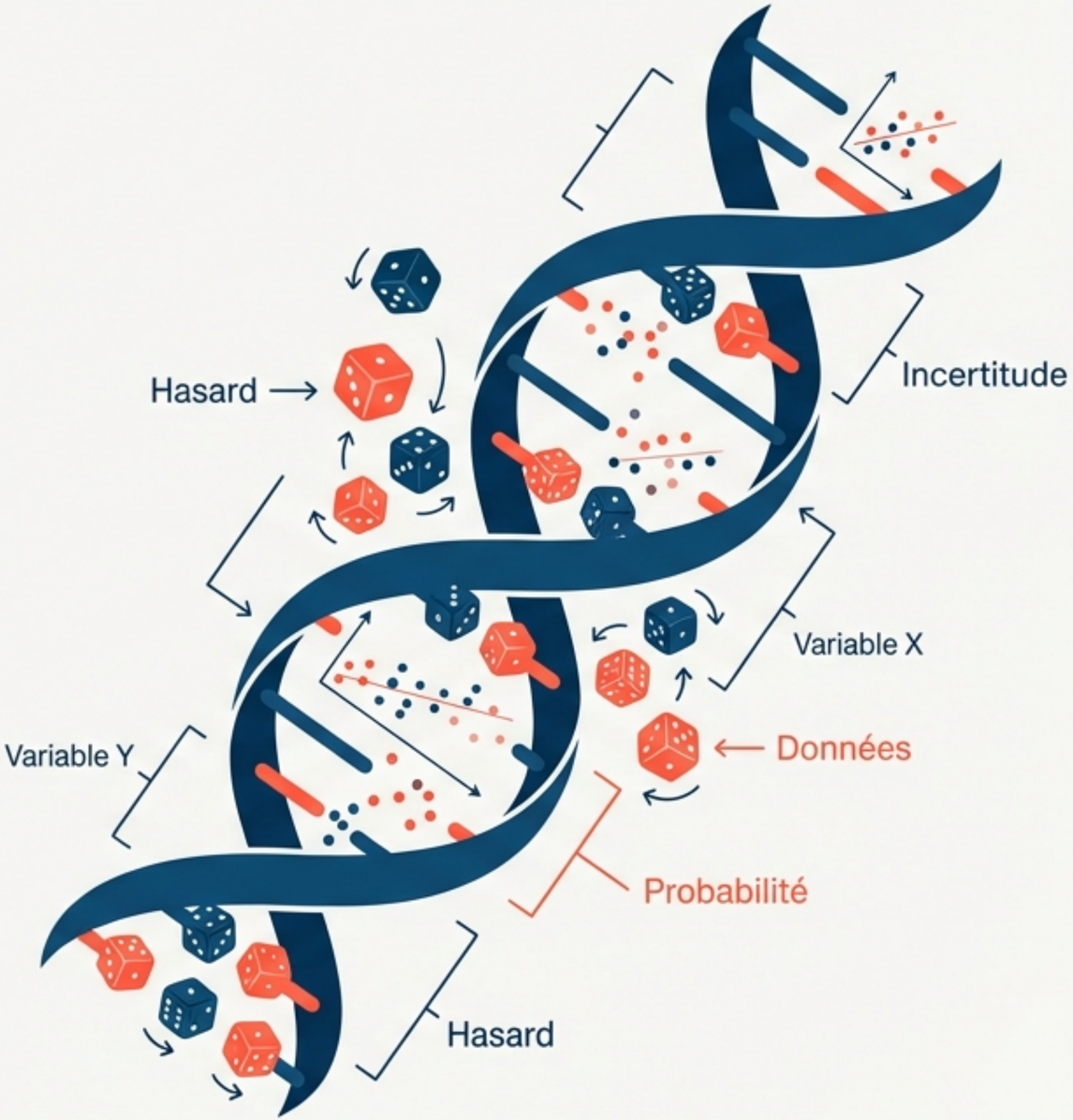
CHAPITRE 6.1 : INTRODUCTION

Probabilités & Décisions : De l’Intuition à la Science des Données

Comprendre l’incertitude en biologie

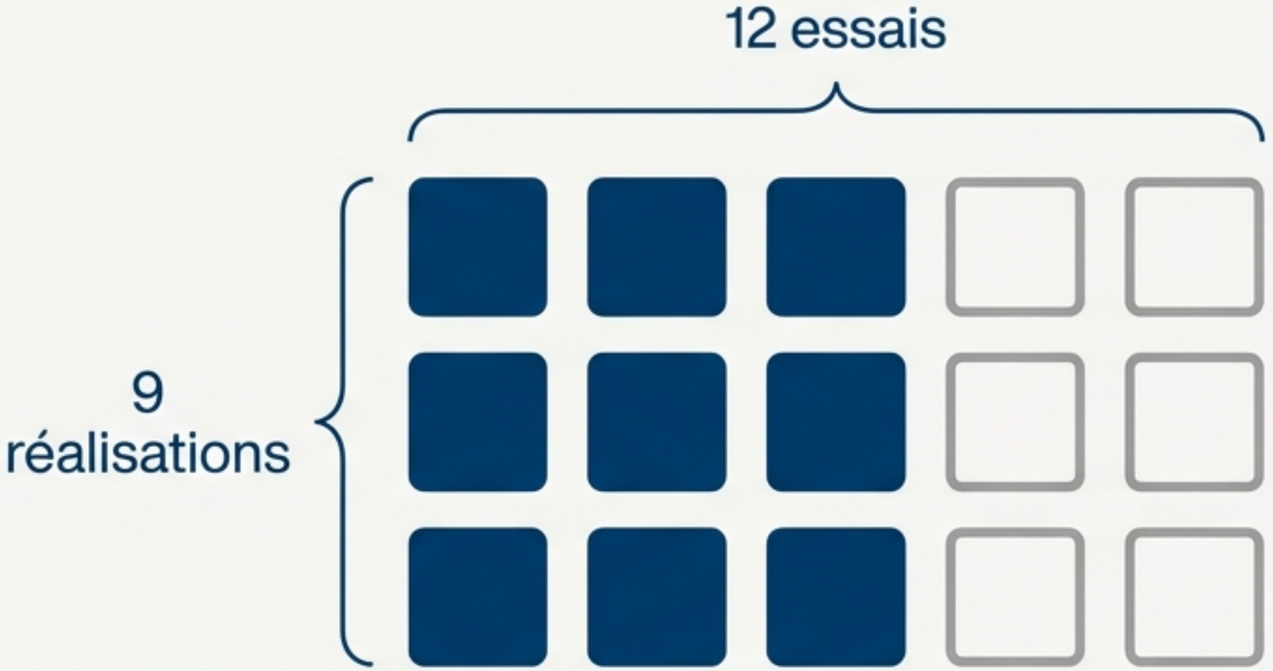
La science des données biologiques ne se limite pas à collecter des chiffres. Elle quantifie l’incertitude. Ce document explore comment nos intuitions nous trompent face au hasard et comment les probabilités nous permettent de prendre des décisions rigoureuses.



NotebookLM

La Mesure de l'Incertitude

Une probabilité quantifie la chance qu'un **évènement** se produise, basée sur l'observation des réalisations passées.



Une probabilité est toujours comprise entre 0 (jamais) et 1 (toujours).

$$0 \leq P(E) \leq 1$$

$$P(E) = \frac{9}{12} = 0,75$$

Basé sur l'exemple du 'petit chat et ses amis' du cours.

NotebookLM

Le Paradoxe du Dépistage

The Context

Imaginez une maladie qui touche **8%** de la population. Un test de dépistage existe avec les performances suivantes :

Sensibilité

Si malade, le test est positif à **95%**.

Taux d'erreur

Si sain, le test se trompe (faux positif) à **10%**.

The Scenario



Le Défi :
Vous êtes testé positif.

The Question

Quelle est la probabilité réelle que vous soyez malade ?

A) 95%

B) 85%

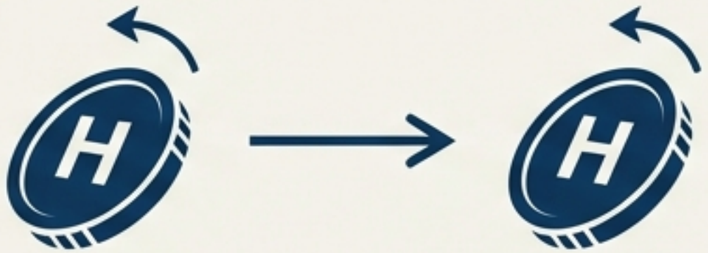
C) Beaucoup moins ?

NotebookLM

La Mécanique des Évènements Successifs

Avant de résoudre le paradoxe, comprenons les règles.

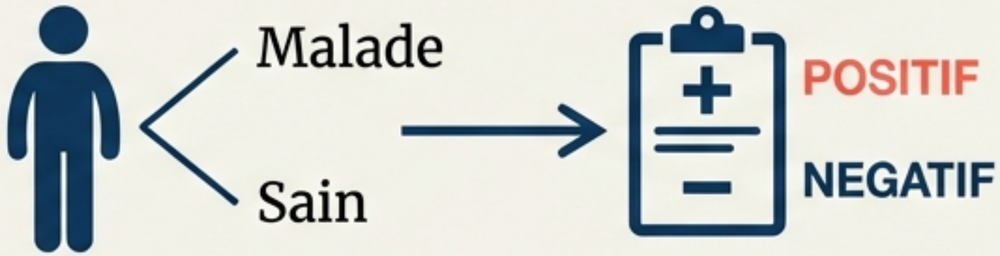
Indépendance



Le second évènement ne dépend pas du premier.

$$P(E_1 \text{ et } E_2) = P(E_1) \times P(E_2)$$

Conditionnalité



Le résultat du test (E_2) dépend de l'état de santé (E_1).

$$P(E_2|E_1)$$

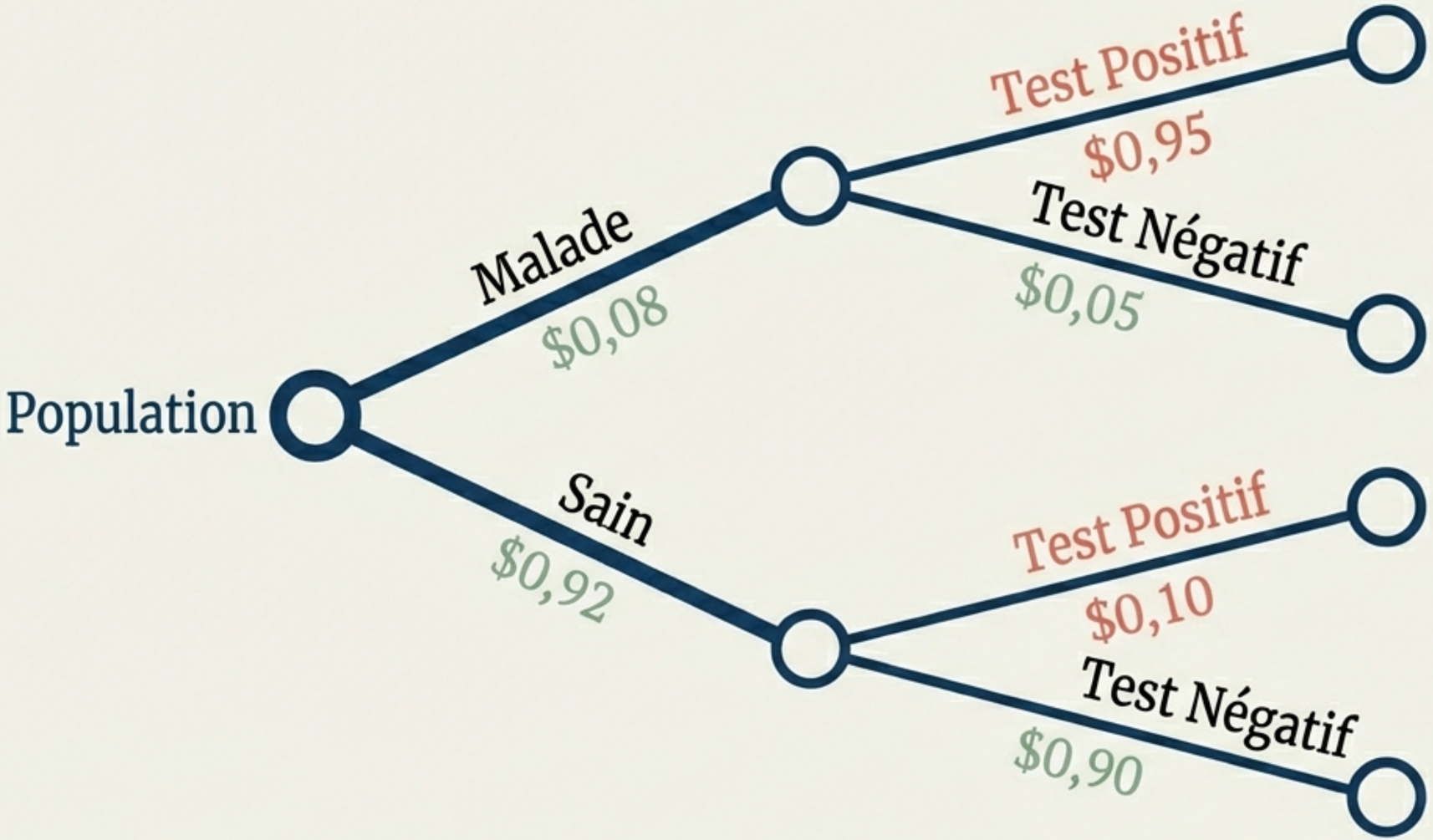
$$P(E_1 \text{ et } E_2) = P(E_1) \times P(E_2|E_1)$$

Basé sur l'exemple du 'petit chat et ses amis' du cours.

NotebookLM

Cartographier le Problème : L'Arbre de Probabilités

Nous visualisons tous les scénarios possibles sous forme d'arbre.

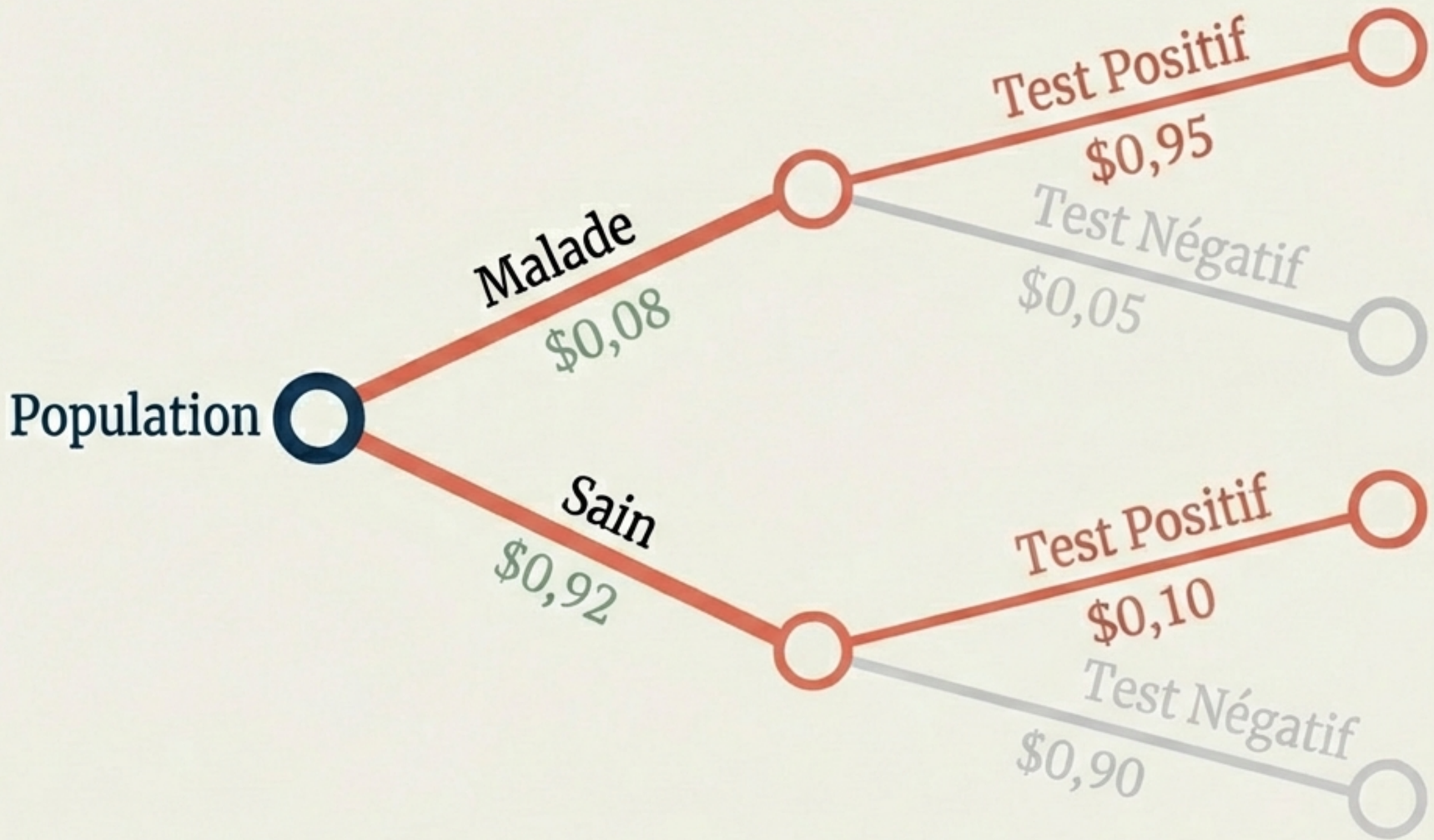


La somme des branches issues d'un même nœud vaut toujours 1.

NotebookLM

Calculer les Réalités Conjointes

Nous isolons uniquement les chemins qui mènent à un résultat **POSITIF**.



Vrais Positifs
(Malade et Detecté)
$0,08 × 0,95 = **0,076** (7,6%)

Faux Positifs (Sain et Erreur)
$0,92 x 0,10 = **0,092** (9,2%)

Total des Tests Positifs

Puisque les évènements sont disjoints :
P(Positif) = 0,076 + 0,092 = **0,168**
(Soit 16,8% de la population totale)

NotebookLM

La Révélation Contre-Intuitive

Si vous êtes positif, quelle est la chance d’être malade ?

The Equation

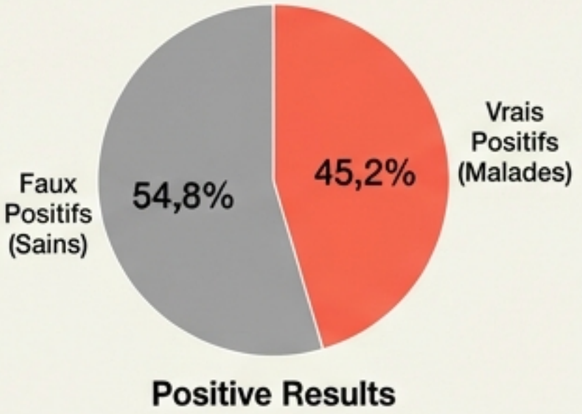
$$\text{Probabilité} = \frac{\text{Vrais Positifs}}{\text{Total des Positifs}}$$

$$= \frac{0,076}{0,168}$$

The Result

45,2%

Bien que le test soit techniquement fiable à 95%, un résultat positif indique que vous êtes malade moins d’une fois sur deux. La majorité des tests positifs (54,8%) sont des fausses alertes.



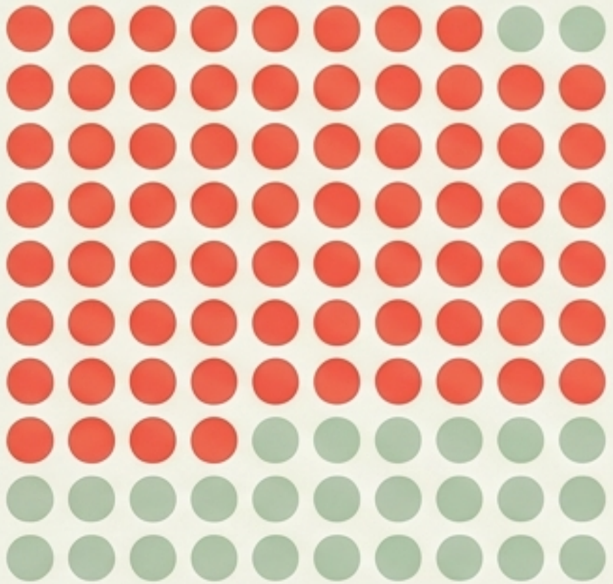
NotebookLM

L’Importance Cruciale de la Prévalence

Pourquoi l’intuition échoue ? Le contexte change tout.

Maladie Commune

Prévalence : 80%

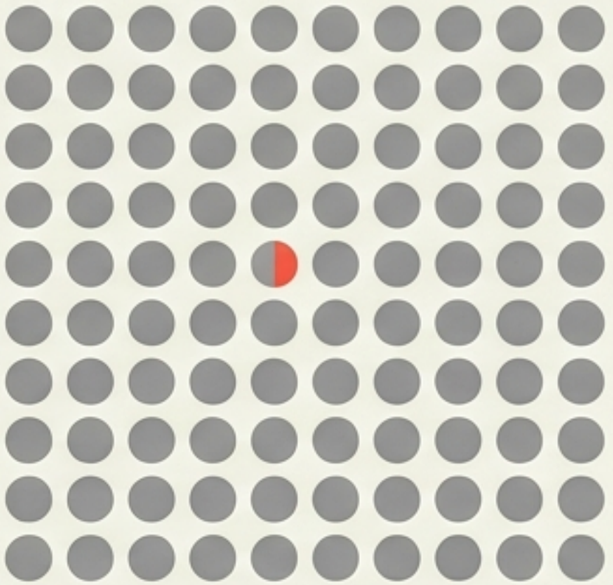


Fiabilité du positif : **97,4%**

Les vrais positifs dominent.

Maladie Très Rare

Prévalence : 0,8%



Fiabilité du positif : **7,1%**

Les faux positifs submergent les vrais cas.

Sans connaître la prévalence (probabilité *a priori*), le résultat d’un test ne veut rien dire.

NotebookLM

Formaliser la Logique : Le Théorème de Bayes



Nous avons résolu le problème visuellement avec un arbre. Mais comment le généraliser ?

Le Problème de l'Inversion :

- Nous connaissons la fiabilité technique : ***P(Positif|Malade)***
- Nous cherchons la réalité du patient : ***P(Malade|Positif)***

La Solution :

Le révérend Thomas Bayes a établi la formule permettant de lier ces deux probabilités conditionnelles inverses. C'est la clé pour mettre à jour nos connaissances ***a priori*** avec de nouvelles données.

NotebookLM

La Formule de Bayes

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

Prévalence (*a priori*)

Sensibilité du test

$$P(\text{Malade}|\text{Positif}) = \frac{P(\text{Positif}|\text{Malade}) \times P(\text{Malade})}{P(\text{Positif})}$$

Total des positifs (Normalisateur)

Résultat : 45,2% (Identique à l'arbre)

NotebookLM

Probabilités sur Données Brutes

En science des données, nous partons souvent d'un **tableau de contingence** plutôt que de pourcentages.

Dataframe : Tabac

	Revenu Faible	Revenu Moyen	Revenu Élevé	Total
Fume	634	332	247	**1213**
Ne fume pas	1846	1622	1868	**5336**
Total	2480	1954	2115	**6549**

Question : Quelle est la probabilité d'être fumeur ?

Calcul : $\frac{\text{Total Fumeurs}}{\text{Population Totale}}$

$$P(\text{Fumeur}) = \frac{1213}{6549} \approx \mathbf{18{,}5\%}$$

NotebookLM

Conditions et Unions

Interroger le tableau avec des conditions complexes.

Probabilité Conditionnelle (SI)

$P(\text{Fumeur} \mid \text{Revenu Élevé})$

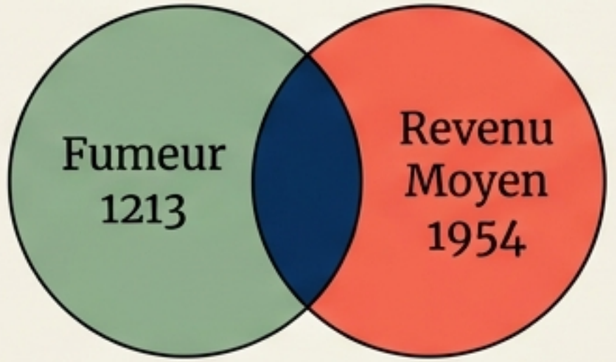
	Revenu Faible	Revenu Élevé	Revenu Moyen	Total
Fumeur	30	117	122	332
Fume	67	247	136	247
Total	142	2115	998	2115

On restreint l'univers aux 2115 personnes riches.

$$\frac{247}{2115} \approx 11{,}7\%$$

Probabilité d'Union (OU)

$P(\text{Fumeur } \mathbf{OU} \text{ Revenu Moyen})$



$$P(A \text{ ou } B) = P(A) + P(B) - P(A \text{ et } B)$$

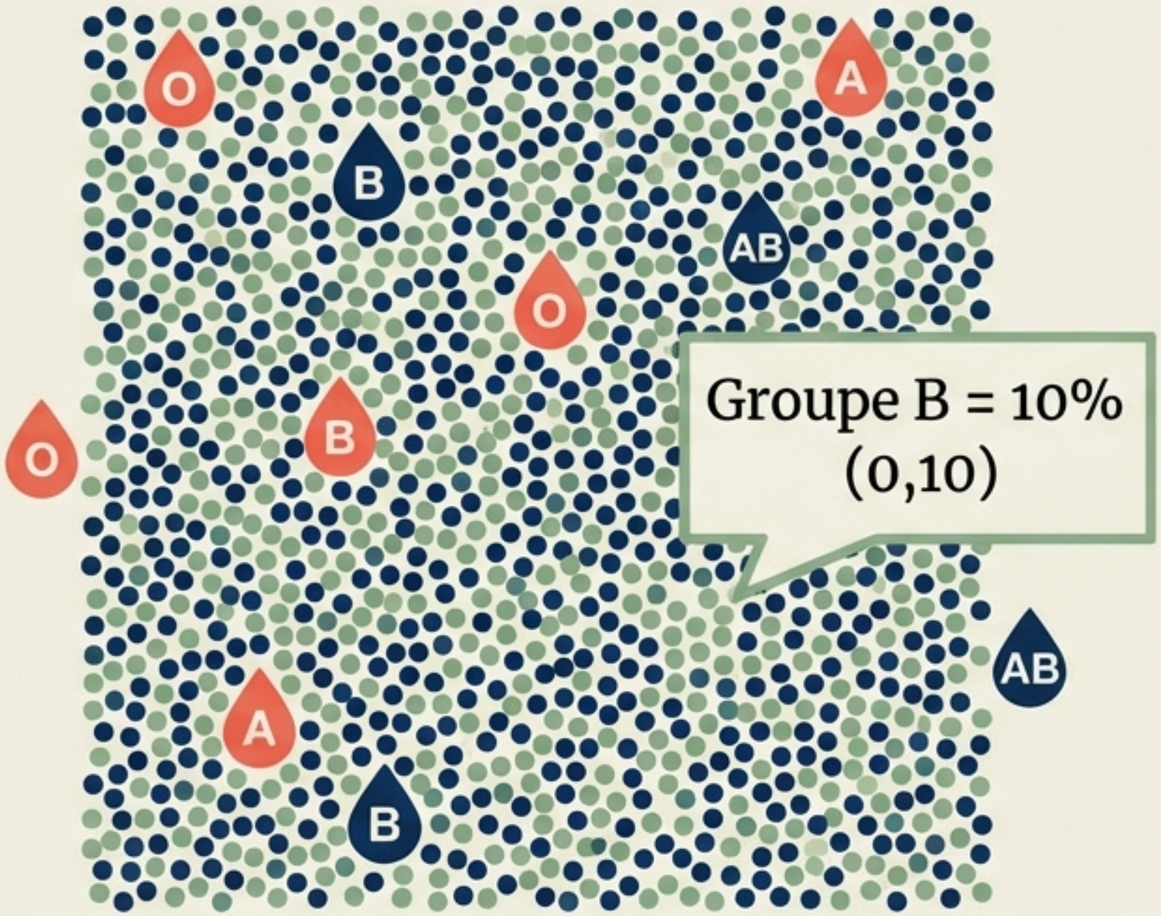
Attention au double comptage !

$$\frac{1213}{6549} + \frac{1954}{6549} - \frac{332}{6549} \approx 43{,}3\%$$

NotebookLM

Échantillonnage : Population Infinie

Tirage avec remise / Population immense



On tire 3 personnes au hasard.

Indépendance : Dans une grande population, tirer une personne ne change pas les chances pour la suivante.

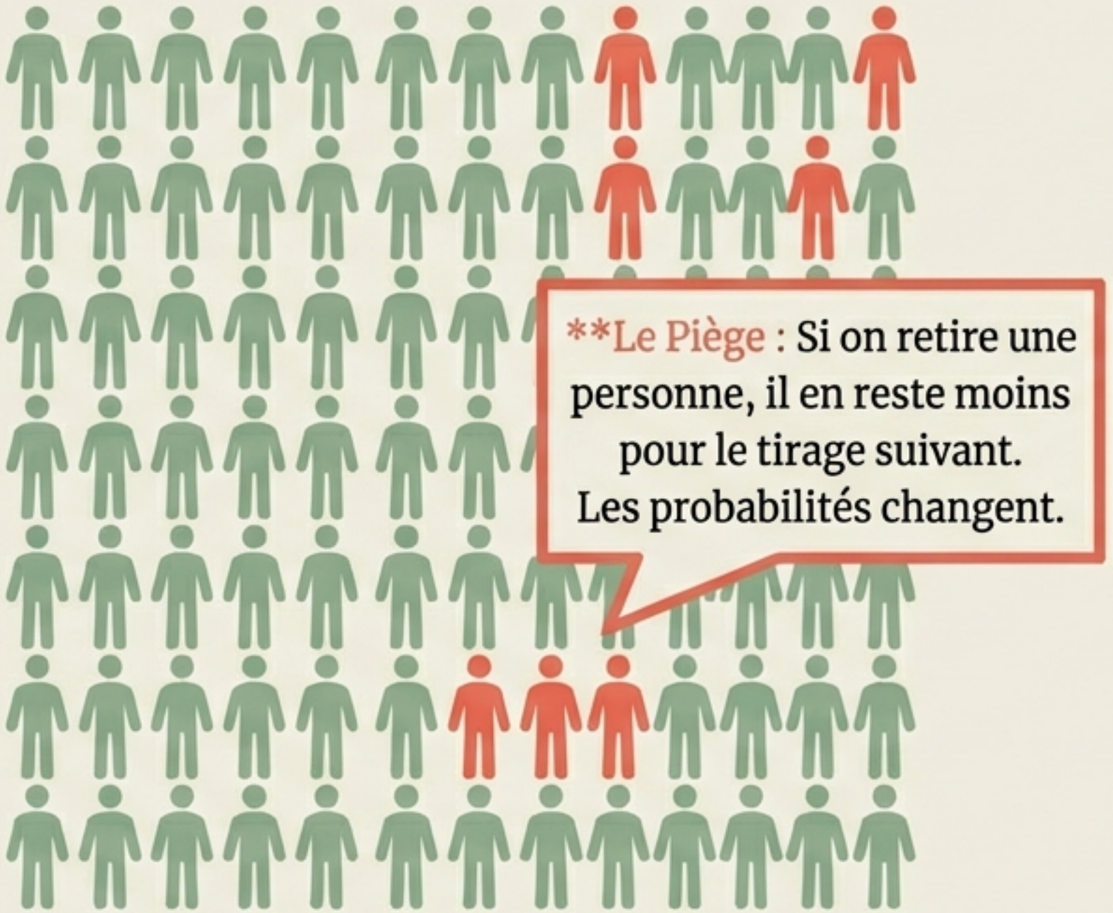
$$P(B \text{ et } B \text{ et } B) = P(B) \times P(B) \times P(B)$$

$$0{,}10 \times 0{,}10 \times 0{,}10 = \mathbf{0{,}001} \ (0{,}1\%)$$

NotebookLM

Échantillonnage : Population Finie

Tirage sans remise / Petite population



Scenario: On tire 3 personnes du groupe B (sans les remettre).

1. Tirage 1 : $\frac{10}{100}$
2. Tirage 2 : $\frac{9}{99}$ (Il reste 9 B sur 99)
3. Tirage 3 : $\frac{8}{98}$ (Il reste 8 B sur 98)

$0{,}10 \times 0{,}091 \times 0{,}081 \approx \mathbf{0{,}00074}$

Note : Nettement plus faible que le cas infini (0,001).

NotebookLM

Ce qu’il Faut Retenir

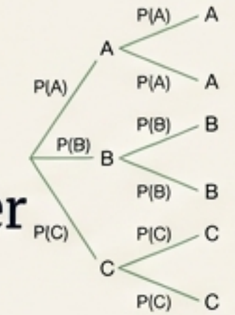
Intuition vs Réalité



La prévalence dicte la fiabilité. Un test positif sur une maladie rare est souvent une fausse alerte.

L’Outil Visuel

L'arbre de probabilités est la meilleure méthode pour démêler les conditions complexes.



L’Outil Formel

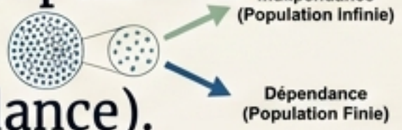
$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

Le théorème de Bayes ($P(A|B)$) permet d'inverser les probabilités et de mettre à jour nos croyances.

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

Le Contexte des Données

Attention à la taille de la population lors de l'échantillonnage (Indépendance vs Dépendance).



À vous de jouer : Pratiquez ces concepts avec les exercices du tutoriel *A06La_proba*.

NotebookLM