

Plan

2/78

Ce diaporama ne constitue pas un
polycopié du cours.
Il accompagne l'enseignement oral et
ne saurait s'y substituer.



Plan

3/78

👉 Intérêt des Bio-Statistiques

👉 Les Probabilités

👉 Applications

👉 Probabilités Conditionnelles

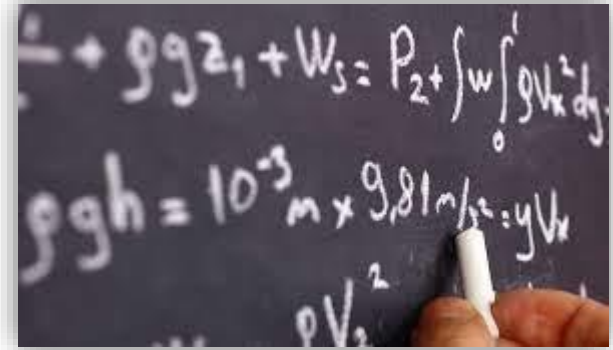
👉 Tableau de Contingence

👉 Courbes ROC

Les Mathématiques et la Santé

5/78

Les bases mathématiques du fonctionnement du scanner :
la transformée de Radon



Prix Nobel de médecine 1979
décerné aux physiciens **G.N. Hounsfield** et **A. M. Cormack**
pour la construction du premier
prototype de scanner

La statistique et les mathématiques sont à la base des innovations médicales majeures.

L'Outil Statistique au Service de la Santé

6/78

La statistique permet de répondre à des questions **centrales et récurrentes** dans la pratique médicale :

1. Définir la normalité

- Quelle est la valeur "normale" d'une grandeur biologique (taille, poids, tension artérielle, glycémie) ?
- Comment distinguer une variation physiologique d'un signe pathologique ?

2. Évaluer la performance des examens

- Quelle est la fiabilité d'un examen complémentaire ?
- Quelle sensibilité (détection des malades) et quelle spécificité (exclusion des non-malades) peut-on attendre ?

3. Apprécier les risques

- Quel est le risque de complication d'une pathologie donnée ?
- Quels sont les bénéfices mais aussi les effets indésirables d'un traitement ?
- Comment estimer la probabilité d'événements rares mais graves ?

4. Comparer les traitements et les stratégies

- Le traitement A est-il réellement plus efficace que le traitement B ?
- Quelle est la taille de l'effet, et est-elle cliniquement pertinente ?
- Qu'apportent les essais cliniques randomisés et les méta-analyses ?

*Sans statistiques, la médecine resterait empirique. Avec elles, elle devient une science fondée sur les preuves
(Evidence-Based Medicine)*

L'Outil Statistique au Service de la Santé

7/78

Outil statistique

- Décrire : moyennes, écarts-types
- Comparer : groupes, hypothèses
- Estimer : p-valeur, IC95%

Observations médicales

- Données brutes : tension, cholestérol, imagerie
- Variabilité naturelle + erreurs de mesure

Décision en santé

- Adapter un traitement (PK-PD)
- Évaluer un test (Se, Sp, VPP, VPN)
- Orienter la santé publique

La statistique transforme des données hétérogènes en connaissances exploitables pour la santé.

Statistiques Vs Variabilité

8/78

Dans la nature la règle c'est la variabilité

Extrait du polycopié de Biostatistique - Carrat, Mallet, Morice, Faculté Pierre

variabilité

nom féminin

Caractère de ce qui est variable.

La variabilité est l'incertitude

Toutes ces questions biologiques qui est (liée au protocole) décomposer la variabilité d'une grandeur mesurée en deux grandes composantes :

ne propriété fondamentale des système la somme d'une variabilité expérimentale proprement biologique. On peut ainsi

variabilité totale = variabilité biologique + variabilité métrologique

La variabilité biologique peut être elle-même décomposée en deux termes : d'une part la variabilité intra-individuelle, qui fait que la même grandeur mesurée chez un sujet donné peut être soumise à des variations aléatoires ; et d'autre part la variabilité inter-individuelle qui fait que cette même grandeur varie d'un individu à l'autre.

variabilité biologique = variabilité intra-individuelle + variabilité inter-individuelle

La variabilité intra-individuelle peut être observée lors de la mesure de la performance d'un athlète qui n'est pas capable des mêmes performances à chaque essai, mais qui se différencie des autres athlètes (variabilité inter-individuelle). En général, la variabilité intra est moindre que la variabilité inter.

La variabilité métrologique peut être elle aussi décomposée en deux termes : d'une part les conditions expérimentales dont les variations entraînent un facteur d'aléas ; et d'autre part les erreurs induites par l'appareil de mesure utilisé.

variabilité métrologique = variabilité expérimentale + variabilité appareil de mesure

La mesure de la pression artérielle peut grandement varier sur un individu donné suivant les conditions de cette mesure ; il est ainsi recommandé de la mesurer après un repos d'au moins 15 minutes, allongé, en mettant le patient dans des conditions de calme maximal. Cette recommandation vise à minimiser la variabilité due aux conditions expérimentales. La précision de l'appareil de mesure est une donnée intrinsèque de l'appareil,

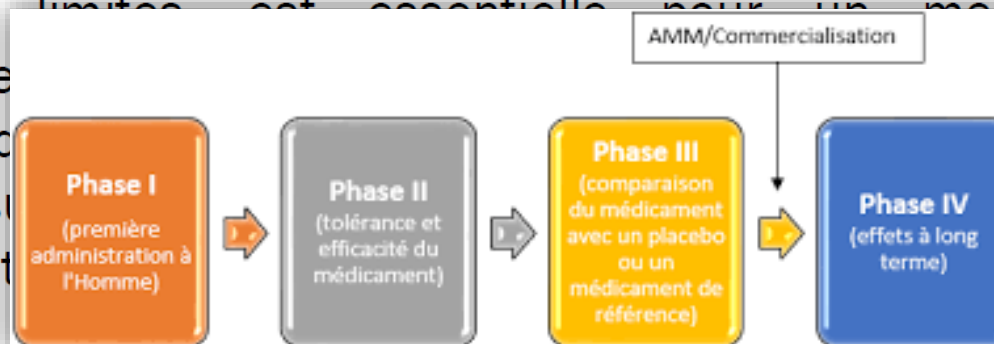
Statistiques Vs Variabilité

9/78

Comment prendre une décision tenant compte de cette variabilité?

Pour prendre une décision diagnostique ou thérapeutique le médecin doit avoir des éléments lui permettant de prendre en compte cette **variabilité** naturelle, pour distinguer ce qui est dans la **norme** de ce qui ne l'est (par exemple ce qui est pathologique) et pour évaluer la qualité d'un nouvel examen, ou d'une nouvelle thérapeutique (décision thérapeutique). De nos jours, une compréhension approfondie des méthodes **statistiques**, de leur potentiel et de leurs limites est essentielle pour un médecin.

Actuellement
à la suite de
s'appuie sur
les résultats

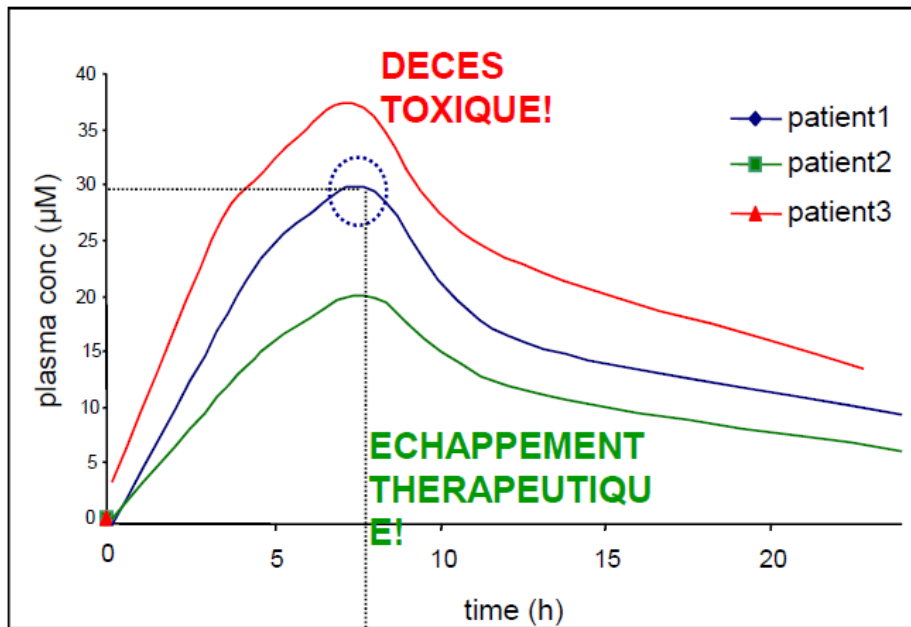


Ajustement Thérapeutique

10/78

☒ SUPPORT PHARMACOCINETIQUE + PILOTAGE DES
POSOLOGIES

Dose standard



Après une **dose standard** administrée en perfusion de 8h, la concentration plasmatique varie fortement d'un patient à l'autre

↔
Perfusion 8 heures

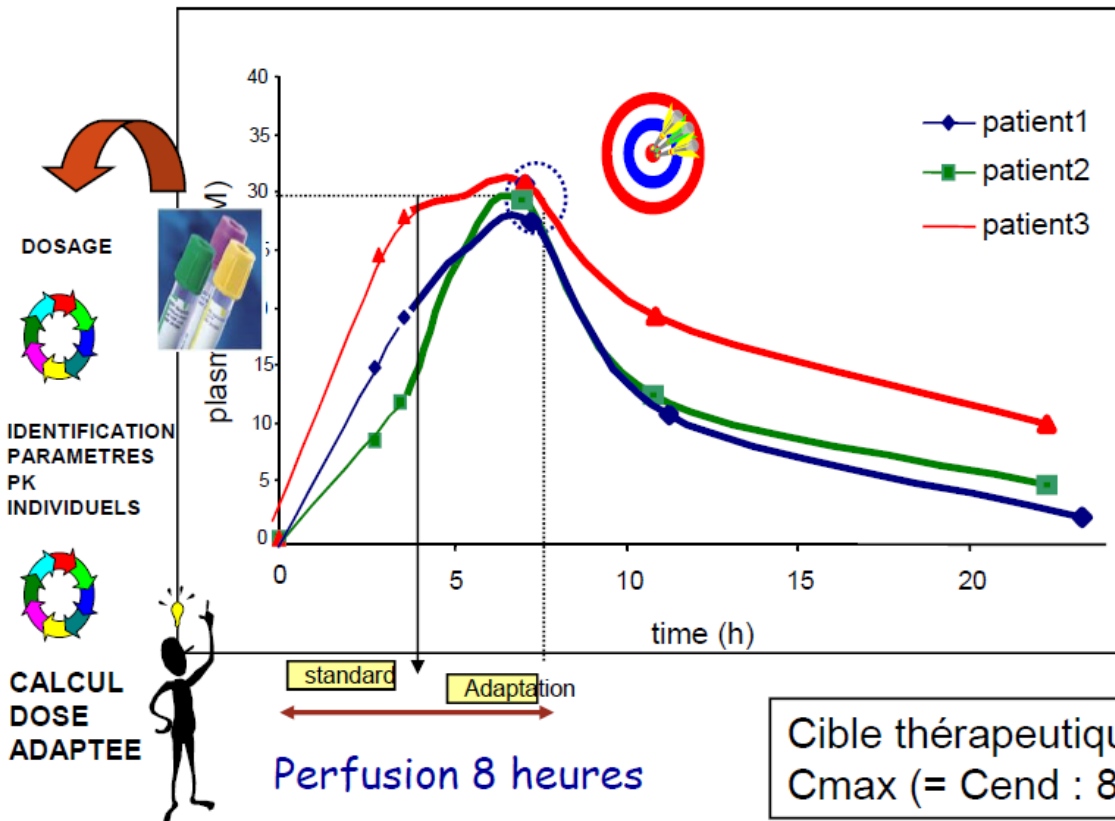
Cible thérapeutique:
 $C_{max} (= C_{end} : 8 \text{ heures})$

L'ajustement est Essentiel

11/78

⊗ SUPPORT PHARMACOCINETIQUE + PILOTAGE DES POSOLOGIES

Dose adaptées



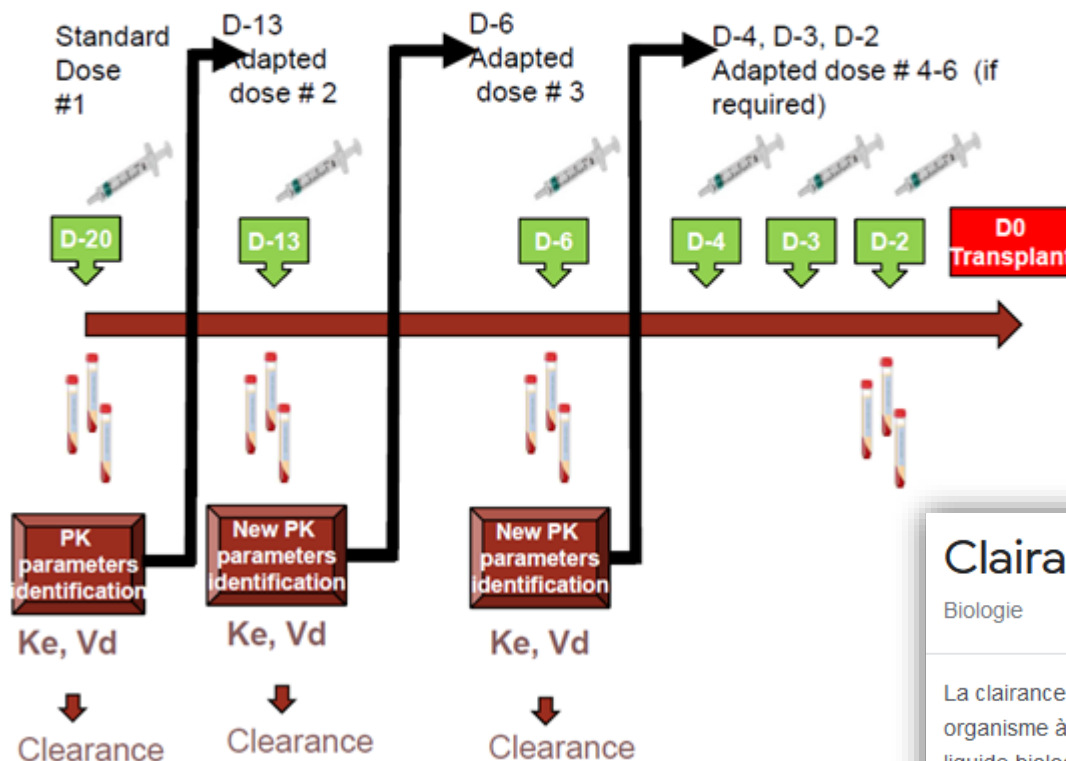
L'effet des médicaments dépend de paramètres pharmacocinétiques (PK) - qui gouvernent son devenir dans l'organisme - et de l'effet sur la cible moléculaire, l'organe et l'organisme entier, qui seront caractérisés par des mesures pharmacodynamiques (PD). C'est donc un modèle PK-PD qui permet d'intégrer l'ensemble des conditions qui prévalent pour observer l'effet d'un médicament.

L'ajustement est Essentiel

12/78

Medicine with HD Busulfan Precision

✓ Phenomenological modeling of Busulfan



Busulfan

Médicament

Le busulfan est un diester sulfonique utilisé comme anticancéreux dans le conditionnement préalable à une greffe de cellules souches hématopoïétiques. Son activité pharmacologique ne s'exerce que sur la moelle osseuse.

Formule : $C_6H_{14}O_6S_2$

Masse molaire : 246,304 g/mol

Numéro CAS : 55-98-1

T° fusion : 287 °C déc

Clairance

Biologie

La clairance est la capacité d'un tissu, organe ou organisme à éliminer une substance donnée d'un liquide biologique.

Objectif de ce Cours

13/78

Introduction aux Biostatistiques

Objectifs :

- Comprendre les méthodes statistiques utilisées
- Interpréter correctement les résultats de recherche
- Adopter un raisonnement scientifique rigoureux
- Aider à la décision dans la pratique médicale

Plan du cours :

- Bases théoriques des probabilités → Compréhension et utilisation des outils
- Applications au diagnostic et au thérapeutique → Sensibilité, spécificité, courbes ROC
- Exemples d'intervention des mathématiques en médecine → Introduction aux tests statistiques
- Statistique inductive → Hypothèses H_0 / H_1 , ex. comparaison de deux traitements

Plan

14/78

👉 Intérêt des Bio-Statistiques

👉 Les Probabilités

👉 Applications

👉 Probabilités Conditionnelles

👉 Tableau de Contingence

👉 Courbes ROC

Probabilités : Corrélation et Causalité

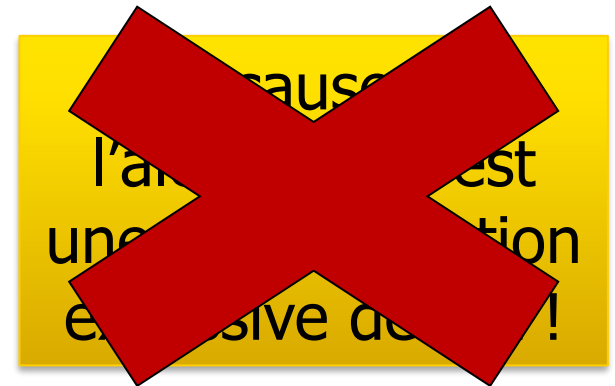
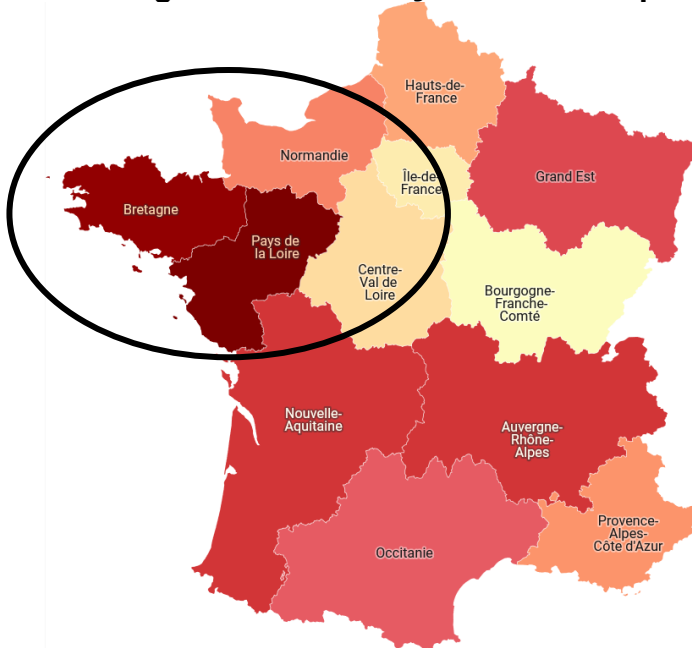
15/78

Régions où il y a le plus de vaches

Trois régions de France détiennent 52% des vaches laitières françaises :

- La **Bretagne** : 1ère région productrice en volume,
- Les **Pays de la Loire**,
- La **Normandie**.

Les régions où les français boivent le plus d'alcool



Interprétation fallacieuse

Probabilités : Les Variables Cachées

16/78

**Les probabilités peuvent « tout expliquer »
mais permettent aussi de « tout remettre en question »**

Ce que l'on sait (INCA)

Le tabac est la cause directe ou un facteur favorisant de nombreux types de cancers. Il est responsable de **plus de 8 cancers du poumon sur 10**, de près de 70 % des cancers des voies aérodigestives supérieures (bouche, larynx, pharynx, œsophage) et de 35 % des cancers de la vessie.



Hypothèse Alternative



**Peut-être existe-t-il un gène responsable de
l'addiction (dont celle au tabac) !!**

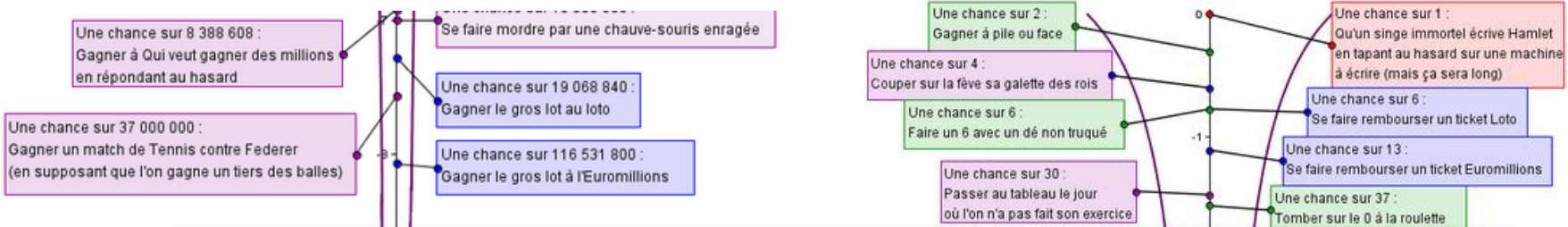
**Peut-être que ce même gène est aussi responsable de la prolifération
incontrôlée de certaines cellules : tumeurs malignes**

Conclusion Erronée

Le tabac n'est pas la cause du cancer du poumon

Probabilités Improbables

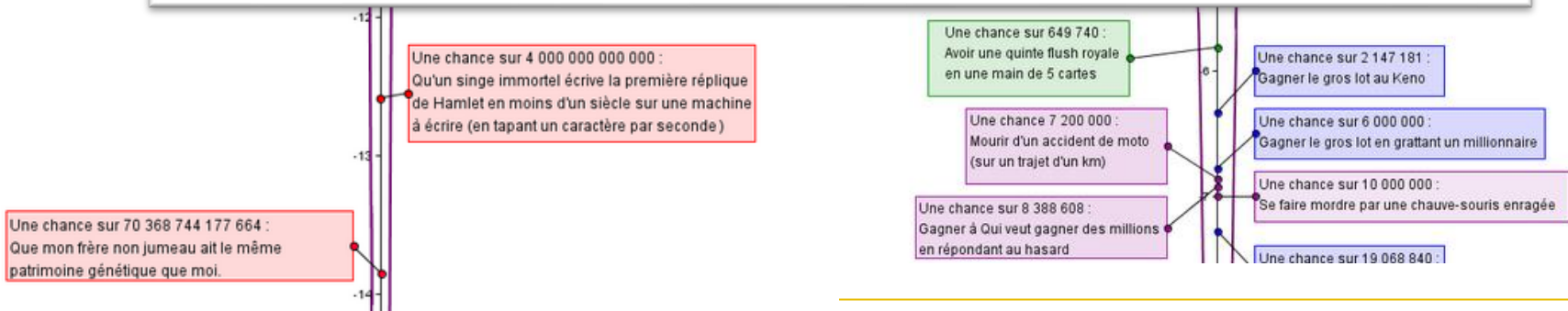
17/78



“Le risque qu’il y ait une bombe dans un avion est de un sur un million. Le risque qu’il y ait deux bombes dans un avion est de un sur cent milliards. La prochaine fois que vous prendrez l’avion, diminuez les risques, emmenez votre bombe !”



De Benny Hill



Probabilités Conditionnelles : Attention

18/78

Paradoxe des 2 Jumeaux (Gardner 1959) où l'importance de la formulation

M. Jones a deux enfants. L'enfant aîné est une fille. Quelle est la probabilité que les deux enfants soient des filles ?

M. Smith a deux enfants. Au moins l'un des deux est un garçon. Quelle est la probabilité que les deux enfants soient des garçons ?

		Sexe du deuxième enfant	
		Fille	Garçon
Sexe du premier enfant	Fille	FF	FG
	Garçon	GF	GG

Non maîtrise des probabilités conditionnelles = sources de confusion!

- (exemple tiré de « petit traité de hasardologie » de Hubert Krivine) :

$P(\text{Noir}/\text{Prison})$

$P(\text{Prison}/\text{Noir})$

- Confusion fréquente entre $P(A/B)$ et $P(B/A)$

Dans la famille de M. Jones, les seuls cas possibles sont notés FF et FG et le seul cas favorable est FF. La probabilité que les deux enfants soient des filles vaut donc **1/2**.

Dans la famille de M. Smith, les cas possibles sont FG, GF et GG, alors que le seul cas favorable est GG. La probabilité que les deux enfants soient des garçons vaut donc **1/3**.

Approche Bayésienne

19/78

Estimation Baysienne dans l'adaptation de posologie

☒ SUPPORT PHARMACOCINETIQUE + PILOTAGE DES POSOLOGIES

☒ Approche Bayesienne

L'approche bayésienne combine les données individuelles (vos prélèvements) et les connaissances de la population (données a priori) pour ajuster la posologie de façon personnalisée.



Information
individuelle
(prélèvements)



Information
A priori
(Matrice de population)



ESTIMATION BAYESIANNE
DES PARAMETRES INDIVIDUELS



CIBLAGE POSOLOGIQUE!

%

Probabilité des
événements
selon une certaine
théorie

Statistiques
Fréquentistes



%

Probabilité des
théories
au vu de certains
événements

Statistiques
Bayésiennes



Expérience Aléatoire

20/78

- Une **expérience** est dite **aléatoire** lorsqu'on ne peut pas prévoir exactement les résultats, car les facteurs influençant le résultat ne sont pas maîtrisés.
- Un **événement aléatoire** est un événement qui peut ou non se produire au cours d'une expérience aléatoire.

Exemple : Dans l'expérience aléatoire de lancer un dé, un événement aléatoire pourrait être « obtenir un 6 ».

- L'**univers** Ω d'une expérience est l'ensemble de tous les résultats possibles. Chaque élément de Ω est appelé **événement élémentaire**.

Idée Intuitive

21/78

- La **probabilité** d'un événement est le rapport k/m : c'est le **nombre de cas favorables** divisé par le **nombre de cas possibles**.

Exemple : Dans un jeu de 52 cartes contenant 13 carreaux, si toutes les cartes ont des chances égales d'être tirées, la probabilité de tirer un carreau est 

- Si m résultats peuvent se produire avec des chances égales et si k résultats correspondent à la réalisation de l'événement, alors la probabilité de cet événement est k/m .

Définition Fréquentielle

22/78

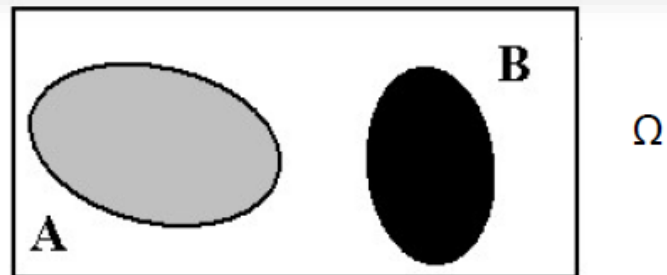
- Si une **expérience** est répétée un grand nombre de fois dans des conditions uniformes, on observe généralement que la **fréquence relative** (pourcentage de réalisation) d'un événement A se stabilise.
 - Ce phénomène est appelé **régularité statistique**.
 - Ce nombre fixe devient par définition la **probabilité mathématique** de l'événement considéré.
-
- La **probabilité**, telle que définie, est une version idéalisée de la **fréquence relative**.
 - Une estimation pratique de la probabilité d'un événement est donnée par la fréquence relative, et la précision de cette estimation peut être indiquée par un **intervalle de confiance** pour un niveau de risque donné.
 - Dans de nombreux cas, la probabilité peut être **modélisée par une loi**.

Expérience Aléatoire

23/78

- Soient Ω l'univers d'une expérience aléatoire, $\mathcal{A}$ une tribu sur Ω , et $(\Omega, \mathcal{A})$ l'espace probabilisable ainsi constitué. On appelle **événement** toute partie de Ω qui appartient à la tribu $\mathcal{A}$.
- Si l'événement est constitué d'un seul élément, on parle alors d'un **événement élémentaire**.

Ω ensemble des événements possibles
 A sous-ensemble de Ω
 B sous-ensemble de Ω



Exemple concret

- $\Omega = \{1, 2, 3, 4, 5, 6\}$ (lancer d'un dé).
- $A = \{2, 4, 6\}$ = "obtenir un nombre pair".
- $B = \{1\}$ = "obtenir un 1" (événement élémentaire).

Événements Incompatibles

24/78

- Les événements A et B ne peuvent se produire simultanément.
 - *Exemple* : extraire un pique ou un cœur d'un paquet de cartes.
- Si deux événements sont exhaustifs et mutuellement **incompatibles** (comme être à l'extérieur ou à l'intérieur d'une pièce), la non-réalisation de l'un implique la réalisation de l'autre.

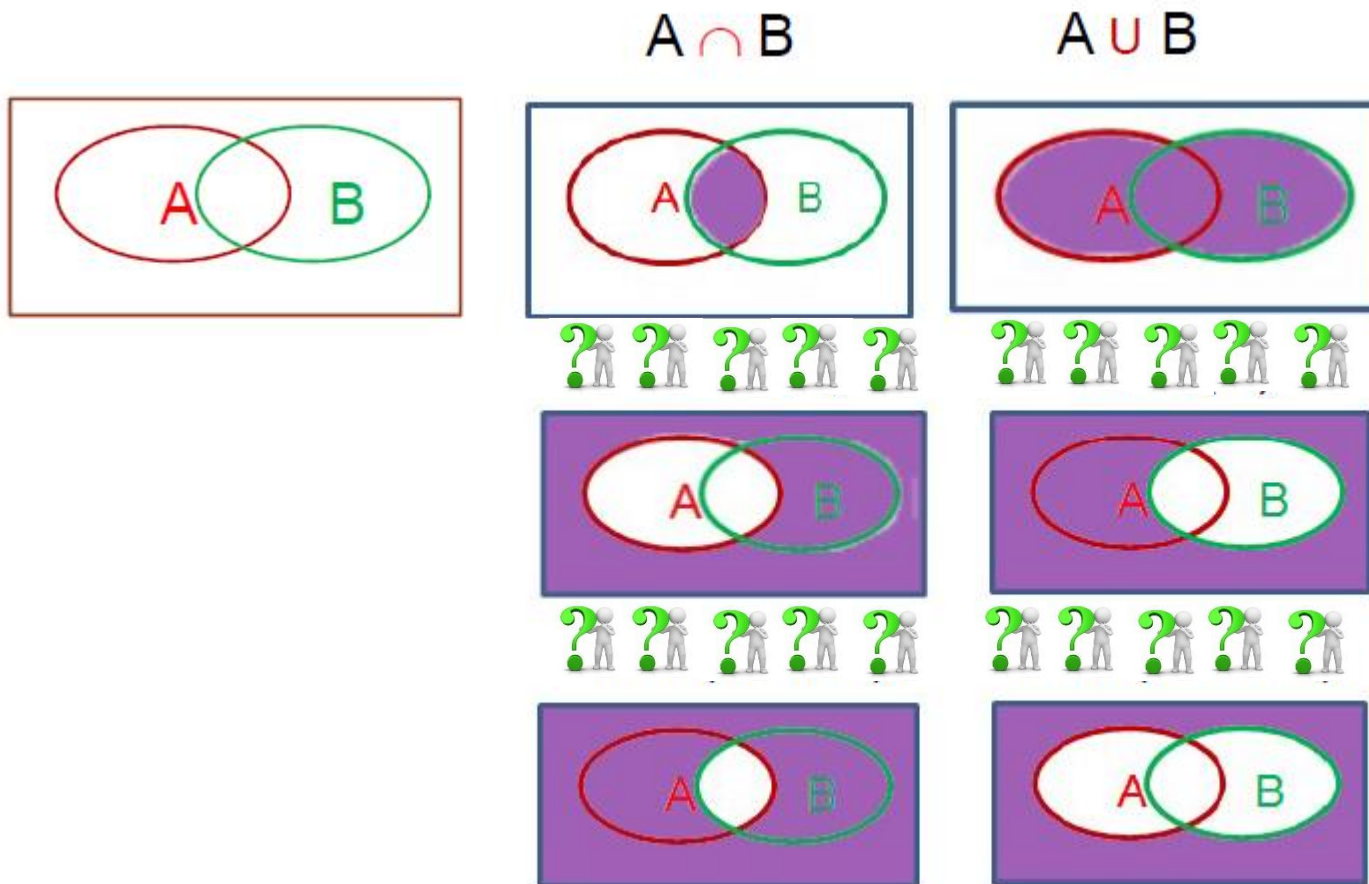
Événements Non-Incompatibles

25/78

- Les événements peuvent se produire simultanément
- L'intersection n'est pas vide.
- *Exemple :*
 - Extraire une dame et un carreau.
 - Avoir du cholestérol et porter un chapeau.
 - Avoir un diabète et une grippe.
- Ne pas confondre événements incompatibles et événements indépendants.

Les Ensembles

26/78



Rappels sur l'intersection et la Réunion

27/78

Définition : Soient E un ensemble et A, B deux éléments de $\mathcal{P}(E)$ (l'ensemble des parties de E).

On définit les parties suivantes de E :

- L'intersection de A et B : $A \cap B = \{x \in E \mid x \in A \text{ et } x \in B\}$
- L'union de A et B : $A \cup B = \{x \in E \mid x \in A \text{ ou } x \in B\}$
- Le complémentaire de A dans E : $C_E(A) = \{x \in E \mid x \notin A\}$

Proposition : Soient E un ensemble et A, B, C trois éléments de $\mathcal{P}(E)$, les sept assertions suivantes sont réalisées :

1. $A \cap B = B \cap A$ et $A \cup B = B \cup A$
2. $A \cap (B \cap C) =$  $A \cup (B \cup C) =$ 
3. $A \cap E =$  $A \cup \emptyset =$ 
4. $A \cap A =$  $A \cup A =$ 
5. $A \cup (B \cap C) =$  et $A \cap (B \cup C) =$ 
6. $C_E(A \cap B) =$  et $C_E(A \cup B) =$ 
7. $C_E(C_E(A)) =$ 

Le Complémentaire

28/78

En théorie des probabilités, le complémentaire de A dans E , $C_E(A) = \{x \in E; x \notin A\}$, est noté $\overline{A}$.

Les propriétés (6) précédentes s'écrivent alors :

$$\overline{A \cap B} = ? ?$$

$$\overline{A \cup B} = ? ?$$

Deux événements disjoints, $A \cap B = \emptyset$, sont dits **incompatibles**.

Définition d'une Probabilité

29/78

Définition

On appelle probabilité sur l'algèbre $\mathcal{A}$ sur Ω une application P de $\mathcal{A}$ dans $[0, 1]$, qui vérifie :

1. $P(\Omega) = 1$
2. Pour tout $A, B \in \mathcal{A}$, avec $A \cap B = \emptyset$,

Si compatible ou incompatible ?

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

Du point de vue terminologique, chaque fois que l'on choisit une algèbre $\mathcal{A}$ et une probabilité P , le triplet $(\Omega, \mathcal{A}, P)$ est appelé **espace probabilisé**.

Premières propriétés

1. $P(\emptyset) = 0$
2. Pour tout $A \in \mathcal{A}$,

$$P(\overline{A}) = 1 - P(A)$$

3. Si $A_1, \dots, A_n$ sont des éléments de $\mathcal{A}$, deux à deux incompatibles,

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$$

Hypothèse d'Equiprobabilité

30/78

Dans le cadre des expériences dont l'univers est fini, il est possible de modéliser une situation probabiliste en posant une hypothèse d'équiprobabilité. Cette hypothèse stipule que chaque événement élémentaire $\{\omega_i\}$ a la même probabilité d'apparition. Autrement dit, pour tout i, j entre 1 et n , nous avons :

$$P(\{\omega_i\}) = P(\{\omega_j\})$$

Sous cette hypothèse, et en prenant en compte que la somme des probabilités de tous les événements élémentaires de l'univers doit être égale à 1 :

$$P(\Omega) = \sum_{i=1}^n P(\{\omega_i\}) = 1$$

on en déduit que pour tout entier i entre 1 et n , la probabilité de chaque événement élémentaire est :

$$P(\{\omega_i\}) = \frac{1}{n} = \frac{1}{\text{card}(\Omega)}$$

Calcul d'une Probabilité

31/78

Hypothèse d'équiprobabilité :

Sous l'hypothèse d'équiprobabilité, la probabilité est définie pour tout événement $A \in \mathcal{A}$ par la formule :

$$P(A) = \frac{p}{n} = \frac{\text{card}(A)}{\text{card}(\Omega)}$$

où :

- p est le nombre de cas favorables (éléments dans A),
- n est le nombre total de cas possibles (éléments dans Ω),
- $\text{card}(A)$ est le cardinal de A (nombre d'éléments dans A),
- $\text{card}(\Omega)$ est le cardinal de Ω (nombre total de résultats possibles).

Cela correspond à la fameuse formule :

$$P(A) = \frac{\text{nombre de cas favorables à l'événement } A}{\text{nombre de cas possibles}}$$

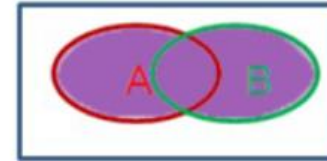
En conclusion, résoudre ce type de problème revient à effectuer un dénombrement des cas favorables et des cas possibles dans l'univers considéré.

Evènements Non-Disjoints

Explication des concepts :

1. Probabilité de l'union de deux événements :

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



Cela signifie que pour obtenir la probabilité que l'un ou l'autre des événements se produise, on additionne les probabilités individuelles puis on soustrait la probabilité de leur intersection (événement commun aux deux).

Si compatible ou incompatible ?

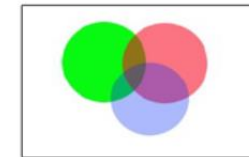
2. Décomposition des ensembles :

Les ensembles peuvent être décomposés de façon à exprimer leurs unions et intersections en utilisant les compléments. Voici les équations corrigées :

$$A \cup B = A \cup (B \cap \overline{A}) \quad \text{avec} \quad A \cap (B \cap \overline{A}) = \emptyset$$

$$A = (A \cap \overline{B}) \cup (A \cap B)$$

$$B = (B \cap \overline{A}) \cup (B \cap A)$$



3. Exercice : Montrer la formule d'union pour trois événements :

$$P(A \cup B \cup C) = ?$$

Plan

33/78

👉 Intérêt des Bio-Statistiques

👉 Les Probabilités

👉 Applications

👉 Probabilités Conditionnelles

👉 Tableau de Contingence

👉 Courbes ROC

Exercices sur le Vocabulaire

34/78

Lors d'un jet de deux dés cubiques, on s'intéresse aux événements suivants :

A : « La somme obtenue est au moins égale à 5 ».

B : « La somme obtenue est au plus égale à 5 ».

C : « La somme obtenue est strictement inférieure à 3 ».

1) A et B sont-ils contraires ?

2) $\bar{B}$ et C sont-ils incompatibles ?

3) Traduire par une phrase $\bar{C}$.

4) A et $\bar{C}$ sont-ils incompatibles ?



1) A et B ne sont pas contraires car une somme égale à 5 les réalise simultanément.

2) $\bar{B}$ et C sont incompatibles car la somme ne peut être simultanément strictement supérieure à 5 (événement $\bar{B}$) et strictement inférieure à 3 (événement C).

3) L'événement $\bar{C}$ est « La somme est supérieure ou égale à 3 ».

4) A et $\bar{C}$ ne sont pas incompatibles car ils sont simultanément réalisés par une somme supérieure ou égale à 5.

Exercices sur le Vocabulaire

35/78

Une urne contient des boules blanches, noires et rouges. On tire une boule de l'urne. On note :

A : « Tirer une boule blanche ».

B : « Tirer une boule ni blanche ni rouge ».

C : Tirer une boule noire ou une boule rouge ».

1) A et B sont-ils incompatibles ?

2) B et C sont-ils incompatibles ?

3) Traduire par une phrase ne comportant pas de négation $\bar{A}$ et $\bar{B}$.



1) A et B sont incompatibles car une boule ne peut être simultanément blanche et non blanche.

2) B et C ne sont pas incompatibles car le tirage d'une boule noire les réalise simultanément.

3) L'événement $\bar{A}$ est « tirer une boule noire ou rouge ».

L'événement $\bar{B}$ est « tirer une boule blanche ou rouge ».

Exercices sur le Vocabulaire

36/78



Dans chacune de situations décrites ci-dessous, énoncer l'événement contraire de l'événement donné.

- 1) Dans une classe, on choisit deux élèves au hasard. A : « Les deux élèves sont des filles ».
- 2) Dans un groupe de suisses et de belges, on discute avec une personne. B : « La personne est un homme belge ».
- 3) Au restaurant, Luc prend un plat et un dessert. C : « Luc prend une viande et une glace ».
- 4) A une loterie, Elise achète 3 billets.
D : « L'un des billets au moins est gagnant » , E : « Deux billets au maximum sont gagnants ».



- 1) L'événement $\overline{A}$ est « au moins un des deux élèves est un garçon ».
- 2) L'événement $\overline{B}$ est « La personne est soit une femme, soit un suisse ».
- 3) L'événement $\overline{C}$ est « Luc ne prend pas de viande ou ne prend pas de glace ».
- 4) L'événement $\overline{D}$ est « aucun billet n'est gagnant ».
- 5) L'événement $\overline{E}$ est « les trois billets sont gagnants ».

Exercices sur le Dénombrement

37/78

Combien de menus différents peut-on composer si l'on a le choix entre 5 entrées, 2 plats et 4 desserts ?

Une femme a dans sa garde-robe 6 pantalons, 5 hauts et 3 vestes. Elle choisit au hasard un pantalon, un haut et une veste. De combien de façons différentes peut-elle s'habiller ?



Un questionnaire à choix multiples, autorisant une seule réponse par question, comprend 20 questions. Pour chaque question, on propose 4 réponses possibles. De combien de façons peut-on répondre à ce questionnaire ?



On peut composer $5 \times 2 \times 4 = 40$ menus différents.

Elle peut s'habiller de $6 \times 5 \times 3 = 90$ façons différentes.

Il y a 4^{20} façons de répondre à ce questionnaire.

Exercices sur le Dénombrement

38/78

Un étudiant s'habille très vite le matin et prend, au hasard dans la pile d'habits, un pantalon, un tee-shirt, une paire de chaussettes ; il y a ce jour-là dans l'armoire 5 pantalons dont 2 noirs, 6 tee-shirt dont 4 noirs, 8 paires de chaussettes, dont 5 paires noires. Combien y-a-t-il de façons de s'habiller ? Quelles sont les probabilités des événements suivants :

1. il est tout en noir ;
2. une seule pièce est noire sur les trois.



Il y a $5 \times 6 \times 8$ façons de s'habiller.

1. Soit A l'événement "il est tout en noir".

$$\mathbb{P}(A) = \frac{2 \times 4 \times 5}{5 \times 6 \times 8}$$

2. Soit B l'événement "une seule pièce est noire sur les trois".

$$\mathbb{P}(B) = \frac{2 \times (6 - 4) \times (8 - 5) + (5 - 2) \times 4 \times (8 - 5) + (5 - 2) \times (6 - 4) \times 5}{5 \times 6 \times 8}$$

Exercices sur le Dénombrement

39/78

Pour se mettre en tenue réglementaire un chirurgien possède :

Sept Callots (3 bleus et 4 blancs)

Six Bavettes (3 bleues et 3 blanches)

Huit Blouses (5 bleues et 3 blanches)

Pour les plus rusés !

Nbre de tenues bleues possibles = $3 \times 3 \times 5 = 45$

Nbre de tenues « non bleues » = $336 - 45 = 291$

- De combien de façons peut-il s'habiller ?
- Le chirurgien refuse de s'habiller tout en bleus, quels est le nombre de tenues qu'il pourra mettre ?



Le chirurgien peut s'habiller de : $7 \times 6 \times 8 = 336$ façons

Puisque le chirurgien refuse de s'habiller tout en bleu, les possibilités sont schématisées comme suit :

Callots bleus	Bavettes bleues	Blouses bleues	Nbre de façons
0	0	0	$4 \times 3 \times 3 = 36$
1	0	0	$3 \times 3 \times 3 = 27$
1	1	0	$3 \times 3 \times 3 = 27$
1	0	1	$3 \times 3 \times 5 = 45$
0	1	0	$4 \times 3 \times 3 = 36$
0	1	1	$4 \times 3 \times 5 = 60$
0	0	1	$4 \times 3 \times 5 = 60$

On obtient 291 façons

Exercices sur le Dénombrement

40/78

On lance un dé équilibré à six faces. Calculer les probabilités d'obtenir un nombre pair, un multiple de 3, un nombre pair et supérieur à 3.



La probabilité d'obtenir un nombre pair est $\frac{3}{6}$. La probabilité d'obtenir un multiple de 3 est $\frac{2}{6}$. La probabilité d'obtenir un nombre pair et supérieur à 3 est $\frac{2}{6}$.

On lance un dé à 6 faces. On note p_i la probabilité de sortie de la face marquée i . Ce dé est truqué de telle sorte que les probabilités de sortie des faces sont : $p_1 = 0,1$; $p_2 = 0,2$; $p_3 = 0,3$; $p_4 = 0,1$; $p_5 = 0,15$.

Quelle est la probabilité de sortie de la face marquée 6 ?

Quelle est la probabilité d'obtenir un nombre pair ?



Si on note p_6 la probabilité d'apparition du chiffre 6, la somme des probabilités des événements élémentaires valant 1, on a $p_6 = 1 - (p_1 + p_2 + p_3 + p_4 + p_5) = 1 - 0,85 = 0,15$.

L'événement A « obtenir un nombre pair » étant $A = \{2; 4; 6\}$, on a $p(A) = p_2 + p_4 + p_6 = 0,2 + 0,1 + 0,15 = 0,45$.



Il ne fallait surtout pas écrire $p(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)} = \frac{3}{6} = \frac{1}{2}$ car il n'y a pas équiprobabilité des faces de dés.

Exercices sur le Dénombrement

41/78

Une maladie génétique autosomale récessive est présente chez 1/2000 enfants nés vivant. Si les parents sont hétérozygotes pour la maladie, il y a une chance sur 4 pour que l'enfant ait la maladie.
Quelle est la probabilité pour un couple d'hétérozygotes d'avoir 2 enfants sains?



Probabilité d'avoir un enfant sain $\Rightarrow 1 - 0,25 = 0,75$
Les deux naissances sont indépendantes, le résultat est donc le produit des 2 événements élémentaires $\Rightarrow 0,75 \times 0,75 = 0,56$

Dans une population, 45% des individus sont vaccinés contre la fièvre jaune, 60% contre l'hépatite B et 30% contre les deux maladies.
Quelle est la probabilité pour qu'un individu choisi au hasard dans cette population ne soit pas vacciné contre aucune de ces maladies ?

A = vacciné contre la fièvre jaune donc $P(A) = 0,45$

B = vacciné contre l'hépatite B donc $P(B) = 0,60$

$P(A \cap B) = 0,30$

La probabilité qu'un individu soit vacciné vaut $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0,75$

La probabilité complémentaire vaut donc $1 - 0,75 = 0,25$



Exercices sur le Dénombrement

42/78

On choisit une carte au hasard dans un jeu de 32 cartes. On note :

A l'événement : "La carte choisie est un pique".

B l'événement : "La carte choisie est rouge (cœur ou carreau)".

C l'événement : "La carte choisie est une figure (valet, dame, roi)".

- 1) Présenter un modèle mathématique décrivant l'expérience aléatoire.
- 2) Déterminer les probabilités des événements $A, B, C, A \cap B, B \cap C, A \cup B, A \cup C$.
- 3) Déterminer la probabilité de "La carte choisie n'est ni un pique ni une figure".



1) On note Ω l'univers des possibles, ensemble des 32 cartes du jeu. Ainsi $\text{Card}(\Omega) = 32$.

Il y a équiprobabilité des tirages de cartes. Ainsi

$$2) \quad p(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)} = \frac{8}{32} = \frac{1}{4}, \quad p(B) = \frac{\text{Card}(B)}{\text{Card}(\Omega)} = \frac{16}{32} = \frac{1}{2}, \quad p(C) = \frac{\text{Card}(C)}{\text{Card}(\Omega)} = \frac{3 \times 4}{32} = \frac{3}{8},$$

$$p(A \cap B) = 0 \text{ car une carte ne peut être simultanément rouge et pique, } p(B \cap C) = \frac{\text{Card}(B \cap C)}{\text{Card}(\Omega)} = \frac{6}{32} = \frac{3}{16}.$$

$$p(A \cup B) = p(A) + p(B) - p(A \cap B) = \frac{1}{4} + \frac{1}{2} - 0 = \frac{3}{4}. \quad p(A \cup C) = p(A) + p(C) - p(A \cap C) = \frac{1}{4} + \frac{3}{8} - \frac{3}{32} = \frac{17}{32}.$$

$$3) \text{ On cherche } p(\overline{A \cup C}) = 1 - p(A \cup C) = 1 - \frac{17}{32} = \frac{15}{32}.$$



Remarque : on a $p(\overline{A \cup C}) = p(\overline{A} \cap \overline{C})$.

Plan

43/78

👉 Intérêt des Bio-Statistiques

👉 Les Probabilités

👉 Applications

👉 Probabilités Conditionnelles

👉 Tableau de Contingence

👉 Courbes ROC

...Sachant que

44/78

Probabilités conditionnelles en médecine :

- En médecine, les probabilités conditionnelles sont couramment utilisées pour évaluer le risque de maladies en fonction de facteurs spécifiques.
 - Par exemple, on dira : « un individu a x fois plus de chances de développer un cancer du poumon s'il fume un paquet de tabac par jour pendant 40 ans que s'il n'a jamais fumé ».
- La connaissance évolue avec l'information :
 - Avant la réalisation d'un test, la probabilité d'avoir une maladie est notée p (ce qu'on appelle la prévalence).
 - Cependant, cette probabilité change dès qu'un test est réalisé et que le résultat est positif. La question est alors : **Quelle est la probabilité d'avoir la maladie sachant que le test est positif ?**

Définition

45/78

Probabilité conditionnelle pour deux événements non exclusifs :

- Soit deux événements A et B qui ne sont pas exclusifs, c'est-à-dire qui peuvent se produire ensemble.
 - On s'intéresse à la probabilité que l'un de ces événements se réalise sachant que l'autre est déjà réalisé.
- **Notation de la probabilité conditionnelle :**
 - $P(A|B)$: Probabilité de A sachant que B est réalisé.
 - Inversement, $P(B|A)$ représente la probabilité de B sachant que A est réalisé.
 - Exemple de question :
 - Quelle est la probabilité d'avoir une douleur de la fosse iliaque gauche sachant que l'on a des diverticules ?
 - Quelle est la probabilité d'avoir des diverticules sachant qu'on observe une douleur dans la fosse iliaque gauche ?

Formulation Mathématique

46/78

Probabilité conditionnelle et Formule de Bayes

- Probabilité de A sachant B :

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

où $P(A \cap B)$ est la probabilité conjointe de A et B .

- Dédution :

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

et de façon symétrique :

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

- Formule de Bayes :

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

- Égalités importantes :

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

$$P(A|B) = 1 - P(\bar{A}|B)$$

Attention : En général, $P(A|\bar{B}) \neq 1 - P(A|B)$.

Exemple 1

47/78

Un laboratoire pharmaceutique a réalisé des tests sur 800 patients atteints d'une maladie. Certains sont traités avec le médicament A, d'autres avec le médicament B. Le tableau présente les résultats de l'étude :

	Médicament A	Médicament B	Total
Guéri	383	291	674
Non guéri	72	54	126
Total	455	345	800

On choisit au hasard un patient et on considère les événements suivants :
A : « Le patient a pris le médicament A. » G : « Le patient est guéri. »

La probabilité qu'un patient soit traité avec le médicament A est égale à $P(A) =$ 

La probabilité qu'un patient soit guéri est égale à $P(G) =$ .

La probabilité qu'un patient soit guéri et qu'il soit traité par le médicament A est égale à $P(G \cap A) =$ .

La probabilité qu'un patient ne soit pas guéri et qu'il soit traité par le médicament A est égale à $P(\bar{G} \cap A) =$ .

Exemple 1Bis

48/78

Un laboratoire pharmaceutique a réalisé des tests sur 800 patients atteints d'une maladie. Certains sont traités avec le médicament A, d'autres avec le médicament B. Le tableau présente les résultats de l'étude :

	Médicament A	Médicament B	Total
Guéri	383	291	674
Non guéri	72	54	126
Total	455	345	800

On choisit au hasard un patient et on considère les événements suivants :
A : « Le patient a pris le médicament A. » G : « Le patient est guéri. »

La probabilité que le patient ait pris le médicament A **sachant qu'il est guéri** se note

$P_G(A)$ et est égale à $P_G(A)$ 

La probabilité que le patient soit guéri **sachant qu'il a pris le médicament B** se note

$P_B(G)$ et est égale à $P_B(G)$ 

Définition :

On appelle **probabilité conditionnelle de B sachant A**, la probabilité que l'événement B se réalise sachant que l'événement A est réalisé. On la note : $P_A(B)$

Exemple 2

49/78

On considère 2 urnes contenant chacune 2 boules blanches et 3 boules noires. On tire au hasard une boule de la première urne et on la met dans la deuxième urne, puis on tire une boule de la deuxième urne.

Question :

Quelle est la probabilité que les deux tirages aient donné des boules blanches ?

- A : La première boule tirée est blanche.
- B : La deuxième boule tirée est blanche.



On cherche $P(A \cap B)$

On utilise la formule des probabilités conditionnelles :

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

- $P(A) = \frac{2}{5}$ (probabilité de tirer une boule blanche au premier tirage).
- $P(B|A) = \frac{1}{2}$ (probabilité de tirer une boule blanche au deuxième tirage, sachant que la première était blanche).

Donc :

$$P(A \cap B) = \frac{2}{5} \times \frac{1}{2} = \frac{1}{5}$$

Résultat :

La probabilité que les deux boules tirées soient blanches est $\frac{1}{5}$.

Exemple 3

50/78



*Quand les événements sont indépendants,
la survie = produit des chances de rater.*

Deux chasseurs aperçoivent simultanément un lapin et tirent en même temps. La probabilité que le premier tue le lapin est $4/5$, celle du second est $3/4$. Quelle est la probabilité que le lapin soit tué ?



Soit A_i l'évènement "le i -ème chasseur tue le lapin". L'énoncé donne $P(A_1) = 4/5$, $P(A_2) = 3/4$ et suggère que A_1 et A_2 sont indépendants. Il en est de même pour les évènements contraires. La probabilité de survie du lapin est donc $P(\bar{A}_1 \cap \bar{A}_2) = P(\bar{A}_1)P(\bar{A}_2) = \frac{1}{5} \times \frac{1}{4} = \frac{1}{20}$, et la probabilité que le lapin soit tué est $1 - \frac{1}{20} = \frac{19}{20}$.

Un Théorème Utile

51/78

Théorème de Probabilité Totale

Théorème :

Si la suite d'événements $A_1, A_2, \dots, A_n$ est un système d'événements complets, alors pour tout événement A , on a la formule suivante, dite **formule des probabilités totales** :

$$P(A) = \sum_i P(A|A_i)P(A_i)$$

Cette formule exprime la probabilité d'un événement A en fonction des probabilités conditionnelles données les différents événements A_i qui forment un système complet (i.e., les événements $A_1, A_2, \dots, A_n$ couvrent l'ensemble de l'espace des événements).

Cas Particulier

Si B est un événement dont la probabilité $P(B) \neq 0$ et $P(B) \neq 1$, le système $\{B, \overline{B}\}$ est complet. Dans ce cas, la formule des probabilités totales devient :

$$P(A) = P(A|B)P(B) + P(A|\overline{B})P(\overline{B})$$

Cela permet de calculer la probabilité de A en fonction de deux scénarios possibles : l'événement B et son complémentaire $\overline{B}$.

Utilisation de ce Théorème 1

52/78

Un Jeu Télévisé

Trois boîtes sont extérieurement identiques. L'une d'elles contient un objet, les deux autres sont vides. Le but du jeu est de trouver la bonne, en respectant la règle suivante :

Le joueur propose une boîte au hasard, et l'animateur lui désigne ensuite une des deux autres boîtes, en précisant qu'elle ne contient pas l'objet. Le joueur peut alors :

- Soit maintenir son choix initial,
- Soit désigner la troisième boîte.

Comparer les trois stratégies suivantes :

1. Le joueur désigne systématiquement la troisième boîte,
2. Le joueur maintient systématiquement son choix initial,
3. Le joueur opte aléatoirement pour l'une ou l'autre des deux attitudes.

Utilisation de ce Théorème : Correction Simple

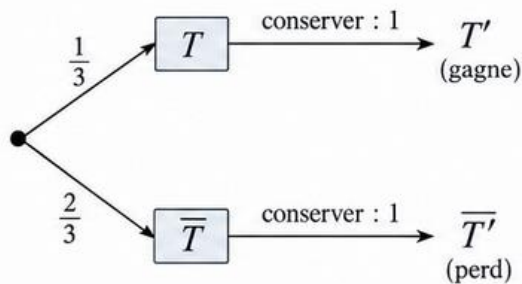
53/78

Jeu télévisé (Monty Hall)

Comparer les trois stratégies par arbres pondérés.

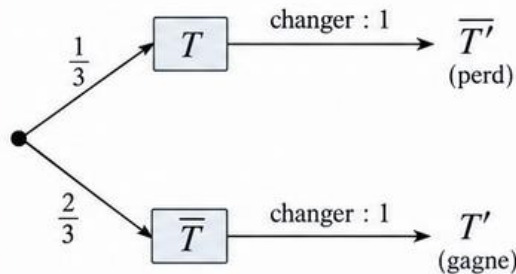
Notations : T = « le premier choix est le bon », $\bar{T}$ = « le premier choix est faux », T' = « le joueur gagne après sa décision finale ».

1. Le joueur conserve son choix



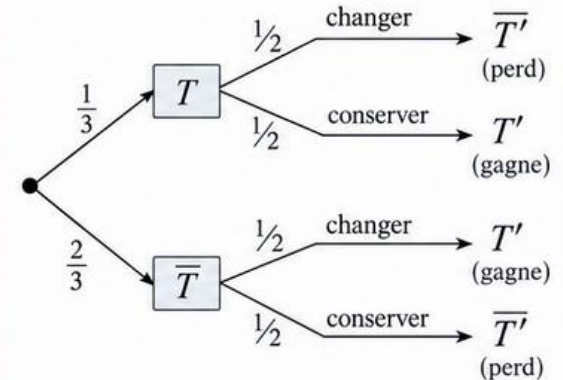
$$P(T') = \frac{1}{3}$$

2. Le joueur change de boîte



$$P(T') = \frac{2}{3}$$

3. Le joueur choisit au hasard



$$P(T') = \frac{1}{3} \times \frac{1}{2} + \frac{2}{3} \times \frac{1}{2} = \frac{1}{2}$$

Conclusion : changer est la meilleure stratégie.

$$P(T' | \text{changer}) = \frac{2}{3} > P(T' | \text{hasard}) = \frac{1}{2} > P(T' | \text{conserver}) = \frac{1}{3}.$$

Utilisation de ce Théorème : Correction plus « math »

54/78

On note T l'événement "le joueur désigne la bonne boîte au premier coup" et T' l'événement "le joueur désigne la bonne boîte au second coup". On cherche à évaluer $P(T')$.

1. Premier cas : Le candidat change son choix après l'indication

Dans ce cas, la probabilité que le joueur choisisse la bonne boîte au second coup est calculée comme suit :

$$P(T') = P(T'|T)P(T) + P(T'|\bar{T})P(\bar{T}) = 0 \times \frac{1}{3} + 1 \times \frac{2}{3} = \frac{2}{3}$$

2. Second cas : Le candidat maintient son choix après l'indication

Si le joueur garde son choix initial, la probabilité d'avoir choisi la bonne boîte au premier coup est simplement :

$$P(T') = P(T'|T)P(T) + P(T'|\bar{T})P(\bar{T}) = 1 \times \frac{1}{3} + 0 \times \frac{2}{3} = \frac{1}{3}$$

3. Troisième cas : Le candidat opte pour l'une des deux attitudes de manière aléatoire

Dans ce cas, la probabilité que le joueur choisisse la bonne boîte au premier coup est la même que celle de choisir aléatoirement entre changer ou maintenir son choix :

$$P(T') = \frac{1}{2}$$

Utilisation de ce Théorème 2

55/78

Des études antérieures ont montré que 25% des infections graves sont dues à des infections nosocomiales, que 15% des infections graves non dues à des infections nosocomiales provoquent un décès tandis que cette proportion s'élève à 30% quand il s'agit d'une infection grave due à une infection nosocomiale.

Dans un service hospitalier, un décès survient suite à une infection grave. Quelle est la probabilité que le décès soit dû à une infection nosocomiale ?

D = Décès après infection grave et IN = Infection Nosocomiale
=> dans la population des patients ayant une infection grave

Traduction de l'énoncé $P(IN) = 0,25$ soit $P(\overline{IN}) = 0,75$

Et $P(D|\overline{IN}) = 0,15$ et $P(D|IN) = 0,3$

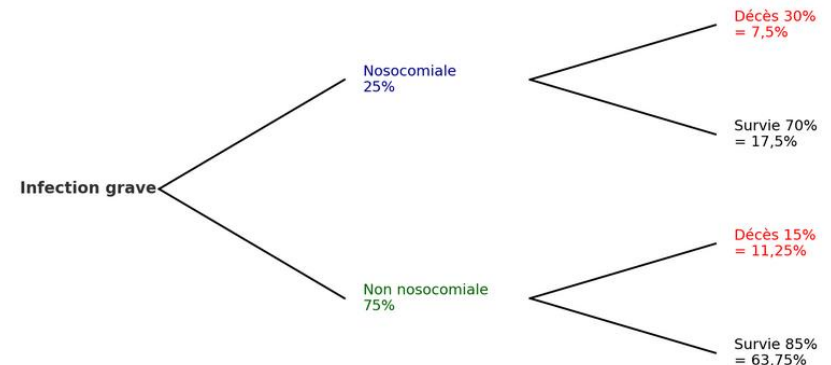
On cherche $P(IN|D)$, l'application du théorème de Bayes

$$P(IN|D) = \frac{P(IN).P(D|IN)}{P(D)} \text{ car } P(IN|D) = P(IN \cap D)/P(D)$$

Le décès pouvant être dû à une infection nosocomiale ou à une infection non nosocomiale, on a :

$$P(D) = P(IN).P(D|IN) + P(\overline{IN}).P(D|\overline{IN})$$

$$\text{On a donc } P(IN|D) = \frac{0,25 \times 0,3}{0,25 \times 0,3 + 0,75 \times 0,15} = 0,4$$



Plan

56/78

👉 Intérêt des Bio-Statistiques

👉 Les Probabilités

👉 Applications

👉 Probabilités Conditionnelles

👉 Tableau de Contingence

👉 Courbes ROC

Le Diagnostic : Approche Bayésienne

57/78

Exemple (Extrait de Biostatistique - Carrat, Mallet, Morice, Faculté Pierre et Marie Curie)

Considérons, pour illustrer notre propos, le problème du diagnostic d'une douleur aiguë de l'abdomen. Il s'agit d'un patient arrivant aux urgences pour un « mal au ventre ».

Si l'on ne sait rien d'autre sur le patient (on n'a pas fait d'examen clinique ou complémentaire), on ne connaît que les probabilités d'avoir tel ou tel diagnostic si on observe une douleur.

Soient D_1 , D_2 et D_3 les 3 diagnostics principaux (il y en a en fait au moins une douzaine) et exclusifs ; par exemple D_1 = appendicite, D_2 = perforation d'ulcère, D_3 = autres diagnostics.

Soit un signe s_1 pour lequel on connaît $Pr(s_1/D_1)$, $Pr(s_1/D_2)$, et $Pr(s_1/D_3)$.

Par exemple, s_1 serait « présence d'une fièvre $\geq 38,5^\circ\text{C}$ » ; $Pr(s_1/D_1) = 0,90$; $Pr(s_1/D_2) = 0,30$; et $Pr(s_1/D_3) = 0,10$.

Ces probabilités peuvent être estimées sur une population de patients en dénombrant le nombre de sujets ayant le diagnostic D_1 et présentant le signe s_1 . De même, on peut connaître $Pr(D_1)$, $Pr(D_2)$ et $Pr(D_3)$.

Le problème diagnostique se pose comme celui de choisir par exemple le diagnostic le plus probable connaissant le signe s_1 ; pour ce faire, on calcule $Pr(D_1/s_1)$, $Pr(D_2/s_1)$, $Pr(D_3/s_1)$ et on retient le diagnostic qui a la plus grande probabilité : c'est l'application de l'approche bayésienne au problème de l'aide au diagnostic.

Les Tableaux souvent Utilisés

58/78

2 tableaux à 4 cases

- En médecine, 2 tableaux à 4 cases sont très utilisés et renvoient au conditionnement.
- Evaluation des signes et examens complémentaires.
- Recherche de facteurs de risque.

	Maladie +	Maladie -	Total
Test +	A (VP)	B (FP)	$T+=A+B=VP+FP$
Test -	C (FN)	D (VN)	$T-=C+D=FN+VN$
Total	$M+ = A+C = VP+FN$	$M- = B+D = FP+VN$	n

$P(T+/M+)$; $P(T-/M-)$; $P(M+/T+)$; $P(M-/T-)$

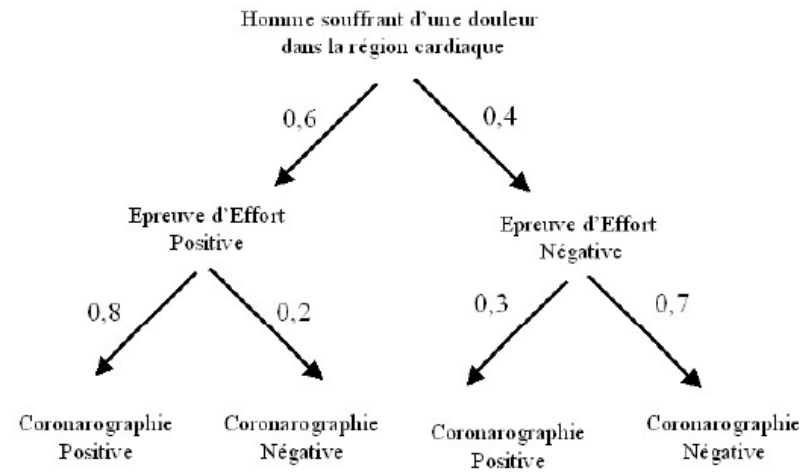
	Maladie +	Maladie -	Total
Exposés	A	B	$E+=A+B$
Non exposés	C	D	$E-=C+D$
	$M+ = A+C$	$M- = B+D$	$A+B+C+D$

$P(M+/E+)$; $P(M+/E-)$; $P(E+/M+)$; $P(E-/M-)$

Une Alternative : Les arbres de Décision

59/78

Un homme se présente aux urgences avec des douleurs thoraciques. Quelle est la probabilité qu'il ait une sténose coronarienne (événement noté Coro+, EE+ événement avoir une épreuve d'effort positive) ?



Prob(Coro+) =



Evaluation des Examens

60/78

Il n'existe pas de signe ou d'examen parfait qui serait toujours présent en cas de présence de la maladie et absent en cas d'absence de la maladie. Nous ne pouvons donc qu'évaluer et fournir des probabilités de présence ou d'absence de maladie.

	Maladie +	Maladie -	Total
Test +	A (VP)	B (FP)	T+=A+B=VP+FP
Test -	C (FN)	D (VN)	T-=C+D=FN+VN
Total	M+ = A+C = VP+FN	M- =B+D=FP+VN	N=A+B+C+D

Fréquence de la maladie = Prévalence = $P(M+) = (VP+FN)/N$

Sensibilité = $P(T+/M+) = VP/(VP+FN)$

Spécificité = $P(T-/M-) = VN/(VN+FP)$

VPP = $P(M+/T+) = VP/(VP+FP)$

VPN = $P(M-/T-) = VN/(VN+FN)$

La Prévalence d'une Maladie

61/78

La prévalence représente la probabilité qu'un individu soit atteint par une maladie avant de subir un test. Elle correspond à la probabilité pré-test, c'est-à-dire avant la réalisation du test de diagnostic.

Facteurs influençant la prévalence :

- **Zone géographique :** Par exemple, le paludisme est beaucoup plus fréquent en Afrique qu'en France.
- **Groupes spécifiques d'individus :** Certains groupes sont plus susceptibles de développer certaines maladies. Par exemple, le sida est plus fréquent chez les toxicomanes.

Sensibilité et Spécificité

62/78

Sensibilité :

La sensibilité d'un signe pour une maladie est la probabilité que le signe soit présent si le sujet est atteint de la maladie considérée. C'est une probabilité conditionnelle :

$$\text{Sensibilité} = Se = P(S|M)$$

Un test diagnostique est d'autant plus sensible que les sujets atteints de la maladie présentent plus souvent le signe S .

Spécificité :

La spécificité d'un signe pour une maladie est la probabilité que le signe soit absent si le sujet n'est pas atteint de la maladie. De manière similaire :

$$\text{Spécificité} = Sp = P(\bar{S}|\bar{M})$$

Un test diagnostique est d'autant plus spécifique que les sujets indemnes de la maladie présentent moins souvent le signe S .

Les Valeurs Prédicatives

63/78

La valeur prédictive positive d'un signe pour une maladie est la probabilité qu'un individu soit atteint de la maladie lorsque le signe est présent.

La valeur prédictive négative d'un signe pour une maladie est la probabilité qu'un individu soit exempt de la maladie lorsque le signe est absent.

Ces paramètres peuvent être notés comme suit :

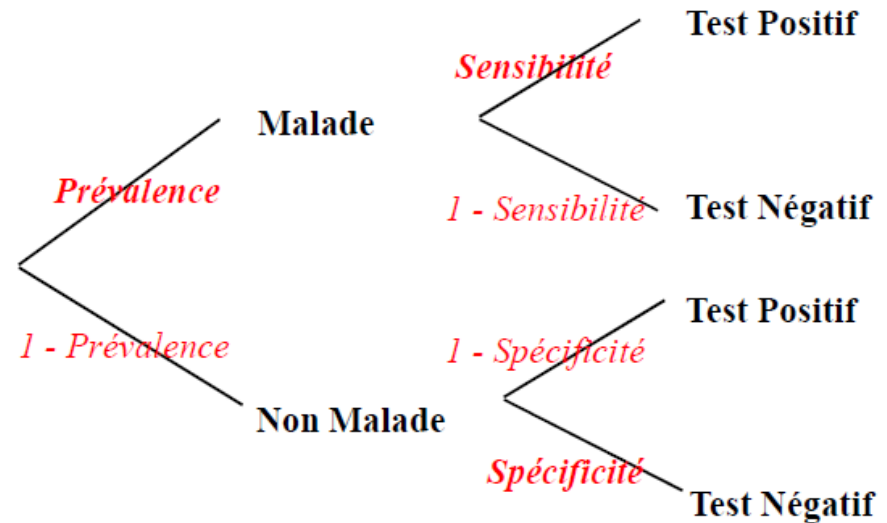
- " $VPP = P(M \mid S)$ "
- " $VPN = P(\overline{M} \mid \overline{S})$ "

Comme pour la sensibilité et la spécificité, les valeurs prédictives positive et négative varient de manière inverse et doivent être prises en compte simultanément.

VPP & VPN : Mode Opérateur

64/78

- VPP et VPN correspondent à l'inversion du conditionnement de la sensibilité et de la spécificité.
- L'arbre des probabilités permet facilement cette opération.



$$VPP = \frac{\text{prévalence} * \text{sensibilité}}{(\text{prévalence} * \text{sensibilité}) + (1 - \text{prévalence}) * (1 - \text{spécificité})}$$

$$VPN = \frac{(1 - \text{prévalence}) * (\text{spécificité})}{(1 - \text{prévalence}) * (\text{spécificité}) + (\text{prévalence}) * (1 - \text{sensibilité})}$$

VPP & VPN : Remarque

65/78

Les valeurs prédictives d'un test dépendent de :

- La sensibilité du test,
- La spécificité du test,
- La prévalence de la maladie dans la population testée.

En conséquence, un même test (ayant la même sensibilité et la même spécificité) peut présenter des valeurs prédictives positives (VPP) et négatives (VPN) très différentes **en fonction de la prévalence de la maladie.**

VPP & VPN : Exemple du Sida

66/78

Contexte :

- Prévalence du SIDA dans une population : 0,1%
- Sensibilité du test : 99% (probabilité d'un test positif si la personne est malade)
- Spécificité du test : 99% (probabilité d'un test négatif si la personne n'est pas malade)

Question :

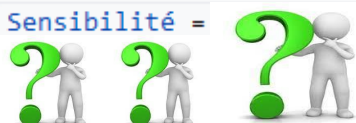
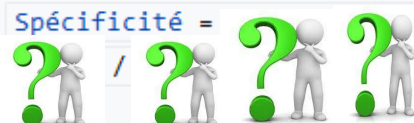
Quelle est la probabilité d'être HIV+ si le test est positif ?



*Entraînez-vous en
calculant la VPN !!!
(0,9999)*

VPP & VPN : Exemple Du Cancer du Colon

67/78

		Patients avec un cancer du colon (tel que confirmé par la coloscopie)		Prévalence = 
		Malades = 30	Non malades = 2000	
Recherche de sang occulte dans les selles (examen diagnostique)	Résultat positif	VP = 20	FP = 180	VPP 
	Résultat négatif	FN = 10	VN = 1820	VPN 
		Sensibilité = 	Spécificité = 	

La sensibilité et la spécificité sont des caractéristiques intrinsèques au test évalué et ne changent pas. Par contre, si l'on augmente la prévalence du cancer du colon dans la population, la VPP augmentera et la VPN diminuera car le nombre de faux positifs et de vrais négatifs vont diminuer.

Valeurs Diagnostiques d'un Test

68/78

Analyse du Test Diagnostique :

Hypothèses en fonction des caractéristiques du test :

- **Test avec une spécificité élevée :** Un **résultat positif** confirme l'hypothèse diagnostique. Cela signifie que le test est très précis pour identifier les personnes qui ne sont pas atteintes de la maladie lorsqu'elles sont réellement en bonne santé.
- **Test avec une sensibilité élevée :** Un **résultat négatif** élimine le diagnostic. Autrement dit, si le test est négatif, cela élimine fortement la possibilité que le patient soit atteint de la maladie, à condition que le test soit sensible.

Gain diagnostique positif :

- **Définition :** Le gain diagnostique positif représente la différence entre la probabilité **pré-test** (la prévalence de la maladie dans la population) et la probabilité **post-test** (la valeur prédictive positive, ou VPP, du test).
- **Formule :**

$$\text{Gain positif} = \text{VPP} - \text{Prévalence}$$

Cela montre l'amélioration de la précision du test pour prédire la maladie après le test, par rapport à la probabilité initiale de la maladie (avant de connaître le résultat du test).

Règles Pratiques

69/78

Privilégier la sensibilité ou la spécificité d'un test :

On privilégie la sensibilité d'un test quand :

- Le but de la démarche diagnostique est d'exclure la maladie.
- La maladie est d'autant plus curable que le traitement est plus précoce.
- Un résultat faussement positif n'a pas de conséquence sérieuse.
- La maladie est rare et on se trouve dans une situation de dépistage.

(Dans ce cas, l'important est d'identifier tous les cas possibles, même à un faible taux de faux positifs, afin de ne pas manquer de cas malades.)

On privilégie la spécificité d'un test quand :

- Le but de la démarche diagnostique est d'affirmer ou de confirmer le diagnostic.
- Le risque d'un résultat faussement positif est grave et n'est pas acceptable.
(Cela est particulièrement important dans les tests où les conséquences d'un diagnostic erroné peuvent être graves, par exemple dans les tests de cancer.)
- La maladie est incurable.

(Dans ce cas, le principal objectif est de réduire les faux positifs pour éviter de diagnostiquer des personnes non malades.)

Exemples

70/78

Dans une population soumise à un test de dépistage du cancer du colon, on estime que la prévalence de la maladie est de 10/1000, pour un sujet cancéreux, le test est positif dans 90% des cas. Pour un sujet non cancéreux, le test est positif dans 35% des cas.

Déterminer la sensibilité et la spécificité du test.

La réponse est directement dans l'énoncé ...
Se = 0,9 et Sp = 0,65



Dans une population, on estime que la prévalence d'une maladie M est de 10/1000. On dispose d'un test diagnostique de sensibilité égale à 98% et de spécificité égale à 90%.

-sur 1000 malades combien ont un test positif ?

-sur 1000 personnes de la population, combien y a-t-il de faux positifs ?



-Sur 1000 malades 98% (980) présentent un test positif (T^+)

-Sur 1000 personnes 10 sont malades (M^+) et 990 sont sains (M^-). On sait que 90% des sains ont un test négatif (=891). Cela signifie qu'il y a $990 - 891 = 99$ faux positifs.

Autre méthode $\Rightarrow$ Sp = 90% donc $P(T^-|M^-) = 0,9$ et $P(T^+|M^-) = 0,1$ donc le nombre de faux positifs correspond au produit de $P(T^+|M^-)$ et du nombre de M^- , soit FP = $0,1 \times 990 = 99$

Exemples

71/78

On dispose d'un examen applicable au diagnostic d'une maladie M qui atteint 20% de la population. Après l'examen, le résultat s'exprime de façon binaire : + ou -. L'évaluation des performances diagnostiques du test dans un échantillon représentatif a conduit aux résultats suivants : Se = 90% et Sp = 85%.

Quelle est la probabilité qu'un individu soit effectivement atteint de la maladie M quand le test est positif ?



Il s'agit de calculer la VPP

$$VPP = \frac{p \cdot Se}{p \cdot Se + (1-p) \cdot (1-Sp)} = \frac{0,2 \times 0,9}{0,2 \times 0,9 + (1-0,2) \times (1-0,85)} = \frac{0,18}{0,3} = 0,6 \Rightarrow 60\%$$

Une maladie affecte 20% des individus. Le test de dépistage utilisé pour un individu malade est positif dans 90% des cas.
Le test de dépistage utilisé pour un individu sain est positif dans 15% des cas.

Quelle est la probabilité qu'un individu soit malade sachant qu'il a un test positif ?

-C'est une autre formulation de la question précédente ... (p = 0,2) (Se = 0,9) et (Sp = 1-0,15 = 0,85) => VPP = 60%



Exemples

72/78

Selon l'OMS, un cholestérol $> 2\text{g/L}$ ou une Apolipoprotéine B $> 1,3\text{g/L}$ sont associés au risque de maladies cardiovasculaires. Voici les trois tableaux de contingence rassemblant ces informations :

A	R+	R-	ToT
Chol >2	VP=28	FP=10	38
Chol <2	FN=2	VN=26	28
ToT	30	36	66

B	R+	R-	ToT
Apo B $>1,3$	VP=22	FP=9	31
Apo B $<1,3$	FN=8	VN=27	35
ToT	30	36	66

C	R+	R-	ToT
Chol >2 et/ou Apo B $>1,3$	VP=29	FP=14	43
Chol <2 et Apo B $<1,3$	FN=1	VN=22	23
ToT	30	36	66

-Calculer Se, Sp, VPP et VPN pour les cas A, B et C
-Conclure sur le type de test à utiliser



-Le test A n'est pas très performant (moyen),
-Le test B admet des VP et une Se très faible (+ faible)
-Le test C propose une meilleure Se (on trouve plus de vrais positifs), la spécificité est plus faible (plus de faux positifs) et VPN est très élevée (les 2 résultats négatifs associent le sujet au groupe R-). L'association des 2 variables est + performante pour discriminer entre les 2 groupes

-A $\Rightarrow$ Se = $28/30 = 0,93$ Sp = $26/36 = 0,72$ VPP = $28/38 = 0,74$ VPN = $26/28 = 0,93$
-B $\Rightarrow$ Se = $0,73$ Sp = $0,75$ VPP = $0,71$ VPN = $0,77$
-C $\Rightarrow$ Se = $0,97$ Sp = $0,61$ VPP = $0,67$ VPN = $0,96$

Plan

73/78

👉 Intérêt des Bio-Statistiques

👉 Les Probabilités

👉 Applications

👉 Probabilités Conditionnelles

👉 Tableau de Contingence

👉 Courbes ROC

Choix d'un Seuil

74/78

Courbe ROC (Receiver Operating Characteristic)

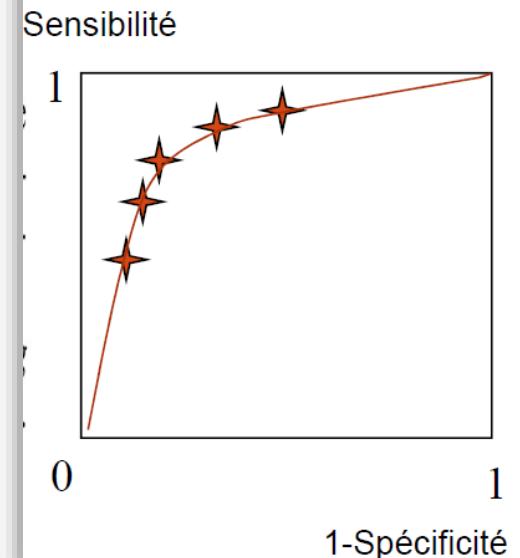
Lorsqu'un examen donne des résultats continus (par exemple, des mesures quantitatives), il est nécessaire de déterminer le seuil optimal qui distingue les valeurs pathologiques (positives) des valeurs normales (négatives). L'idéal serait d'obtenir une sensibilité et une spécificité égales à 1, ce qui signifierait que le test est parfait. Cependant, ce n'est généralement pas possible, et il faut chercher à maximiser ces deux paramètres.

La courbe ROC permet de visualiser ce compromis entre la sensibilité (probabilité d'identifier correctement un malade) et la spécificité (probabilité de ne pas identifier faussement un sain). La courbe trace la sensibilité (Se) en fonction de $1 - \text{Spécificité}$ (le taux de faux positifs).

Interprétation de la courbe ROC :

- L'axe horizontal représente $1 - \text{Spécificité}$, c'est-à-dire la proportion de faux positifs.
- L'axe vertical représente la **sensibilité**, c'est-à-dire la proportion de vrais positifs.

Les points sur la courbe représentent différents seuils possibles du test, où chaque seuil donne une combinaison différente de sensibilité et de spécificité. Le but est d'identifier le seuil qui maximise à la fois la sensibilité et la spécificité, idéalement en haut à gauche du graphique (sensibilité élevée et spécificité élevée).



L'Aire Sous Courbe (AUC)

75/78

Aire sous la courbe : AROC (Area Under the Receiver Operating Characteristic Curve)

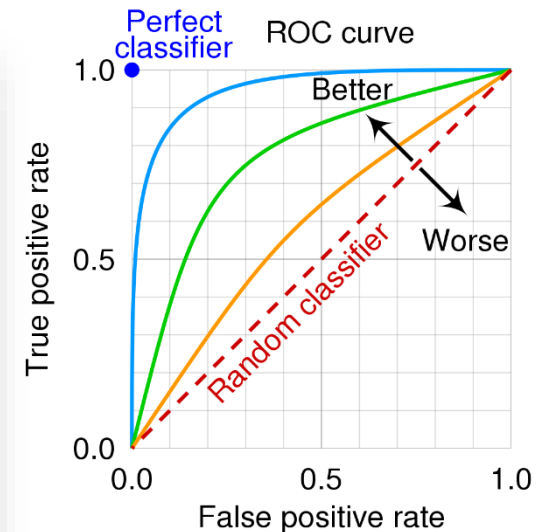
- **Aire sous la courbe (AUC) :** L'AUC est une mesure de la performance d'un test diagnostique. Elle est comprise entre 0,5 et 1.
 - 0,5 indique un test sans pouvoir discriminant (examen aléatoire, comme un tirage au sort, similaire à pile ou face).
 - 1 indique un test parfait qui distingue parfaitement les malades des non-malades.
- **Instrument d'évaluation et de comparaison des performances diagnostiques :** L'AUC est un instrument privilégié pour évaluer et comparer les performances des tests diagnostiques ou des examens complémentaires. Elle permet de mesurer la capacité d'un test à distinguer les deux états (par exemple, malade vs non malade).
- **Choix du meilleur test :** Lorsque plusieurs tests sont disponibles, celui ayant l'AUC maximale est généralement considéré comme le meilleur test en termes de discrimination des résultats.

Exemple : Comparaison de 2 Tests

76/78

L'aire sous la courbe ROC est définie comme la probabilité qu'un marqueur distingue correctement entre un sujet malade et un non-malade, sélectionnés au hasard. Cette probabilité reflète la capacité du marqueur à attribuer une valeur plus élevée aux malades qu'aux non-malades.

- **Interprétation de l'AUC :**
 - **AUC = 0,5 (50%) :** Cela signifie que le marqueur n'a aucune capacité discriminatoire. Il ne distingue pas mieux les malades des non-malades qu'un tirage au sort.
 - **AUC = 1,0 (100%) :** Cela représente un test parfait, capable de distinguer sans erreur les malades des non-malades.
- **Augmentation de l'AUC :** Une augmentation de l'AUC indique une amélioration de la capacité du test à distinguer les différentes classes (malades et non-malades).



Quel est le meilleur test ?

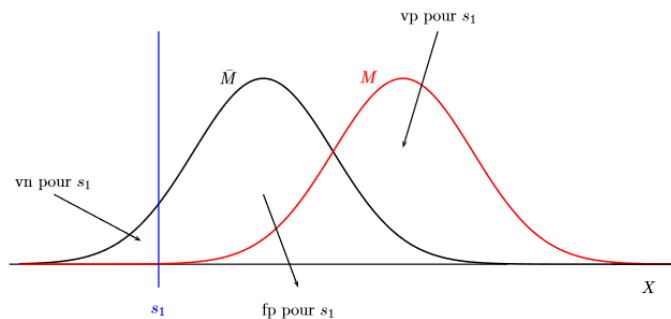
Le Test Bleu > Vert > Jaune

Exemple : Choix du Seuil

77/78

On dispose d'une variable quantitative X (ex : dosage biologique). On souhaite :

- Déterminer le seuil optimal (pour séparer les M des $\bar{M}$)
- Quantifier le pouvoir diagnostic de X

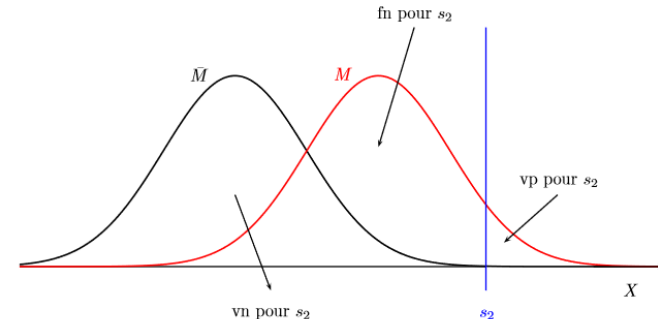


Seuil s_1 :

- Si $X < s_1$ alors $\bar{M}$ (pas de fn)
- Si $X \geq s_1$ alors M et $\bar{M}$ (bcp de fp)

⇒ $Se = 1$ mais Sp mauvaise

	M	$\bar{M}$
$T^+(X \geq s_1)$	vp	fp
$T^-(X < s_1)$	0	vn



Seuil s_2 :

- Si $X < s_2$ alors M et $\bar{M}$ (bcp de fn)
- Mais si $X \geq s_2$ alors M (pas de fp)

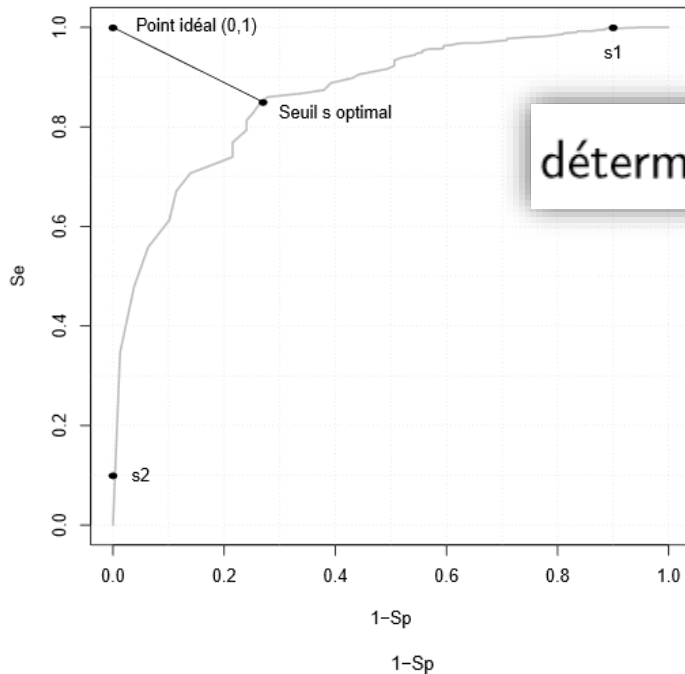
⇒ $Sp = 1$ mais Se mauvaise

	M	$\bar{M}$
$T^+(X \geq s_2)$	vp	0
$T^-(X < s_2)$	fn	vn

→ Nécessité de trouver un compromis !! ←

Exemple : Choix du Seuil

78/78



déterminer le seuil s qui sépare au mieux les M des $\bar{M}$

AUC	Discrimination
0.5	Nulle
0.7 - 0.8	Acceptable
0.8 - 0.9	Excellente
> 0.9	Exceptionnelle

Remarques :

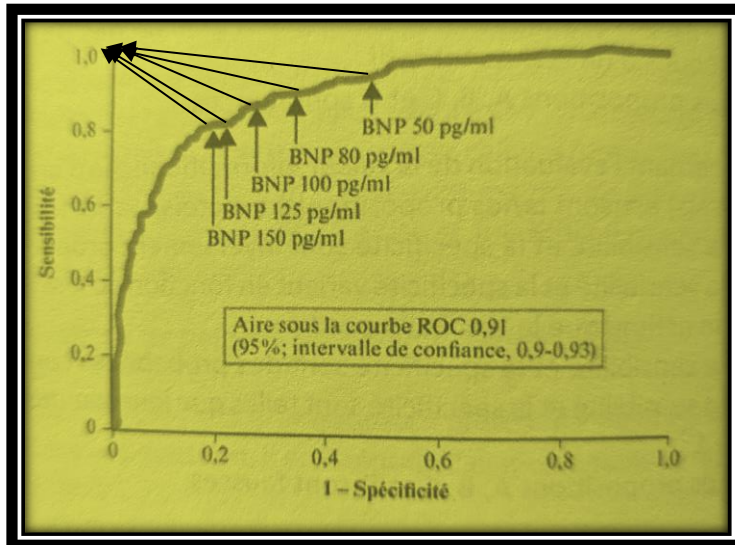
- Si $AUC = 0.5$ alors on classe de manière complètement aléatoire les observations
- Si $AUC > 0.9$ le classement est très bon, voire trop bon, il faut évaluer s'il y a *overfitting*

déterminer le seuil s qui minimise la distance euclidienne du point $(0, 1)$

$$d((0, 1), s) = \sqrt{(0 - x_s)^2 + (1 - y_s)^2}$$

Exercice

79/78



Le BNP est un peptide sécrété par les myocytes du ventricule gauche pour induire une relaxation vasculaire systémique et artérielle pulmonaire. Une étude a permis de déterminer les différentes valeurs-seuils du BNP dans le diagnostic de l'insuffisance cardiaque aigue gauche en situation d'urgence,

-Est-ce que le test a une utilité clinique ?

-Pour un seuil à 150 pg/ml, que vaut la spécificité ?

-Quel est le seuil offrant le meilleur compromis Se/Sp



-Le test a évidemment une utilité clinique car l'aire sous courbe est statistiquement différente de 0,5

-Une lecture du graph montre $1 - Sp = 20\% \Rightarrow Sp = 80\%$

-Il suffit de comparer les distances $\Rightarrow 150pg/ml$ semble être le seuil optimal

MERCI POUR VOTRE ATTENTION

