Chapitre 1

Probabilités

David FOUCHET
david.fouchet@univ-lyon1.fr

Introduction

David Fouchet

Mathématicien de formation (L1 au M1)

Laboratoire de Biométrie et Biologie Evolutive (M2 à aujourd’hui)

Enseignant-chercheur (maître de conférence)

Thème de recherche: épidémiologie mathématique des maladies infectieuses (ex: analyse de la dynamique de contamination du FIV chez le chat)

$\Rightarrow$ Statistiques

$\Rightarrow$ Modélisation

Introduction

Un exemple simple :

On a découvert un nouveau traitement contre une maladie mortelle

$\Rightarrow$ Etape 1: Mise en évidence de l'effet positif du traitement chez la souris

Sans traitement: 1 chance sur 2 de survie pour les souris

Que fait-on maintenant ?

Introduction

On va prendre un certain nombre de souris, ici N=20, à qui on donne le traitement

On observe 13 souris survivantes

Le traitement a-t-il un effet bénéfique ?
(augmentation de la probabilité de survie des souris?)

Introduction

Aurait-on conclu à un effet bénéfique si on avait observé

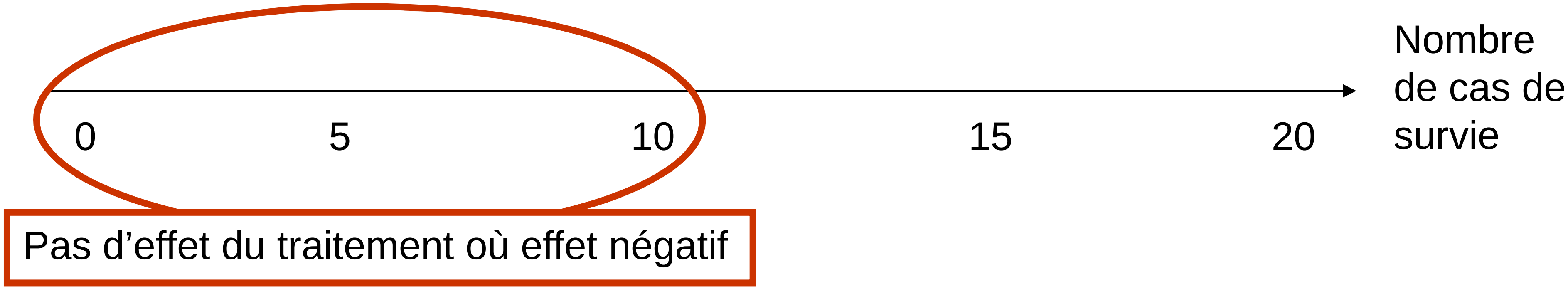
11 survivants?
12 survivants? 14? 15?...

11 12 13 14 15 16 17 18 19 20

Où se situe la limite?

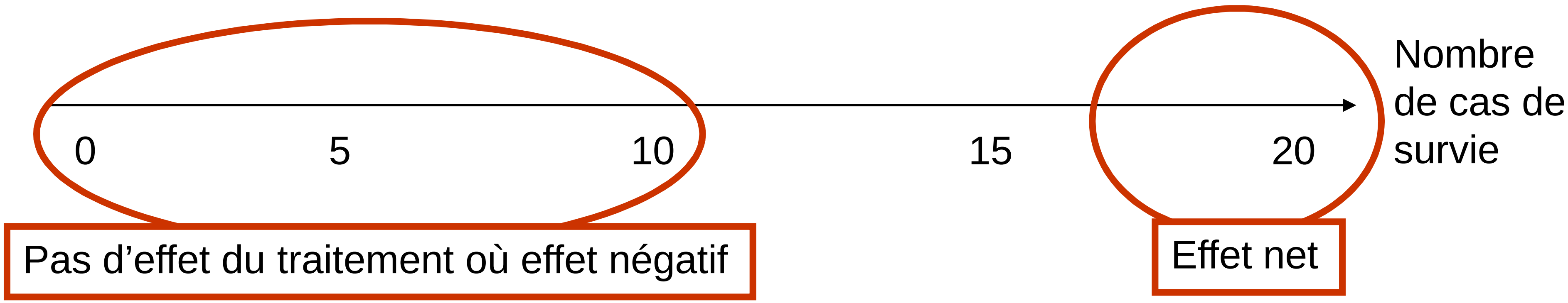
Introduction

En s’appuyant sur du **calcul de probabilité**, les **statistiques** permettent de répondre à cette question



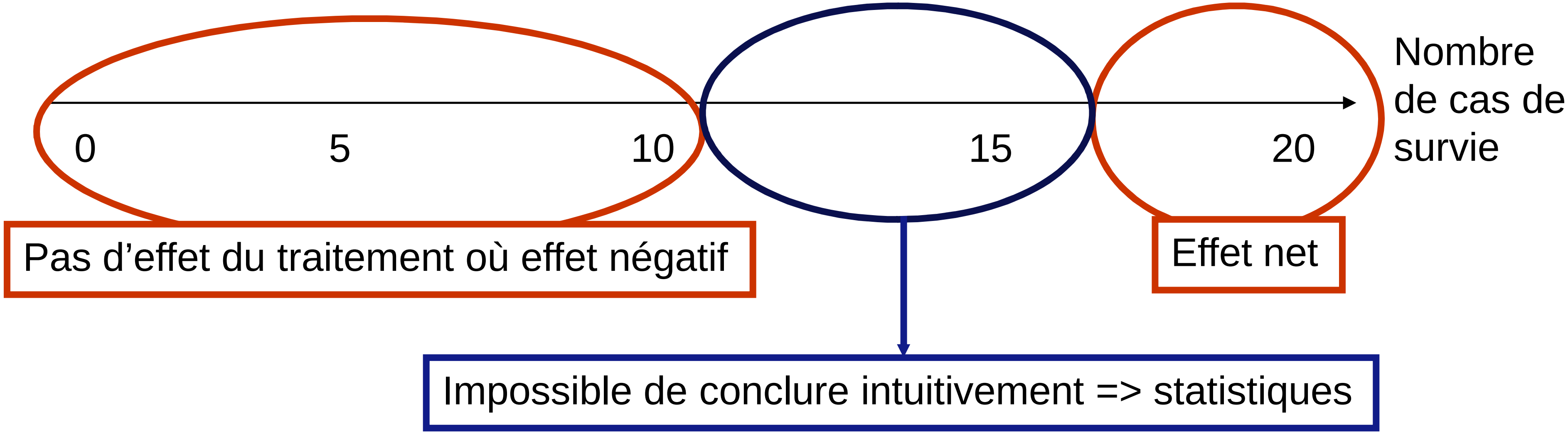
Introduction

En s'appuyant sur du **calcul de probabilité**, les **statistiques** permettent de répondre à cette question



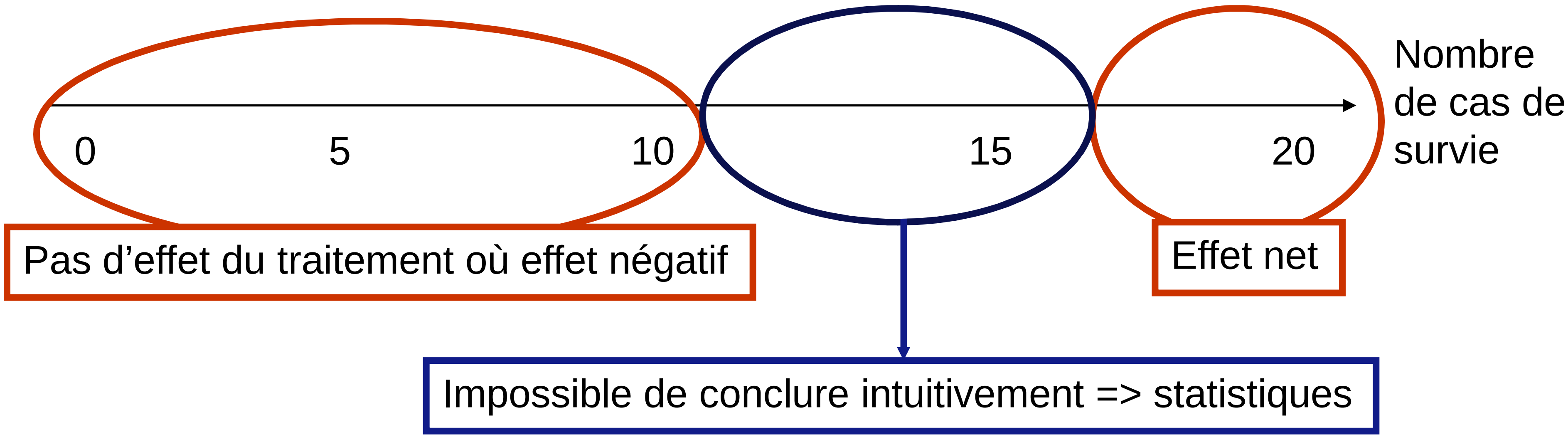
Introduction

En s'appuyant sur du **calcul de probabilité**, les **statistiques** permettent de répondre à cette question



Introduction

En s'appuyant sur du **calcul de probabilité**, les **statistiques** permettent de répondre à cette question



<u>MORALITE</u>: les statistiques permettent de tirer le maximum d'informations sur nos données

Introduction

Sur quel critère décide-t-on que 13 est un résultat probant?

Introduction

Sur quel critère décide-t-on que 13 est un résultat probant?

On souhaiterait connaître la probabilité qu'un placebo donne de si bons résultats (cad avoir au moins 13 survivants).

Comment calculer cette probabilité?

Il faudrait faire l'expérience un très grand nombre de fois pour voir à quelle fréquence on atteint « par hasard » un résultat aussi bon que celui du traitement.

<u>Problèmes:</u>

- c'est long!
- ça coûte cher!

Introduction

<u>Solution alternative:</u>

Sans traitement les chances de survie sont les mêmes qu'en jouant à pile ou face

$\Rightarrow$ Modèle de la pièce: au lieu d'inoculer la maladies à 20 souris on lance une pièce 20 fois

Introduction

Solution alternative:

Sans traitement les chances de survie sont les mêmes qu'en jouant à pile ou face

$\Rightarrow$ Modèle de la pièce: au lieu d'inoculer la maladies à 20 souris on lance une pièce 20 fois

Avantages:
- ça ne coûte pas cher
- c'est moins long

Introduction

<u>Solution alternative:</u>

Sans traitement les chances de survie sont les mêmes qu'en jouant à pile ou face

$\Rightarrow$ Modèle de la pièce: au lieu d'inoculer la maladies à 20 souris on lance une pièce 20 fois

Avantages:
- ça ne coûte pas cher
- c'est moins long

Inconvénients:
- c'est quand même un peu long
- ça marche parce que la probabilité est de 50%

Introduction

Dernière option: les mathématiques

On calcule directement la probabilité à l'aide de la loi binomiale (cf plus loin)

$\Rightarrow$ p = 0.1316

Avantages:

- ça ne coûte pas cher
- c'est rapide
- on a accès à la probabilité exacte
- ce sont des maths et on aime ça!

Objectifs de la partie proba-stats

Contrairement à la science physique, la biologie doit gérer en permanence avec la variabilité individuelle.

Exemple:

- Chute d'un corps: il suffit de mesurer l'écart du point de chute à la verticale pour connaître le sens et la force du vent

- effet du tabagisme sur le taux de cholestérol: il ne suffit pas de comparer le taux de cholestérol chez un fumeur et un non-fumeur

$\Rightarrow$ l'objectif de cette partie est d'initier à l'analyse de données en biologie.

Objectifs de la partie proba-stats

- exemple tabagisme/cholestérol: mesure du taux de cholestérol chez 100 individus

F	2.11	**NF**	1.57	**F**	2.27	**F**	1.48	**NF**	0.99	**NF**	1.65	**NF**	1.73	**NF**	1.99	**NF**	2.15	**NF**	2.13
NF	1.49	**F**	1.87	**NF**	1.89	**NF**	2.51	**F**	1.75	**F**	1.83	**F**	2.05	**NF**	1.87	**F**	2.24	**F**	2.52
NF	1.62	**F**	2.33	**F**	0.97	**F**	1.80	**F**	2.23	**NF**	1.60	**NF**	1.59	**NF**	2.19	**NF**	2.63	**NF**	2.31
NF	2.54	**NF**	1.57	**F**	2.06	**NF**	1.30	**NF**	1.83	**F**	2.29	**NF**	2.34	**F**	1.86	**F**	1.72	**F**	1.12
F	1.93	**F**	1.39	**NF**	2.79	**NF**	2.15	**NF**	2.61	**NF**	1.87	**NF**	0.81	**F**	1.16	**NF**	2.13	**F**	2.34
F	2.19	**F**	1.94	**F**	2.50	**NF**	2.77	**F**	1.68	**F**	2.24	**F**	2.49	**NF**	1.48	**NF**	1.99	**NF**	2.40
NF	2.04	**NF**	1.96	**NF**	1.20	**F**	2.35	**F**	0.83	**F**	2.33	**NF**	2.10	**F**	2.12	**NF**	1.70	**F**	2.31
F	1.68	**F**	2.24	**F**	1.96	**F**	2.97	**NF**	1.38	**NF**	1.96	**NF**	2.13	**NF**	1.37	**F**	3.06	**NF**	2.65
NF	1.72	**NF**	1.7	**NF**	1.65	**F**	2.25	**F**	2.52	**F**	2.44	**F**	2.60	**F**	1.82	**F**	1.87	**F**	2.16
F	2.22	**NF**	1.92	**NF**	1.48	**NF**	2.93	**NF**	1.94	**NF**	3.15	**F**	1.86	**F**	1.52	**NF**	1.29	**NF**	1.66
NF	1.52	**NF**	1.78	**F**	1.38	**F**	1.83	**F**	2.18	**NF**	2.26	**NF**	1.93	**NF**	1.41	**F**	2.88	**F**	1.92
NF	2.39	**F**	1.96	**F**	2.14	**F**	1.43	**NF**	2.47	**NF**	1.99	**F**	1.36	**F**	1.48	**NF**	2.16	**NF**	0.77
NF	2.28	**F**	2.76	**NF**	1.78	**NF**	1.89	**NF**	0.93	**F**	2.45	**NF**	1.16	**NF**	1.79	**F**	1.44	**F**	2.23
F	1.58	**NF**	1.69	**F**	2.02	**F**	2.59	**NF**	1.67	**F**	2.02	**F**	1.64	**NF**	2.08	**NF**	2.31	**F**	2.05
F	1.86	**F**	1.32	**NF**	1.81	**NF**	1.44	**F**	1.64	**NF**	1.44	**NF**	2.14	**NF**	1.94	**NF**	2.63	**NF**	1.70
NF	1.40	**NF**	2.23	**F**	1.76	**NF**	2.31	**NF**	1.49	**NF**	2.24	**NF**	1.72	**F**	2.53	**NF**	1.55	**F**	1.67
F	0.89	**F**	1.54	**NF**	2.18	**NF**	1.69	**F**	1.90	**NF**	1.99	**NF**	1.33	**F**	1.87	**F**	2.06	**NF**	1.45
NF	2.49	**F**	2.01	**NF**	2.36	**F**	2.27	**NF**	2.76	**NF**	1.86	**F**	2.53	**NF**	1.24	**F**	1.93	**F**	1.97
F	1.74	**NF**	1.68	**NF**	3.05	**F**	1.45	**NF**	1.98	**F**	2.63	**NF**	1.64	**NF**	2.18	**NF**	1.41	**NF**	2.18
F	2.16	**NF**	2.26	**F**	1.32	**NF**	2.04	**NF**	2.61	**F**	2.93	**F**	1.99	**F**	2.03	**F**	2.59	**F**	1.83

Objectifs de la partie proba-stats

On ne peut pas laisser les données comme ça (manque de visibilité)

$\Rightarrow$ Besoin d'outils de synthèse (moyenne, graphique,…)

STATISTIQUES DESCRIPTIVES

Objectifs de la partie proba-stats

On aimerait savoir s'il existe une différence liée au tabagisme **au-delà de la variabilité individuelle**

STATISTIQUES INFERENTIELLES

Mise en équation du problème biologique (modélisation statistique) → Calcul de probabilités → Tests statistiques/ estimation

Objectifs de la partie proba-stats

Programme des cours/TD/TT:

- Calcul de probabilités

- Statistiques descriptives: résumer des données par des indicateurs (moyennes, variances,…) et les représenter par des graphes (histogrammes,…)

- Statistiques inférentielles:

  1) estimation d’une moyenne et d’une proportion et incertitude liée à ces estimations

  2) ma moyenne/fréquence est-elle conforme à ce que j’attends?

  3) y a-t-il un écart de moyenne/fréquence entre deux groupes?

  4) liaison entre variables (ex: réponse à un traitement / stade d’un cancer)

Evaluation

- **2 QCM**
  - 2*10% de la note finale
- **1 problème**: 25% de la note finale

- **TTs** 20% de la note finale en comptant la partie analyse sur les bactéries

Introduction

De manière générale:

Les modèles mathématiques permettent de modéliser des expériences aléatoires (cad dont le résultat ne peut être prévu avec une totale certitude).

Quelques exemples :

Situation mathématique	Résultat observable	Expérience	Question
Pièce à pile ou face	Séquence de P et F (ou 0 et 1)	Prélèvement de *n* individus dans une population	Sex-ratio = 1 ?
		Nombre de survivants	Taux de succès d'une thérapie?
		Présence/absence d'une espèce	Habitat de l'espèce

Situation mathématique	**Résultat observable**	**Expérience**	**Question**
Lancer d'un dé	Un entier k $\in$ {1,...,6}	Période de reproduction d'une espèce (6 périodes de 2 mois)	Saison de reproduction priviligiée pour l'espèce?
Positionnement aléatoire de points dans un carré	*n* positions	Distribution spatiale de *n* individus (GPS)	Agrégation dans certains milieux ?

La notion d'ensemble

<u>Expérience aléatoire</u>: expérience dont on ne peut pas prédire le résultat de manière certaine

<u>Ensemble **Ω**</u> = tous les résultats possibles d'une expérience aléatoire donnée.

<u>Ex</u>: lancé d'un dé: **Ω** = {1, 2, 3, 4, 5, 6}

La notion d'événement

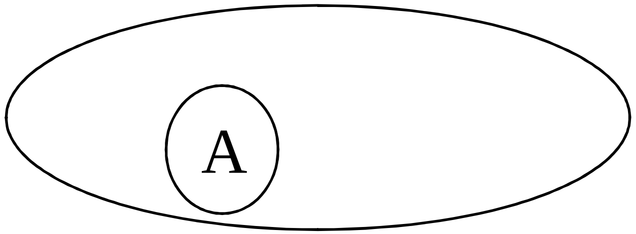
Déterminé par un critère de réussite de l'expérience

<u>Exemple</u>: on jette un dé
« tirer 1» = événement élémentaire
(marche aussi avec 2, 3, 4, 5 ou 6)

« tirer un nombre impair » = événement composé
On notera cet événement A = {1, 3, 5} $\in$ **Ω**.

Représentation sous forme de patate

Ensemble des événements possibles Ω



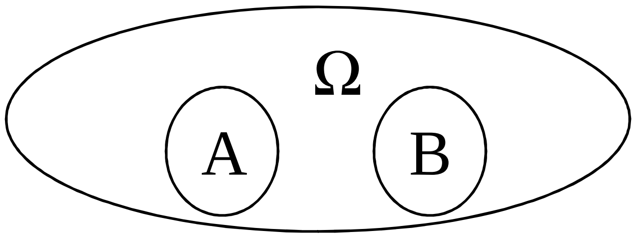
Exemple : dé à six faces

Ω : {1,2,3,4,5,6}

A : {1,3,5}

Quelques notations

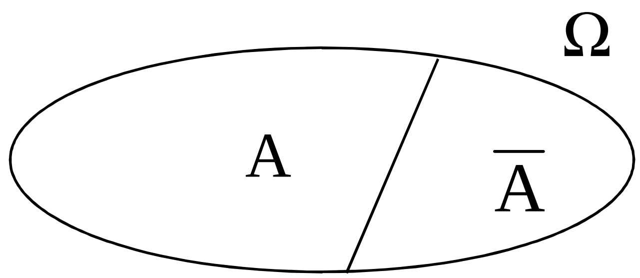
- événement élémentaire A = {3},
- événement impossible A = $\emptyset$,
- événement certain A = Ω,
- événements incompatibles A = {1,5} disjoints de B = {2,6}



Evénements complémentaires

$\overline{A}$: Partie de Ω complémentaire de A

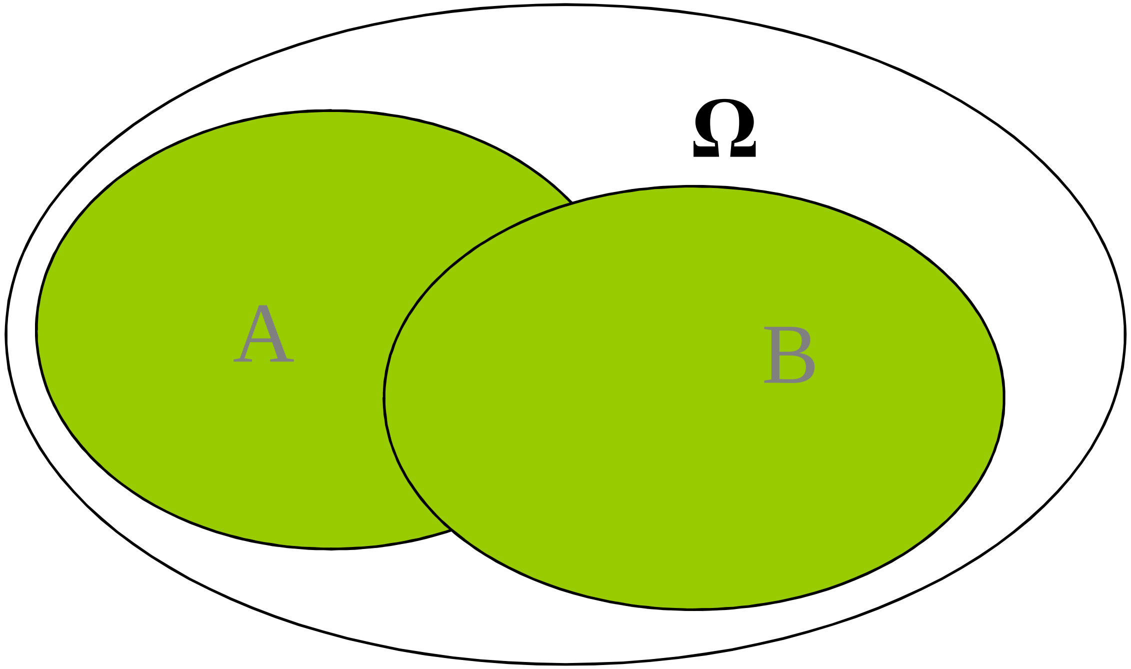
$A = \{2,4,6\}$ paire $\Longrightarrow$ $\overline{A} = \Omega - \{2,4,6\} = \{1,3,5\}$ impaire



Opérations

A = {2,4,6} (paire) ; B={3,6} (multiples de 3)

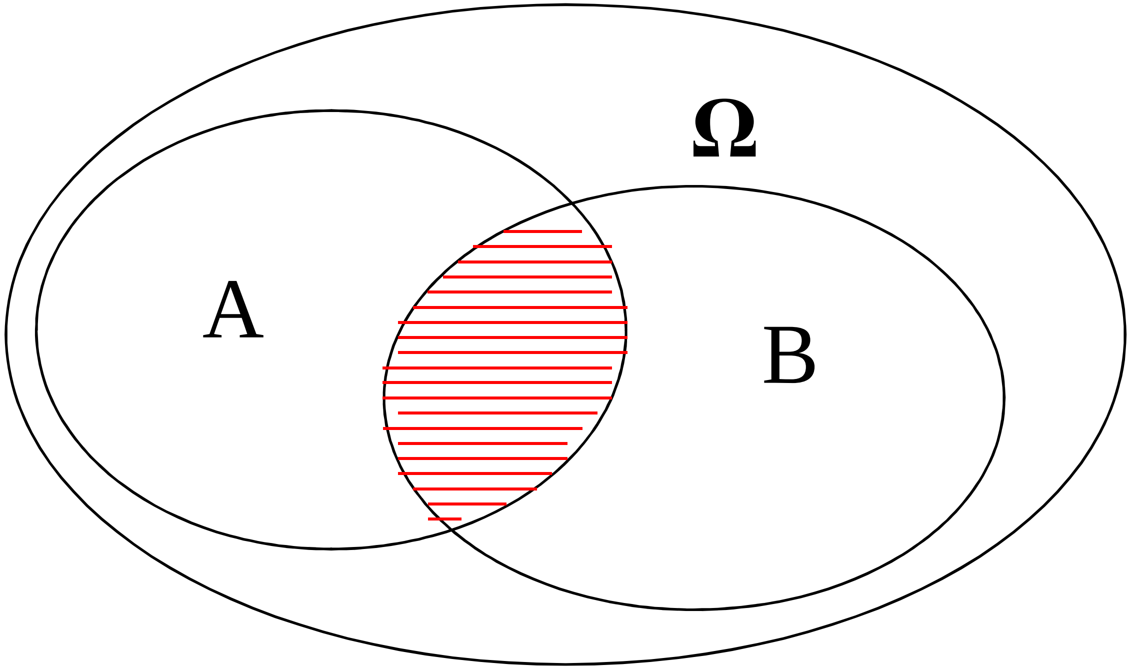
$\cup$: Réunion $\qquad$ $A \cup B = \{2, 3, 4, 6\}$



Opérations

A = {2,4,6} (paire) ; B={3,6} (multiples de 3)

$\cap$: Intersection $A \cap B = \{6\}$



Définition et propriétés d’une probabilité

- une probabilité est toujours comprise entre 0 et 1
- positivité : $p(X) \geq 0$ par exemple: $p(\{1,3\}) \geq 0$
- échelle : $p(\varnothing) = 0,\ p(\Omega) = 1$
- additivité : si $A \cap B = \varnothing$ (incompatibles) alors $p(A \cup B) = p(A) + p(B)$

$$p(\{3\} \cup \{2\}) = p(\{3\}) + p(\{2\})$$

d’où $p(A) + p(\overline{A}) = 1$ $p(\{1,3,5\}) + p(\{2,4,6\}) = 1$

Evénements équiprobables

<u>Exemple du lancé de dé:</u>

6 résultats élémentaires équiprobables

$$\Omega = \{1, .., 6\}$$

$$p(\{1\}) = \ldots = p(\{6\})$$

Or $$p(\{1\}) + \ldots + p(\{6\}) = 1$$

Donc $$p(\{1\}) = \ldots = p(\{6\}) = \frac{1}{6}$$

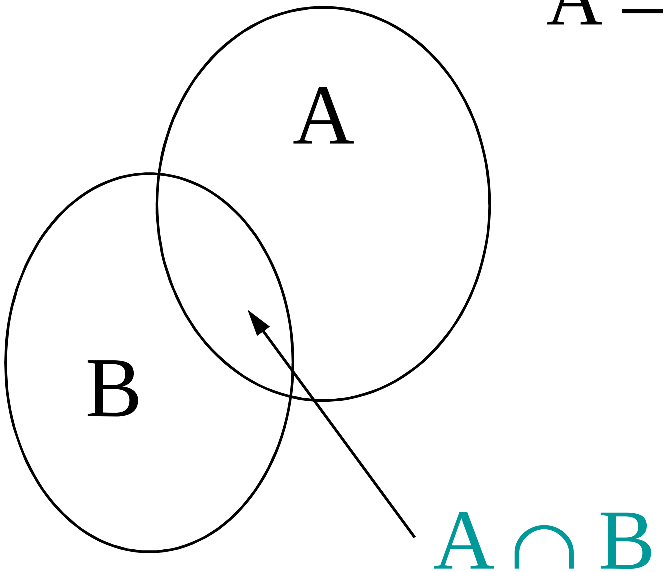
<u>Généralisation facile:</u>

P(résultat élémentaire) = 1 / nombre de résultats possibles

Théorème des probabilités totales

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

$$A = \{1,2,6\};\ B = \{2,4\} \qquad A \cup B = \{1,2,4,6\}$$



$$p(A \cup B) = \frac{3}{6} + \frac{2}{6} - \frac{1}{6} = \frac{4}{6}$$

Probabilités conditionnelles

Exemple:

Population d'élevage avec 30 mâles et 25 femelles. 18 mâles et 4 femelles ont un poids supérieur à 3 Kg.

Je considère une femelle au hasard dans ma population. Quelle est la probabilité que son poids soit supérieur à 3 Kg?

Raisonnement faux:
55 individus dont 22 font plus de 3 Kg donc P = 22/55 = 0.4

Ce raisonnement est faux car il ne prend pas en compte l'information sur le sexe. On parle alors de **probabilités conditionnelles**.

Probabilités conditionnelles

Intuitivement, cette probabilité vaut

$$P = \frac{4}{25} = \frac{\text{Nombre de femelles dont le poids est supérieur à 3 Kg}}{\text{Nombre de femelles total}}$$

En divisant en haut et en bas par le nombre total d'individus, et en écrivant F=« l'individu est une femelle », X = « avoir un poids supérieur à 3 Kg » :

$$\boxed{P = \frac{p(X \; et \; F)}{p(F)} = \frac{p(X \cap F)}{p(F)}}$$

On parle alors de la probabilité d' « avoir un poids > 3 Kg (X) » sachant que (ou conditionnellement au fait que) « l'individu est une femelle (F) »

$$\boxed{p(X/F) = \frac{p(X \cap F)}{p(F)}}$$

Probabilités conditionnelles

On peut retourner la formule:

$$p(A/B)=\frac{p(A\cap B)}{p(B)} \qquad \Leftrightarrow \qquad p(A\cap B)=p(B)\times p(A/B)$$

$$\Leftrightarrow \qquad p(A\cap B)=p(A)\times p(B/A)$$

Jeu de 32 cartes : Probabilité de tirer successivement 2 valets ?

A : 1ère carte est un valet

B : 2ème carte est un valet

$$p(A\cap B)=p(A)\times p(B/A) \quad \text{avec} \quad p(A)=\frac{4}{32} \quad p(B/A)=\frac{3}{31}$$

$$p(A\cap B)=\frac{4}{32}\times\frac{3}{31}=\frac{3}{248}\approx 0{,}012=1{,}2\%$$

Evénements indépendants

$$p(A/B) = p(A)$$

B réalisé ne modifie rien

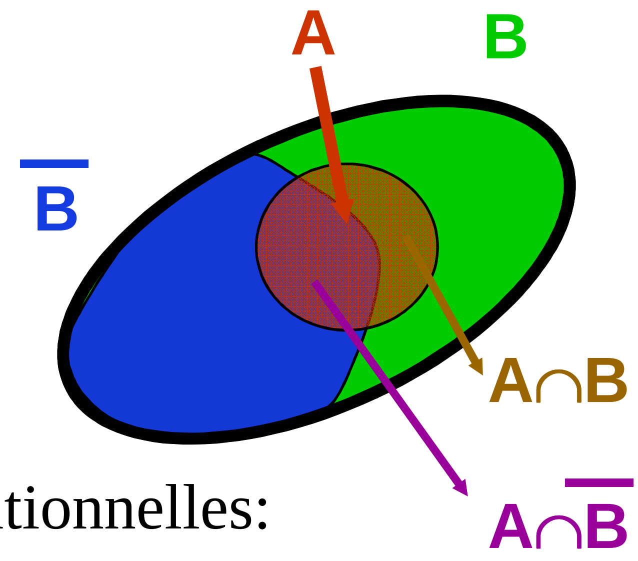
$$p(A \cap B) = p(A) \times p(B)$$

Exemple de valet mais **avec remise de la 1ère carte**

$$p(A \cap B) = p(A) \times p(B) = \frac{4}{32} \times \frac{4}{32} = \frac{1}{64} \approx 0{,}0156 = 1{,}56\%$$

Décomposition de la probabilité d'un événement

Mathématiquement:
(A et B deux événements)



$$p(A) = p(A \cap B) + p(A \cap \overline{B})$$

En utilisant la formule des probabilités conditionnelles:

$$\boxed{p(A) = p(A/B)\,p(B) + p(A/\overline{B})\,p(\overline{B})}$$

<u>Généralisation:</u>

$B_1,\ldots, B_n$ n événements disjoints et donc la réunion vaut Ω

$$\boxed{p(A) = p(A/B_1)\,p(B_1) + p(A/B_2)\,p(B_2) + \ldots + p(A/B_n)\,p(B_n)}$$

Décomposition de la probabilité d'un événement

Un exemple: dans un lycée, on sait que

- 30% de Term. L, 25% d' S et 45% de ES.
- 65% des L font du grec
- 15% des S font du grec
- 5% des ES font du grec

Probabilité qu'un élève pris au hasard fasse du grec?

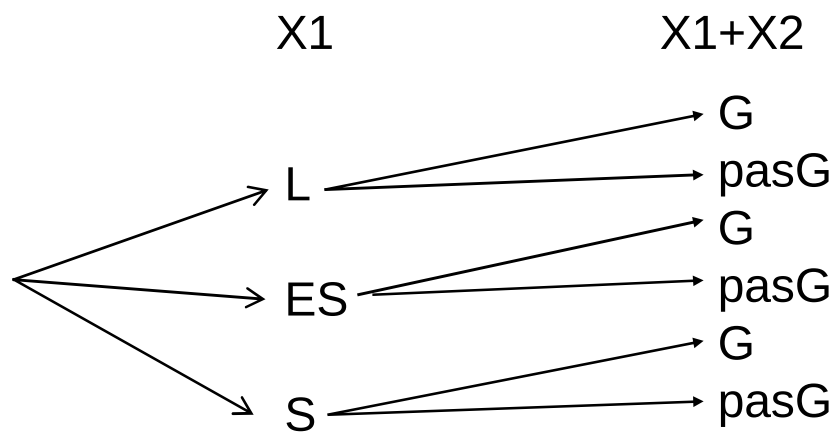
G = événement « faire du grec »

$$
\begin{aligned}
P(G) &= P(G/L)\ P(L) + P(G/S)\ P(S) + P(G/ES)\ P(ES) \\
&= 0.65*0.30+0.15*0.25+0.05*0.45 \\
&= 0.255
\end{aligned}
$$

Décomposition de la probabilité d'un événement

Remarque:

C'est la formalisation mathématique de ce que l'on fait intuitivement en faisant des arbres



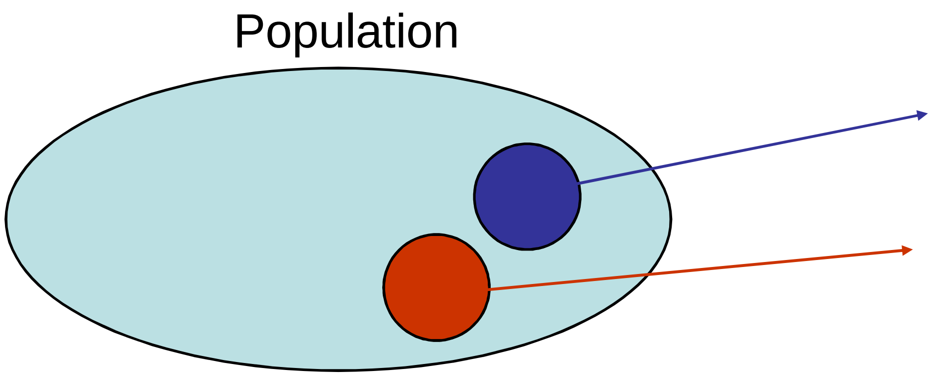
Théorème de Bayes

Exemple: test sanguin pour un virus (prévalence = 0.01 %) :

- positif dans 99% des cas si infecté

- 1% de faux positifs

Je viens de faire les test, il est positif: quelle est la probabilité que je sois porteur du virus ?



Deux types de positifs:

Vrais positifs

Faux positifs

$\Rightarrow$ combien de chaque ?

Théorème de Bayes

V = {la personne testée est porteuse du virus}
T = {la personne testée a un test positif}

Mise en équation de l'énoncé:
P(T / V) = 0.99
P(T / nonV) = 0.01
P(V) = 0.0001

Population = 1 000 000
Nombre de vrais positifs = 1 000 000 * 0.0001 * 0.99 = 99
Nombre de faux positifs = 1 000 000 * 0.9999 * 0.01 = 9999

Proportion de vrais positifs parmi les positifs = 99 / (99 + 9999) = 0.01

Conclusion: comme il y a environ 100 fois plus de faux positifs que de vrais positifs, même si je suis positif au test je n'ai qu'une chance sur 100 d'être malade!

Théorème de Bayes

Ce calcul découle du théorème de Bayes:

Soient A et B deux événements, alors

$$\boxed{p(B/A)=\frac{p(A\cap B)}{p(A)}=\frac{p(A\cap B)}{p(A\cap B)+p(A\cap \overline{B})}=\frac{p(A/B)\,p(B)}{p(A/B)\,p(B)+p(A/\overline{B})\,p(\overline{B})}}$$

Théorème de Bayes généralisé

$B_1,\ldots, B_n$ n événements disjoints et donc la réunion vaut Ω

$$\boxed{p(B_1/A)=\frac{p(A/B_1)\times p(B_1)}{\sum_{i=1}^{n} p(A/B_i)\times p(B_i)}}$$