Module 6 – Probabilité et échantillonnage

MQT 1001
Mathématiques appliquées
à la gestion

Houda Affes

UNIVERSITÉ TÉLUQ

Table des matières

Section 1 : l'analyse combinatoire

Faire l'étude de la probabilité sans au préalable introduire des notions d'analyse combinatoire serait comme tenter de résoudre des expressions algébriques sans connaître la notion de polynôme ou sans savoir résoudre une équation simple. L'*analyse combinatoire* est l'étude du dénombrement des dispositions qu'il est possible de former à partir d'un nombre fini d'éléments. Dénombrer ne veut pas dire énumérer car, dans beaucoup de cas, l'énumération s'avère impossible étant donné le nombre élevé de possibilités. Par exemple, combien peut-on former de plaques d'immatriculation d'autos différentes avec trois lettres et trois chiffres? Quelle est la probabilité de gagner à la 6/49 en achetant une combinaison à un dollar? Certains avancent le chiffre d'une chance sur 14 000 000. D'où vient ce chiffre? Quelle est la probabilité d'obtenir un bon rendement d'un fonds mutuel?

La raison d'être de l'analyse combinatoire est justement de connaître la réponse exacte à ces questions, de dénombrer les dispositions ou les possibilités qu'un événement se produise lorsque ces dispositions ou ces possibilités sont élevées.

Pour faire l'étude de l'analyse combinatoire, il est indispensable de connaître une notation mathématique particulière : la factorielle. Nous commencerons cette section par l'étude de cette notation afin de la maîtriser parfaitement, car elle est omniprésente en analyse combinatoire. Puis nous verrons les règles de l'analyse combinatoire et les différents types de disposition.

La factorielle

6! (lire 6 factorielle). Ce point d'exclamation placé à droite du nombre 6 n'a rien de mystérieux. C'est l'expression du produit de tous les entiers positifs de ce nombre jusqu'à un. Il est donc essentiel que *n* soit un nombre entier positif. La notation n'est possible que pour les nombres naturels.

$$6! = 6\times5\times4\times3\times2\times1 = 720$$

La formule générale de la factorielle est la suivante :

$$n! = n\times(n-1)\times(n-2)\times(n-3)\times\ldots\times3\times2\times1$$

EXEMPLES :

$5! = 5\times 4\times 3\times 2\times 1 = 120$

$4! = 4\times 3\times 2\times 1 = 24$

$3! = 3\times 2\times 1 = 6$

$2! = 2\times 1 = 2$

$1! = 1$

$0! = 1$,

En effet, vous pouvez vérifier que quelle que soit la valeur de n , l'équation

$(n+1)! = (n+1)\times n!$ est toujours vraie. (vous pouvez vérifier cette équation avec les exemples précédents).

En particulier, si $n = 0$, alors $(0+1)! = (0+1)\times 0! \Rightarrow 1! = 1\times 0! \Rightarrow 1 = 0!$.

Retenez bien : $\mathbf{(n+1)! = (n+1)\, n!}$**.** Cette équation vous servira beaucoup dans les calculs d'analyse combinatoire.

Les expressions utilisant la notation factorielle commandent des opérations de multiplication et de division qui sont souvent longues et fastidieuses, à moins de posséder une calculatrice scientifique[1]. La formule permettant de connaître le nombre de combinaisons possibles à la 6/49 est la suivante : $\frac{49!}{(49-6)!\times 6!}$. Auriez-vous la patience de calculer cette expression sans l'aide d'une calculatrice? Il existe heureusement une solution plus simple qui consiste à simplifier l'expression à calculer :

$$\frac{49!}{(49-6)!\times 6!} = \frac{49\times 48\times 47\times 46\times 45\times 44\times 43!}{43!\times 6\times 5\times 4\times 3\times 2\times 1}$$

Ces changements dans l'équation sont directement basés sur l'équation que nous avons notée juste un peu plus tôt : $(n+1)! = (n+1)n!$

À cause de cela, $49! = 49\times 48!$ Et comme $48! = 48\times 47!$, alors $49! = 49\times 48\times 47!$.

Et comme $47 = 47\times 46!$, alors $49! = 49\times 48\times 47\times 46!$. Et on continue de la même façon jusqu'à ce que l'on obtienne une factorielle qui sera la même qu'une factorielle du dénominateur.

À cette étape du calcul, on élimine les nombres 43! qui se trouvent au numérateur et au dénominateur. Par la suite, on peut réduire certains chiffres du numérateur en les divisant par les diviseurs du dénominateur (par exemple, 48 divisé par 6). Le reste du calcul se fera facilement avec la plupart des

1. La calculatrice scientifique limite toutefois l'affichage des résultats à la factorielle 69.

calculatrices. Si le produit est encore trop élevé, vous devrez vous rabattre sur la bonne vieille méthode du calcul manuel.

EXEMPLE :

Simplifiez l'expression $\frac{6!}{3!}$

SOLUTION

$$\frac{6!}{3!}=\frac{6\times5\times4\times3!}{3!}=6\times5\times4=120$$

EXEMPLE :

Simplifiez l'expression $\frac{9!\times5!}{6!\times3!}$

SOLUTION

$$\frac{9!\times5!}{6!\times3!}=\frac{(9\times8\times7\times6!)\times(5\times4\times3!)}{6!\times3!}=9\times8\times7\times5\times4=10\ 080$$

EXEMPLE :

Simplifiez l'expression $\frac{49!}{(49-6)!\times6!}$

SOLUTION

$$\frac{49!}{(49-6)!\times6!}=\frac{49\times48\times47\times46\times45\times44\times43!}{43!\times6\times5\times4\times3\times2\times1}=\frac{49\times48\times47\times46\times45\times44}{6\times5\times4\times3\times2\times1}=13\,983\,816$$

Les règles de l'analyse combinatoire

Il existe deux règles fondamentales en analyse combinatoire : la règle de la multiplication et la règle de l'addition. Voici les énoncés de ces deux règles, suivis d'un exemple.

Règle de la multiplication

Si une opération A peut se faire de m façons et s'il y a n façons d'accomplir une opération B, alors il y a $m\times n$ façons différentes d'effectuer les deux opérations A et B.

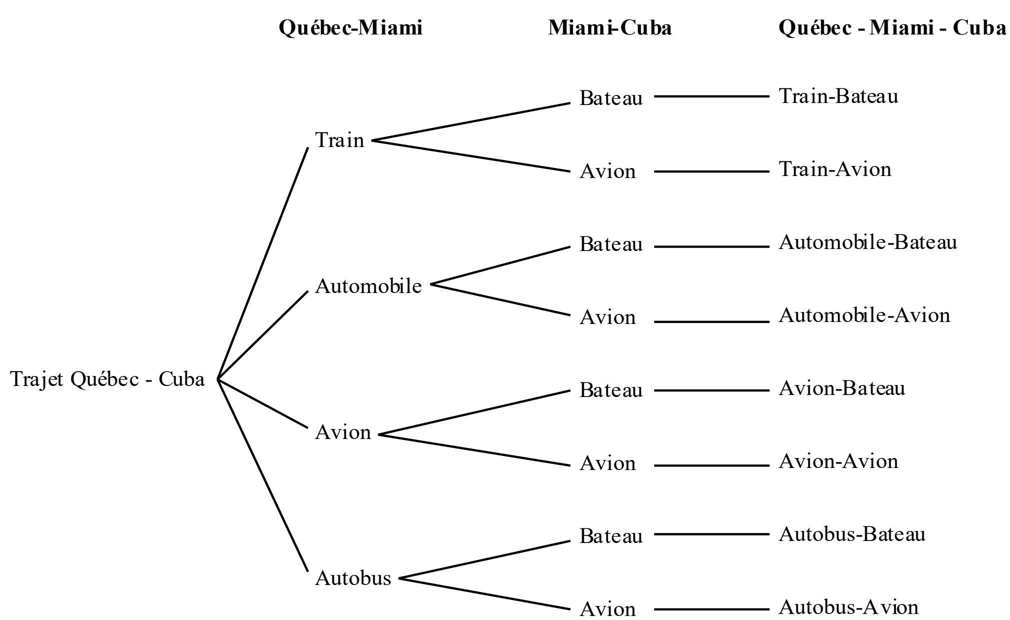
Règle de l'addition

Si une opération A peut se faire de m façons et qu'il existe n façons d'accomplir une opération B, alors il y a $m+n$ façons d'obtenir l'un ou l'autre résultat.

Un résident de Québec désire se rendre à Miami puis, après une semaine de vacances à cet endroit, il désire se rendre à Cuba pour le reste de ses vacances. Pour se rendre à Miami, les moyens de transport sont l'avion, l'automobile, le train et l'autobus. Pour se rendre à Cuba à partir de Miami, les moyens de transport possibles sont limités au bateau et à l'avion. De combien de façons différentes peut-il se rendre à Cuba?

Il existe quatre façons de faire le trajet Québec-Miami. Pour chacune de ces façons, il y a deux façons de faire le trajet Miami-Cuba. Pour mieux visualiser cette situation, construisons un diagramme des différentes façons d'effectuer le trajet Québec-Cuba. Ce graphique, appelé diagramme en arbre, facilite la compréhension des problèmes de dénombrement qui contiennent un petit nombre de variables. Il est évidemment inutilisable dans le cas du dénombrement de dispositions tel le choix des six chiffres à la 6/49.

Figure 6.1
Diagramme en arbre du trajet Québec-Cuba



Il existe donc huit façons différentes de faire le voyage Québec-Cuba en passant par Miami : quatre façons de faire le trajet Québec-Miami *et* deux façons de faire le trajet Miami-Cuba. Il a donc 4×2 façons différentes de faire le voyage complet. Cette solution, imagée par un diagramme en arbre, est l'application de la règle de la multiplication. La règle de la multiplication est associée à la conjonction *et*.

On peut illustrer la règle de l'addition en utilisant le même exemple, mais en supposant cette fois que le voyageur est un résident de Miami et qu'il a le choix entre le voyage à Cuba *ou* le voyage à Québec. Étant donné qu'il peut choisir l'un *ou* l'autre voyage, il peut se rendre à Québec de quatre façons ou bien il peut se rendre à Cuba de deux façons. Il a donc six possibilités quant au moyen de transport, soit $4+2$ possibilités. La règle de l'addition est associée à la conjonction *ou*.

EXEMPLE :

Deux amis ont leur chalet de pêche de chaque côté d'un lac. Ils communiquent à l'aide de cinq drapeaux de couleur différente. Il leur est possible de former des phrases avec différentes dispositions des drapeaux. Par exemple, un seul drapeau bleu signifie « Viens-tu à la pêche ce matin? »; un drapeau rouge veut dire « Non, pas aujourd'hui. » alors qu'un drapeau vert et un drapeau jaune veulent dire « Oui, je te rejoins. ». Combien de signaux différents peuvent-ils se lancer?

SOLUTION

Ici, les deux règles s'appliquent : la règle de l'addition, car chacun peut utiliser un *ou* deux *ou* trois *ou* quatre *ou* cinq drapeaux; la règle de la multiplication car, par exemple, avec trois drapeaux, l'un des deux amis peut choisir le rouge *et* le vert *et* le noir. Voyons d'abord, d'après la règle de multiplication, le nombre de possibilités selon le nombre de drapeaux utilisés :

- avec 1 drapeau : 5 possibilités[2];
- avec 2 drapeaux : $5\times4=20$ possibilités;
- avec 3 drapeaux : $5\times4\times3=60$ possibilités;
- avec 4 drapeaux : $5\times4\times3\times2=120$ possibilités;
- avec 5 drapeaux : $5\times4\times3\times2\times1=120$ possibilités.

Étant donné qu'il est possible de faire des signaux avec 1 *ou* 2 *ou* 3 *ou* 4 *ou* 5 drapeaux (règle de l'addition), il y aura $5+20+60+120+120=325$ possibilités.

Il y a donc 325 messages différents possibles.

Les dispositions

Nous avons défini l'analyse combinatoire comme étant l'étude du dénombrement des dispositions qu'il est possible de former à partir d'un nombre fini d'éléments. Cette étude du dénombrement porte sur deux grandes catégories de dispositions : les *dispositions ordonnées* et les *dispositions non ordonnées*. Ces deux catégories se subdivisent en *dispositions sans répétition* ou *dispositions avec répétition*.

2. Nous verrons, dans la partie suivante, que ce type de disposition porte le nom d'arrangement.

Dans une *disposition ordonnée*, la place occupée par chacun des éléments de l'ensemble a de l'importance. C'est le cas des plaques d'immatriculation. Deux plaques peuvent contenir les mêmes symboles (par exemple XSS 271 et SXS 127), mais être différentes. Les *arrangements* et les *permutations* font partie de cette catégorie.

Dans une *disposition non ordonnée* ou sans ordre, la place occupée par chacun des éléments de l'ensemble n'a pas d'importance. Ainsi, dans l'exemple de la 6/49, les nombres qui forment la combinaison gagnante ne sortent pas dans l'ordre. Le terme *combinaison* est utilisé pour décrire cette catégorie.

La *disposition sans répétition* (aussi appelée *tirage sans remise*), ordonnée ou non, est une disposition dans laquelle un élément ne peut se répéter. Par exemple, assigner des places à 30 élèves dans une classe où il y a 30 bureaux. Un autre exemple est le choix des six chiffres à la 6/49 : une boule extraite du boulier ne peut être remise dans le boulier.

Un élément peut au contraire figurer plusieurs fois dans l'ensemble des éléments d'une *disposition avec répétition* (aussi appelée *tirage avec remise*), qu'elle soit ordonnée ou non. C'est le cas si on lance plusieurs fois un dé ou une pièce de monnaie. C'est aussi le cas lorsqu'on choisit un numéro d'identification personnel (NIP) pour l'utilisation d'une carte de guichet automatique. On pourrait choisir les chiffres 121212. Les chiffres 1 et 2 se répètent, mais ce n'est pas important, car on accepte la répétition.

Les dispositions sont prises à partir d'un ensemble d'éléments. Ces éléments sont de deux ordres : *indiscernables* ou *discernables*[3]. S'ils sont égaux, ils sont indiscernables. Par exemple (5, 5, 5, 5, 5, 5) est un ensemble formé de six éléments indiscernables. Par contre, l'ensemble (1, 2, 3, 4, 5, 6) est formé de six éléments discernables ou distincts.

Les dispositions ordonnées

Une disposition ordonnée est une disposition d'éléments dans laquelle la place occupée par chacun des éléments est importante. Les dispositions ordonnées peuvent être soit sans répétition, soit avec répétition.

3. L'expression « indiscernable » tirerait son origine des premières expériences de l'analyse combinatoire, alors qu'on utilisait abondamment la pige de boules dans une urne pour valider cette théorie mathématique. On aurait alors utilisé soit des boules de même grosseur et de même poids, des boules « indiscernables au toucher », et des boules de grosseur et de poids différents, « discernables au toucher ».

Les dispositions ordonnées sans répétition

Une disposition ordonnée sans répétition est une disposition de tous les éléments d'un ensemble dans lequel un élément ne peut se répéter. Dans cette catégorie, on retrouve l'arrangement et deux types de permutation.

L'*arrangement* est une disposition ordonnée de n éléments choisis parmi m éléments discernables. Le nombre d'arrangements différents de *n* éléments choisis parmi m $(n \leq m)$ éléments est noté A_n^m et est égal à :

$$\boxed{A_n^m = \frac{m!}{(m-n)!}}$$

EXEMPLE :

Dans une classe de 20 personnes, on désire former un comité d'étudiants formé de quatre membres : un président, un vice-président, un trésorier et un secrétaire. De combien de façons peut-on le faire?

SOLUTION

Quatre membres (n) doivent être choisis parmi m éléments (les 20 personnes) discernables : 20, 19, 18, 17, 16, 15, etc. Il s'agit d'une disposition sans répétition, car une personne est choisie une seule fois.

$$A_n^m = \frac{m!}{(m-n)!}$$

$$A_4^{20} = \frac{20!}{(20-4)!} = \frac{20!}{16!} = \frac{20 \times 19 \times 18 \times 17 \times 16!}{16!} = 20 \times 19 \times 18 \times 17 = 116\ 280$$

Il y a donc 116 280 façons différentes de former ce comité.

EXEMPLE :

Un directeur de production livre ses produits dans cinq villes différentes. Il désire comparer les distances – et ultimement les coûts de transport – de tous les circuits routiers qu'il doit emprunter quotidiennement pour livrer ses produits dans chacune des cinq villes. Combien de circuits routiers différents devra-t-il comparer?

SOLUTION

Le nombre total de circuits routiers est égal au nombre *n* de destinations (5) qui doivent être choisis parmi *m* villes (5) distinctes. Il s'agit d'une disposition sans répétition, car une ville est « traversée » une seule fois.

$$A_n^m = \frac{m!}{(m-n)!}$$

$$A_5^5 = \frac{5!}{(5-5)!} = \frac{5!}{0!} = \frac{5\times4\times3\times2\times1}{1} = 120$$

Il y a donc 120 circuits routiers différents.

EXEMPLE :

Une firme de sondage est chargée d'identifier, parmi les dix marques de détersif les plus connues, les trois marques les plus utilisées dans l'ordre. Combien y a-t-il de classements possibles?

SOLUTION

Trois marques de détersif doivent être choisies parmi dix marques distinctes. Il s'agit d'une disposition sans répétition, car une marque de détersif ne peut être choisie qu'une seule fois.

$$A_n^m = \frac{m!}{(m-n)!}$$

$$A_3^{10} = \frac{10!}{(10-3)!} = \frac{10\times9\times8\times7!}{7!} = 10\times9\times8 = 720$$

Il y a donc 720 classements possibles.

La *permutation* est un type particulier d'arrangement dans lequel on choisit tous les éléments. Chercher les permutations possibles, c'est donc mettre les éléments de l'ensemble dans tous les ordres possibles. Dans une disposition sans répétition, deux cas de permutation peuvent se présenter : les *permutations avec tous les éléments discernables* et les *permutations circulaires*.

Lorsque les éléments sont tous discernables, le nombre de permutations sans répétition équivaut à :

$$\boxed{P_n = n!}$$

Par exemple, si on désire trouver le nombre d'anagrammes qu'il est possible de faire avec le mot AVRIL, la solution consiste à permuter les lettres de ce mot, étant entendu que les lettres sont discernables et toutes les lettres ne sont prises qu'une seule fois.

$P_n = n! \Rightarrow P_5 = 5! = 5\times4\times3\times2\times1 = 120$ anagrammes différentes.

Le second type de permutation, plus rare, est la permutation circulaire. Il s'agit d'une permutation où la position d'un élément est toujours fixe. Cette façon particulière de disposer les éléments fait que les permutations peuvent se faire ainsi :

$$\frac{n!}{n} \text{ ou } (n-1)!$$

EXEMPLE :

Combien de permutations peut-on faire avec 12 chiffres si le chiffre 12 occupe une position fixe?

SOLUTION

Les chiffres de 1 à 12 étant disposés circulairement, et le chiffre 12 étant fixe, on aura $\frac{n!}{n}$ ou $(n-1)!$ permutations.

$(12-1)!=11!=11\times 10\times 9\times 8\times 7\times 6\times 5\times 4\times 3\times 2\times 1=39\ 916\ 800$ permutations possibles.

Les dispositions ordonnées avec répétition

Une disposition ordonnée avec répétition est une disposition de m éléments dans laquelle un élément peut figurer plusieurs fois. Deux cas peuvent se produire, selon que les éléments sont discernables ou indiscernables.

Lorsque les éléments d'une disposition ordonnée avec répétition sont discernables, on utilise la formule de calcul suivante :

$$\boxed{m^n}$$

EXEMPLE :

Combien peut-on former de nombres en utilisant simultanément les chiffres 1, 2 et 3 si les chiffres peuvent être répétés?

SOLUTION

Les éléments (les chiffres 1, 2 et 3) ne sont pas égaux, donc discernables.

$m^n \ \Rightarrow\ 3^3 = 3\times 3\times 3 = 27$

Il y a 27 façons de former des nombres avec les chiffres 1, 2 et 3, car il y a trois façons de choisir le premier chiffre, trois façons de choisir le deuxième, et finalement trois façons de choisir le troisième.

EXEMPLE :

Pour ouvrir un cadenas de bicyclette à cinq chiffres, on doit aligner 5 des chiffres de 0 à 9. Combien de dispositions ordonnées avec répétition est-il possible de dénombrer?

SOLUTION

Les dix symboles décimaux (les chiffres de 0 à 9) sont possibles et sont tous discernables. De plus, la répétition est permise. On a donc :

$m^n \ \Rightarrow\ 10^5 = 10\times 10\times 10\times 10\times 10 = 100\,000$ possibilités différentes.

Mais il peut aussi arriver que certains éléments d'une disposition ordonnée avec répétition soient indiscernables. On parlera alors de *permutation avec des éléments qui sont indiscernables*. Par exemple, on peut se demander combien d'anagrammes peuvent être formées à partir du mot MISSISSIPPI. On sait que si toutes les lettres étaient discernables, comme c'est le cas dans le mot AVRIL, on aurait tout simplement 11! anagrammes différentes. Mais dans le mot MISSISSIPPI, il y a quatre lettres S, quatre lettres I, deux lettres P et une lettre M. Si on permute les S, par exemple, cela ne changera pas l'anagramme. On sait par ailleurs qu'avec les quatre lettres S le nombre de permutations possibles est 4! = 24. Par conséquent, à cause de ces quatre lettres S non discernables, nous ferons 24 fois moins d'anagrammes. Il en est ainsi pour les I et les P. La formule utilisée pour calculer la permutation avec des éléments qui sont indiscernables est la suivante :

$$\frac{m!}{n_1!\times n_2!\times n_3!\times\ldots\times n_i!}\text{, où } m=n_1+n_2+n_3+\ldots+n_i\,.$$

Nous pouvons donc comprendre que le nombre d'anagrammes sera :

$$\frac{11!}{4!\times4!\times2!\times1!}=\frac{11\times10\times9\times8\times7\times6\times5\times4!}{4!\times\left(4\times3\times2\times1\right)\times\left(2\times1\right)\times1}=34\ 650\text{ anagrammes.}$$

EXEMPLE :

Un entrepreneur en construction a sous sa charge douze menuisiers. Il désire en envoyer trois sur un premier chantier, quatre sur un second et cinq sur un troisième. De combien de façons peut-il répartir ses employés?

SOLUTION

Dans cet exemple, *m* est le nombre de menuisiers, et nous avons $n_1=3$, $n_2=4$ et $n_3=5$. Alors, les possibilités d'assigner les menuisiers seront au nombre de :

$$\frac{m!}{n_1!\times n_2!\times n_3!\times\ldots\times n_i!}$$

$$\frac{12!}{3!\times4!\times5!}=\frac{12\times11\times10\times9\times8\times7\times6\times5!}{\left(3\times2\times1\right)\times\left(4\times3\times2\times1\right)\times5!}=\frac{12\times11\times10\times9\times8\times7\times6}{\left(3\times2\times1\right)\times\left(4\times3\times2\times1\right)}=27\ 720$$

Il y a donc 27 720 possibilités d'assignation des menuisiers.

Les dispositions non ordonnées

Une disposition non ordonnée, ou *combinaison*, est une disposition d'éléments dans laquelle la place occupée par chacun des éléments d'un ensemble n'a pas d'importance. Dans cette catégorie, on distingue les combinaisons sans répétition et les combinaisons avec répétition.

Les combinaisons sans répétition

Les combinaisons sans répétition sont des dispositions non ordonnées de n éléments choisis parmi m éléments, dispositions dans lesquelles un élément ne peut se répéter. Le symbole utilisé pour représenter une combinaison sans répétition est C_n^m. La formule de calcul est la suivante :

$$\boxed{C_n^m = \frac{m!}{(m-n)! \times n!}}$$

EXEMPLE :

Combien de combinaisons de deux chiffres pouvons-nous faire avec les chiffres 1, 2, 3 et 4 si la répétition n'est pas permise?

SOLUTION

$$C_n^m = \frac{m!}{(m-n)! \times n!} \Rightarrow C_2^4 = \frac{4!}{(4-2)! \times 2!} = \frac{4!}{2! \times 2!} = \frac{4 \times 3 \times 2!}{2! \times 2 \times 1} = \frac{12}{2} = 6$$

Cela revient à faire des nombres de 12 à 43, mais si 12 est pris, alors 21 ne doit pas être choisi (sans ordre). Bien sûr, 22 ne peut être choisi (sans répétition). Les nombres possibles sont 12, 13, 14, 23, 24 et 34. Il y a donc six combinaisons possibles.

Comparons, dans l'exemple précédent, le nombre de combinaisons au nombre d'arrangements.

$$A_n^m = \frac{m!}{(m-n)!} \Rightarrow A_2^4 = \frac{4!}{(4-2)!} = \frac{4!}{2!} = \frac{4 \times 3 \times 2!}{2!} = 12 \text{ arrangements différents.}$$

Il y a deux fois moins de combinaisons que d'arrangements. En effet, pour tenir compte de l'ordre dans les groupes de deux chiffres, il faut les permuter. Par conséquent, on dira qu'une combinaison est un arrangement dans lequel on a enlevé l'ordre. En d'autres termes, une combinaison est un arrangement divisé par la possibilité de permuter les éléments. On peut représenter cette assertion par la formule :

$$C_n^m = \frac{A_n^m}{n!} = \frac{\frac{m!}{(m-n)!}}{n!} = \frac{m!}{(m-n)!} \times \frac{1}{n!} = \frac{m!}{(m-n)! \times n!}$$

Et si on traduit cette équation avec les données de l'exemple qui précède, on obtient :

$$C_2^4 = \frac{A_2^4}{2!} = \frac{\frac{4!}{(4-2)!}}{2!} = \frac{4!}{2!} \times \frac{1}{2!} = \frac{4!}{2! \times 2!} = \frac{4 \times 3 \times 2!}{2! \times 2 \times 1} = \frac{12}{2} = 6$$

EXEMPLE :

Le nombre de combinaisons possibles à la 6/49 est un exemple type de combinaisons sans répétition. L'achat d'un billet de loterie nous donne une chance sur 14 000 000 de gagner le gros lot. C'est en effet ce chiffre astronomique dont tous les journaux font état. Mais ce nombre de combinaisons est approximatif. Quel en est le nombre exact?

SOLUTION

Il y a 49 choix possibles ($m=49$) et on choisit six numéros ($n=6$). Le soir du tirage, l'unique boulier tourne le temps réglementaire de brassage. La première boule indique 16. Sur le billet que vous avez acheté, le 16 est le troisième nombre. Cela n'a pas d'importance, car ce tirage est sans ordre. Le nombre 16 ne reviendra plus (sans répétition). C'est donc une combinaison sans répétition. Appliquons la formule :

$$C_6^{49}=\frac{49!}{(49-6)!\times 6!}=\frac{49!}{43!\times 6!}=\frac{49\times 48\times 47\times 46\times 45\times 44\times 43!}{43!\times 6!}=\frac{49\times 48\times 47\times 46\times 45\times 44}{6\times 5\times 4\times 3\times 2\times 1}=13\ 983\ 816$$

Vous avez une chance sur 13 983 816 de gagner le gros lot de la 6/49. C'est déjà mieux que ce que racontent les journalistes, ne trouvez-vous pas?

EXEMPLE :

Dans une classe de 20 personnes, on désire former un comité comptant quatre personnes. Combien de dispositions de quatre personnes est-il possible de faire, sachant que ces quatre personnes n'auront pas de postes attitrés et qu'une même personne ne peut être choisie deux fois?

SOLUTION

Les quatre personnes n'ont pas de postes attitrés : il s'agit donc d'une disposition sans ordre (donc, d'une combinaison). Il s'agit aussi d'une disposition sans répétition, puisqu'une même personne ne peut être choisie deux fois.

$$C_4^{20}=\frac{20!}{(20-4)!\times 4!}=\frac{20!}{16!\times 4!}=\frac{20\times 19\times 18\times 17\times 16!}{16!\times 4!}=\frac{20\times 19\times 18\times 17}{4\times 3\times 2\times 1}=4\ 845$$

Il y a 4 845 combinaisons différentes alors que, vous vous rappelez peut-être, il y avait 116 280 façons différentes de former ce comité dans une disposition où on aurait tenu compte de l'ordre.

Le nombre de combinaisons est associé au triangle de Pascal[4] présenté ci-dessous :

Ligne no	1	1	1								
Ligne no	2	1	2	1							
Ligne no	3	1	3	3	1						
Ligne no	4	1	4	6	4	1					
Ligne no	5	1	5	10	10	5	1				
Ligne no	6	1	6	15	20	15	6	1			
Ligne no	7	1	7	21	35	35	21	7	1		
Ligne no	8	1	8	28	56	70	56	28	8	1	
Ligne no	9	1	9	36	64	126	126	64	36	9	1

Ce triangle est formé de la façon suivante : Chaque ligne commence et finit par 1.

Tous les autres nombres sont obtenus en additionnant les deux nombres qui sont au-dessus de celui que l'on va placer. Par exemple, le 2 de la 2e ligne est obtenu en additionnant les deux 1 de la première ligne. De même, le 6 de la 4e ligne est la somme des deux 3 de la ligne précédente. On pourrait prolonger le triangle autant qu'on le désire. Les lignes de ce triangle contiennent les valeurs de C_n^m. La première ligne contient la valeur de C_1^0 et C_1^1, la deuxième ligne les valeurs de $C_0^2, C_1^2 \ et \ C_2^2$, la troisième ligne celles de $C_0^3, C_1^3, C_2^3 \ et \ C_3^3$, la ligne 6 celles de $C_0^6, C_1^6, C_2^6, C_3^6, C_4^6, C_5^6 \ et \ C_3^6$ etc. C'est la même chose pour les autres lignes.

Mais ce n'est pas tout, ce triangle contient en même temps tous les coefficients des polynômes que l'on obtient en élevant le binôme ($a+b$) à différentes puissances :

$(a+b)^1 = a+b$ (les coefficients sont ceux de la première ligne)

$(a+b)^2 = a^2+2ab+c^2$ (ceux de la 2e ligne)

$(a+b)^3 = a^3+3a^2b+3ab^2+b^3$ (ceux de la 3e ligne)

$(a+b)^5 = a^5+5a^4b+10a^3b^2+10a^2b^3+5ab^4+b^5$ (ceux de la 5e ligne)

Et ainsi de suite pour toutes les puissances de ($a+b$). Ce binôme est souvent appelé le binôme de Newton[5], mais vous n'avez pas à mémoriser cela.

4. Philosophe et mathématicien français du 17e siècle, auteur des *Pensées.*
5. Physicien, mathématicien et théologien anglais du 17e siècle, auteur de la théorie de la gravitation universelle (l'histoire de la pomme qui tombe de l'arbre).

Les combinaisons avec répétition

Dans un problème de dénombrement où l'on ne tient pas compte de l'ordre et dans lequel un élément peut figurer plusieurs fois, il y a un nombre de dispositions possibles correspondant à la formule suivante :

$$\boxed{K_n^m = \frac{(m+n-1)!}{(m-1)! \times n!}}$$

Cette formule est mise en évidence par un exemple simple.

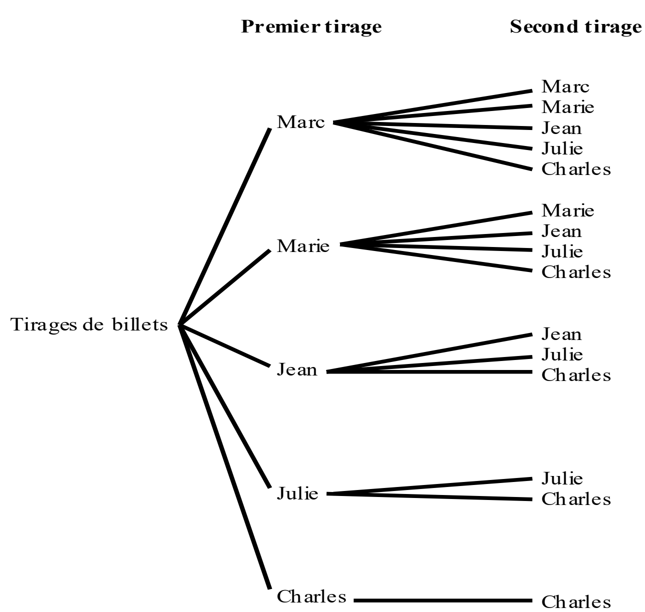
EXEMPLE :

On fait tirer deux billets de cinéma parmi cinq personnes. On accepte que la même personne puisse gagner les deux billets. Le nom du premier gagnant est donc remis dans la boîte pour le deuxième tirage. Combien peut-on former de paires différentes de gagnants si on ne tient pas compte de l'ordre dans l'énumération des gagnants?

SOLUTION

Construisons d'abord un diagramme en arbre pour représenter les données du problème. Dans ce diagramme, m représente le nombre de branches au départ de l'arbre et n, le nombre d'étapes. Remarquez que les couples qui se répètent (Marc-Marie et Marie-Marc) n'apparaissent qu'une seule fois; et que Marc, par exemple, peut gagner les deux billets, car la répétition est permise. On constate qu'il y a quinze dispositions de résultats possibles.

Figure 6.2
Diagramme en arbre d'un tirage de deux billets de cinéma



Le calcul des dispositions donnera le même résultat avec la formule de la combinaison avec répétition. Dans cette formule, m représente le nombre de personnes et n, le nombre de billets.

$$K_n^m = \frac{(m+n-1)!}{(m-1)! \times n!} \Rightarrow K_2^5 = \frac{(5+2-1)!}{(5-1)! \times 2!} = \frac{6!}{4! \times 2!} = \frac{6 \times 5 \times 4!}{4! \times 2!} = \frac{6 \times 5}{2 \times 1} = \frac{30}{2} = 15$$

Si nous avions à résoudre le même problème, mais avec quatre billets à faire tirer dans une classe de 32 étudiants, nous ferions mieux d'oublier la construction d'un diagramme en arbre. Mais la formule est infaillible pour trouver le résultat.

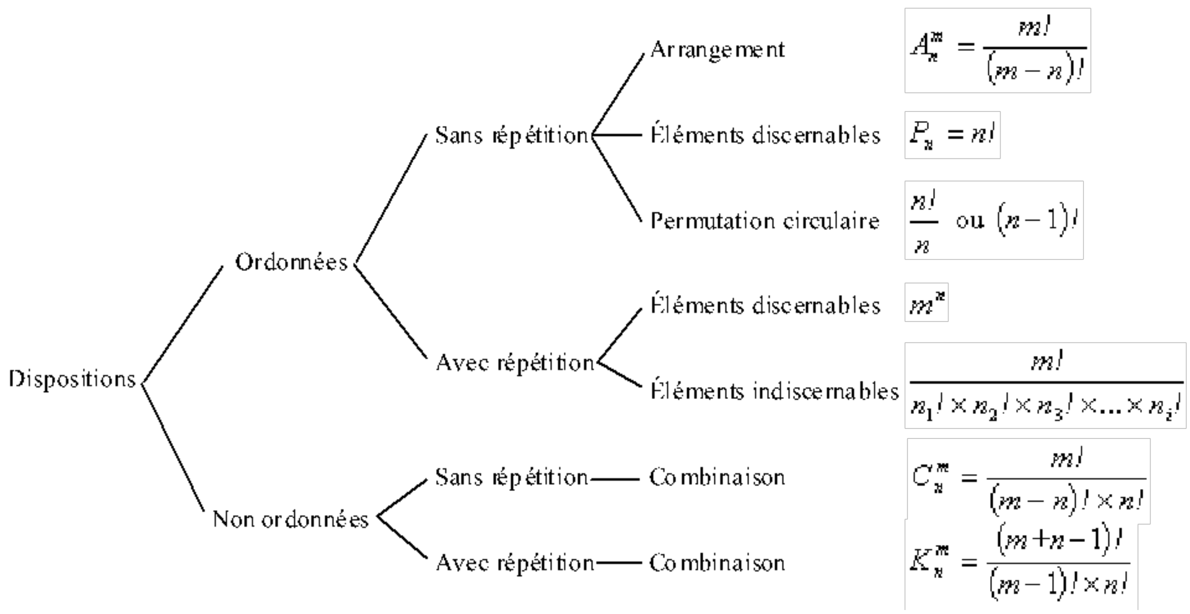
$$K_n^m = \frac{(m+n-1)!}{(m-1)! \times n!} \Rightarrow K_4^{32} = \frac{(32+4-1)!}{(32-1)! \times 4!} = \frac{35!}{31! \times 4!} = \frac{35 \times 34 \times 33 \times 32 \times 31!}{31! \times 4!} = \frac{35 \times 34 \times 33 \times 32}{4 \times 3 \times 2 \times 1} = 52\ 360$$

La résolution de problèmes de dénombrement

Résumons la stratégie à adopter pour résoudre un problème de dénombrement.

1. Lire attentivement le problème pour bien comprendre le sens de la question.
2. Identifier les éléments m et n : s'agit-il d'éléments discernables (distincts)? indiscernables?
3. Analyser ce problème de dénombrement et le ramener à une des catégories suivantes : disposition ordonnée avec répétition, disposition ordonnée sans répétition, disposition non ordonnée avec répétition, disposition non ordonnée sans répétition. Si c'est une disposition ordonnée, s'agit-il d'un arrangement ou d'une permutation?
4. Si le problème comporte peu de résultats, construire un diagramme en arbre, faire un tableau des résultats ou énumérer les résultats. Si le nombre de résultats possibles est élevé, utiliser une des formules de calcul. Le diagramme en arbre de la figure 6.3 résume ces formules.

Figure 6.3
Les formules de calcul des dispositions



EXEMPLE :

Une revue d'affaires bien connue a récemment atteint le chiffre de 50 000 abonnés. Pour souligner l'événement, elle désire organiser le tirage d'un prix alléchant parmi ses abonnés. Un membre du conseil d'administration soumet l'idée d'une loterie semblable à la 6/49 : inspiré par le numéro civique de l'établissement (le 440), il propose la 4/40. Chaque abonné recevra une combinaison différente de la 4/40; la personne qui détiendra la bonne combinaison, déterminée par un tirage au boulier, gagnera le gros lot. Compte tenu du nombre d'abonnés et des combinaisons possibles, le choix de cette loterie vous apparaît-il judicieux? Y aura-t-il forcément un gagnant?

SOLUTION

1. La question à résoudre est de savoir si la 4/40 offre le nombre de possibilités approximativement similaires au nombre d'abonnés, de manière qu'il y ait suffisamment de billets pour chaque abonné, mais aussi que le tirage permette à coup sûr de déclarer un gagnant.
2. Identifiez d'abord les éléments m et n. Dans notre exemple, n est le nombre de numéros choisis (4) parmi m choix possibles (40). Les numéros sont discernables ou distincts.
3. L'ordre dans lequel les numéros sont inscrits sur le billet n'a pas d'importance, car le tirage du numéro gagnant se fera, selon la méthode du boulier, sans ordre. Il s'agit donc d'une disposition non ordonnée ou combinaison. La loterie 4/40 est sans répétition, car il n'y a qu'un seul gagnant.

4. Étant donné le grand nombre de résultats possibles, utilisez la formule de calcul des combinaisons sans répétition, soit :

   $$C_n^m = \frac{m!}{(m-n)! \times n!}$$

   $$C_4^{40} = \frac{40!}{(40-4)! \times 4!} = \frac{40 \times 39 \times 38 \times 37 \times 36!}{36! \times 4 \times 3 \times 2 \times 1} = \frac{40 \times 39 \times 38 \times 37}{4 \times 3 \times 2 \times 1} = 91\,390 \text{ combinaisons.}$$

   Le nombre de combinaisons dépassant largement celui des abonnés, le choix de la 4/40 n'assure pas qu'il y aura un gagnant.

NOTE : Faites les exercices de la section 1 dans le *Recueil des activités pratiques* avant de continuer la lecture.

Section 2 : la probabilité

On se souvient de toutes les interrogations sur les probabilités que Charles se posait en lisant son journal quotidien. Bien sûr, il s'agissait là d'une mise en situation, d'un prétexte pour susciter votre intérêt.

Il y a autour de nous des phénomènes fortuits, indépendants de toute volonté humaine : les tremblements de terre, les raz de marée, les éruptions volcaniques ou les tornades dévastatrices en sont des exemples. Cependant, il en est autrement de certains types d'événements comme le nombre de décès sur les routes du Québec à chaque année. Des éléments d'information de base tels que le nombre de conducteurs sur les routes, leur âge et leur sexe sont autant d'informations recueillies par le statisticien qui sont ensuite utilisées pour calculer la probabilité qu'un accident survienne. En physique, en chimie, en biologie, en psychologie, en sciences humaines et dans bien d'autres disciplines, ces phénomènes aléatoires sont omniprésents.

Le calcul de la probabilité sert essentiellement à résoudre des problèmes liés aux phénomènes ou expériences aléatoires. Le terme expérience prend ici un sens plus large que celui des expériences menées en laboratoire, là où on s'imagine des chercheurs en sarrau penchés sur des microscopes ou manipulant des tubes remplis de produits chimiques de toutes sortes. Les expériences dont il est question ici sont, par exemple : le recueil des opinions politiques d'un citoyen ou des préférences d'un consommateur, le dénombrement des erreurs dans un inventaire, l'observation des prix de la Bourse à la fermeture des marchés.

Dans les pages qui suivent, vous verrez que les connaissances déjà acquises en analyse combinatoire vous seront utiles pour résoudre plusieurs problèmes. Mais définissons d'abord les notions d'expérience aléatoire, d'ensemble fondamental et d'événement.

L'expérience aléatoire

Le terme *aléatoire* tire ses racines du latin *aleatorius*, de *alea*, qui signifie aléa. Un aléa est un événement imprévisible, fruit du hasard. Une expérience aléatoire – en statistique, on utilise également le terme épreuve – dépend donc du hasard. De fait, pour qu'une expérience puisse être considérée comme étant aléatoire, trois conditions doivent être respectées :

1. Il faut que le hasard détermine son résultat.
2. Tous les résultats possibles doivent être connus.
3. L'expérience doit pouvoir être répétée à volonté.

Prenons par exemple l'action de lancer un dé à jouer[6]. Si le dé est équilibré, c'est le hasard qui déterminera le résultat obtenu parmi tous les résultats (les six chiffres possibles) et cette expérience peut être répétée à volonté. Il s'agit donc d'une expérience aléatoire.

Cette notion de hasard, conjuguée aux autres conditions qui font qu'une expérience est considérée comme aléatoire, sert d'assise aux principes gouvernant la probabilité et l'échantillonnage.

Les composantes d'une expérience aléatoire

Toute expérience aléatoire comporte deux éléments : l'*ensemble fondamental* et l'*événement*

L'ensemble fondamental

La première composante d'une expérience aléatoire est un énoncé de tous les résultats possibles. Le symbole utilisé pour représenter cet ensemble est la lettre grecque Ω (*oméga* majuscule), parfois la lettre S. On appelle parfois cet ensemble un univers des possibles ou un espace échantillonnal. Par

6. Pour comprendre les aspects mathématiques du calcul de la probabilité et pour appliquer ces notions à des situations reliées aux phénomènes de la vie courante, nous utiliserons abondamment des exemples tirés de la théorie des jeux (dés, cartes, etc.). Cette théorie des jeux, à cause de son aspect plus concret, fournit des exemples en quantité suffisante pour illustrer l'étude de l'ensemble des expériences aléatoires.

exemple, si on lance une pièce de monnaie équilibrée, les résultats possibles sont au nombre de deux, soit pile ou face :

Ω = {Pile, Face}, ou en notation abrégée : $\Omega = \{P, F\}$

Pour un dé à jouer à six faces, les résultats possibles sont au nombre de six :

Ω = {1, 2, 3, 4, 5, 6}

Pour deux dés lancés successivement, on aura, si on tient compte de l'ordre, 36 possibilités, soit :

Ω = {(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), …,(6,6)}.

Il arrive parfois qu'un ensemble fondamental soit un sous-ensemble d'un autre ensemble. Par exemple, si on s'intéresse à la possibilité d'obtenir une paire en lançant deux dés, on a :

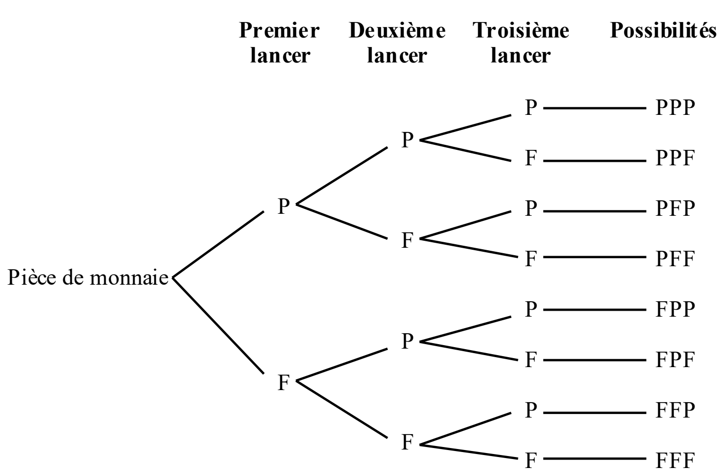
Ω = {(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)}

C'est pour cette raison qu'il est important de déterminer à l'avance ce qui nous intéresse dans le résultat de l'expérience.

L'événement

La deuxième composante d'une expérience aléatoire est l'événement. Un événement est un résultat possible qui est pris à l'intérieur de l'ensemble fondamental. C'est ce qui caractérise l'issue d'une expérience aléatoire. Prenons l'exemple d'une pièce de monnaie qui est lancée trois fois : construisons d'abord le diagramme en arbre de l'ensemble fondamental Ω.

Figure 6.4
Diagramme en arbre de lancers d'une pièce de monnaie



À l'aide du diagramme en arbre de la figure 6.4, recherchez l'événement « obtenir au moins deux faces ». L'événement « obtenir au moins deux faces » est un sous-ensemble de l'ensemble Ω, il comprend les possibilités qui contiennent au moins deux faces (FF). Ces possibilités, au nombre de quatre, seront notées $E=\{\mathrm{FFF}, \mathrm{FFP}, \mathrm{FPF}, \mathrm{PFF}\}$. *E* représente l'ensemble des événements.

Il existe différentes catégories d'événements autres qu'un événement possible. Un événement peut être impossible, c'est-à-dire qu'il n'est jamais réalisable : on lance une pièce de monnaie trois fois et on cherche à obtenir $E=\{\mathrm{FFF}\}$. Un événement peut être certain : il se réalise toujours, par exemple, lorsqu'on lance une pièce de monnaie trois fois et que l'événement cherché est obtenir au moins pile ou face. Un événement peut enfin être contraire ou complémentaire à un autre événement. Le complément d'un événement, noté E' (lire E prime) ou $\overline{E}$ (lire E barre), se réalise si et seulement si l'événement E ne se réalise pas : en lançant notre pièce de monnaie trois fois, si on considère l'événement obtenir trois fois le même résultat ($E=\{\mathrm{FFF}, \mathrm{PPP}\}$), l'événement contraire sera $E'=\{\mathrm{PPF}, \mathrm{PFP}, \mathrm{PFF}, \mathrm{FPP}, \mathrm{FPF}, \mathrm{FFP}\}$.

Les règles de l'expérience aléatoire

Les exemples que nous avons étudiés jusqu'à présent étaient des exemples d'événements simples. Mais, le plus souvent, deux ou plusieurs événements peuvent se combiner. On parle alors d'événements composés, c'est-à-dire d'un ensemble comprenant deux ou plusieurs événements simples. Deux règles gouvernent ces combinaisons d'événements : *l'union* et *l'intersection d'événements*.

L'union de deux événements

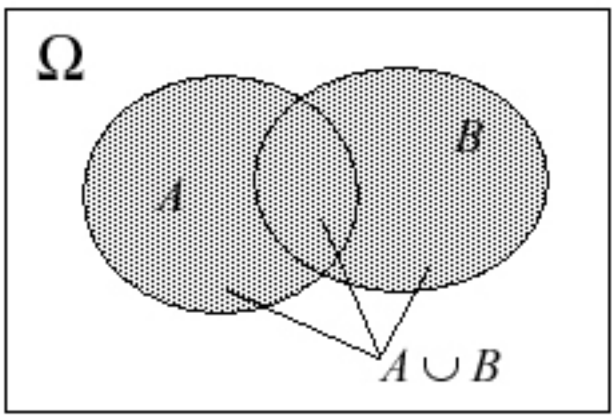
Soit l'événement E et deux événements A et B. L'union de deux événements A et B, qui s'écrit $A\cup B$ et se lit « l'union des ensembles A et B », comprend les événements appartenant soit à A, soit à B, soit à A et B à la fois. L'événement E se réalisera *si et seulement si* au moins un des événements se réalise. La façon d'exprimer cette situation est :

$$E=(A\cup B)$$

L'union des ensembles *A* et *B* se compose donc de trois termes : les événements appartenant à A, ceux appartenant à B et, finalement, les événements communs à A et à B. Dans notre exemple du lancer de deux dés, si l'événement A égale « la somme des résultats est supérieure à 9 » et que l'événement B égale « la somme égale 10 », alors $E=(A\cup B)=\{(4,\ 6),\ (5,\ 5),\ (5,6),\ (6,\ 4),\ (6,\ 5),\ (6,\ 6)\}$.

Les diagrammes de Venn utilisés dans la théorie des ensembles sont des outils précieux pour visualiser et résoudre ce type de problèmes de probabilité. Voici par exemple le diagramme de Venn de l'événement $E = (A \cup B)$.

Figure 6.5
Diagramme de Venn de l'événement *A* ∪ *B*



L'intersection de deux événements

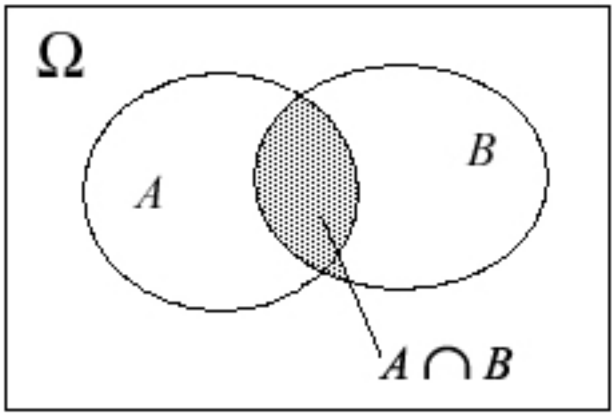
Soit l'événement E et deux événements A et B. L'intersection de deux événements A et B, qui s'écrit $A \cap B$ et se lit « l'intersection des ensembles A et B », se compose de tous les événements simples qui réalisent à la fois A et B. L'événement *E* se réalisera *si et seulement si* A et B se réalisent simultanément. La façon d'exprimer cette situation est :

$$E = (A \cap B)$$

Reprenons l'exemple du lancer de deux dés. Comme nous l'avons déjà constaté, il y a 36 résultats possibles si l'on tient compte de l'ordre. Si l'événement A égale « la somme des résultats est supérieure à 8 » et que l'événement B égale « la somme des résultats est inférieure à 11 », alors $E = (A \cap B) = \{(5,\ 4),\ (4,\ 5),\ (6,\ 3),\ (3,\ 6),\ (6,\ 4),\ (4,\ 6),\ (5,\ 5)\}$.

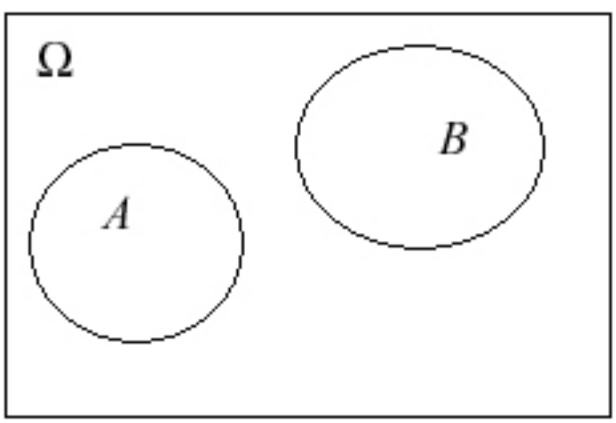
Le diagramme de Venn de la figure 6.6 rend compte de l'intersection de deux événements.

Figure 6.6
Diagramme de Venn de l'événement $A \cap B$



Deux événements peuvent parfois être incompatibles : ce sont des événements qui ne peuvent se réaliser simultanément. On dit de ces événements qu'ils sont mutuellement exclusifs. Par exemple, au lancer des deux dés, l'événement A est d'obtenir une paire et l'événement B égale « la somme des chiffres est impair », alors on a : $A \cap B = \varnothing$, où $\varnothing$ est le symbole d'un ensemble vide. On utilise également, parfois, $\{\ \}$.

Figure 6.7
Diagramme de Venn de l'événement $A \cap B = \varnothing$



La notion de probabilité

La probabilité qu'un événement à caractère aléatoire se réalise est égale au quotient du nombre des résultats favorables à l'événement E divisé par le nombre de résultats possibles. En termes algébriques, la probabilité d'un événement E est notée $P(E)$, ou p, et se formule :

$$\boxed{P(E)=\frac{N(E)}{N(\Omega)}=\frac{\text{Nombre de résultats favorables à l'événement } E}{\text{Nombre de résultats possibles}}}$$

Nous avons vu précédemment, dans la description de l'expérience aléatoire, les notions d'ensemble fondamental et d'événement. Dans la formule précédente, Ω représente cet ensemble fondamental, alors que *E* représente l'événement de cet ensemble dont on cherche la probabilité.

Quelle est la probabilité que la somme des résultats obtenus en lançant deux dés soit 8? En appliquant cette formule, on obtient $P(E)=\frac{5}{36}=0{,}138=13{,}8\,\%$. En effet, sur les 36 résultats possibles, les cas favorables sont au nombre de cinq : $E=\{(6,\ 2),\ (5,\ 3),\ (4,\ 4),\ (3,\ 5),\ (2,\ 6)\}$.

La probabilité qu'un événement se réalise est toujours un nombre entre 0 (événement impossible) et 1 (événement certain) tel que : $0 \le P(E) \le 1$. De plus, la somme des probabilités de tous les événements est égale à 1, ou $P(E)=1$.

Notez que le résultat d'une probabilité peut s'exprimer de trois façons : en nombre fractionnaire inférieur ou égal à 1, en nombre décimal ou sous forme de pourcentage. Une probabilité ne peut évidemment pas dépasser 100 %.

EXEMPLE

On pige simultanément deux cartes dans un jeu de 52 cartes. Quelle est la probabilité d'obtenir deux cartes de cœur?

SOLUTION

$P(E)=\frac{N(E)}{N(\Omega)}$, où $N(E)$ est le nombre de possibilités d'obtenir deux cartes de cœur.

C'est dans des cas comme celui-ci qu'on doit recourir aux notions apprises en analyse combinatoire. Le nombre de résultats possibles, « piger deux cartes dans un jeu de 52 cartes », est une combinaison sans ordre et sans répétition :

$N(\Omega)=C_2^{52}=\frac{52!}{(52-2)!\times 2!}=\frac{52\times 51\times 50!}{50!\times 2\times 1}=\frac{52\times 51}{2\times 1}=1\ 326$ façons de choisir 2 cartes parmi les 52 du jeu de cartes.

Le nombre de résultats favorables à l'événement *E* (« piger deux cartes de cœur ») est aussi une combinaison sans ordre et sans répétition. Cette combinaison se traduit par :

$N(E)=C_2^{13}=\dfrac{13!}{(13-2)!\times 2!}=\dfrac{13\times 12\times 11!}{11!\times 2\times 1}=\dfrac{13\times 12}{2\times 1}=78$ façons de choisir 2 cartes parmi les 13 cartes de cœur.

La probabilité de piger deux cartes de cœur dans un jeu de 52 cartes est donc la suivante :

$$P(E)=\frac{N(E)}{N(\Omega)}=\frac{78}{1\ 326}=\frac{1}{17}=5{,}88\ \%$$

Il y a une chance sur 17 ou 5,88 % de chance que cette éventualité survienne.

Comme le laissait entendre la mise en situation du module, la probabilité peut se décliner de trois manières : la probabilité déductive (ou *a priori*), la probabilité « fréquentiste » (ou probabilité de la fréquence relative) et la probabilité subjective.

La *probabilité déductive* est le quotient du nombre des résultats favorables à l'événement E divisé par le nombre de résultats possibles. Dans le cas de la probabilité déductive, les événements ont les mêmes probabilités de se produire : on dit qu'ils sont équiprobables. Pour être valide, chacun des résultats possibles doit en effet avoir la même probabilité de se produire.

Cette approche de la probabilité est dite *a priori* car, si nous assumons par exemple qu'un dé n'est pas truqué, nous pouvons prévoir à l'avance un résultat. Au lancer d'un dé, la probabilité d'obtenir un 2 sera toujours d'une chance sur six (à moins de truquer le dé).

Utile lorsqu'il s'agit de prévoir le résultat au lancer d'un dé ou au tirage d'une carte, cette approche présente de sérieux inconvénients lorsqu'elle s'applique aux décisions de gestion. Le monde de la gestion n'est en effet ni aussi parfait, structuré ou ordonné que ne le laisse entendre cette approche. De nombreuses expériences, de nombreux tests et moult sondages doivent être menés avant tout calcul de probabilité. C'est ici le domaine de l'approche statistique basée sur la fréquence relative.

La *probabilité « fréquentiste »* (ou probabilité de la fréquence relative) est, tout comme la probabilité déductive, le quotient du nombre des résultats favorables à l'événement E divisé par le nombre de résultats possibles[7]. La différence tient aux types d'événements en cause. Dans le cas de la probabilité « fréquentiste », les événements ne sont pas équiprobables. Prenons l'exemple suivant : au cours de la prochaine année, un conducteur d'auto de 25 ans de la région de Québec sera impliqué dans un accident ou n'aura aucun accident. Les deux événements ne sont pas également probables. Pour

7. Selon la formule $P(E)=\dfrac{N(E)}{N(\Omega)}$. On peut également rencontrer la formule $P(E)=\dfrac{n_E}{N}$, où n_E est le nombre d'apparitions de l'événement E et N est une épreuve répétée N fois.

calculer la probabilité que ce conducteur de 25 ans ait un accident, il faut se baser sur une étude statistique de cet événement. Si sur 10 000 conducteurs de 25 ans, un an plus tard, 9 920 de ces conducteurs n'ont eu aucun accident alors :

$P(E)=\frac{10\ 000\text{-}9\ 920}{10\ 000}=\frac{80}{10\ 000}=0{,}008=0{,}8\ \%$, où $P(E)$ est la probabilité « fréquentiste » qu'un événement E se produise.

Prenons un autre exemple : une entreprise de votre région fabrique des puces électroniques. Elle en fabrique 20 000 par jour. Chaque jour, aléatoirement, on en vérifie 1 000 et on en trouve quatre défectueuses. La probabilité « fréquentiste » qu'une puce soit défectueuse est donc de : $P(E)=\frac{4}{1\ 000}=0{,}004=0{,}4\ \%$. Dans bien des situations, le nombre 1 000 est suffisant comme échantillon.

Quant à la *probabilité subjective*, c'est une probabilité qui se base sur l'intuition ou l'expérience d'une ou de plusieurs personnes dans un domaine donné. Elle aura une crédibilité dans la mesure où la ou les personnes qui l'expriment sont considérées comme compétentes dans le domaine concerné. Dans les mises en situation du début de ce module, une vieille dame affirmait qu'il y a eu, au Mexique, beaucoup de tremblements de terre et qu'il y en aura d'autres. Son expérience de vie particulière accorde à cette personne un bon degré de confiance. L'invention d'un médicament contre l'obésité fait dire au spécialiste en chef de l'équipe de chercheurs que l'utilisation de ce médicament selon les règles prescrites fera perdre à son utilisateur 30 % de son surplus de poids en un an. La valeur de cette affirmation se fonde sur la qualité et la crédibilité de l'équipe de chercheurs. C'est ce que l'on appelle une probabilité subjective.

Les lois de la probabilité

Au même titre que l'événement, la probabilité a ses lois lorsqu'il s'agit d'union ou de croisement de probabilités. Nous vous présentons les trois principales lois qui gouvernent la probabilité: la loi de l'union, la probabilité conditionnelle et la loi de l'intersection.

La loi de l'union

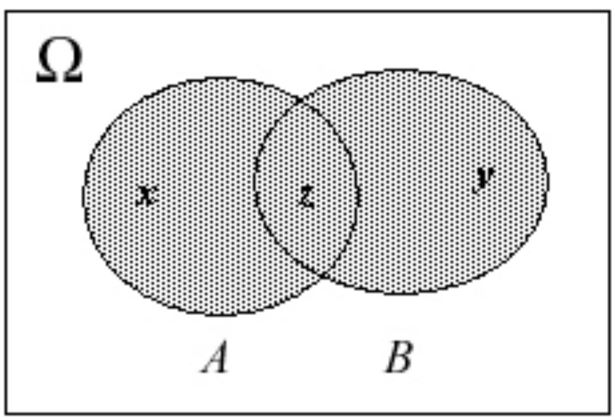
Soit deux événements A et B, la probabilité qu'au moins un des deux événements se réalise est égale à la somme des probabilités respectives de A et de B dont il faut soustraire la probabilité de leur réalisation simultanée. En termes mathématiques :

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

Cette équation est la règle d'addition des probabilités, aussi appelée la loi de l'union ou le théorème des probabilités totales.

Considérons le diagramme de Venn des deux événements compatibles A et B. Les trois régions de ce diagramme sont identifiées par les lettres x, y, z.

Figure 6.8
Diagramme de Venn des événements compatibles *A* et *B*



À partir de ce diagramme, on peut former les égalités suivantes :

$N(A) = x + z$, où $N(A)$ se lit « le nombre d'événements A »;

$N(B) = y + z$;

$N(A \cap B) = z$;

$N(A \cup B) = x + y + z$, ou encore :

$N(A \cup B) = N(A) + N(B) - N(A \cap B)$, car $(x+z)+(y+z)-z = x+y+z$.

EXEMPLE :

On pige une carte au hasard dans un jeu de 52 cartes. Quelles sont les probabilités respectives de piger[8] :

a) Une carte de trèfle?

b) Un valet?

c) Une figure?

d) Une figure et un trèfle?

e) Une figure ou un trèfle?

SOLUTION

a) Une carte de trèfle : $P(\text{Trèfle}) = \dfrac{13}{52} = 25\ \%$.

b) Un valet : $P(\text{Valet}) = \dfrac{4}{52} = 7{,}7\ \%$.

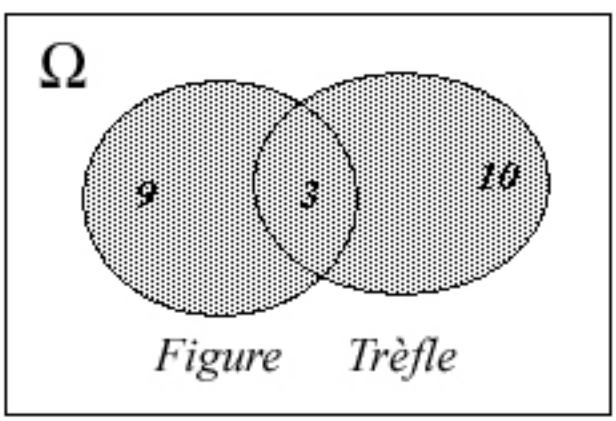
c) Une figure : $P(\text{Figure}) = \dfrac{12}{52} = \dfrac{3}{13} = 23{,}08\ \%$.

d) Une figure et un trèfle : $P(\text{Figure} \cap \text{Trèfle}) = \dfrac{3}{52} = 5{,}77\ \%$.

e) Une figure ou un trèfle : $P(\text{Figure} \cup \text{Trèfle}) = ?$

P (Figure $\cup$ Trèfle) se calcule à partir du diagramme de Venn suivant :

Figure 6.9
Diagramme de Venn de l'événement P(Figure $\cup$ Trèfle)



D'où $N(\text{Figure} \cup \text{Trèfle}) = 9 + 3 + 10 = 22$ et $P(\text{Figure} \cup \text{Trèfle}) = \dfrac{22}{52} = 42{,}31\ \%$.

8. Un jeu de cartes conventionnel comporte 4 couleurs (pique, cœur, carreau et trèfle), comprenant chacune 13 cartes dont 3 figures (roi, dame, valet).

Il est aussi possible d'utiliser la formule du théorème des probabilités totales :

P (Figure $\cup$ Trèfle) = P (Figure) + P (Trèfle) – P (Figure $\cap$ Trèfle)

P (Figure $\cup$ Trèfle) = 23,08 % + 25 % – 5,77 % = 42,31 %

EXEMPLE

Les dossiers d'un hôpital montrent que 12 % de tous les patients sont admis pour un traitement chirurgical, que 16 % sont admis en obstétrique et que 2 % reçoivent à la fois des traitements chirurgicaux et d'obstétrique. Si un nouveau patient est admis, quelle est la probabilité qu'il soit admis pour un traitement chirurgical, en obstétrique ou les deux à la fois?

SOLUTION

Les deux événements sont les suivants :
$A =$ {Un patient est admis pour un traitement chirurgical};
$B =$ {Un patient est admis en obstétrique}.

Selon les dossiers de l'hôpital, la probabilité qu'il soit admis pour un traitement chirurgical est $P(A) = 0{,}12$; la probabilité qu'il soit admis pour un traitement en obstétrique est $P(B) = 0{,}16$; et la probabilité que le patient reçoive à la fois un traitement chirurgical et en obstétrique est $P(A \cap B) = 0{,}02$.

Il s'ensuit alors que la probabilité qu'un patient soit admis pour un traitement chirurgical, en obstétrique ou les deux est égale à :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$= 0{,}12 + 0{,}16 - 0{,}02 = 0{,}26 \text{ ou } 26\,\%$$

La probabilité conditionnelle

C'est la probabilité qu'un événement A se produise, sachant qu'un événement B s'est produit. La probabilité conditionnelle s'écrit $P(A \mid B)$ et se lit « la probabilité de A étant donné B ». La formule utilisée pour rendre compte de la probabilité conditionnelle est :

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

EXEMPLE :

Trouvez la probabilité d'obtenir deux faces si vous lancez successivement trois pièces de monnaie et qu'au premier lancer vous avez obtenu face.

SOLUTION

L'ensemble fondamental de cet énoncé est :
$N(\Omega)=\{PPP, PPF, PFP, PFF, FPF, FPP, FFP, FFF\}$.

Dans cet énoncé, $P(A)$ égale la probabilité d'obtenir exactement deux faces alors que $P(B)$ égale la probabilité d'avoir obtenu face au premier lancer.

Le nombre de possibilités d'obtenir exactement deux faces est :

$N(A)=\{FPF, FFP, PFF\}$. Alors $P(A)=\dfrac{N(A)}{N(\Omega)}=\dfrac{3}{8}$. On a ici la probabilité d'obtenir deux faces.

Le nombre de possibilités d'obtenir face au premier lancer est :
$N(B)=\{FFF, FPP, FFP, FPF\}$

Donc $P(B)=\dfrac{N(A)}{N(\Omega)}=\dfrac{4}{8}=\dfrac{1}{2}$

L'intersection des deux possibilités se traduit par $N(A\cap B)=\{FPF, FFP\}$.

Donc, $P(A\cap B)=\dfrac{N(A\cap B)}{N(\Omega)}=\dfrac{2}{8}=\dfrac{1}{4}$

Mais ce que nous cherchons, c'est la probabilité d'obtenir deux faces étant donné que l'événement B s'est déjà réalisé. Si B s'est réalisé, alors l'espace échantillonnal $N(\Omega)$ est modifié. On aura un nouvel espace échantillonnal dont l'ensemble des résultats possibles devra comporter une face au premier lancer (première lettre) :

$N(\Omega)=\{FFF, FPF, FPP, FFP\}$; donc, $N(\Omega)=N(B)=4$
et :
$P(A|B)=\dfrac{P(A\cap B)}{P(B)}=\dfrac{2}{4}=\dfrac{1}{2}$

Une autre façon de faire est de calculer à partir des probabilités déjà calculées :

$$P(A\,|\,B)=\frac{P(A\cap B)}{P(B)}=\frac{\frac{1}{4}}{\frac{1}{2}}=\frac{1}{4}\times\frac{2}{1}=\frac{2}{4}=\frac{1}{2}$$

NOTES : La probabilité de A, comme telle, n'intervient pas dans le calcul de la probabilité conditionnelle.

La probabilité placée au dénominateur est la probabilité de la condition (celle de l'événement qui suit le « étant donné » ou le « si »).

EXEMPLE :

Un manufacturier d'automobiles a mené une vaste enquête scientifique portant sur les plaintes des acheteurs. Le résultat de cette enquête est reproduit au tableau suivant. D'après ce tableau, si un client appelle pour se plaindre, quelle est la probabilité que la raison de son appel soit l'aspect extérieur de l'automobile, sachant que la plainte survient durant la période de garantie du manufacturier?

Tableau 6.1
Les motifs de plainte des acheteurs

Période	Motif de la plainte			Total
	Mécanique	Circuit électrique	Aspect extérieur	
Durant la période de garantie	18 %	13 %	32 %	63 %
Après la période de garantie	12 %	22 %	3 %	37 %
Totaux	30 %	35 %	35 %	100 %

SOLUTION

Les deux événements sont les suivants :

$A =$ {Le motif de la plainte est l'aspect extérieur};
$B =$ {La plainte survient durant la période de garantie}.

La lecture de la dernière colonne du tableau montre que 63 % des plaintes surviennent durant la période de garantie; alors $P(B) = 0{,}63$. De plus, le pourcentage des plaintes concernant l'aspect extérieur de l'automobile survenues durant la période de garantie (l'événement $A \cap B$) est de 32 %; alors $P(A \cap B) = 0{,}32$. Avec ces deux éléments en mains, calculons la probabilité que la raison de l'appel du client soit l'aspect extérieur de l'automobile, sachant que la plainte survient durant la période de garantie du manufacturier :

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0{,}32}{0{,}63} = 0{,}51$$

Il s'ensuit que la probabilité qu'un client se plaigne de l'aspect extérieur de son automobile compte pour plus de la moitié des plaintes formulées durant la période de garantie.

La loi de l'intersection

La probabilité de l'intersection de deux événements A et B s'obtient par la multiplication de l'un des deux événements par la probabilité conditionnelle de l'autre. La loi de l'intersection découle directement de la formule qui définit la probabilité conditionnelle. Il suffit d'isoler $P(A \cap B)$ dans la formule de probabilité conditionnelle en multipliant les deux membres par $P(B)$.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A|B) \times P(B) = \frac{P(A \cap B)}{P(B)} \times P(B) \quad P(A|B) \times P(B) = P(A \cap B)$$

On obtient ainsi la loi de l'intersection ou le théorème des probabilités composées :

$$P(A \cap B) = P(B) \times P(A|B)$$

ou

$$P(A \cap B) = P(A) \times P(B|A)$$

Attention : la probabilité de l'intersection n'est pas le produit des probabilités des 2 événements simples, mais le produit d'une probabilité conditionnelle par une probabilité d'un événement simple.

EXEMPLE :

On pige simultanément deux cartes dans un jeu de 52 cartes. Quelle est la probabilité d'obtenir deux as?

SOLUTION

Deux méthodes peuvent être utilisées pour résoudre ce problème. La première méthode utilise la loi de l'intersection que nous venons de voir : si A (La première carte est un as) et B (La deuxième carte est un as), sachant que A s'est produit, alors :

$$P(A \cap B) = P(A) \times P(B|A) = \frac{4}{52} \times \frac{3}{51} = \frac{12}{2\ 652} = 0{,}00452 = 0{,}452\ \%$$

Il est aussi possible d'appliquer la formule de l'analyse combinatoire. On sait que la probabilité d'obtenir deux as est :

$$P(2\ \text{As}) = \frac{\text{Nombre de résultats favorables à l'événement « Obtenir deux as »}}{\text{Nombre de résultats possibles}}$$

Le nombre de résultats possibles est le dénombrement de toutes les dispositions possibles de deux cartes parmi les 52 cartes du jeu. Le nombre de tous les résultats possibles équivaut donc à trouver les combinaisons de deux cartes parmi 52 :

$$C_2^{52} = \frac{52!}{(52-2)! \times 2!} = \frac{52 \times 51 \times 50!}{50! \times 2 \times 1} = \frac{52 \times 51}{2 \times 1} = 1326 \text{ combinaisons.}$$

Par ailleurs, le nombre de résultats favorables à l'événement « obtenir deux as » sur les quatre que comporte un jeu de cartes, également une combinaison, est égal à :

$C_2^4 = \frac{4!}{(4-2)! \times 2!} = \frac{4 \times 3 \times 2!}{2! \times 2 \times 1} = \frac{4 \times 3}{2 \times 1} = 6$ combinaisons.

La probabilité d'obtenir deux as est donc égale à :

$$P(2\ \text{As}) = \frac{C_2^4}{C_2^{52}} = \frac{\cancel{\frac{4!}{(4-2)! \times 2!}}}{\cancel{\frac{52!}{(52-2)! \times 2!}}} = \frac{4 \times 3 \times 2! / \cancel{2! \times 2 \times 1}}{52 \times 51 \times 50! / \cancel{50! \times 2 \times 1}} = \frac{4 \times 3 / \cancel{2 \times 1}}{52 \times 51 / \cancel{\times 2 \times 1}} = \frac{6}{1\ 326} = 0{,}00452 = 0{,}452\ \%$$

Méthodes de résolution de problèmes de probabilité

Tout au long de l'étude de la probabilité, certains outils ont été suggérés pour faciliter la résolution des problèmes de probabilité : la construction d'un diagramme en arbre, l'utilisation de l'analyse combinatoire ou la mise en forme des données dans un tableau. Faisons un bref rappel de ces outils.

Construction d'un diagramme en arbre

Lorsque le nombre de possibilités est petit, il est souvent souhaitable de débuter la résolution d'un problème par la construction d'un diagramme en arbre. Le diagramme en arbre permet de représenter visuellement toutes les réponses possibles.

EXEMPLE :

Une urne contient six boules rouges, trois boules noires et cinq boules vertes. On tire au hasard et sans remise deux boules. Quelle est la probabilité d'en piger au moins une verte?

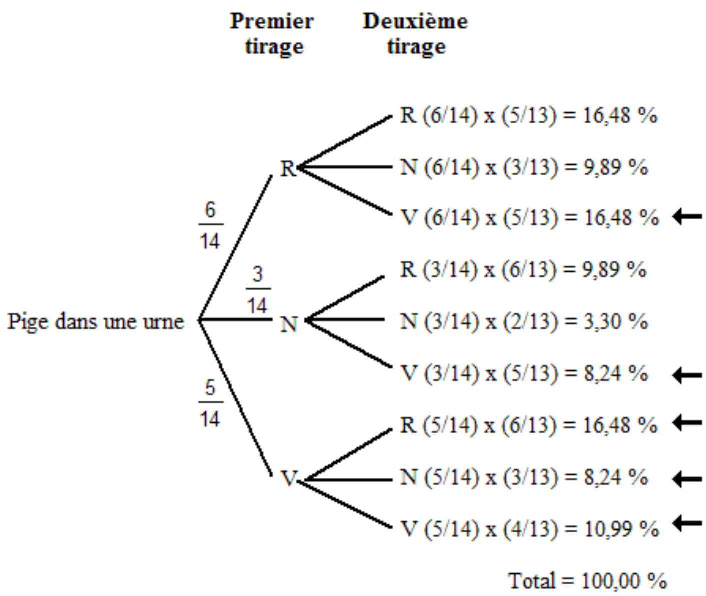
SOLUTION

Il y a quatorze boules au total. La probabilité d'en piger une verte au premier tirage est de 5/14. Une autre possibilité est de tirer en premier une boule rouge (6/14) et ensuite une boule verte (5/13). Finalement, on peut tirer en premier une boule noire (3/14) et en deuxième une boule verte (5/13).

$$P(\text{Piger au moins une verte}) = \left(\frac{5}{14}\right) + \left(\frac{6}{14} \times \frac{5}{13}\right) + \left(\frac{3}{14} \times \frac{5}{13}\right) = \frac{110}{182} = 60{,}43\%$$

Voici le diagramme en arbre des probabilités :

Figure 6.10
Diagramme en arbre de piges dans une urne



Dans ce problème, piger au moins une boule verte signifie additionner les cinq embranchements contenant une boule verte (V) :

$$(R \cap V) \text{ ou } (N \cap V) \text{ ou } (V \cap V) \text{ ou } (V \cap R) \text{ ou } (V \cap N)$$

$$16,48\ \% + 8,24\ \% + 10,99\ \% + 16,48\ \% + 8,24\ \% = 60,43\ \%$$

EXEMPLE :

Examinons maintenant, à partir de l'exemple qui précède, la probabilité de tirer simultanément (ou sans remise) deux boules de même couleur.

SOLUTION

À partir du diagramme en arbre des probabilités, on peut établir que :

$$P(\text{2 boules de même couleur}) = P(\text{2R}) + P(\text{2N}) + P(\text{2V}) = 16,48\ \% + 3,30\ \% + 10,99\ \% + = 30,77\ \%$$

Utilisation de l'analyse combinatoire

Dans la plupart des problèmes de probabilité à résoudre, l'utilisation du diagramme en arbre n'est que de peu d'utilité. La raison est fort simple et tient dans le trop grand nombre de possibilités à manipuler. Si tel est le cas, l'utilisation de l'analyse combinatoire est conseillée. Par exemple, à partir des connaissances acquises en analyse combinatoire, nous pouvons facilement résoudre le problème décrit précédemment.

SOLUTION

Nous cherchons ici la probabilité de piger deux boules rouges *ou* deux boules noires *ou* deux boules vertes. Le nombre de cas possibles de cette pige est de 91. Il s'agit d'une combinaison de deux parmi quatorze.

$$C_2^{14} = \frac{14!}{(14-2)!\times 2!} = \frac{14\times 13\times 12!}{12!\times 2\times 1} = \frac{14\times 13}{2\times 1} = 91$$

Les cas favorables sont respectivement :

$$C_2^{6} = \frac{6!}{(6-2)!\times 2!} = \frac{6\times 5\times 5!}{4!\times 2\times 1} = \frac{6\times 5}{2\times 1} = 15 \text{, d'où } P(2\text{R}) = \frac{15}{91} = 16{,}48\%$$

$$C_2^{3} = \frac{3!}{(3-2)!\times 2!} = \frac{3\times 2!}{1!\times 2!} = \frac{3}{1} = 3 \text{, d'où } P(2\text{N}) = \frac{3}{91} = 3{,}30\%$$

$$C_2^{5} = \frac{5!}{(5-2)!\times 2!} = \frac{5\times 4\times 3!}{3!\times 2\times 1} = \frac{5\times 4}{2} = 10 \text{, d'où } P(2\text{V}) = \frac{10}{91} = 10{,}99\%$$

Si on fait la somme des trois cas précédents, on obtient 30,77 %. On a la même réponse en utilisant un diagramme en arbre et l'analyse combinatoire.

Utilisation d'un tableau

Dans certains problèmes de probabilité, il arrive parfois que les données relatives à une situation soient nombreuses. On a alors recours à un tableau pour les classer en effectifs ou en fréquences.

EXEMPLE :

Afin de planifier une campagne de recrutement, on fait un sondage auprès des 382 étudiants d'une école secondaire. On demande à chaque étudiant le nombre d'enfants dans sa famille. Voici les résultats obtenus :

Tableau 6.2
Le nombre d'enfants dans les familles des étudiants d'une école secondaire

Nombre d'enfants	1	2	3	4	5	6
Nombre d'étudiants	116	127	85	32	17	5

Si on choisit au hasard l'un des 382 étudiants, quelle est la probabilité que cette personne ait trois frères ou sœurs?

SOLUTION

On cherche donc la probabilité $P(E)$ qu'il y ait quatre enfants dans cette famille.

$$P(E)=\frac{N(E)}{N(\Omega)}=\frac{32}{382}=0,08376=8,38\ \%$$

NOTE : Faites les exercices de la section 2 dans le *Recueil des activités pratiques* avant de continuer la lecture

Section 3 : la distribution normale

Lawrence-T. Dayhaw, dans son *Manuel de statistique* (1979), résume ainsi la petite histoire de la distribution (ou courbe) normale :

« À la suite de nombreuses mesures de traits physiques ou de traits mentaux, de rendements scolaires, de phénomènes économiques, biologiques, physiques, etc., divers auteurs, travaillant en divers pays et employant diverses échelles de mesure, sont parvenus à des répartitions numériques et graphiques de leurs résultats qui présentaient toutes une frappante ressemblance entre elles. Ces répartitions tendaient à se rapprocher d'une courbe dont la forme suggérait le profil d'une cloche.

[Le mathématicien] Gauss arriva à une solution mathématique [pour expliquer ces ressemblances] et formula une loi qui permettait de calculer la probabilité des diverses variations de grandeur du phénomène. […] En rapprochant les deux genres de données, les chercheurs furent saisis de l'analogie frappante entre la courbe théorique de Gauss et les histogrammes obtenus de fait. […] Ils décidèrent alors de substituer aux histogrammes empiriques la courbe mathématique de Gauss […]. »

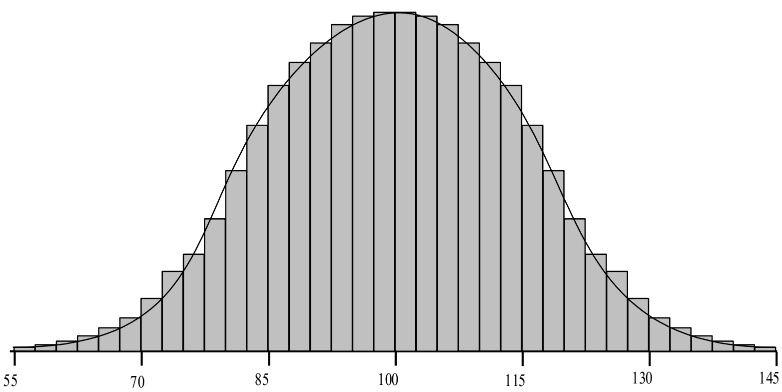
Plusieurs phénomènes génèrent des variables avec des distributions de données qui se rapprochent d'une distribution normale : distribution de probabilités des pluies annuelles pour une région donnée, résultats à un test d'aptitude ou à un test psychologique, caractéristiques biologiques (taille, poids) de différentes populations, humaines ou animales.

L'intérêt que nous portons à la distribution normale, à l'instar de Gauss, tient essentiellement à ce qu'elle permet de calculer la probabilité des diverses variations de grandeur d'un phénomène aléatoire. Nous verrons un peu plus loin quelle est la procédure à suivre pour effectuer ces calculs. Mais d'abord, qu'est-ce qu'une distribution normale?

La distribution normale : définition

Une distribution normale est une distribution de probabilités d'un caractère (ou variable) aléatoire continu dont la dispersion des valeurs autour de la moyenne est parfaitement symétrique. Les données d'une telle distribution, données qui proviennent d'une expérience aléatoire, se présenteraient sous la forme graphique suivante :

Figure 6.11
La distribution des résultats à un test d'intelligence (exemple fictif)

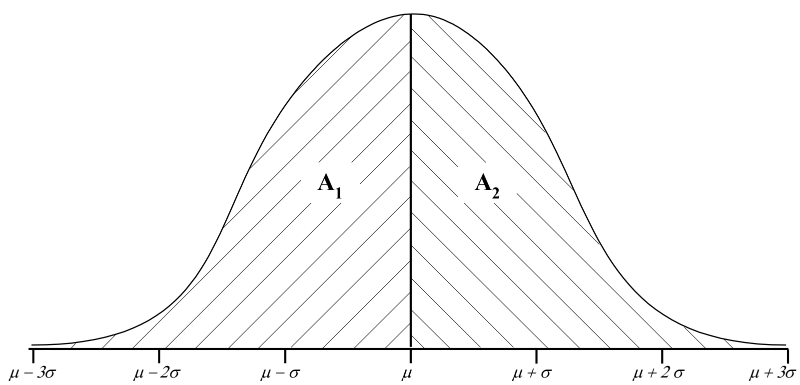


On constate que cette courbe, plus élevée au centre, descend symétriquement à gauche et à droite, de manière plus prononcée au début, puis plus lentement aux extrémités. Épousant la forme d'une cloche, la courbe d'une distribution normale se poursuit théoriquement jusqu'à l'infini de chaque côté. Dans cette distribution, la moyenne, la médiane et le mode ont la même valeur et ces trois mesures de tendance centrale se superposent au centre de la distribution. Dans l'exemple ci-dessus, cette valeur moyenne est de 100 alors que l'écart type est de 15.

Le mathématicien Gauss n'a pas été sans remarquer que les propriétés de cette courbe, particulièrement deux, la rendaient intéressante pour le calcul des probabilités :

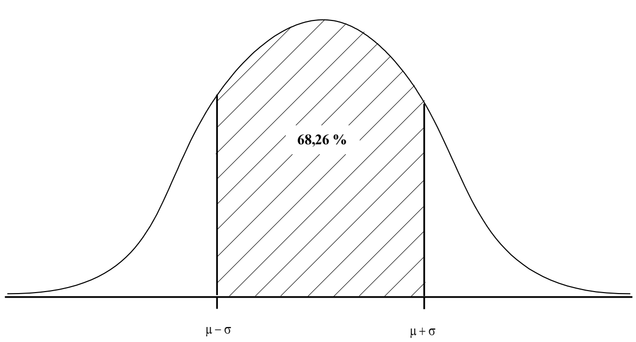
1. La courbe normale est dite **symétrique par rapport à la moyenne**. La moyenne sépare en effet la courbe en deux parties égales. La totalité (100 %) des valeurs se distribuant sous cette courbe, les aires à gauche et à droite de la moyenne équivalent donc chacune à la moitié de l'aire (ou 50 % ou 0,5, selon que l'on calcule la surface totale sous la courbe en pourcentage ou en nombre décimal). En raison de cette symétrie, l'aire située d'un côté de la moyenne (A_1) est une **image en miroir** de l'aire située de l'autre côté de la moyenne (A_2).

Figure 6.12
La symétrie de la courbe normale

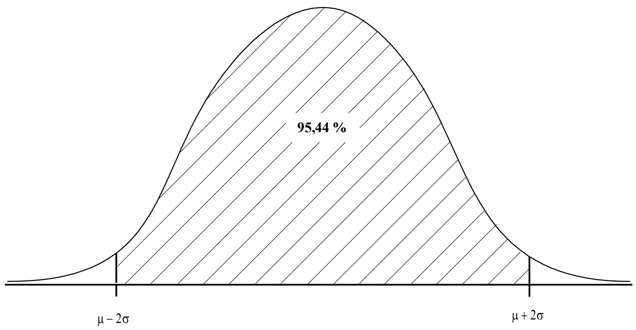


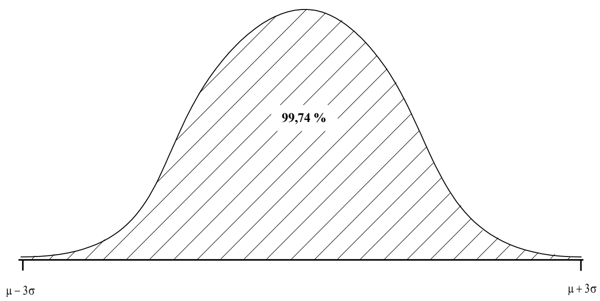
2. Comme la distribution des variables aléatoires est symétrique et en forme de cloche, environ **68 %** des données de l'ensemble se retrouveront à l'intérieur d'un intervalle compris entre « un écart type de la moyenne[9] » ($\mu + \sigma$) et « moins un écart type de la moyenne » ($\mu - \sigma$); environ **95 %** des données de l'ensemble se retrouveront à l'intérieur d'un intervalle compris entre « deux écarts types de la moyenne » ($\mu + 2\sigma$) et « moins deux écarts types de la moyenne » ($\mu - 2\sigma$); près de **100 %** des données de l'ensemble se retrouveront à l'intérieur d'un intervalle compris entre « trois écarts types de la moyenne » ($\mu + 3\sigma$) et « moins trois écarts types de la moyenne » ($\mu - 3\sigma$). Les trois graphiques de la figure 6.13 rendent compte de cette propriété appelée *règle empirique*.

Figure 6.13
Les probabilités totales en fonction de l'écart à la moyenne



9. Pour un rappel des notions de moyenne et d'écart type, veuillez vous référer au module 5, section 4.





Ces propriétés font que le mathématicien Gauss a pu proposer une formule qui, après transformation des données, permet de calculer l'aire sous n'importe quelle partie de la courbe et ainsi déterminer les probabilités d'occurrence des événements. Pourquoi transformer les données? Parce que les valeurs changent d'une courbe normale à l'autre selon les valeurs de la moyenne et de l'écart type. Ainsi, dans l'exemple de la figure 6.11, la moyenne est égale à 100 et l'écart type est de 15. Trouver les valeurs des aires sous la distribution de ces données implique d'effectuer une série de calculs complexes (à l'aide du calcul intégral) et de construire des tables pour les valeurs recherchées. La situation se complique encore du fait qu'il y a une infinité de courbes normales possibles. On contourne ce problème en ramenant toutes les courbes normales à une seule courbe par la *standardisation* (ou *centralisation*) des données.

Cette standardisation ou transformation des données consiste à ramener toutes les mesures à une moyenne μ égale à 0 et à un écart type σ égale à 1. On nomme *distribution centrée réduite* la distribution qui en résulte. La transformation linéaire des données s'effectue en soustrayant la moyenne μ de chaque mesure individuelle puis en divisant le résultat obtenu par l'écart type σ :

$z = \dfrac{x-\mu}{\sigma}$, (notez que si $x = \mu, \text{alors } z = 0$)

Dans l'exemple de la distribution des résultats à un test d'intelligence (figure 6.11), le z de la valeur 125 serait :

$$z = \frac{125-100}{15} = \frac{25}{15} = 1{,}67$$

Ce symbole, z, est appelé la cote z ou la cote standard. Ainsi, si x est une variable aléatoire normalement distribuée de moyenne μ et d'écart type σ, alors $z=\frac{x-\mu}{\sigma}$ se distribuera normalement avec une moyenne de 0 et un écart type de 1. Grâce à ces transformations, il suffit de construire une seule table d'aires sous la surface de la courbe normale, celle de la variable centrée réduite z. Pour faciliter le calcul (et nous faciliter la vie), les mathématiciens ont conçu une table d'aire pour la courbe normale centrée réduite. Cette table est reproduite en annexe à la fin de ce module. Vous remarquerez que la table en annexe ne donne que les valeurs de la moitié de droite de la distribution, la moitié de gauche étant équivalente (à cause de la symétrie de la courbe normale). Ainsi, dans l'exemple précédent, la cote z de 125 a la même valeur dans la table que la cote z de 75, car l'aire comprise entre 0 et 1,33 est la même que celle comprise entre 0 et –1,33.

La transformation des données en cote z a deux avantages : elle permet la comparaison des données de distribution numériquement différentes et elle simplifie le calcul des probabilités; en effet, lorsque les résultats x sont distribués normalement, les cotes z prennent toujours des valeurs autour d'une moyenne égale à 0 et d'un écart type égal à 1.

Le calcul des probabilités pour une distribution normale

Le calcul des probabilités utilisant les propriétés de la distribution normale est simple. Il tient en trois étapes :

1. À l'aide de la formule $z=\frac{x-\mu}{\sigma}$, calculez la cote z de la ou des variables x à transformer.
2. Référez-vous à la table de l'annexe 1[10]. Trouvez d'abord, dans la colonne de gauche, la valeur de la cote z correspondant à l'unité et à la première décimale (par exemple, 1,7 pour la cote $z=1{,}75$). Repérez ensuite, sur la ligne du haut, le chiffre correspondant à la seconde décimale de la cote z (par exemple, 0,05 pour la cote $z=1{,}75$). Le nombre qui coupe ces deux valeurs représente la probabilité que la mesure se situe entre la moyenne μ et la variable x.
3. Complétez vos calculs en additionnant ou en soustrayant, selon le cas, les valeurs obtenues.

NOTE : la probabilité totale (aire sous la courbe) de la loi normale est toujours de 1. La moitié (50 %) de cette aire est sous la moyenne et l'autre moitié est située au-dessus de la moyenne. Avec une courbe centrée réduite, il y a donc 50 % de probabilité d'avoir une valeur entre 0 et l'infini. La table de l'annexe 1 présente la probabilité (l'aire) entre 0 et z. Ainsi, si l'aire entre 0 et un z donné est de

10. L'annexe qui vous est présentée cumule la courbe à partir de la moyenne 0. D'autres tables cumulent les courbes de différentes façons, parfois à partir de l'extrémité gauche de la courbe.

0,3, la probabilité que la valeur soit supérieure à z est de 0,5 – 0,3 = 0,2. En utilisant la symétrie et le fait qu'il y ait 50 % de l'aire de part et d'autre de la moyenne, toutes les possibilités peuvent être calculées.

L'exemple qui suit démontre, graphiques à l'appui, les principaux types de calcul pouvant être obtenus lors du calcul des probabilités pour une distribution normale.

EXEMPLE :

Une entreprise fabrique des calculatrices de poche munies de piles rechargeables. Une série de tests sur 1 000 calculatrices prises au hasard a démontré qu'il fallait en moyenne 50 heures (avec un écart type de 15) avant qu'une pile de calculatrice ne doive être rechargée.

a) Quelle est la probabilité qu'une calculatrice fonctionne entre 50 et 70 heures (avant qu'elle ne doive être rechargée)?
b) Quelle est la probabilité qu'une calculatrice fonctionne entre 30 et 50 heures?
c) Quelle est la probabilité qu'une calculatrice fonctionne entre 30 et 70 heures?
d) Quelle est la probabilité qu'une calculatrice fonctionne moins de 70 heures?
e) Quelle est la probabilité qu'une calculatrice fonctionne plus de 70 heures?
f) Quelle est la probabilité qu'une calculatrice fonctionne moins de 30 heures?
g) Quelle est la probabilité qu'une calculatrice fonctionne moins de 30 ou plus de 70 heures?

SOLUTIONS

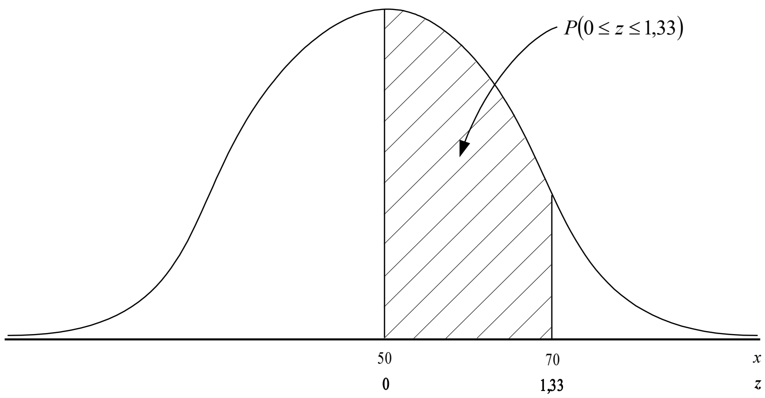
a) Mathématiquement, le problème à résoudre se lit comme suit, x étant la durée de fonctionnement de la calculatrice :

$P(50 \leq x \leq 70)$, $x \sim N(50;\ 15)$, qui se lit « la probabilité qu'une calculatrice fonctionne entre 50 et 70 heures, étant donné que x suit[11] une loi normale de moyenne 50 et d'écart type 15 ». Calculez d'abord les cotes z centrées réduites :

$$P\left(\frac{50-50}{15} \leq z \leq \frac{70-50}{15}\right),\ z \sim N(0;\ 1) \Rightarrow P\left(0 \leq z \leq \frac{20}{15}\right) \Rightarrow P(0 \leq z \leq 1{,}33)$$

11. Le symbole ~ se lit « suit ».

Figure 6.14 $P(0 \le z \le 1{,}33)$
La probabilité



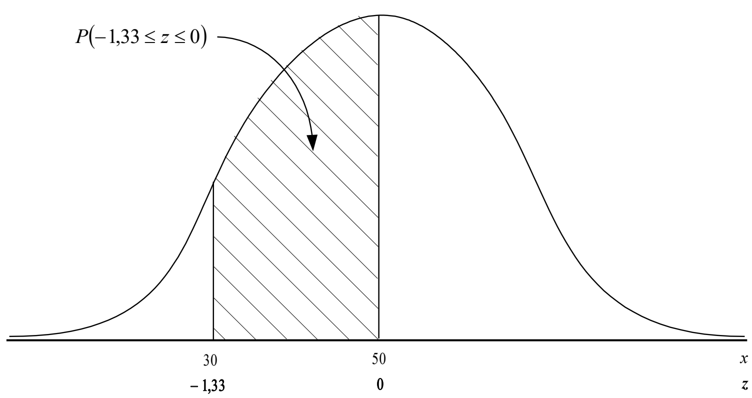
Cherchez ensuite, à l'aide de la table en annexe 1, l'aire sous la courbe (la probabilité) associée à cette valeur de z. Cette valeur est égale à 0,4082.

La probabilité qu'une calculatrice fonctionne entre 50 et 70 heures, $P(50 \le x \le 70)$, est donc égale à 0,4082 ou 40,82 %.

b) $P(30 \le x \le 50)$, $x \sim N(50;\ 15)$

$$\Rightarrow P\left(\frac{30-50}{15} \le z \le \frac{50-50}{15}\right),\ z \sim N(0;\ 1) \Rightarrow P\left(-\frac{20}{15} \le z \le 0\right) \Rightarrow P(-1{,}33 \le z \le 0)$$

Figure 6.15
La probabilité $P(-1{,}33 \le z \le 0)$



Déterminez la probabilité associée à la valeur de $z = -1{,}33$. En raison de la symétrie de la courbe, cette valeur équivaut à $z = -1{,}33$ et correspond, dans la table, à 0,4082.

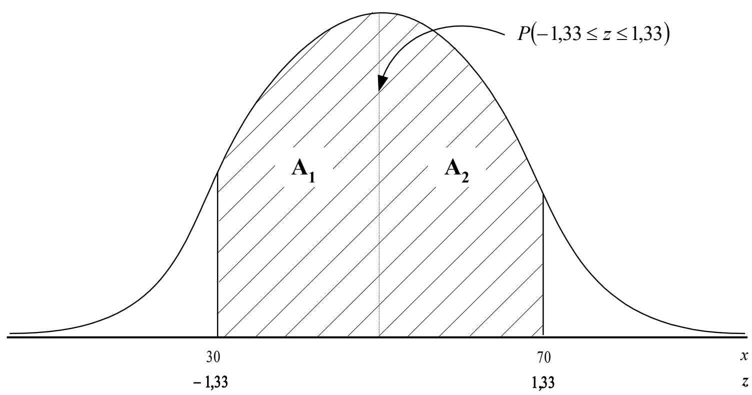
La probabilité qu'une calculatrice fonctionne entre 30 et 50 heures, $P(30 \le x \le 50)$, est donc égale à 0,4082 ou 40,82 %. Comme la loi normale est symétrique, la probabilité que la pile dure 20 heures **de plus** que la moyenne est égale à la probabilité qu'elle dure 20 heures **de moins** que la moyenne.

c) $P(30 \le x \le 70)$, $x \sim N(50;\ 15)$

$\Rightarrow P\left(\frac{30-50}{15} \le z \le \frac{70-50}{15}\right)$, $z \sim N(0;\ 1) \Rightarrow P\left(-\frac{20}{15} \le z \le \frac{20}{15}\right)$

$\Rightarrow P(-1{,}33 \le z \le 1{,}33)$

Figure 6.16
La probabilité $P(-1{,}33 \le z \le 1{,}33)$

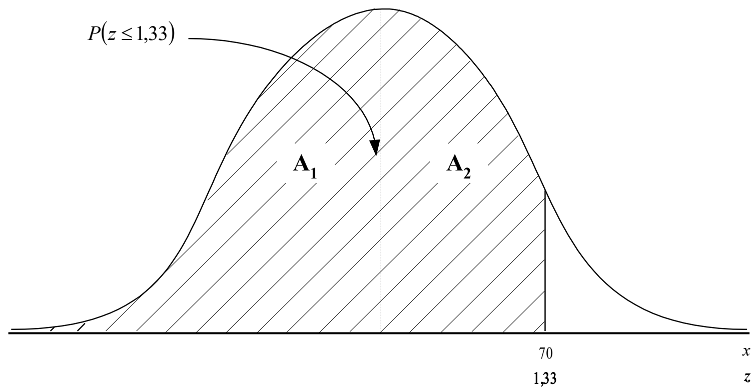


Nous cherchons donc l'aire sous la courbe qui est comprise entre –1,33 et 1,33. Le tableau de l'annexe 1 indique que l'aire entre 0 et 1,33 est de 0,4082, et que l'aire entre 0 et –1,33 (ou 1,33) est de 0,4082. L'aire totale sera donc de 0,8164 ($A_1 + A_2 = 0{,}4082 + 0{,}4082$). La probabilité qu'une calculatrice fonctionne entre 30 et 70 heures, $P(30 \le x \le 70)$, est donc égale à 0,8164 ou 81,64 %.

d) $P(x \le 70)$, $x \sim N(50;\ 15)$

$\Rightarrow P\left(z \le \frac{70-50}{15}\right)$, $z \sim N(0;\ 1) \Rightarrow P\left(z \le \frac{20}{15}\right) \Rightarrow P(z \le 1{,}33)$

Figure 6.17
La probabilité $P(z \le 1{,}33)$



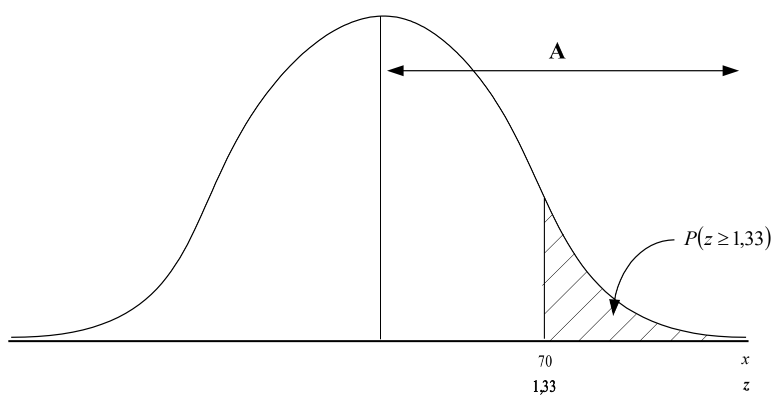
Ici, nous avons un cas spécial, car nous ne cherchons pas l'aire située entre 0 et 1,33, mais l'aire totale sous 1,33. Sachant que l'aire sous la courbe à gauche de la moyenne (A_1) est égale à 0,5 (50 %), et que l'aire sous la courbe associée à la valeur de $z = 1{,}33$ (A_2) est de 0,4082, il suffit d'additionner les deux aires : 0,5 + 0,4082 = 0,9082.

La probabilité qu'une calculatrice fonctionne moins de 70 heures[12], $P(x \le 70)$, est donc égale à 0,9082 ou 90,82 %.

e) $P(x \ge 70)$, $x \sim N(50;\ 15)$

$$\Rightarrow P\left(z \ge \frac{70-50}{15}\right),\ z \sim N(0;\ 1) \Rightarrow P\left(z \ge \frac{20}{15}\right) \Rightarrow P(z \ge 1{,}33)$$

Figure 6.18
La probabilité $P(z \ge 1{,}33)$



12. Dans le calcul des probabilités, les signes > et ≥ ainsi que < et ≤ sont équivalents.

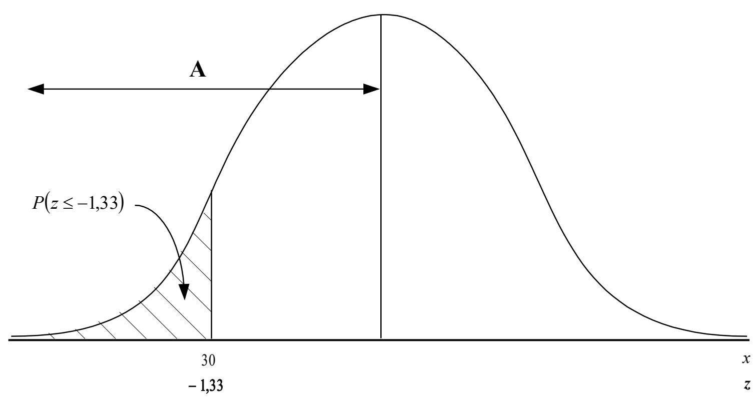
Cherchez l'aire sous la courbe (la probabilité) associée à cette valeur de z. Cette valeur est égale à 0,4082. Nous savons que l'aire totale de A est de 0,5. Nous savons également que l'aire entre 0 et 1,33 est de 0,4082. L'aire recherchée est donc de 0,5 –0,4082 = 0,0918.

La probabilité qu'une calculatrice fonctionne plus de 70 heures, $P(x \geq 70)$, est donc égale à 0,0918 ou 9,18 %.

f) $P(x \leq 30)$, $x \sim N(50;\ 15)$

$$\Rightarrow\ P\left(z \leq \frac{30-50}{15}\right),\ z \sim N(0;\ 1)\ \Rightarrow\ P\left(z \leq -\frac{20}{15}\right)\ \Rightarrow\ P(z \leq -1{,}33)$$

Figure 6.19
La probabilité $P(z \leq -1{,}33)$



Cherchez l'aire sous la courbe (la probabilité) associée à cette valeur de $z = -1{,}33$ (ou 1,33). Cette valeur est égale à 0,4082. La loi normale étant symétrique, nous savons que l'aire totale de A est de 0,5. Nous savons également que l'aire entre 0 et –1,33 est de 0,4082. L'aire recherchée est donc de 0,5 –0,4082 = 0,0918.

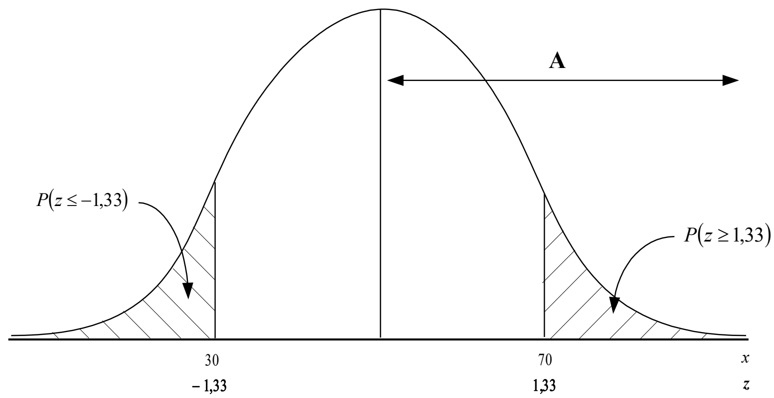
La probabilité qu'une calculatrice fonctionne moins de 30 heures, $P(x \leq 30)$, est donc égale à 0,0918 ou 9,18 %.

g) $P(x \leq 30 \,\text{ou}\, x \geq 70)$, $x \sim N(50;\ 15)$

$$\Rightarrow\ P\left(z \leq \frac{30-50}{15}\right)\text{ou}\left(z \geq \frac{70-50}{15}\right),\ z \sim N(0;\ 1)\ \Rightarrow\ P\left(z \leq -\frac{20}{15}\right)\text{ou}\left(z \geq \frac{20}{15}\right)$$

$$\Rightarrow\ P(z \leq -1{,}33 \,\text{ou}\, z \geq 1{,}33)$$

Figure 6.20 $P(z \leq -1{,}33 \text{ ou } z \geq 1{,}33)$
La probabilité



Cherchez d'abord la première portion, soit l'aire sous la courbe (la probabilité) associée à la valeur de $z = 1{,}33$. Cette valeur est égale à 0,4082. Sachant que l'aire totale de A est de 0,5 et que l'aire entre 0 et 1,33 est de 0,4082, cette première portion est donc de 0,5 –0,4082 = 0,0918. Par la règle de symétrie, nous pouvons déduire la seconde portion de l'aire recherchée. L'addition des deux portions donne 0,1836. La probabilité qu'une calculatrice fonctionne moins de 30 heures ou plus de 70 heures est donc égale à 0,1836 ou 18,36 %.

EXEMPLE :

Une ampoule électrique a une durée de vie moyenne de 1 000 heures. La durée de vie de 10 000 de ces ampoules est considérée comme étant distribuée suivant la loi normale et l'écart type est de 200 heures.

a) Quelle est la probabilité qu'une ampoule tirée au hasard de ce lot de 10 000 ampoules ait une durée de vie supérieure à 1 300 heures?

b) Quelle est la probabilité qu'une de ces ampoules tirée au hasard ait une durée inférieure à 650 heures?

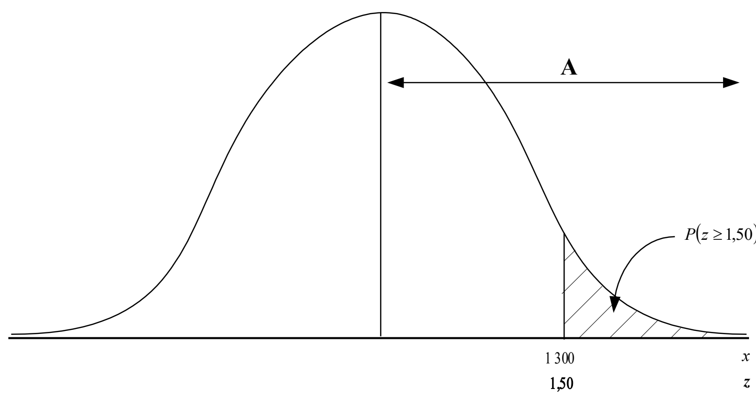
c) Quelle est la probabilité qu'une de ces ampoules tirée au hasard ait une durée de vie entre 650 heures et 1 300 heures?

SOLUTIONS

a) $P(x \geq 1\,300)$, $x \sim N(1\,000;\ 200)$

$$\Rightarrow P\left(z \geq \frac{1\,300 - 1\,000}{200}\right),\ z \sim N(0;\ 1) \Rightarrow P\left(z \geq \frac{300}{200}\right) \Rightarrow P(z \geq 1{,}50)$$

Figure 6.21
La probabilité $P(z \geq 1{,}50)$

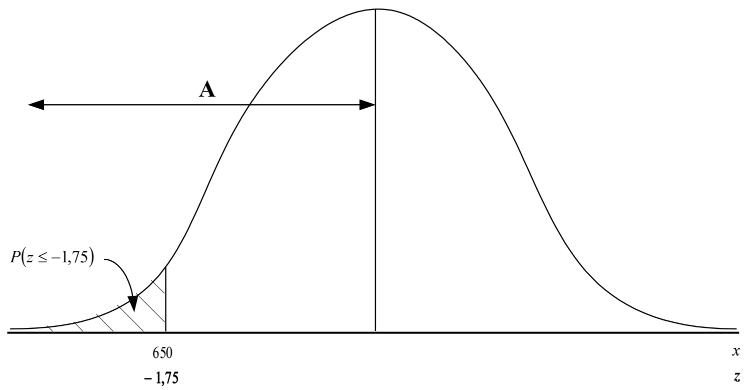


Cherchez l'aire sous la courbe (la probabilité) associée à cette valeur de z. Cette valeur est égale à 0,4332. Nous savons que l'aire totale de A est de 0,5. Nous savons également que l'aire entre 0 et 1,50 est de 0,4332. L'aire recherchée est donc de 0,5 –0,4332 = 0,0668. La probabilité qu'une ampoule tirée au hasard de ce lot de 10 000 ampoules ait une durée de vie supérieure à 1 300 heures, $P(x \geq 1\,300)$, est donc égale à 0,0668 ou 6,68 %.

b) $P(x \leq 650)$, $x \sim N(1\,000;\ 200)$

$$\Rightarrow P\left(z \leq \frac{650 - 1\,000}{200}\right),\ z \sim N(0;\ 1) \Rightarrow P\left(z \leq -\frac{350}{200}\right) \Rightarrow P(z \leq -1{,}75)$$

Figure 6.22
La probabilité $P(z \leq -1{,}75)$



Cherchez l'aire sous la courbe (la probabilité) associée à cette valeur de z. Cette valeur est égale à 0,4599. Elle représente la probabilité d'être entre la moyenne et $z = 1{,}75$ (ou $z = -1{,}75$). La loi normale étant symétrique, nous savons que l'aire totale de A est de 0,5. Nous savons également que l'aire entre 0 et –1,75 est de 0,4599. L'aire recherchée est donc de 0,5 –0, 4599 = 0, 0401.

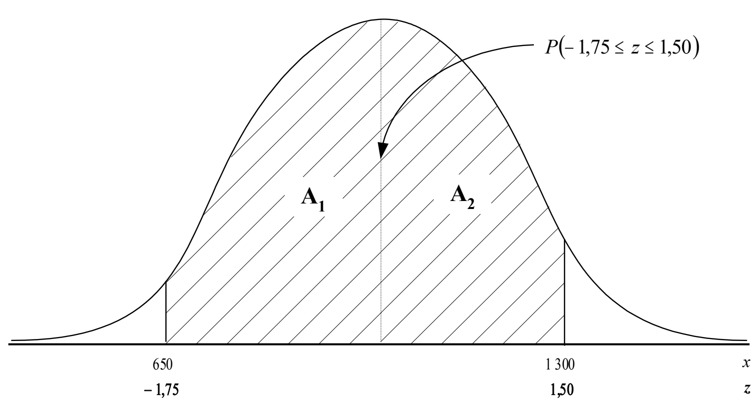
La probabilité qu'une de ces ampoules tirée au hasard ait une durée inférieure à 650 heures, $P(x \leq 650)$, est donc égale à 0,0401 ou 4,01 %.

c) $P(650 \leq x \leq 1\,300)$, $x \sim N(1\,000;\ 200)$

$$\Rightarrow P\left(\frac{650 - 1\,000}{200} \leq z \leq \frac{1\,300 - 1\,000}{200}\right),\ z \sim N(0;\ 1) \Rightarrow P\left(-\frac{350}{200} \leq z \leq \frac{300}{200}\right)$$

$$\Rightarrow P(-1{,}75 \leq z \leq 1{,}50)$$

Figure 6.23
La probabilité $P(-1{,}75 \leq z \leq 1{,}50)$



Nous cherchons donc l'aire sous la courbe qui est comprise entre –1,75 et 1,50. Comme le tableau de l'annexe 1 nous dit que l'aire entre 0 et 1,50 est de 0,4332 et que l'aire entre 0 et –1,75 (ou 1,75) est de 0,4599, l'aire totale sera de 0,8931 ($A_1 + A_2 = 0,4599 + 0,4332$).

La probabilité qu'une de ces ampoules tirée au hasard ait une durée de vie entre 650 heures et 1 300 heures, $P(650 \leq x \leq 1\,300)$, est donc égale à 0,8931 ou 89,31 %.

Nous ne saurions terminer sans faire la mise en garde suivante : rappelez-vous toujours que ces calculs valent uniquement pour les valeurs distribuées normalement. Les distributions « anormales » (la distribution des salaires, la distribution des ventes d'une entreprise, la distribution de la durée de vie de composantes électroniques ou mécaniques) sont légion; elles ne peuvent donc utiliser les propriétés de calcul des probabilités de la distribution normale.

Autre remarque : si vous voulez que la recherche des valeurs soit facile, faites toujours une esquisse du graphique de la courbe et déterminez la région dont vous devez déterminer l'aire pour connaître la probabilité de la variable aléatoire. Il n'est aucunement nécessaire que ce graphique soit précis, mais il doit permettre d'identifier l'aire à déterminer.

NOTE : Faites les exercices de la section 3 dans le *Recueil des activités pratiques* avant de continuer la lecture.

Section 4 : l'échantillonnage

L'échantillonnage est l'ensemble des opérations à faire pour déterminer un échantillon. Échantillonner, c'est choisir aléatoirement, dans un univers (de personnes, d'animaux, d'objets) un sous-ensemble de manière à obtenir un échantillon représentatif de cet univers. Par exemple, lorsqu'il s'agit de l'ensemble des consommateurs d'une grande ville, l'échantillon est constitué d'un certain nombre d'individus pris au hasard parmi ces consommateurs. Pour un test de qualité ou de fiabilité d'un produit, l'échantillon peut être un ensemble d'objets comme 1 000 ampoules de 100 watts tirées aléatoirement d'une production quotidienne de 100 000 ampoules.

L'échantillon ne représente qu'une petite partie de l'univers qui doit être étudié par les statisticiens. Il est donc important de préciser dès le départ qu'une étude statistique ne peut jamais être aussi précise qu'un recensement, car un recensement inclut tous les individus d'une population. Mais plusieurs raisons font qu'on préférera utiliser un échantillon plutôt que la population entière :

1. La trop grande taille de la population à étudier.
2. Les coûts et les contraintes financières reliés à cette étude.
3. Le temps alloué à cette étude.
4. Le degré de précision désiré.
5. Le test peut être destructif (aucune entreprise d'ampoules n'acceptera qu'on teste tout son stock d'ampoules).

Un collège d'enseignement privé de niveau secondaire désire se relocaliser en banlieue d'une grande ville, car la grande majorité de sa clientèle provient de ce milieu. Avant d'investir des millions de dollars, le conseil d'administration de cet établissement commande une étude statistique portant sur la rentabilité d'un tel projet.

Si la taille de la population concernée est d'environ 200 000 habitants, il serait long et coûteux de téléphoner ou d'expédier un questionnaire aux 200 000 personnes : un échantillon s'impose. Cet échantillon pourrait comporter 1 000 individus (le nombre d'individus sélectionnés dépend du degré de précision désirée). En choisissant de faire porter l'étude sur 1 000 habitants, il est évident que les coûts reliés à cette étude seront plus raisonnables. Une étude statistique doit souvent se faire dans des délais très courts. Donc l'échantillonnage devient une priorité.

La constitution d'un échantillon

Pour constituer un échantillon satisfaisant, il est avant tout important d'établir les objectifs que l'on désire atteindre. En d'autres mots, il faut fixer des limites de précision à partir des contraintes de temps et des contraintes financières. Malgré ces contraintes, l'échantillon doit demeurer de qualité, c'est-à-dire être représentatif de la population étudiée. On construit donc cet échantillon en respectant les étapes suivantes.

1. Il faut bien définir la population (l'univers) à étudier. Dans notre exemple, la banlieue constitue l'ensemble des municipalités autour de cette grande ville. Si cet établissement désire s'installer au nord-ouest de la ville dans une de ces municipalités, alors les deux municipalités voisines de celle-ci devraient s'ajouter à l'étude si les distances sont raisonnables et les transports publics présents. La population de ces trois municipalités deviendra l'univers idéal à étudier.

2. Dans cet ensemble universel, il existe un univers opérationnel. Idéalement, l'univers opérationnel comprend l'ensemble des familles habitant ces trois municipalités. Mais dans les faits, à cause des contraintes précitées, il faut restreindre cet univers en acceptant certains compromis et en tentant d'éviter d'introduire des biais qui pourraient résulter de la perte d'information en passant de la population à l'échantillon. La qualité de l'échantillon est plus importante que sa taille. Il faut donc définir un cadre échantillonnal. Dans cet exemple, le cadre échantillonnal pourrait être les familles qui ont au moins un enfant âgé de 8 à 16 ans.

3. La troisième étape est la détermination de la taille de l'échantillon. L'ordre de grandeur dépend du degré de précision désiré. Sauf pour une très petite population, il n'existe aucun lien proportionnel direct entre la taille de la population et la taille que doit prendre l'échantillon. Un exemple éloquent de cette affirmation est le fait que, lors d'un sondage préélectoral municipal, provincial ou fédéral, l'échantillon utilisé dépasse rarement le millier de personnes. Tout statisticien sait que 1 000 éléments assurent une marge d'erreur très faible, de l'ordre de 2 % ou 3 %[13].

Les méthodes d'échantillonnage

Pour obtenir un sondage fiable, le statisticien doit s'assurer d'utiliser une méthode d'échantillonnage adéquate. Voyons les méthodes qui sont les plus couramment utilisées.

Il existe deux méthodes principales d'échantillonnage, les méthodes non probabilistes et les méthodes probabilistes. Les méthodes non probabilistes sont des méthodes dites non scientifiques. Elles sont fort utiles pour effectuer une recherche exploratoire; elles présentent l'avantage d'être simples à utiliser et coûtent moins cher que les méthodes probabilistes. Les trois principales méthodes non probabilistes sont l'échantillonnage de commodité, l'échantillonnage discrétionnaire et l'échantillonnage par quotas. Par exemple, lors d'un débat public télévisé où les auditeurs peuvent se prononcer par téléphone sur le projet d'implantation d'un établissement d'enseignement secondaire privé, les résultats obtenus seraient une bonne source de renseignements. Dans ce cas, il s'agit d'un échantillonnage de commodité, c'est-à-dire que les répondants sont ceux qui sont disponibles au moment de l'émission. Voici un tableau de classification des méthodes d'échantillonnage.

---

13. Il existe des façons de calculer avec précision le degré de représentativité d'un échantillon ainsi que la marge d'erreur associée à cette représentativité. Cependant, ces notions dépassent largement le cadre de ce cours.

Tableau 6.3
Les méthodes d'échantillonnage

Méthodes d'échantillonnage probabilistes	Méthodes d'échantillonnage non probabilistes
Échantillonnage aléatoire simple	Échantillonnage de commodité
Échantillonnage systématique	Échantillonnage discrétionnaire
Échantillonnage stratifié	Échantillonnage par quotas
Échantillonnage en grappes	

La suite du texte sera consacrée à la présentation des méthodes d'échantillonnage probabilistes.

Les méthodes d'échantillonnage probabilistes

Il existe quatre méthodes probabilistes : l'échantillonnage aléatoire simple, l'échantillonnage systématique, l'échantillonnage stratifié et l'échantillonnage en grappes. Ces méthodes ont une caractéristique commune : la sélection aléatoire des éléments qui composent l'échantillon.

L'échantillonnage aléatoire simple

Cette méthode est probablement la plus utilisée des quatre méthodes. Dans notre exemple, supposons que le cadre échantillonnal est une liste de toutes les familles comprenant au moins un enfant âgé entre 8 ans et 16 ans. Cette liste est extraite d'un récent recensement électoral dans ces trois municipalités. Si 10 000 familles font partie de cette liste, il suffit dans un premier temps de les numéroter de 1 à 10 000, puis à l'aide d'une des nombreuses techniques de choix de nombres aléatoires (table des nombres aléatoires, fonction *rand* (pour *random*) d'une calculatrice ou d'un logiciel statistique), on constitue notre échantillon des 1 000 familles qui seront interrogées.

L'échantillonnage systématique

Cette méthode est aussi valide et encore plus simple que la première méthode. À partir de cette liste de 10 000 familles, si on désire un échantillon de 1 000 familles, il suffit de diviser 10 000 par 1 000. Le nombre 10 obtenu constitue le pas de l'échantillon. En utilisant par exemple la méthode par calculatrice, on pige une première famille à l'aide de la fonction *rand* et si le nombre obtenu est 0,2732, la 2 732^{e} famille de la liste est le point de départ, la suivante étant la 2 742^{e} et ainsi de suite jusqu'à 1a fin de la liste qui sera la 9 992^{e}. Ensuite, on continue en revenant au début de la liste toujours en faisant plus 10. Le suivant est donc le deuxième et ainsi de suite jusqu'au 1 000^{e} élément de l'échantillon.

L'échantillonnage stratifié

Cette méthode d'échantillonnage est un peu plus complexe que les deux autres méthodes, mais elle a l'avantage de réduire l'erreur aléatoire liée à la taille de l'échantillon. Dans notre exemple, le choix des parents pour une école privée pourrait dépendre d'une autre variable comme le sexe de l'enfant. Donc si au niveau provincial, on découvre qu'il y a plus de filles que de garçons dans les établissements secondaires privés, il convient de découper l'échantillon en strates. Si au niveau provincial, il y a 40 % de garçons et 60 % de filles, il conviendra de choisir 400 familles ou 40 % de familles avec des garçons ou une majorité de garçons et 600 familles ou 60 % de familles avec des filles ou une majorité de filles.

L'échantillonnage en grappes

Par quartiers de chacune des trois municipalités. Supposons que chacune de ces municipalités ait 10 quartiers, on aura alors 30 grappes. Les familles choisies dans chacune de ces grappes doivent l'être aléatoirement. Le processus d'échantillonnage par grappes comporte deux étapes. Premièrement, on choisit par une méthode aléatoire simple ou une méthode systématique un certain nombre de grappes. Deuxièmement, on échantillonne au hasard un certain nombre de familles à l'intérieur de chaque grappe. Cette méthode n'est valable que dans la mesure où il existe une forte homogénéité entre les groupes. Ce qui n'est certainement pas le cas dans notre exemple. Certains quartiers sont mal desservis par le circuit d'autobus, d'autres ont déjà une école qui donne un excellent service et d'autres n'ont aucun service de ce genre et en désirent un. Si cette condition d'homogénéité n'existe pas, on obtient un échantillon qui comporte une marge d'erreur très élevée.

Résumé

L'analyse combinatoire est l'étude du dénombrement des dispositions qu'il est possible de former à partir d'un nombre fini d'éléments. Cette étude du dénombrement porte sur deux grandes catégories de dispositions, les dispositions ordonnées et les dispositions non ordonnées. Les dispositions ordonnées sont des dispositions d'éléments dans lesquelles la place occupée par chacun des éléments est importante : l'arrangement et les permutations font partie de cette catégorie. Les dispositions non ordonnées sont des dispositions d'éléments dans lesquelles la place occupée par chacun des éléments d'un ensemble n'a pas d'importance : les combinaisons avec ou sans répétition forment cette catégorie.

Aidés par les règles de l'analyse combinatoire, le calcul des probabilités s'applique essentiellement à résoudre des problèmes liés aux expériences aléatoires. Une expérience aléatoire comporte deux éléments : un ensemble fondamental et un événement qui peut être simple ou se combiner à d'autres événements. La combinaison d'événements est soumise aux règles d'union et d'intersection d'événements.

Trois types de probabilités ont été définis dans ce module : la probabilité déductive, la probabilité « fréquentiste » et la probabilité subjective. La probabilité « fréquentiste », celle qui est la plus utilisée en gestion, se définit par la formule suivante :

$$P(E)=\frac{N(E)}{N(\Omega)}=\frac{\text{Nombre de résultats favorables à l'événement } E}{\text{Nombre de résultats possibles}}$$

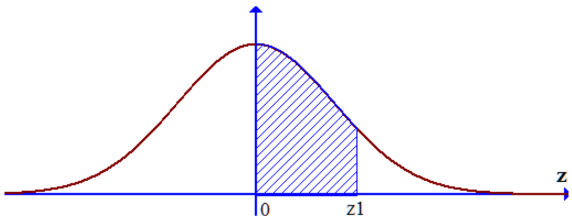
Les trois principales lois qui gouvernent la probabilité sont la loi de l'union, la probabilité conditionnelle et la loi de l'intersection.

Une distribution normale est une distribution de probabilités d'un caractère (ou variable) aléatoire continu dont la dispersion des valeurs autour de la moyenne est parfaitement symétrique. Les propriétés associées à la distribution normale (symétrie, règle empirique) et la standardisation des données permettent de calculer la probabilité des diverses variations de grandeur d'un phénomène lorsque les valeurs sont distribuées normalement.

Échantillonner c'est choisir aléatoirement, parmi un ensemble fondamental, un sous-ensemble de cet ensemble de manière à en obtenir un échantillon représentatif. Les statisticiens ont le choix d'une panoplie de méthodes d'échantillonnage, probabilistes ou non probabilistes. Parmi ces méthodes, l'échantillonnage aléatoire simple est la méthode la plus couramment utilisée.

Annexe 1 : La table de la loi normale centrée réduite

L'aire sous la courbe normale centrée réduite : $P(0 < z < z1)$



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998

3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000