Today*, 24 septembre 2012). L'enquête a également révélé que 33 % des travailleurs américains envisagent de prendre une retraite anticipée et 67 % attendent 65 ans ou plus pour prendre leur retraite. Supposez que le tableau des probabilités jointes suivant soit issu des résultats de l'enquête.

		Retraite anticipée		
		Oui	Non	
Carrière	Identique	0,20	0,21	0,41
	Différente	0,13	0,46	0,59
		0,33	0,67	

a) Quelle est la probabilité qu'un travailleur choisisse la même carrière ?

b) Quelle est la probabilité qu'un travailleur qui aurait choisi la même carrière, envisage de prendre une retraite anticipée ?

c) Quelle est la probabilité qu'un travailleur qui aurait choisi une carrière différente, envisage de prendre une retraite anticipée ?

d) Que suggèrent les probabilités conditionnelles des questions (b) et (c) quant aux raisons que les travailleurs pourraient avancer pour justifier qu'ils choisiraient la même carrière ?

38. Un institut de recherche basé à Washington, the Institute for Higher Education Policy, a étudié le remboursement des prêts étudiants contractés par 1,8 million d'étudiants qui ont commencé à rembourser leur prêt il y six ans (*The Wall Street Journal*, 27 novembre 2012). L'étude a montré que 50 % des prêts étudiants étaient remboursés de façon satisfaisante alors que 50 % étaient non remboursés. Le tableau des probabilités jointes suivant indique les probabilités que le prêt soit remboursé ou non et que l'étudiant soit diplômé ou non.

		Diplôme obtenu		
		Oui	Non	
Prêt	Remboursé	0,26	0,24	0,50
	Non remboursé	0,16	0,34	0,50
		0,42	0,58	

a) Quelle est la probabilité qu'un étudiant qui a contracté un prêt étudiant, ait obtenu son diplôme ?

b) Quelle est la probabilité qu'un étudiant qui a contracté un prêt étudiant, n'ait pas obtenu son diplôme ?

c) Sachant que l'étudiant est diplômé, quelle est la probabilité qu'il ne rembourse pas son prêt ?

d) Sachant que l'étudiant n'est pas diplômé, quelle est la probabilité qu'il ne rembourse pas son prêt ?

e) Quel est l'impact de ne pas avoir obtenu son diplôme pour les étudiants qui ont contracté un prêt étudiant ?

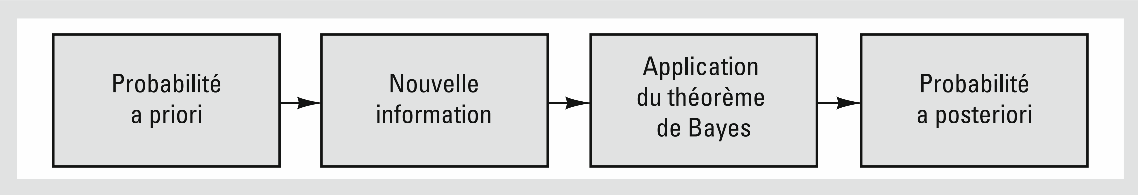

Figure 4.9 *Révision des probabilités en utilisant le théorème de Bayes*

4.5 LE THÉORÈME DE BAYES

Dans la discussion sur les probabilités conditionnelles, nous avons indiqué que la révision des probabilités, suite à l'obtention de nouvelles informations, est une phase importante de l'analyse probabiliste. Souvent, on commence l'analyse avec des **probabilités** initiales ou **a priori** concernant les différents événements en question. Ensuite, on obtient des informations supplémentaires sur ces événements grâce à un échantillon, un rapport spécial ou un test de production. Étant données ces informations, on révise les valeurs des probabilités a priori en calculant des probabilités révisées, dites **probabilités a posteriori**. Le **théorème de Bayes** permet d'effectuer ces calculs. La figure 4.9 illustre les étapes du processus de révision des probabilités.

Considérons, pour illustrer le théorème de Bayes, une entreprise manufacturière qui possède deux fournisseurs différents. Soient A_1 l'événement « la pièce est fournie par le fournisseur 1 » et A_2 l'événement « la pièce est fournie par le fournisseur 2 ». Actuellement, 65 % des pièces achetées par l'entreprise proviennent du fournisseur 1 et les 35 % restant proviennent du fournisseur 2. Par conséquent, si une pièce est sélectionnée aléatoirement, on assigne les probabilités a priori suivantes aux deux événements : $P(A_1) = 0{,}65$ et $P(A_2) = 0{,}35$.

La qualité des pièces achetées varie en fonction du fournisseur. Les données historiques révèlent les niveaux de qualité présentés dans le tableau 4.6. Soient B l'événement « la pièce est de bonne qualité » et M l'événement « la pièce est défectueuse ». Les informations contenues dans le tableau 4.6 permettent de calculer les probabilités conditionnelles suivantes :

$$P(B|A_1) = 0{,}98 \qquad P(M|A_1) = 0{,}02$$

$$P(B|A_2) = 0{,}95 \qquad P(M|A_2) = 0{,}05$$

Tableau 4.6 *Niveaux de qualité historiques des deux fournisseurs*

	Pourcentage de pièces de bonne qualité	Pourcentage de pièces défectueuses
Fournisseur 1	98	2
Fournisseur 2	95	5

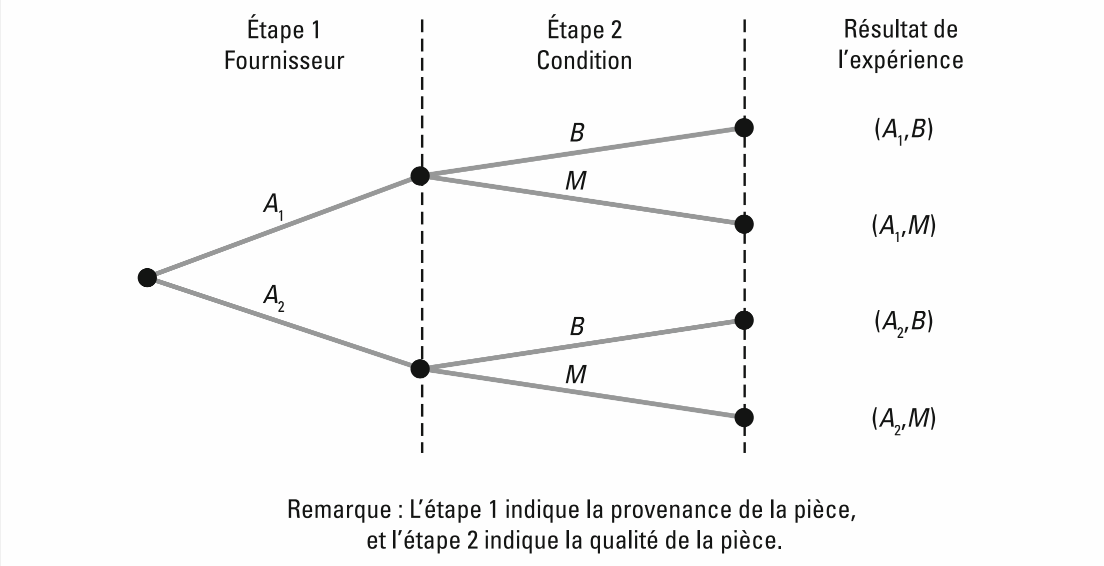

Figure 4.10 ***Diagramme arborescent associé à l'exemple des deux fournisseurs***

Le diagramme arborescent de la figure 4.10 décrit le processus de réception d'une pièce de l'un des deux fournisseurs et de contrôle de sa qualité, comme une expérience en deux étapes. Quatre résultats sont possibles : deux correspondent à une pièce de bonne qualité et deux correspondent à une pièce de mauvaise qualité.

Chacun des résultats possibles de l'expérience est l'intersection de deux événements ; nous pouvons donc utiliser la loi de la multiplication pour calculer les probabilités. Par exemple,

$$P(A_1, B) = P(A_1 \cap B) = P(A_1)P(B|\,A_1)$$

Le processus de calcul de ces probabilités jointes est décrit par ce qui est appelé un arbre des probabilités (cf. figure 4.11). À l'étape 1, les probabilités de chaque branche correspondent aux probabilités a priori ; à l'étape 2, les probabilités de chaque branche correspondent aux probabilités conditionnelles. Pour obtenir les probabilités de chaque résultat possible de l'expérience, on multiplie simplement les probabilités se trouvant sur chaque branche conduisant au résultat considéré. Chacune de ces probabilités jointes sont indiquées à la figure 4.11.

Supposons maintenant que les pièces des deux fournisseurs soient utilisées dans le système de production de l'entreprise et que l'une des machines tombe en panne à cause d'une pièce défectueuse. Sachant que la pièce est défectueuse, quelle est la probabilité qu'elle provienne du fournisseur 1 ? Du fournisseur 2 ? Avec les informations contenues dans l'arbre des probabilités (figure 4.11), le théorème de Bayes permet de répondre à ces questions.

Nous cherchons à déterminer les probabilités a posteriori $P(A_1|\,M)$ et $P(A_2|\,M)$, où M correspond à l'événement « la pièce est défectueuse ». Par la loi des probabilités

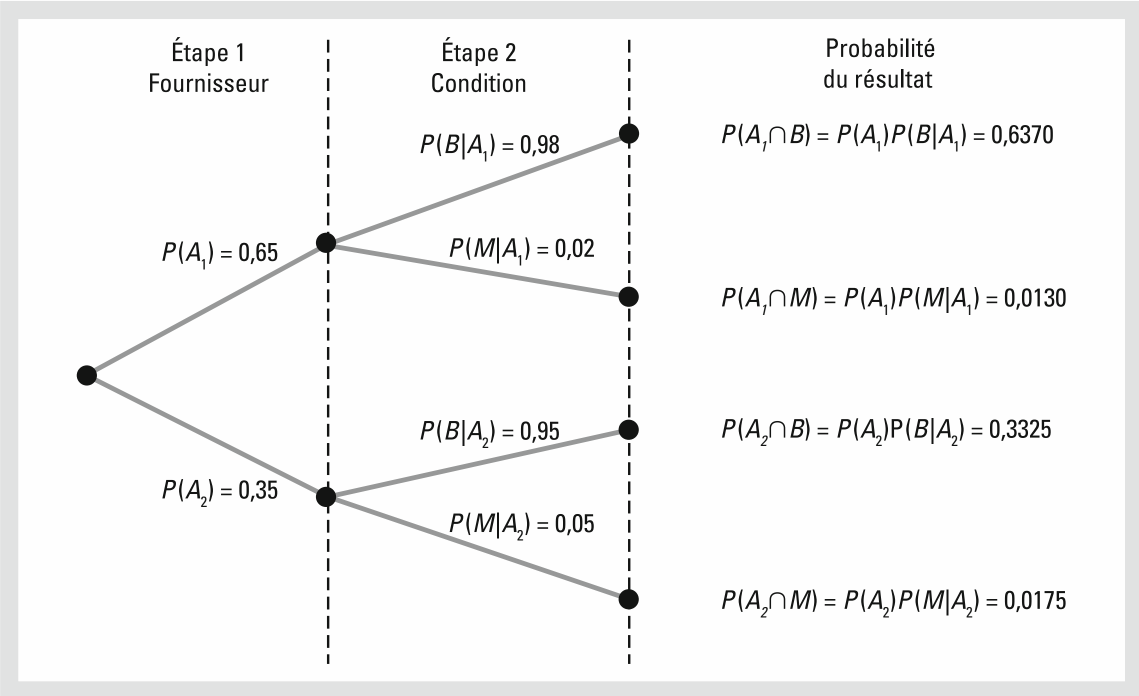

Figure 4.11 ***Arbre des probabilités pour l'exemple des deux fournisseurs***

conditionnelles, nous savons que

$$P(A_1 | M) = \frac{P(A_1 \cap M)}{P(M)} \tag{4.14}$$

En se référant à l'arbre des probabilités, on note que

$$P(A_1 \cap M) = P(A_1)P(M | A_1) \tag{4.15}$$

Pour trouver $P(M)$, notez que l'événement M ne se produit que dans deux cas : $(A_1 \cap M)$ et $(A_2 \cap M)$. Par conséquent,

$$P(M) = P(A_1 \cap M) + P(A_2 \cap M) = P(A_1)P(M | A_1) + P(A_2)P(M | A_2) \tag{4.16}$$

En substituant les équations (4.15) et (4.16) dans l'équation (4.14) et en suivant le même raisonnement pour calculer $P(A_2 | M)$, on obtient le théorème de Bayes dans le cas de deux événements.

► **Théorème de Bayes (cas de deux événements)**

$$P(A_1|M) = \frac{P(A_1)P(M|A_1)}{P(A_1)P(M|A_1) + P(A_2)P(M|A_2)} \tag{4.17}$$

$$P(A_2|M) = \frac{P(A_2)P(M|A_2)}{P(A_1)P(M|A_1) + P(A_2)P(M|A_2)} \tag{4.18}$$

> Les travaux du révérend Thomas Bayes (1702-1761), un pasteur presbytérien, sont supposés être à l'origine de la version actuelle du théorème de Bayes.

En utilisant la formule (4.17) et les valeurs des probabilités fournies dans l'exemple,

$$\begin{aligned}P(A_1 \mid M) &= \frac{P(A_1)P(M \mid A_1)}{P(A_1)P(M \mid A_1) + P(A_2)P(M \mid A_2)} \\ &= \frac{0{,}65 \times 0{,}02}{(0{,}65 \times 0{,}02) + (0{,}35 \times 0{,}05)} = \frac{0{,}0130}{0{,}0130 + 0{,}0175} \\ &= \frac{0{,}0130}{0{,}0305} = 0{,}4262\end{aligned}$$

De plus, en utilisant la formule (4.18), on obtient $P(A_2 \mid M)$.

$$\begin{aligned}P(A_2 \mid M) &= \frac{0{,}35 \times 0{,}05}{(0{,}65 \times 0{,}02) + (0{,}35 \times 0{,}05)} \\ &= \frac{0{,}0175}{0{,}0130 + 0{,}0175} = \frac{0{,}0175}{0{,}0305} = 0{,}5738\end{aligned}$$

Notez que, dans cet exemple, nous avons commencé avec une probabilité égale à 0,65 qu'une pièce, aléatoirement sélectionnée, provienne du fournisseur 1. Cependant, sachant que la pièce est défectueuse, la probabilité que la pièce provienne du fournisseur 1 chute à 0,4262. En fait, si la pièce est défectueuse, il y a plus d'une chance sur deux qu'elle provienne du fournisseur 2 ; en effet, $P(A_2 \mid M) = 0{,}5738$.

Le théorème de Bayes est applicable lorsque les événements pour lesquels nous voulons calculer les probabilités a posteriori, sont mutuellement exclusifs ; leur union correspond alors à l'espace-échantillon entier[1]. Le théorème de Bayes peut être étendu au cas de n événements mutuellement exclusifs $A_1, A_2, \ldots, A_n$, dont l'union correspond à l'espace-échantillon entier. Dans un tel cas, le théorème de Bayes permettant de calculer la probabilité a posteriori $P(A_i \mid B)$ a la forme suivante :

► **Théorème de Bayes**

$$P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + \ldots + P(A_n)P(B \mid A_n)} \tag{4.19}$$

En utilisant les probabilités a priori $P(A_1), P(A_2), \ldots, P(A_n)$ et les probabilités conditionnelles appropriées $P(B \mid A_1), P(B \mid A_2), \ldots, P(B \mid A_n)$, l'équation (4.19) permet de calculer les probabilités a posteriori des événements $A_1, A_2, \ldots, A_n$.

[1] Si l'union des événements correspond à l'espace-échantillon entier, les événements sont dits collectivement exhaustifs.

4.5.1 L'approche tabulaire

Une approche tabulaire est utile pour effectuer les calculs du théorème de Bayes. Une telle approche est présentée dans le tableau 4.7, dans le cadre du problème concernant les pièces livrées par deux fournisseurs. Les calculs sont obtenus en suivant les étapes présentées ci-dessous.

Étape 1. Préparer les trois colonnes suivantes :

Colonne 1 – Les événements mutuellement exclusifs A_i pour lesquels on souhaite obtenir les probabilités a posteriori.

Colonne 2 – Les probabilités a priori $P(A_i)$ des événements.

Colonne 3 – Les probabilités conditionnelles $P(B|A_i)$ des nouvelles informations B *sachant* chaque événement.

Étape 2. Dans la colonne 4, calculer les probabilités jointes $P(A_i \cap B)$ de chaque événement et de la nouvelle information B, en utilisant la loi de la multiplication. Ces probabilités jointes sont obtenues en multipliant les probabilités a priori de la colonne 2 par les probabilités conditionnelles correspondantes de la colonne 3 ; c'est-à-dire, $P(A_i \cap B) = P(A_i)P(B|A_i)$.

Étape 3. Additionner les probabilités jointes dans la colonne 4. La somme correspond à la probabilité de la nouvelle information, $P(B)$. Ainsi, nous voyons que, dans l'exemple précédent, l'événement « pièce défectueuse et fournisseur 1 » a une probabilité de 0,0130 ; l'événement « pièce défectueuse et fournisseur 2 » a une probabilité de 0,0175. Puisqu'une pièce défectueuse ne peut être obtenue que deux façons, la probabilité de trouver une pièce défectueuse parmi toutes les pièces livrées (par les deux fournisseurs) est égale à 0,0305 (0,0130+0,0175).

Étape 4. Dans la colonne 5, calculer les probabilités a posteriori en utilisant la relation des probabilités conditionnelles.

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)}$$

Tableau 4.7 *Approche tabulaire du théorème de Bayes appliqué au problème des deux fournisseurs*

(1)	(2)	(3)	(4)	(5)
Événements **A_i**	**Probabilités a priori** **$P(A_i)$**	**Probabilités conditionnelles** **$P(B\|A_i)$**	**Probabilités jointes** **$P(A_i \cap B)$**	**Probabilités a posteriori** **$P(A_i\|B)$**
A_1	0,65	0,02	0,0130	0,0130/0,0305 = 0,4262
A_2	0,35	0,05	0,0175	0,0175/0,0305 = 0,5738
	1,00		$P(B)$ = 0,0305	1,0000

Notez que les probabilités jointes $P(A_i \cap B)$ sont énumérées dans la colonne 4 et la probabilité $P(B)$ correspond à la somme de la colonne 4.

REMARQUES

1. Le théorème de Bayes est beaucoup utilisé dans l'analyse décisionnelle. Les probabilités a priori correspondent souvent à des estimations subjectives faites par un responsable. Une fois qu'il a obtenu des informations à partir d'un échantillon par exemple, il peut calculer les probabilités a posteriori, pour déterminer sa stratégie.
2. Un événement et son complément sont mutuellement exclusifs et leur union correspond à l'espace-échantillon entier. Par conséquent, le théorème de Bayes est toujours applicable lorsqu'il s'agit de calculer les probabilités a posteriori d'un événement et de son complément.

EXERCICES

Méthode

39. Les probabilités a priori des événements A_1 et A_2 sont $P(A_1) = 0{,}40$ et $P(A_2) = 0{,}60$. On sait également que $P(A_1 \cap A_2) = 0$. Supposez que $P(B \mid A_1) = 0{,}20$ et $P(B \mid A_2) = 0{,}05$.

a) Les événements A_1 et A_2 sont-ils mutuellement exclusifs ? Pourquoi ?

b) Calculer $P(A_1 \cap B)$ et $P(A_2 \cap B)$.

c) Calculer $P(B)$.

d) Appliquer le théorème de Bayes pour calculer $P(A_1 \mid B)$ et $P(A_2 \mid B)$.

40. Les probabilités a priori des événements A_1, A_2 et A_3 sont $P(A_1) = 0{,}20$, $P(A_2) = 0{,}50$ et $P(A_3) = 0{,}30$. Les probabilités conditionnelles de l'événement B sachant A_1, A_2 et A_3 sont $P(B \mid A_1) = 0{,}50$, $P(B \mid A_2) = 0{,}40$ et $P(B \mid A_3) = 0{,}30$.

a) Calculer $P(B \cap A_1)$, $P(B \cap A_2)$ et $P(B \cap A_3)$.

b) Appliquer le théorème de Bayes, équation (4.19), pour calculer la probabilité a posteriori $P(A_2 \mid B)$.

c) Utiliser l'approche tabulaire pour appliquer le théorème de Bayes afin de calculer $P(A_1 \mid B)$, $P(A_2 \mid B)$ et $P(A_3 \mid B)$.

Applications

41. Une entreprise de conseil a fait une offre pour un important projet de recherche. Initialement, la direction de la firme pensait avoir une chance sur deux de remporter le marché. Cependant, l'agence à laquelle l'offre a été soumise, a demandé des informations supplémentaires sur l'offre. L'expérience passée indique que lorsque l'agence a demandé des informations supplémentaires, dans 75 % des cas, les offres ont finalement été acceptées et dans 40 % des cas, elles ont été rejetées.

a) Quelle est la probabilité a priori que l'offre soit acceptée (c'est-à-dire, avant la demande d'informations supplémentaires) ?

b) Quelle est la probabilité conditionnelle d'une demande d'informations supplémentaires sachant que l'offre sera finalement acceptée ?

c) Calculer la probabilité a posteriori que l'offre soit acceptée sachant que des informations supplémentaires ont été demandées.

42. Une banque locale révise sa politique de carte de crédit avec un rappel d'une partie de celles-ci. Par le passé, environ 5 % des détenteurs d'une carte de crédit se sont révélés insolvables et la banque a été incapable de recouvrer les soldes impayés. Par conséquent, la direction a estimé égale à 0,05 la probabilité qu'un détenteur d'une carte de crédit soit insolvable. La banque a également découvert que la probabilité de ne pas honorer un prélèvement mensuel est de 0,20 pour les clients solvables. Bien entendu, la probabilité de ne pas honorer un prélèvement mensuel pour les clients insolvables est de 1.

a) Sachant qu'un client n'a pas honoré un prélèvement mensuel, calculer la probabilité a posteriori que le client soit insolvable.

b) La banque voudrait reprendre sa carte de crédit si la probabilité qu'un client soit insolvable est supérieure à 0,20. La banque devrait-elle reprendre sa carte de crédit si le client n'honore pas un prélèvement mensuel ? Pourquoi ?

43. En août 2012, la tempête tropicale Isaac s'est formée dans les Caraïbes et a touché le Golfe du Mexique. Il y avait initialement une probabilité de 0,69 qu'Isaac se transforme en ouragan avant d'atteindre le Golfe du Mexique (site Internet du Centre national des ouragans, 21 août 2012).

a) Quelle était la probabilité qu'Isaac ne se transforme pas en ouragan mais reste une tempête tropicale en atteignant le Golfe du Mexique ?

b) Deux jours plus tard, le Centre national des ouragans anticipait qu'Isaac passerait sur Cuba avant d'atteindre le Golfe du Mexique. Comment le fait de passer sur Cuba altère la probabilité qu'Isaac ne se transforme en ouragan avant qu'il n'atteigne le Golfe du Mexique ? Utiliser les probabilités suivantes pour répondre à cette question. Les ouragans qui atteignent le Golfe du Mexique ont une probabilité de 0,08 de passer sur Cuba. Les tempêtes tropicales qui atteignent le Golfe du Mexique ont une probabilité de 0,20 de passer sur Cuba.

c) Comment évolue la probabilité de se transformer en ouragan lorsqu'une tempête tropicale passe par une bande de terre comme Cuba ?

44. ParFore a créé un site Internet pour vendre des équipements et des vêtements de golf. Les responsables voudraient faire apparaître une publicité spéciale pour les femmes visitant le site et une publicité différente pour les hommes. À partir d'un échantillon de visiteurs qui ont visité le site par le passé, les responsables de ParFore ont appris que 60 % des visiteurs étaient des hommes et 40 % des femmes.

a) Quelle est la probabilité qu'un visiteur soit une femme ?

b) Supposez que 30 % des femmes qui visitent le site de ParFore aient préalablement visité le site Internet du magasin Dillard et que ce pourcentage s'élève à 10 % pour les hommes. Si la personne qui visite actuellement le site de ParFore a préalablement visité le site de Dillard, quelle est la probabilité révisée qu'il s'agisse d'une femme ? Le site ParFore devrait-il faire apparaître la publicité visant les femmes ou celle visant les hommes ?

45. Deux professeurs de Wharton ont analysé 1 613 234 putts effectués par des golfeurs lors du championnat de l'association des golfeurs professionnels (PGA) et ont trouvé que 983 764 de ces putts ont été réussis et 629 470 ont été ratés (*Is Tiger Woods Loss Averse ? Persistent Bias in the Face of Experience, Competition and High Stakes*, American Economic Review, février 2011).

a) Quelle est la probabilité qu'un joueur du championnat PGA réussisse un putt ? Le rate ?

b) Supposez qu'un joueur du championnat PGA puisse tenter un par putt. On sait que parmi les putts réussis, 64,0 % sont des par putt alors que parmi les putts ratés, 20,3 % sont des par putt. Quelle est la probabilité révisée que le joueur réussisse son putt sachant qu'il a l'occasion de faire un par putt ?

c) Un joueur fait un birdie lorsqu'il réussit un putt avec un coup de moins qu'un par. Supposez qu'un joueur du championnat PGA puisse tenter un birdie putt. On sait que parmi les putts réussis, 18,8 % sont des birdie alors que parmi les putts ratés, 73,4 % sont des birdie. Quelle est la probabilité révisée de faire un putt sachant que le joueur a l'occasion de faire un birdie putt ?

d) Commenter la différence entre les probabilités calculées aux questions (b) et (c) ?

RÉSUMÉ

Dans ce chapitre, nous avons introduit des concepts probabilistes fondamentaux et illustré l'utilisation de l'analyse probabiliste dans le but d'obtenir des informations utiles au processus de décision. Nous avons interprété les probabilités comme une mesure numérique de la vraisemblance qu'un événement se produise. De plus, nous avons vu que la probabilité d'un événement peut être calculée en sommant les probabilités des résultats possibles (des points d'échantillon) qui constituent l'événement ou en utilisant les formules des lois de la somme, de la multiplication ou des probabilités conditionnelles. Dans les cas où l'on peut obtenir des informations supplémentaires, le théorème de Bayes permet d'obtenir des probabilités révisées ou a posteriori.

GLOSSAIRE

PROBABILITÉ. Mesure numérique de la vraisemblance qu'un événement se produise.

EXPÉRIENCE. Processus qui génère des résultats bien définis.

ESPACE-ÉCHANTILLON. Ensemble de tous les résultats possibles de l'expérience.

POINT D'ÉCHANTILLON. Élément de l'espace-échantillon. Un point d'échantillon représente un résultat possible de l'expérience.

EXPÉRIENCE À PLUSIEURS ÉTAPES. Expérience qui peut être décrite par une séquence d'étapes. Si une expérience à plusieurs étapes a *k* étapes avec n_1 résultats possibles à la première étape, n_2 résultats possibles à la seconde étape et ainsi de suite, alors le nombre total de résultats possibles de l'expérience est égal à $(n_1)(n_2)\ldots(n_k)$.

DIAGRAMME ARBORESCENT. Représentation graphique utile pour définir les points d'échantillon d'une expérience en plusieurs étapes.

COMBINAISON. Dans une expérience, nous pouvons être intéressés par le nombre de façons de sélectionner n objets parmi N quel que soit l'ordre de tirage de ces n objets. Chaque tirage de n objets est appelé une combinaison et le nombre total de combinaisons de n objets sélectionnés parmi N est égal à $C_n^N = \begin{pmatrix} N \\ n \end{pmatrix} = \dfrac{N!}{n!(N-n)!}$ pour $n = 0, 1, 2, \ldots, N$.

PERMUTATION. Dans une expérience, nous pouvons être intéressés par le nombre de façons de sélectionner n objets parmi N dans un ordre de tirage précis. Chaque tirage ordonné de n objets est appelé une permutation et le nombre total de permutations de n objets sélectionnés parmi N est égal à $P_n^N = n!\begin{pmatrix} N \\ n \end{pmatrix} = \dfrac{N!}{(N-n)!}$ pour $n = 0, 1, 2, \ldots, N$.

CONDITIONS FONDAMENTALES DES PROBABILITÉS. Deux conditions qui restreignent la manière d'assigner des probabilités :

(1) Pour tout résultat possible E_i, on doit avoir $0 \leq P(E_i) \leq 1$.

(2) Considérant tous les résultats possibles de l'expérience, on doit avoir $\sum P(E_i) = 1$.

MÉTHODE CLASSIQUE. Méthode de détermination des probabilités appropriée lorsque les résultats possibles de l'expérience sont équiprobables.

MÉTHODE DE LA FRÉQUENCE RELATIVE. Méthode de détermination des probabilités appropriée lorsque les données disponibles permettent d'estimer la proportion de fois où le résultat de l'expérience se produira si l'expérience est répétée un grand nombre de fois.

MÉTHODE SUBJECTIVE. Méthode de détermination des probabilités basée sur le jugement.

ÉVÉNEMENT. Collection de points d'échantillon.

COMPLÉMENT DE L'ÉVÉNEMENT *A*. Événement contenant tous les points d'échantillon qui ne constituent pas A.

DIAGRAMME DE VENN. Représentation graphique de l'espace-échantillon et des opérations impliquant des événements dans laquelle l'espace-échantillon est représenté par un rectangle et les événements par des cercles.

UNION DES ÉVÉNEMENTS *A* ET *B*. Événement contenant tous les points d'échantillon qui appartiennent à A, à B ou aux deux. L'union est notée $A \cup B$.

INTERSECTION DE *A* ET *B*. Événement contenant tous les points d'échantillon qui appartiennent à la fois à A et à B. L'intersection est notée $A \cap B$.

LOI DE LA SOMME. Loi de probabilité utilisée pour calculer la probabilité de l'union de deux événements :
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
Pour des événements mutuellement exclusifs, puisque $P(A \cap B) = 0$, elle se réduit à $P(A \cup B) = P(A) + P(B)$.

ÉVÉNEMENTS MUTUELLEMENT EXCLUSIFS. Événements qui n'ont aucun point d'échantillon en commun ; c'est-à-dire, $A \cap B$ est vide et $P(A \cap B) = 0$.

PROBABILITÉ CONDITIONNELLE. Probabilité d'un événement sachant qu'un autre événement s'est déjà produit. La probabilité conditionnelle de A sachant B est donnée par $P(A | B) = \dfrac{P(A \cap B)}{P(B)}$.

PROBABILITÉ JOINTE. Probabilité que deux événements surviennent ; en d'autres termes, il s'agit de la probabilité de l'intersection de deux événements.

PROBABILITÉ MARGINALE. Valeurs situées dans les marges d'un tableau des probabilités jointes, correspondant aux probabilités de chaque événement séparément.

ÉVÉNEMENTS INDÉPENDANTS. Deux événements A et B tels que $P(A|B) = P(A)$ ou $P(B|A) = P(B)$; en d'autres termes, les événements n'ont aucune influence l'un sur l'autre.

LOI DE LA MULTIPLICATION. Loi de probabilité utilisée pour calculer la probabilité de l'intersection de deux événements : $P(A \cap B) = P(A)P(B|A)$ ou $P(A \cap B) = P(B)P(A|B)$. Pour des événements indépendants, la loi se réduit à $P(A \cap B) = P(A)P(B)$.

PROBABILITÉS A PRIORI. Estimation initiale des probabilités des événements.

PROBABILITÉS A POSTERIORI. Probabilités révisées des événements, basées sur des informations supplémentaires.

THÉORÈME DE BAYES. Méthode utilisée pour calculer des probabilités a posteriori.

FORMULES CLÉ

Règle de comptage par combinaisons

$$C_n^N = \binom{N}{n} = \frac{N!}{n!(N-n)!} \tag{4.1}$$

Règle de comptage par permutations

$$P_n^N = n!\binom{N}{n} = \frac{N!}{(N-n)!} \tag{4.2}$$

Calculer une probabilité en se servant de son complément

$$P(A) = 1 - P(A^c) \tag{4.5}$$

Loi de la somme

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{4.6}$$

Probabilité conditionnelle

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{4.7}$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \tag{4.8}$$

Loi de la multiplication

$$P(A \cap B) = P(B)P(A|B) \tag{4.11}$$

$$P(A \cap B) = P(A)P(B \mid A) \tag{4.12}$$

Loi de la multiplication pour événements indépendants

$$P(A \cap B) = P(A)P(B) \tag{4.13}$$

Théorème de Bayes

$$P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + \ldots + P(A_n)P(B \mid A_n)} \tag{4.19}$$

EXERCICES SUPPLÉMENTAIRES

46. Lors d'une enquête menée par les croisières Princess auprès d'adultes de 18 ans et plus, la question suivante était posée : en vacances, combien de jours vous faut-il pour vous sentir réellement détendu (*USA Today*, 24 août 2011). Les réponses ont été les suivantes : 422 – un jour ou moins ; 181 – 2 jours ; 80 – 3 jours ; 121 – 4 jours ou plus et 201 – ne se sent jamais détendu.

a) Combien d'adultes ont participé à l'enquête des croisières Princess ?

b) Quelle réponse a la plus forte probabilité de survenir ? Quelle est la probabilité de cette réponse ?

c) Quelle est la probabilité qu'une personne ne se sente jamais réellement détendue en vacances ?

d) Quelle est la probabilité qu'il faille deux jours ou plus à une personne pour se sentir réellement détendue ?

47. Un responsable financier a fait deux nouveaux investissements – l'un dans l'industrie pétrolière, l'autre dans les titres municipaux. Après une période d'un an, chacun des deux investissements sera reconnu comme un succès ou un échec. Considérez la réalisation de ces deux investissements comme une expérience.

a) Combien existe-t-il d'éléments d'échantillon pour cette expérience ?

b) Construire un diagramme arborescent et énumérer les éléments de l'échantillon.

c) Soit P l'événement « l'investissement dans l'industrie pétrolière est un succès » et M l'événement « l'investissement dans les titres municipaux est un succès ». Énumérer les éléments de l'échantillon qui constituent les événements P et M.

d) Énumérer les éléments de l'échantillon qui composent l'union des événements $(P \cup M)$.

e) Énumérer les éléments de l'échantillon qui composent l'intersection des événements $(P \cap M)$.

f) Les événements P et M sont-ils mutuellement exclusifs ? Expliquer.

48. Quarante-trois pourcent des Américains utilisent les réseaux sociaux et autres sites internet pour donner leur opinion sur les programmes télévisés (*The Huffington Post*, 23 novembre 2011). Ci-dessous sont donnés les résultats d'une enquête menée auprès de 1 400 individus à qui on a demandé s'ils utilisaient les réseaux sociaux et autres sites internet pour donner leur opinion sur les programmes télévisés.

	Utilise les réseaux sociaux et autres sites internet pour donner son opinion sur les programmes télévisés	N'utilise pas les réseaux sociaux et autres sites internet pour donner son opinion sur les programmes télévisés
Femme	395	291
Homme	323	355

a) Construire le tableau des probabilités jointes.

b) Quelle est la probabilité qu'une personne interrogée soit une femme ?

c) Quelle est la probabilité conditionnelle qu'une personne interrogée utilise les réseaux sociaux et autres sites internet pour donner son opinion sur les programmes télévisés, sachant qu'il s'agit d'une femme ?

d) Soit F l'évènement « la personne interrogée est une femme » et A l'évènement « la personne interrogée utilise les réseaux sociaux et autres sites internet pour donner son opinion sur les programmes télévisés ». Les évènements F et A sont-ils indépendants ?

49. Une étude des 31 000 admissions hospitalières de l'État de New York estime à 4 % le nombre des admissions qui sont suivies d'infections, dues aux traitements. Un septième de ces infections ont causé le décès du malade et un quart ont été faites par négligence. Dans un cas sur 7,5 impliquant des négligences, une plainte pour faute professionnelle est déposée et des dédommagements financiers sont obtenus une fois sur deux.

a) Quelle est la probabilité qu'une personne admise à l'hôpital souffre d'une infection à la suite de négligences ?

b) Quelle est la probabilité qu'une personne admise à l'hôpital meure suite à une infection ?

c) Dans le cas d'une négligence, quelle est la probabilité qu'une plainte pour faute professionnelle aboutisse au paiement de dédommagements financiers ?

50. Un sondage par téléphone a été mené auprès de téléspectateurs pour évaluer une nouvelle émission. Les données suivantes ont été obtenues.

Évaluation	Fréquence
Mauvaise	4
En-dessous de la moyenne	8
La moyenne	11
Au-dessus de la moyenne	14
Excellente	13

a) Quelle est la probabilité qu'un téléspectateur sélectionné aléatoirement donne une note supérieure ou égale à la moyenne à la nouvelle émission ?

b) Quelle est la probabilité qu'un téléspectateur sélectionné aléatoirement donne une note inférieure à la moyenne (en-dessous de la moyenne ou mauvaise) à la nouvelle émission ?

51. La tabulation croisée ci-dessous présente les revenus des ménages par niveau d'études des chefs de famille (*Statistical Abstract of the United States, 2008*).

	Revenu des ménages (en milliers de dollars)					
Niveau d'études	**Inférieur à 25**	**25,0-49,9**	**50,0-74,9**	**75,0-99,9**	**100 ou plus**	**Total**
Non bachelier	4 207	3 459	1 389	539	367	9 961
Bachelier	4 917	6 850	5 027	2 637	2 668	22 099
Niveau universitaire	2 807	5 258	4 678	3 250	4 074	20 067
Licence	885	2 094	2 848	2 581	5 379	13 787
Maîtrise et au-delà	290	829	1 274	1 241	4 188	7 822
Total	13 106	18 490	15 216	10 248	16 676	73 736

a) Construire un tableau des probabilités jointes.

b) Quelle est la probabilité qu'un chef de famille n'ait pas le baccalauréat ?

c) Quelle est la probabilité qu'un chef de famille ait un diplôme supérieur ou égal à la licence ?

d) Quelle est la probabilité qu'un ménage ayant à sa tête une personne diplômée d'une licence gagne au moins 100 000 dollars ?

e) Quelle est la probabilité qu'un ménage ait un revenu inférieur à 25 000 dollars ?

f) Quelle est la probabilité qu'un ménage ayant à sa tête une personne diplômée d'une licence gagne moins de 25 000 dollars ?

g) Le revenu du ménage est-il indépendant du niveau d'études ?

52. Une étude sur les nouveaux inscrits dans une école de commerce a révélé les données suivantes sur 2 018 étudiants.

		Candidat dans plus d'une école	
		Oui	**Non**
Groupe d'âge	**Au plus 23 ans**	207	201
	24-26	299	379
	27-30	185	268
	31-35	66	193
	Au moins 36 ans	51	169

a) Pour un étudiant en école de commerce choisi aléatoirement, construire le tableau des probabilités jointes de l'expérience qui consiste à observer l'âge de l'étudiant et le fait qu'il ait postulé dans une ou plusieurs écoles.

b) Quelle est la probabilité qu'un candidat sélectionné aléatoirement ait au plus 23 ans ?

c) Quelle est la probabilité qu'un candidat sélectionné aléatoirement ait plus de 26 ans ?

d) Quelle est la probabilité qu'un candidat sélectionné aléatoirement postule dans plus d'une école ?

53. Reprendre les données de l'étude sur les nouveaux étudiants, de l'exercice 52.

a) Sachant qu'une personne postule dans plusieurs écoles, quelle est la probabilité que cette personne ait entre 24 et 26 ans ?

b) Sachant qu'une personne a au moins 36 ans, quelle est la probabilité que cette personne postule dans plusieurs écoles ?

c) Quelle est la probabilité qu'une personne ait entre 24 et 26 ans ou postule dans plusieurs écoles ?

d) Supposez que l'on sache qu'une personne ne postule que dans une seule école. Quelle est la probabilité que cette personne ait au moins 31 ans ?

e) Est-ce que le nombre de candidatures déposées est indépendant de l'âge ? Expliquer.

54. En février 2012, dans le cadre du projet « Internet et la vie américaine », le centre de recherche Pew a mené une enquête dans laquelle étaient posées plusieurs questions sur le ressenti des internautes vis-à-vis des moteurs de recherche et autres sites qui collectent des données personnelles et utilisent ces informations pour améliorer les résultats de la recherche ou proposer des publicités ciblées (Centre de recherche Pew, 9 mars 2012). En particulier, une des questions posées était la suivante : « Si un moteur de recherche conservait des traces de ce que vous recherchez et utilisait ensuite cette information pour personnaliser vos futurs résultats de recherche, que ressentiriez-vous ? » Les personnes interrogées pouvaient indiquer « qu'elles ne seraient pas d'accord avec cette pratique, considérée comme une atteinte à la vie privée » ou « qu'elles n'y verraient pas d'inconvénient même si cela nécessite la collecte d'informations personnelles ». Les probabilités jointes des réponses et des groupes d'âge sont résumées dans le tableau ci-dessous.

Âge	Pas d'accord	D'accord
18-29	0,1485	0,0604
30-49	0,2273	0,0907
50 et plus	0,4008	0,0723

a) Quelle est la probabilité qu'une personne interrogée ne soit pas d'accord avec cette pratique ?

b) Sachant que la personne interrogée a entre 30 et 49 ans, quelle est la probabilité qu'elle soit d'accord avec cette pratique ?

c) Sachant que la personne interrogée n'est pas d'accord avec cette pratique, quelle est la probabilité qu'elle est au moins 50 ans ?

d) L'attitude envers cette pratique est-elle indépendante de l'âge ? Pourquoi ?

e) L'attitude envers cette pratique diffère-elle selon que les personnes interrogées ont entre 18 et 29 ans ou plus de 50 ans ?

55. Une importante société de biens de consommation a développé un spot publicitaire pour l'un de ses savons. Une enquête a été menée. Sur la base de cette enquête, les probabilités suivantes ont été attribuées aux événements A « l'individu a acheté le produit », S « l'individu se souvient avoir vu la publicité » et $A \cap S$ « l'individu a acheté le produit et se souvient avoir vu la publicité » : $P(A) = 0,20$, $P(S) = 0,40$ et $P(A \cap S) = 0,12$.

a) Quelle est la probabilité qu'un individu ait acheté le produit, sachant qu'il se souvient avoir vu la publicité ? Est-ce que le fait d'avoir vu la publicité accroît la probabilité d'achat du produit ? À la place du responsable, recommanderiez-vous de poursuivre la campagne publicitaire (dans la mesure où son coût est raisonnable) ?

b) Supposez que les individus qui n'achètent pas le produit de la société en question achètent celui de concurrents. Quelle serait votre estimation de la part de marché de la société ? Pensez-vous que poursuivre la campagne publicitaire permettrait d'augmenter cette part de marché ? Pourquoi ?

c) La société a également essayé une autre publicité et lui a attribué les probabilités suivantes : $P(S) = 0{,}30$ et $P(A \cap S) = 0{,}10$. Quelle est la valeur de $P(A \mid S)$ pour cette autre publicité ? Quelle publicité semble avoir le plus d'effet sur les achats des consommateurs ?

56. Cooper Realty est une petite agence immobilière implantée à Albanie, dans l'État de New York, spécialisée dans les annonces de vente de propriétés résidentielles. L'agence a récemment cherché à déterminer la probabilité que l'une de ses propriétés soit vendue en un certain nombre de jours. Une analyse des 800 ventes de l'agence réalisées les années précédentes a fourni les données suivantes.

		Nombre de jours durant lesquels l'annonce de la vente de la résidence est en agence avant la vente			
		Inférieur à 30	**Entre 31 et 90**	**Supérieur à 90**	**Total**
Prix initialement affiché	**Inférieur à 150 000 dollars**	50	40	10	100
	Entre 150 000 et 199 999 dollars	20	150	80	250
	Entre 200 000 et 250 000 dollars	20	280	100	400
	Supérieur à 250 000 dollars	10	30	10	50
	Total	100	500	200	800

a) Si A correspond à l'événement « l'annonce est passée pendant plus de 90 jours avant la vente », estimer la probabilité de A.

b) Si B correspond à l'événement « le prix initialement affiché est inférieur à 150 000 dollars », estimer la probabilité de B.

c) Quelle est la probabilité de $A \cap B$?

d) En supposant qu'un contrat vienne juste d'être signé pour faire paraître l'annonce d'une résidence vendue à un prix initial inférieur à 150 000 dollars, quelle est la probabilité que Cooper Realty mette plus de 90 jours pour la vendre ?

e) Les événements A et B sont-ils indépendants ?

57. Une société a étudié le nombre d'accidents survenus dans son usine de Brownsville, dans l'État du Texas. Les données historiques ont révélé que 6 % des employés avaient eu des accidents l'année précédente. La direction pense qu'un programme de sécurité spécial réduira le nombre d'accidents à 5 % cette année. De plus, on estime à 15 % le nombre d'employés qui, ayant eu un accident l'an passé, auront un accident cette année.

a) Quel est le pourcentage d'employés qui auront eu des accidents au cours des deux années ?

b) Quel est le pourcentage d'employés qui auront eu au moins un accident au cours des deux années ?

58. Selon le rapport Open Doors, 9,5 % des étudiants américains à temps complet étudient à l'étranger (Institut de l'éducation internationale, 14 novembre 2011). Supposez que 60 % des étudiants qui étudient à l'étranger sont des femmes et que 49 % des étudiants qui n'étudient pas à l'étranger sont des femmes.

a) Sachant qu'il s'agit d'une femme, quelle est la probabilité qu'elle étudie à l'étranger ?

b) Sachant qu'il s'agit d'un homme, quelle est la probabilité qu'il étudie à l'étranger ?

c) Quel est le pourcentage global d'étudiants qui sont des femmes ? Quel est le pourcentage global d'étudiants qui sont des hommes ?

59. Une compagnie pétrolière a posé une option sur l'achat d'une terre en Alaska. Les études géologiques préliminaires ont attribué les probabilités a priori suivantes :

P(pétrole de haute qualité) = 0,50
P (pétrole de qualité moyenne) = 0,20
P (pas de pétrole) = 0,30

a) Quelle est la probabilité de trouver du pétrole ?

b) Après avoir foré un premier puits à 200 mètres sous terre, un test du sol est effectué. Les probabilités de trouver un type particulier de sol, identifiées par le test, sont les suivantes :

P(sol | pétrole de haute qualité) = 0,20
P(sol | pétrole de qualité moyenne) = 0,80
P(sol | pas de pétrole) = 0,20

Comment la compagnie doit-elle interpréter ce test ? Quelles sont les probabilités a posteriori ? Quelle est la nouvelle probabilité de trouver du pétrole ?

60. Les cinq mots les plus fréquents apparaissant dans des spams sont *livraison* !, *aujourd'hui* !, *ici* !, *disponible* et *à porter de main* ! (Andy Greenberg, « The Most Common Words in Spam Email », site Internet de *Forbes*, 17 mars 2010). De nombreux filtres anti-spam séparent les spam des autres emails en appliquant le théorème de Bayes. Supposez que pour un compte de messagerie, un message sur dix soit un spam et que la proportion de spams qui contiennent les cinq mots les plus fréquents soit donnée ci-dessous.

livraison !	0,051
aujourd'hui !	0,045
ici !	0,034
disponible	0,014
à porter de main !	0,014

Supposez également que les proportions de messages contenant ces mots qui ne sont pas des spams soient

livraison !	0,0015
aujourd'hui !	0,0022
ici !	0,0022
disponible	0,0041
à porter de main !	0,0011

a) Si un message contient le mot *livraison !*, quelle est la probabilité qu'il s'agisse d'un spam ? Si un message contient le mot *livraison !*, quelle est la probabilité qu'il ne s'agisse pas d'un spam (mais d'un email désiré) ?

b) Si un message contient le mot *aujourd'hui !*, quelle est la probabilité qu'il s'agisse d'un spam ? Si un message contient le mot *ici !*, quelle est la probabilité qu'il s'agisse d'un spam ? Lequel de ces deux mots est un meilleur indicateur de spam ? Pourquoi ?

c) Si un message contient le mot *disponible*, quelle est la probabilité qu'il s'agisse d'un spam ? Si un message contient le mot *à porter de main !*, quelle est la probabilité qu'il s'agisse d'un spam ? Lequel de ces deux mots est un meilleur indicateur de spam ? Pourquoi ?

d) Quelles indications fournissent les réponses aux questions (b) et (c) concernant ce qui permet à un filtre anti-spam basé sur le théorème de Bayes de fonctionner correctement ?

PROBLÈME *Les juges du comté de Hamilton*

Les juges du comté de Hamilton instruisent des milliers d'affaires par an. Dans une majorité écrasante des cas jugés, le verdict rendu est appliqué. Cependant, certaines affaires sont renvoyées en appel et parfois le jugement est annulé. Kristen DelGuzzi, journaliste au *Cincinnati Enquirer*, a effectué une étude sur les affaires traitées par les juges du comté de Hamilton sur une période de trois ans. Les résultats de l'étude sur les 182 908 affaires traitées par les 38 juges de la Cour des Plaids communs, du Tribunal des affaires familiales et du Tribunal municipal sont présentés dans le tableau 4.8 (fichier en ligne Juge). Deux juges (Dinkelacker et Hogan) n'ont pas exercé dans le même tribunal pendant les trois années de l'étude.

L'objectif de l'étude du journal était d'évaluer les performances des juges. Les appels sont souvent le résultat d'erreurs commises par les juges et le journal voulait savoir quels juges faisaient du bon travail et quels juges faisaient beaucoup d'erreurs. On vous demande d'aider à analyser les données. Utilisez vos connaissances sur les probabilités et les probabilités conditionnelles pour évaluer les juges. Vous devriez également être capable d'analyser la probabilité de renvoi en appel et d'annulation du jugement dans les différents tribunaux.

Tableau 4.8 ***Nombre total d'affaires jugées, renvoyées en appel et révisées dans les tribunaux du comté de Hamilton***

Juge

Juge	Affaires jugées	Affaires renvoyées en appel	Affaires révisées	Tribunal
Fred Cartolano	3 037	137	12	Plaids communs
Thomas Crush	3 372	119	10	Plaids communs
Patrick Dinkelacker	1 258	44	8	Plaids communs
Timothy Hogan	1 954	60	7	Plaids communs
Robert Kraft	3 138	127	7	Plaids communs
William Mathews	2 264	91	18	Plaids communs
William Morrissey	3 032	121	22	Plaids communs
Norbert Nadel	2 959	131	20	Plaids communs
Arthur Ney Jr.	3 219	125	14	Plaids communs
Richard Niehaus	3 353	137	16	Plaids communs
Thomas Nurre	3 000	121	6	Plaids communs
John O'Connor	2 969	129	12	Plaids communs
Robert Ruehlman	3 205	145	18	Plaids communs
J. Howard Sundermann Jr.	955	60	10	Plaids communs
Ann Marie Tracey	3 141	127	13	Plaids communs
Ralph Winkler	3 089	88	6	Plaids communs
Penelope Cunningham	2 729	7	1	Affaires familiales
Patrick Dinkelacker	6 001	19	4	Affaires familiales
Deborah Gaines	8 799	48	9	Affaires familiales
Ronald Panioto	12 970	32	3	Affaires familiales
Mike Allen	6 149	43	4	Municipal
Nadine Allen	7 812	34	6	Municipal
Timothy Black	7 954	41	6	Municipal
David Davis	7 736	43	5	Municipal
Leslie Isaiah Gaines	5 282	35	13	Municipal
Karla Grady	5 253	6	0	Municipal
Deidra Hair	2 532	5	0	Municipal
Dennis Helmick	7 900	29	5	Municipal
Timothy Hogan	2 308	13	2	Municipal
James Patrick Kenney	2 798	6	1	Municipal
Joseph Luebbers	4 698	25	8	Municipal
William Mallory	8 277	38	9	Municipal
Melba Marsh	8 219	34	7	Municipal
Beth Mattingly	2 971	13	1	Municipal
Albert Mestemaker	4 975	28	9	Municipal
Mark Painter	2 239	7	3	Municipal
Jack Rosen	7 790	41	13	Municipal
Mark Schweikert	5 403	33	6	Municipal
David Stockdale	5 371	22	4	Municipal
John A. West	2 797	4	2	Municipal

Rapport

Préparer un rapport sur votre évaluation des juges. Inclure également une analyse de la probabilité qu'un jugement soit renvoyé en appel et annulé, dans les trois tribunaux. Votre rapport doit au moins contenir :

1. La probabilité qu'une affaire soit renvoyée en appel et le jugement annulé dans les trois tribunaux ;
2. La probabilité qu'une affaire soit renvoyée en appel, pour chaque juge ;
3. La probabilité que le jugement d'une affaire soit annulé, pour chaque juge ;
4. La probabilité que le jugement d'une affaire soit annulé sachant qu'elle a été renvoyée en appel, pour chaque juge ;
5. Le classement des juges dans chaque tribunal. Expliquez le choix du critère que vous avez utilisé.