Eric Depiereux, Benoît DeHertogh, Grégoire Vincke

www.fundp.ac.be/biostats **Module 30**

30 Probabilités

30.1 Introduction

Dans le premier chapitre, nous avons parcouru une partie importante des statistiques descriptives, et nous pouvons manipuler l'essentiel des techniques les plus communément employées.
Cependant, l'intérêt du chercheur ne se limite généralement pas à décrire un échantillon de valeurs expérimentales. Intéressé par le phénomène naturel qu'il cherche à décrire, son but est de pouvoir étendre ses conclusions à l'ensemble de la population qu'il étudie.

Imaginons qu'avant de se lancer dans la conception d'un analgésique, une firme pharmaceutique souhaite déterminer la proportion π d'une population qui souffre de migraines.

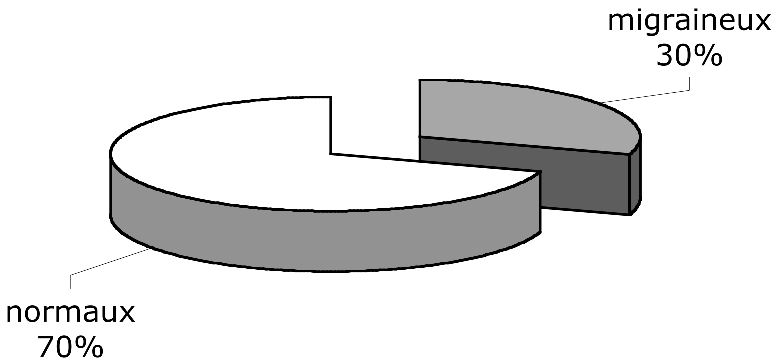

Figure 30 -1 Proportion de migraineux dans une population fictive.

Il est peu réaliste de vouloir interroger tous les individus de la population. Il faudra généralement se limiter à observer un échantillon, en étant attentif à ce qu'il représente bien la population au sujet de laquelle on souhaite tirer des conclusions.

Les observations qui seront réalisées dans l'échantillon devront donc être extrapolées à l'ensemble de la population. Cette démarche est celle de l'inférence statistique, très fréquente dans l'expérimentation scientifique.

Nous avons déjà défini la population et l'échantillon. Tout le monde admettra que la proportion de migraineux observée dans l'échantillon ne sera pas exactement égale à la proportion dans l'ensemble de la population. Si je prends plusieurs fois un échantillon dans la population, les valeurs de p seront toutes différentes. Je peux seulement espérer que les caractéristiques de mon échantillon soient telles que ces deux valeurs soient relativement proches :

π = p[1] ± une « petite erreur ».

Cependant le fait d'observer une variabilité dans ces valeurs ne veut pas dire que l'on puisse obtenir « n'importe quoi ». Le hasard suit certaines règles et, dans certaines circonstances, il nous sera possible de répondre aux questions suivantes :

imprécision

quelle est cette petite erreur ?

incertitude

quel est le risque de se tromper en affirmant que

π = p ± cette « petite erreur »?

Pour aborder ces questions, il est nécessaire d'introduire quelques éléments de la théorie des probabilités. Ceci nous permettra de suivre une démarche déductive.

Notez qu'il est préférable de parler de l'imprécision (qui est inévitable) que d'erreur (qui est supposée évitable).

La déduction est la démarche qui consiste à prédire, à partir d'une population connue, quelles seront les caractéristiques des échantillons qui y seront prélevés. Ensuite, cette connaissance sera utilisée pour suivre une démarche dite *inductive* (ou *inférentielle*). L'induction est la démarche suivant laquelle on peut prédire les caractéristiques d'une population inconnue à partir des statistiques calculées dans un échantillon.

30.2 Notions d'épreuve et d'événement

L' épreuve est une expérience dont le résultat n'est pas prévisible, et pour laquelle on peut définir l'ensemble des résultats possibles.

Si je prends un individu au hasard dans la population, je ne peux pas déterminer a priori[2] quel est son groupe sanguin: le résultat n'est pas

[1] *Dans notre notation, P représente le mot Probabilité, p sa valeur observée (en fait, une fréquence relative) et π sa valeur théorique.*

prévisible. Par contre, je peux déterminer l'ensemble des possibilités : A, B, AB, O. Cette expérience est une épreuve.

Si mon expérience consiste à questionner les gens dans la rue : "A quoi avez -vous pensé en premier lieu ce matin à votre réveil?", le résultat n'est pas prévisible mais je ne peux pas non plus définir l'ensemble des résultats possibles… Cette expérience n'est pas une épreuve.

Ajoutons à la définition de l'épreuve que cette expérience peut être reproduite dans les mêmes conditions autant de fois que l'on veut.

Ceci n'est pas le cas si l'expérience consiste à déterminer le taux de C^{14} *dans une relique unique pour en déterminer l'âge.*

- L'événement est un sous -ensemble des résultats possibles de l'épreuve.

*Je peux définir l'événement **A** comme "l'individu est migraineux". Lors d'une réalisation de l'épreuve (prendre un individu au hasard et déterminer s'il est migraineux) on dira que l'événement **A** est, ou n'est pas, réalisé, si l'individu est, ou n'est pas, sujet à la migraine.*

*Je peux aussi définir l'événement **B** comme "l'individu est du groupe sanguin A ou B". Cet événement sera réalisé*[3] *si l'individu est soit du groupe sanguin A, soit du groupe sanguin B.*

- Des événements incompatibles sont des événements tels que la réalisation de l'un entraîne la non réalisation de l'autre. Les événements **A** "l'individu est du groupe sanguin A" et **B** "l'individu est du groupe sanguin O" sont *incompatibles*.

- Des compositions d'événements sont des événements définis à partir d'autres événements.

L'union de deux événements, **A** ∪ **B**, est un événement qui est réalisé lorsque **A** ou **B** est réalisé.

*Dans notre exemple, **A** ∪ **B** correspond à l'événement : l'individu est du groupe A ou du groupe B.*

[2] a priori *: avant de faire la mesure, l'expérience*
[3] *On utilise parfois le mot occurrence pour désigner le fait d'observer la réalisation d'un événement. Ce mot est emprunté à la linguistique : on parle de l'occurrence d'un mot déterminé dans un texte. De même on parle de l'occurrence d'un événement dans un espace -temps déterminé.*

A * est défini comme le sous -ensemble de tous les résultats de l'épreuve qui entraînent la non réalisation de **A.**

*Des événements contraires, **A** et **A** * sont des événements incompatibles dont l'union correspond à l'ensemble des résultats possibles.*

*Classons tous les individus de la population soit dans la catégorie "migraineux", soit dans la catégorie "non migraineux", sans aucune ambiguïté. Si **A** représente les migraineux, **A** * représente les non migraineux.*

L'intersection de deux événements, **A ∩ B** est un événement qui est réalisé lorsque **A** et **B** sont réalisés simultanément.

*Si **A** représente « les individus du groupe sanguin A ou O » et **B** « les individus du groupe sanguin B ou O », alors **A ∩ B** correspond à l'événement : l'individu est du groupe sanguin O.*

L'intersection de deux événements incompatibles est un ensemble vide.

*Si **A** représente les individus du groupe A et **B** les individus du groupe sanguin B alors **A ∩ B** ne représente plus personne (il n'est pas possible d'être A **et** B à la fois).*

L'événement (**A ∩ B**) * est identique à l'événement **A * ∪ B *** et l'événement (**A ∪ B**) * est identique à l'événement **A * ∩ B *.**

Cette propriété est particulièrement utile pour résoudre certains problèmes en transformant une intersection inconnue en union connue et réciproquement.

30.3 Notion de probabilité

Reprenons le problème posé au départ : on entreprend une recherche dont le but est de déterminer la proportion de migraineux dans la population. Dans la terminologie que nous venons de définir , le problème pourrait être posé ainsi :
- soit une **épreuve** qui consiste à prendre un individu au hasard dans la population
- soit l'**événement** A : l'individu est migraineux,
- quelle chance a -t -on de réaliser cet événement lors d'une épreuve ? Ou, en d'autres termes :

Quelle est la *probabilité* de réalisation de l'événement A?

30.3.1 La probabilité définie à partir de l'observation.

*Prenons une pièce de monnaie équilibrée et lançons -la un certain nombre de fois, n. Notons la fréquence de réalisation de l'événement **A** "la face apparente est pile", n_A et calculons la fréquence relative de réalisation n_A/n. Si je recommence cette opération pour un nombre n de plus en plus grand, et que je porte les résultats obtenus sur un graphique, je vais m'apercevoir que la fréquence relative, très variable lorsque n est petit, se stabilise lorsque n augmente, et tend vers une valeur constante lorsque n tend vers l'infini. Cette valeur constante est la probabilité de réalisation de l'événement A.*

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n}$$ **Équation 30-1**

Nous écrirons : P(A) = π,

P( ) représentant la fonction et π la valeur théorique d'une probabilité.

C'est ainsi que la probabilité de réalisation de l'événement **A** : l'individu est du groupe sanguin A peut être estimée par la fréquence relative de réalisation dans un échantillon de taille n : P(A) ≈ n_A/n. Cette estimation sera d'autant meilleure que le nombre de répétitions de l'épreuve n sera grand.

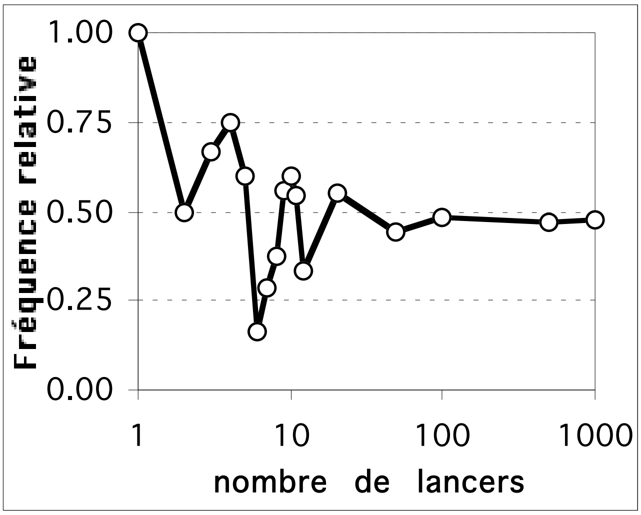

Figure 30 -2 Evolution de la fréquence relative vers une probabilité lorsque le nombre de réalisations de l'épreuve augmente.

30.3.2 Propriétés des probabilités

$$0 \leq P(\mathbf{A}) \leq 1$$

événement certain ⇔ probabilité 1

événement impossible = > probabilité 0

Est considéré comme événement impossible un événement qui pratiquement ne se réalisera jamais, même si on peut l'envisager comme théoriquement possible, par exemple une pièce de monnaie qui retombe en équilibre sur la tranche, ne donnant ainsi ni pile ni face.

Malgré l'abus de langage, l'usage est souvent de parler d'une probabilité en terme de pourcentage. p = 0,8 ou p = 80%.

30.3.3 La probabilité définie a priori

Dans certains cas, et <u>moyennant certaines hypothèses</u>, la probabilité de réalisation d'un événement peut être déterminée sans réaliser l'épreuve, sur base de sa seule définition.

Par exemple, je n'ai pas besoin de lancer 1.000.000 de fois une pièce de monnaie pour savoir que la probabilité d'obtenir pile est de 0,5. Notons les hypothèses sur lesquelles cette valeur de probabilité repose :

la pièce est équilibrée : P (pile) = P (face)

la pièce ne retombe jamais sur la tranche : P (pile) + P(face) = 1
par conséquent : P (pile) = P (face) = 0,5 .

Puisque les probabilités définies a priori reposent toujours sur certaines hypothèses, il importe toujours de les énoncer puisque la validité des calculs de probabilité dépend de ces hypothèses. Par exemple, en génétique, on définira a priori la probabilité d'obtenir un individu homozygote récessif aa à partir d'un croisement d'hétérozygotes Aa est de 25%.

Aa		Aa	
A	a	A	a
AA	aA	Aa	aa
25%	25%	25%	25%

*Tableau 30 -30-1 Séparation des allèles **A** et **a** à la méiose[4]. Recombinaison de paires d'allèles à la fusion des gamètes.*

[4] Méiose : division de la cellule aboutissant à la réduction de moitié du nombre des chromosomes, se produisant au moment de la formation des

L'hypothèse est que la viabilité des individus de génotypes **AA Aa (**ou **aA)** et **aa** est identique.

30.4 Lois de probabilités

30.4.1 Tables de contingence

Considérons la prévalence d'un cancer en fonction du sexe des individus. Les événements et leur probabilité sont définis de la façon suivante :

A : cancéreux	P(A) = 0,10
A * : non cancéreux	P(A *) = 0,90
B : homme	P(B) = 0,60
B * : femme	P(B *) = 0,40

Tableau 30 -30-2 Nomenclature et probabilités des événements.

La probabilité qu'un individu pris au hasard dans la population soit un homme cancéreux est P(**A** ∩ **B)**. Cette intersection peut être établie pour chaque combinaison d'événements et les résultats présentés dans une table à deux entrées :

Cancer	Sexe B	B *	Total
A	0,06	0,04	0,10
A *	0,54	0,36	0,90
Total	0,60	0,40	1,00

Tableau 30 -30-3 Table de contingence[5] représentant la probabilité des événements et de leur intersection.

30.4.2 Probabilités conditionnelles

La prévalence peut être établie séparément pour les deux sexes :

Cancer	Sexe B	B *
A	0,10	0,10
A *	0,90	0,90
Total	1,00	1,00

Tableau 30 -30-4 Table de contingence présentée par sexe.

cellules reproductrices (gamètes).

[5] *Contingence : éventualité que quelque chose arrive ou non.*

La probabilité qu'une femme ait ce cancer (0,10) est différente de celle qu'un individu soit une femme cancéreuse (0,04).

La probabilité qu' « une femme ait ce cancer » peut être exprimée plus explicitement comme la probabilité qu'un individu ait ce cancer « sous condition que ce soit une femme ».

Cette probabilité s'appelle la probabilité conditionnelle et se note P(**A/B** *), probabilité d'avoir le cancer, *sachant que* la personne concernée est une femme.
Dans l'autre sens, la répartition des sexes peut être établie séparément pour les cancéreux et les non cancéreux :

Cancer	Sexe B	B *	Total
A	0,60	0,40	1,00
A *	0,60	0,40	1,00

Tableau 30 -30-5 Table de contingence présentée par état de santé,

La probabilité qu'un cancéreux soit une femme (0,4) est différente de celle qu'un individu soit une femme cancéreuse (0,04) ou encore qu'une femme soit cancéreuse (0,10).

Soit je fais référence à tous les individus de la population, ce qui correspond à la définition de l'épreuve suivante : si j'interroge un individu *au hasard,* quelle est la probabilité que ce soit une femme ET qu'elle soit cancéreuse : P(**B** * ∩ **A**) = 0,04.

Soit, je pose une condition : l'événement "être une femme" (**B** *) doit d'abord être réalisé (je ne compte les cancers que chez les femmes); ensuite, sachant que c'est une femme, je détermine la probabilité qu'elle ait le cancer P(**A/B** *). Etre une femme est un contexte particulier, une condition préalable, d'où l'expression "probabilité conditionnelle".

Il faut bien voir que **A/B *** n'est pas le même événement que **B */A** : dans ce cas, la condition est d'être cancéreux. Parmi les cancéreux, je détermine la proportion de femmes P(**B */A**) = 0,4.

30.4.3 Indépendance

Dans les tables ci -dessus, nous pouvons remarquer que la prévalence du cancer considéré est la même chez les hommes et chez les femmes :

P(**A/B**) = P(**A/B** *) = P(**A**)

On dira dans ce cas que le contexte B (ici le sexe) n'apporte aucune information susceptible de modifier la probabilité de réalisation de A, ou l'on peut dire que ce cancer est indépendant du sexe.

Ce n’est pas toujours le cas. La prévalence du cancer du sein est quasi nulle chez les hommes, et celle du cancer de la prostate est nulle chez la femme.

Dans de nombreuses situations, la probabilité de réalisation de A dépend du contexte B :

P(**A/B**)≠P(**A/B** *)≠P(**A**)

La probabilité de faire un accident dépend des circonstances : la probabilité de déraper s’il pleut est différente de la probabilité de déraper s’il ne pleut pas : le risque d’accident n’est pas indépendant de la météo.

Reprenons les tables de contingence pour un cancer dépendant du sexe :

Cancer	**Sexe B**	**B ***	
A	0,08[1]	0,02	0,10[2]
A *	0,52	0,38	0,90
	0,60	0,40	1,00

Tableau 30 -30-6 Table de contingence représentant la probabilité des événements et de leur intersection.

Cancer	**Sexe B**	**B ***
A	0,13[3]	0,05
A *	0,87	0,95
	1,00	1,00

Tableau 30 -30-7 Table de contingence présentée par sexe.

Cancer	**Sexe B**	**B ***	
A	0,80	0,20	1,00
A *	0,58	0,42	1,00

Tableau 30 -30-8 Table de contingence présentée par état de santé.

Nous constatons donc que
P(**A** ∩ **B**) = 0,08(1)
P(**A**) = 0,10(2)
P(**A/B**) = 0,13(3)

30.4.4 Loi des probabilités composées

Recherchons maintenant la probabilité de l'intersection en fonction des conditionnelles. L'événement **A** ∩ **B** peut être décomposé en deux étapes : **A** se réalise puis **B** se réalise sachant que **A** est réalisé ou réciproquement : **B** se réalise puis **A** se réalise sachant que **B** est réalisé.

P(**B**) = 0,60
P(**A/B**) = 0,13(3)
et P(**A** ∩ **B**) = 0,60x0,13 = 0,08(1)

ou réciproquement :

P(**A**) = 0,10(2)
P(**B/A**) = 0,80
et P(**A** ∩ **B**) = 0,80x0,10 = 0,08(1)

Loi des probabilités composées

P(A ∩ B) **= P**(A) **P**(B/A) **= P**(B) **P**(A/B) **Équation 30-2**

30.4.5 Condition d'indépendance

Deux événements sont indépendants si le contexte de **B** ne modifie pas la probabilité de réalisation de **A**, et réciproquement.

Ce qui s'écrit en termes de probabilités conditionnelles :

P(**A/B)** = P(**A/B** *) = P(**A**)

Et en termes de probabilités composées :

P(**A** ∩ **B**) = P(**A**) P(**B**)

La table de contingence permet donc de vérifier l'indépendance :

Cancer	Sexe B	B *	Total
A	0,06	0,04	0,10
A *	0,54[1]	0,36	0,90[3]
Total	0,60[2]	0,40	1,00

Tableau 30 -30-9 Table de contingence concernant un cancer indépendant du sexe.

Cancer	Sexe B	B *	Total
A	0,08	0,02	0,10
A *	0,52	0,38	0,90[3]
Total	0,60[2]	0,40	1,00

Tableau 30 -30-10 Table de contingence concernant un cancer dépendant du sexe.

Dans le premier cas, nous notons que
$0{,}54^1 = 0{,}60^2 \times 0{,}9^3 \Leftrightarrow P(\mathbf{A^*} \cap \mathbf{B}) = P(\mathbf{A^*})\, P(\mathbf{B})$
Cette égalité se vérifie pour toutes les combinaisons **de A,B, A *** et **B *.**

Dans le second cas, nous notons que
$0{,}08^{(1)} \neq 0{,}10^{(2)} \times 0{,}60^{(3)} \Leftrightarrow P(\mathbf{A} \cap \mathbf{B}) \neq P(\mathbf{A})\, P(\mathbf{B})$
Cette inégalité se vérifie pour toutes les combinaisons **de A,B, A *** et **B *,**

Nous en déduisons que le contexte « sexe » influence la prévalence du second cancer, mais pas celle du premier.

Dans une famille, le sexe d'un enfant à naître est indépendant des enfants nés : le fait que le premier enfant soit un garçon ne renseigne en rien sur la probabilité que le prochain enfant soit une fille …

Le lancement d'un dé (équilibré) est indépendant du lancement suivant : le fait d'obtenir 4 au premier lancement n'apporte aucune information sur la probabilité de faire un lancer particulier au coup suivant.

Dans les jeux de hasard, on exploite souvent l'impression -fausse - que la probabilité d'un événement peut se modifier en fonction de la réalisation d'autres événements indépendants. Imaginons par exemple que je réussisse en 10 lancers à obtenir 10 fois "pile" avec une pièce de monnaie. A ce moment je peux faire parier sur le fait que je vais réussir un 11ème lancer "pile"… Cela donne l'impression d'être très difficile à réussir. Cependant la probabilité de réussir ce lancer est de 1/2, comme pour chacun des lancers. Ce qui est difficile, c'est de dire a priori que l'on va réaliser 11 lancers "pile", et de le réussir .

30.4.6 Loi des probabilités totales

Considérons une enquête réalisée chez des enfants d'âge scolaire, définissant si l'usage d'une substance fluorée protégeant les dents, le Zymafluor, dépend du niveau socio -culturel de l'enfant, représenté ici par la profession du père.

Profession du père	Zymafluor		
	Non	Oui	(T)
Sans	200	60	260
Agriculteur	100	20	120
Prof. Libérale	400	80	480
Artisan	600	40	640
Cadre Moyen	400	120	520
Employé	1000	120	1120
Ouvrier	800	60	860
	3500	500	4000

Tableau 30 -30-11 Table de fréquences. Répartition de 4000 réponses concernant l'utilisation de Zymafluor par l'enfant en fonction de la profession du père.

Nous considérerons que l'échantillon est suffisamment grand pour que les fréquences relatives puissent être assimilées à des probabilités

Profession du père	Zymafluor		
	Non	Oui	(T)
Sans	0,05	0,015	0,065
Agriculteur	0,025	0,005	0,03
Prof, Libérale	0,1	0,02	0,12
Artisan	0,15	0,01	0,16
Cadre Moyen	0,1	0,03	0,13
Employé	0,25	0,03	0,28
Ouvrier	0,2	0,015	0,215
	0,875	0,125	1

Tableau 30 -30-12 Table de fréquences relatives, assimilées ici à des probabilités.

Cette table permet de lire directement la probabilité des événements tels que :

***A :** l'enfant prend régulièrement du zymafluor (P = 0,125)*
***A * :** l'enfant ne prend pas de zymafluor (P = 0,875)*
***B :** le père est sans profession (P = 0,065)*
***A ∩ B** : l'enfant prend régulièrement du zymafluor **et** le père est sans profession (P = 0,015).*

Recherchons maintenant la probabilité de l'événement "le père est employé **ou** cadre moyen" soit la composition A ∪ B. Cette probabilité est représentée par le rapport entre le nombre total d'employés (A) **ou** de cadres moyens (B) divisé par le nombre total d'enfants, soit (1120 +520)/4000 = 0,410 = 0,280 +0,130 =

P(A) +P(B). Nous pouvons généraliser cette observation à l'ensemble des événements incompatibles :

Si deux événements sont incompatibles,

$$P(\mathbf{A} \cup \mathbf{B}) = P(\mathbf{A}) + P(\mathbf{B})$$

Recherchons maintenant la probabilité de l'événement "le père est employé **ou** l'enfant prend régulièrement du zymafluor".
En dénombrant le nombre total d'employés plus le nombre total d'enfants zymafluor (1120 +500) on comptera deux fois les enfants zymafluor dont le père est employé (120)! Il faut donc les retirer une fois de ce total (1120 +500 - 120 = 1500). Exprimé en terme de probabilité, cela donnera : 1500/4000 = 0,375 = 0,280 +0,125 -0,030.

Loi des probabilités totales :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$ **Équation 30-3**

Cette loi englobe la précédente, puisqu'en cas d'événement incompatible[6], par définition $P(A \cap B)=0$.

30.4.7 Reconstruction de P(A)

La combinaison de la loi des probabilités totales et de la loi des probabilités composées donne la propriété suivante :

$P(\mathbf{A}) = P((\mathbf{A} \cap \mathbf{B}) \cup (\mathbf{A} \cap \mathbf{B}^*))$
$P(\mathbf{A}) = P(\mathbf{A} \cap \mathbf{B}) + P(\mathbf{A} \cap \mathbf{B}^*)$
car ils sont incompatibles

$$P(A) = P(A/B)\, P(B) + P(A/B^*)\, P(B^*)$$ **Équation 30-4**

[6] Il convient de ne pas confondre les mots incompatible (la réalisation de A empêche la réalisation de B) et indépendant (la réalisation de A ne modifie pas la probabilité de réalisation de B). Deux événements incompatibles sont donc dépendants (au plus fort sens du terme).

Ceci permet de retrouver la probabilité de A à partir de ses conditionnelles.

Parmi plusieurs centaines d'enfants traités par un vaccin anti -grippe, 20% ont fait une poussée grippale durant les deux mois qui ont suivi l'administration du vaccin. Durant la même période, pour les enfants non vaccinés, 46% ont fait une poussée grippale. Quelle est la probabilité qu'un enfant soit grippé, sachant que 25% des enfants sont vaccinés?

*A : Vaccin; A * pas de vaccin*
*B : grippé; B * : pas grippé*

On assimile ici les fréquences relatives aux probabilités : on observe

*P(B/A) = 0,2 et P(B/A *) = 0,46*
*P(A) = 0,25; P(A *) = 0,75*
*P(B) = P(A)P(B/A) +P(A *)P(B/A *) = 0,25 *0,2 +0,75 *0,46 = 0,395*

30.5 Exemples

30.5.1 Probabilité a priori[7] et a posteriori[8]

En appliquant la loi des probabilités composées, la probabilité a priori de réussir 11 lancers "pile" consécutivement (c'est -à -dire : parier le résultat avant de commencer) est de $(0,5)^{11} = 4,88\ 10^{-4}$. La probabilité de réussir un 11° lancer pile, après avoir réussi l'exploit d'en réussir 10, est de … 0,5 ! En effet, il s'agit de P(A/B) avec A : obtenir pile B : contexte : nous venons d'obtenir 10 fois pile. Sous l'hypothèse raisonnable que les lancers sont indépendants, le 11° lancer est indépendant du contexte et P(A/B) = P(A) = 0,5,

Le dernier numéro gagnant du Lotto, a exactement la même chance de sortir au prochain tirage que n'importe quelle combinaison. Par contre, le dernier tiercé gagnant me fournit des informations sur les chances des chevaux à la prochaine course. Le contexte m'apporte une information et P(A/B)≠ P(A).

30.5.2 Petites et grandes populations

Exemple 3.5.2.1. Imaginons une grande population animale dans laquelle le sexe -ratio (nombre de mâles / nombre de femelles) est de 1.5. Si je capture de façon indépendante 2 individus au hasard dans cette population, quelle est la probabilité d'obtenir un couple ?

---

[7] A priori : (locution latine, en partant de ce qui vient avant).« en se fondant sur des données admises avant toute expérience »

[8] A posteriori (locution latine, en partant de ce qui vient après). « en se fondant sur l'expérience, sur les faits constatés »

> *La population est grande : cela me permet de dire que le fait de capturer un premier individu mâle ne modifie pas la probabilité de capturer un second individu mâle, a condition que mon procédé de capture permette des captures indépendantes les unes des autres.*

épreuve : capturer deux individus au hasard dans la population
événements :
A : le premier est un mâle, **A** * : le premier est une femelle
B : le second est un mâle, **B** * : le second est une femelle

composition d'événements :

$(A \cap B^*) \cup (A^* \cap B)$

> *union de deux événements incompatibles*
> *intersections de deux événements indépendants*

probabilités :

$P(A) = P(B) = 0{,}6$
$P(\mathbf{A}^*) = P(\mathbf{B}^*) = 0{,}4$
$P((\mathbf{A} \cap \mathbf{B}^*) \cup (\mathbf{A}^* \cap B)) = P(\mathbf{A} \cap \mathbf{B}^*) + P(\mathbf{A}^* \cap \mathbf{B}) =$
$P(\mathbf{A}) . P(\mathbf{B}^*) + P(\mathbf{A}^*).P(\mathbf{B}) = 0{,}6 \text{ x}0{,}4 + 0{,}4 \text{ x } 0{,}6 = 0{,}48$

Exemple 3.5.2.2. Une ASBL comprenant 10 membres, dont 6 hommes et 4 femmes, forme par tirage au sort un conseil d'administration de 3 personnes. Quelle est la probabilité que trois hommes soient désignés?

> *La population envisagée ici est de taille limitée. Les trois personnes désignées doivent être des personnes différentes. L'événement "la première personne désignée est un homme" n'est donc pas indépendant de l'événement "la deuxième personne désignée est un homme".*

épreuve : désigner trois personnes au hasard. Les **événements sont :**
A : le premier est un homme, **A** * : le premier est une femme
B : le second est un homme, **B** * : le second est une femme
C : le troisième est un homme, **C** * : le troisième est une femme

composition d'événements recherchée :
$(\mathbf{A} \cap \mathbf{B} \cap \mathbf{C})$
intersection de trois événements non indépendants.
probabilités :

$P(A) = P(B) = P(C) = 6/10$
$P(A^*) = P(B^*) = P(C^*) = 4/10$
en effet :

$P(B) = P(B/A).P(A) + P(B/A^*).P(A^*) =$
$5/9 \times 6/10 + 6/9 \times 4/10 = 54/90 = 6/10$

$P(A \cap B \cap C) = P(A).P(B/A).P(C/A \cap B)$
$= 6/10 \times 5/9 \times 4/8 = 120/720 = 1/6$

30.6 Faux négatif et faux positif.

Les probabilités conditionnelles mettent en évidence un problème connu sous le nom du problème du faux positif (ou du faux négatif) particulièrement crucial dans le cas des analyses de dépistage de maladie.

On effectue sur une large échelle le dépistage du Sida. Dans le test, nous pouvons définir deux événements distincts : R : le test est positif, S : la personne est séropositive. On peut évidemment souhaiter que ces événements coïncident, mais il faut admettre que même si le test est performant, certaines personnes séropositives puissent donner par erreur un test négatif (c'est le faux négatif) et certaines personnes séronégatives puissent donner un test positif (c'est le faux positif). Dans le premier cas, on ne détectera pas un porteur contagieux, dans le second cas, une personne saine se croira atteinte. De ces deux situations à éviter, la première est vraiment la plus grave, car dans le second cas de figure, on finira par établir que la personne est saine.

Soient les événements :

A : séropositif, **A *** : séronégatif
B : test positif, **B * :** test négatif

Pour tester la méthode de dépistage, des tests cliniques sont réalisés sur des personnes diagnostiquées séropositives avec certitude : la réaction est positive dans 99% des cas.

$P(B/A) = 0{,}99$ et $P(B^*/A) = 0{,}01$

Sur des personnes diagnostiquées saines avec certitude : le test est négatif dans 98% des cas.

$P(B^*/A^*) = 0{,}98$ et $P(B/A^*) = 0{,}02$

On en déduit donc que le taux d'erreur est très faible... Attention, car les probabilités qui nous intéressent sont les conditionnelles **A*/B** et **A/B*** qui sont très différentes. En effet, lors du dépistage, l’observation sera un test soit positif soit négatif, et les questions sont de déterminer quelle est :

- la probabilité que l'individu soit sain lorsque le test est positif ? $P(A^*/B)$

- la probabilité qu'il soit porteur lorsque le test est négatif ? $P(A/B^*)$

Le résultat dépendra dramatiquement de la prévalence de la maladie.

Supposons : P(**A**) = 0,001

Remarquez que cette prévalence est relativement élevée et que ce qui suit sera d'autant plus étonnant que P(A) sera petit.

La table de contingence est établie en utilisant la loi des probabilités composées, pour des événements dépendants

P(**A** ∩ **B**) = P(**A**)P(**B/A**)

	B	**B ***	
A	0,00099	0,00001	0,001
A *	0,01998	0,97902	0,999
	0,02097	0,97903	

Tableau 30 -30-13 Table de contingence calculée à partir des tests cliniques et de l'estimation de la prévalence de la maladie.

et les conditionnelles réciproques calculées par la relation inverse :

P(**A/B**) = P(**A** ∩ **B**)/P(**B**), ce qui donne

	B	**B ***
A	0,04721	0,00001
A *	0,95279	0,99999
	1	1

Tableau 30 -30-14 Table de contingence calculée pour les tests positifs et pour les tests négatifs.

P(**A** */**B** *) = 0,97902/0,97903 = 0,99999
P(**A**/**B** *) = 0,00001/0,97903 = 0,00001

en cas de réaction négative, tout se passe donc bien.

P(**A** */**B**) = 0,01998/0,02097 = 0,95279

(95% de faux positifs!!)

P(**A**/**B**) = 0,00099/0,02097 = 0,04721