
MTH101 · Séance 2

Le nom de l'agence · Le théorème de Bayes

Adrien Chiodo

Vendredi 9 octobre 2026, matin



OUVERTURE

La journée

- Pour commencer : l'essentiel de la séance 1, et deux de ses exercices.
- Ce matin : le nom de l'agence, le retour sur la commande 1, puis le théorème de Bayes.
- Cet après-midi : les lois de Bernoulli, binomiale et de Poisson ; espérance et variance ; un notebook de simulation.
- En fin de journée : les quatre missions de l'agence, puis le tirage au sort.



LA SÉANCE 1, L'ESSENTIEL

Le vocabulaire

- Une expérience aléatoire : on connaît les résultats possibles, pas celui qui se produira.
- L'univers Ω : tous les résultats possibles. Un événement : une partie de Ω .
- $\bar{A}$, le contraire de A : A ne se réalise pas.
- A et B : les deux se réalisent. A ou B : au moins l'un des deux.
- A et B sont incompatibles s'ils ne peuvent pas se réaliser ensemble.

Les règles de calcul

- Pour tout événement A : $0 \leq P(A) \leq 1$; et $P(\Omega) = 1$.
- $P(\bar{A}) = 1 - P(A)$.
- $P(A \text{ ou } B) = P(A) + P(B) - P(A \text{ et } B)$.
- Résultats équiprobables : $P(A) = \text{cas favorables} / \text{cas possibles}$.
- Sur un grand nombre de répétitions, la fréquence de A se stabilise autour de $P(A)$.

La probabilité conditionnelle

- $P(A | B)$: la probabilité de A, sachant que B est réalisé.
- Dans un tableau croisé, on ne lit que la ligne, ou la colonne, de B.
- $P(A | B) = P(A \text{ et } B) / P(B)$; donc $P(A \text{ et } B) = P(B) \times P(A | B)$.
- $P(A | B)$ et $P(B | A)$ répondent à deux questions différentes.
- Indépendance : savoir B ne change rien, $P(A | B) = P(A)$.

Deux exercices du 7/10

- Fascicule du 7/10 : exercices 2 et 6.
- En binôme, 12 minutes.
- Le 2 : des événements, et la règle du « ou ». Le 6 : un tableau croisé.
- La suite du fascicule se termine à la maison, dans `exercices/seance-01.md`.



LE NOM DE L'AGENCE



RETOUR SUR LA COMMANDE 1

Les trois probabilités

- $P(\text{annulée} \mid \text{groupe}) = 9\,623 / 13\,975 \approx 0,689$.
- $P(\text{annulée} \mid \text{agence en ligne}) = 14\,491 / 38\,748 \approx 0,374$.
- $P(\text{agence en ligne} \mid \text{annulée}) = 14\,491 / 33\,102 \approx 0,438$.
- Les deux premières se lisent dans une ligne, la troisième dans une colonne.

Source : Hotel booking demand, Antonio et al., 2019, licence CC BY 4.0.

Le taux et la part

- Les groupes annulent le plus souvent : 69 % de leurs réservations.
- Mais ils ne font que 29 % des annulations.
- Les agences en ligne annulent moins que la moyenne : 37 % contre 42 %.
- Par leur volume, elles font 44 % des annulations.
- Deux questions différentes : le risque d'une réservation, le poids d'un segment.

Une recommandation qui tient

- Elle choisit une priorité, et dit pourquoi.
- Elle s'appuie sur les chiffres, avec la bonne lecture : taux ou part.
- Elle rappelle d'où viennent les données : un hôtel comparable, pas le client.
- Trois lignes suffisent.

La question 3, c'était déjà Bayes

- On connaissait $P(\text{annulée} \mid \text{agence en ligne})$: le taux d'un segment.
- On cherchait $P(\text{agence en ligne} \mid \text{annulée})$: la part d'un segment.
- Le tableau donnait les deux.
- Sans tableau, avec seulement des pourcentages : c'est le théorème de Bayes.



LE THÉORÈME DE BAYES

Renverser une probabilité conditionnelle

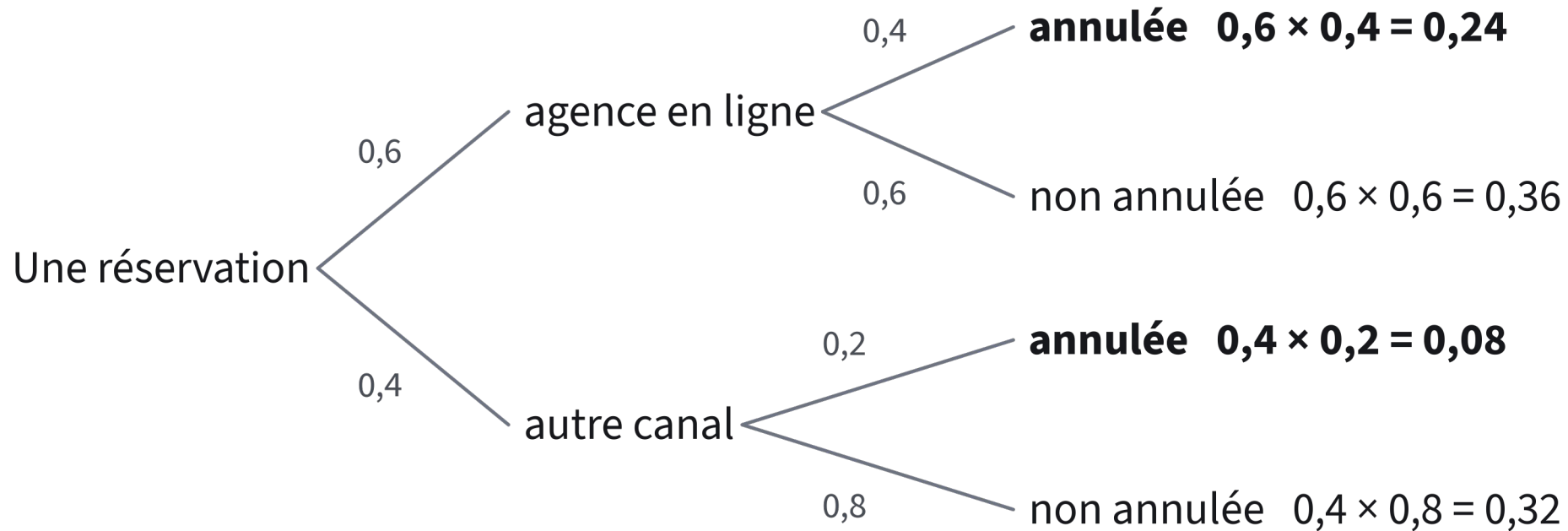
- On connaît souvent $P(B | A)$: le taux d'annulation d'un canal, la fiabilité d'un contrôle.
- On cherche $P(A | B)$: d'où vient une annulation, si une alerte est fondée.
- Ces deux probabilités sont différentes (séance 1).
- Le théorème de Bayes passe de l'une à l'autre.

L'exercice 5 du 7/10

- 60 % des réservations d'un hôtel passent par une agence en ligne.
- 40 % d'entre elles sont annulées.
- 20 % des autres réservations sont annulées.
- Parmi les réservations annulées, quelle part vient d'une agence en ligne ?

Chiffres fictifs. C'est l'exercice 5 du fascicule du 7/10.

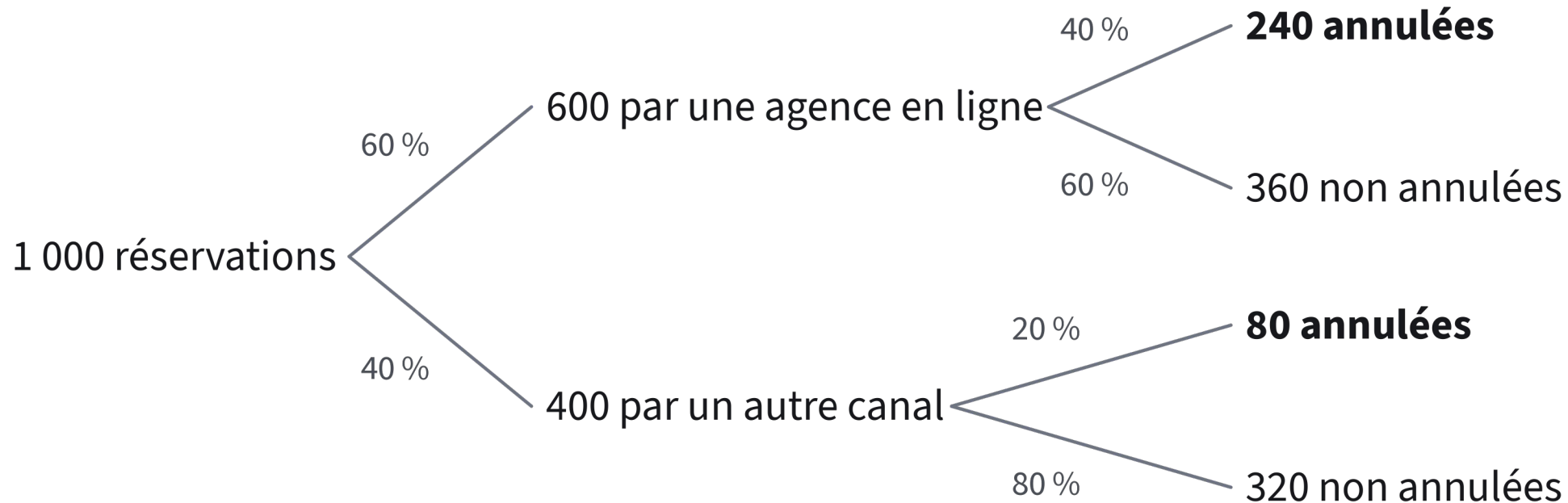
L'arbre des probabilités



Lire l'arbre

- Multiplier le long d'un chemin : $P(\text{agence et annulée}) = 0,6 \times 0,4 = 0,24$.
- Additionner les chemins qui mènent à « annulée » : $P(\text{annulée}) = 0,24 + 0,08 = 0,32$.
- Diviser : $P(\text{agence} \mid \text{annulée}) = 0,24 / 0,32 = 0,75$.
- Trois annulations sur quatre viennent d'une agence en ligne.

Les mêmes chiffres, en nombres



Pourquoi compter

- 1 000 réservations imaginées : on ne manipule que des nombres entiers.
- On voit qui est dans le groupe « annulées » : 240 et 80.
- La réponse est une part de ce groupe : 240 sur 320.
- Choisir un total qui tombe juste : 1 000, 10 000, 100 000.

La formule

$$P(A \mid B) = \frac{P(B \mid A) P(A)}{P(B)}$$

P(B), par les deux chemins

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

La formule, sans la retenir par cœur

- Numérateur : le chemin qui nous intéresse.
- Dénominateur : tous les chemins qui mènent à ce qu'on a observé.
- Un arbre, ou 1 000 cas imaginés, redonne toujours la formule.

Exercices 1 à 3

- Fascicule de la journée, exercices 1, 2 et 3.
- En binôme, 25 minutes.
- L'exercice 3 refait la question 3 de la commande 1 avec la formule.



PAUSE · 10 MINUTES



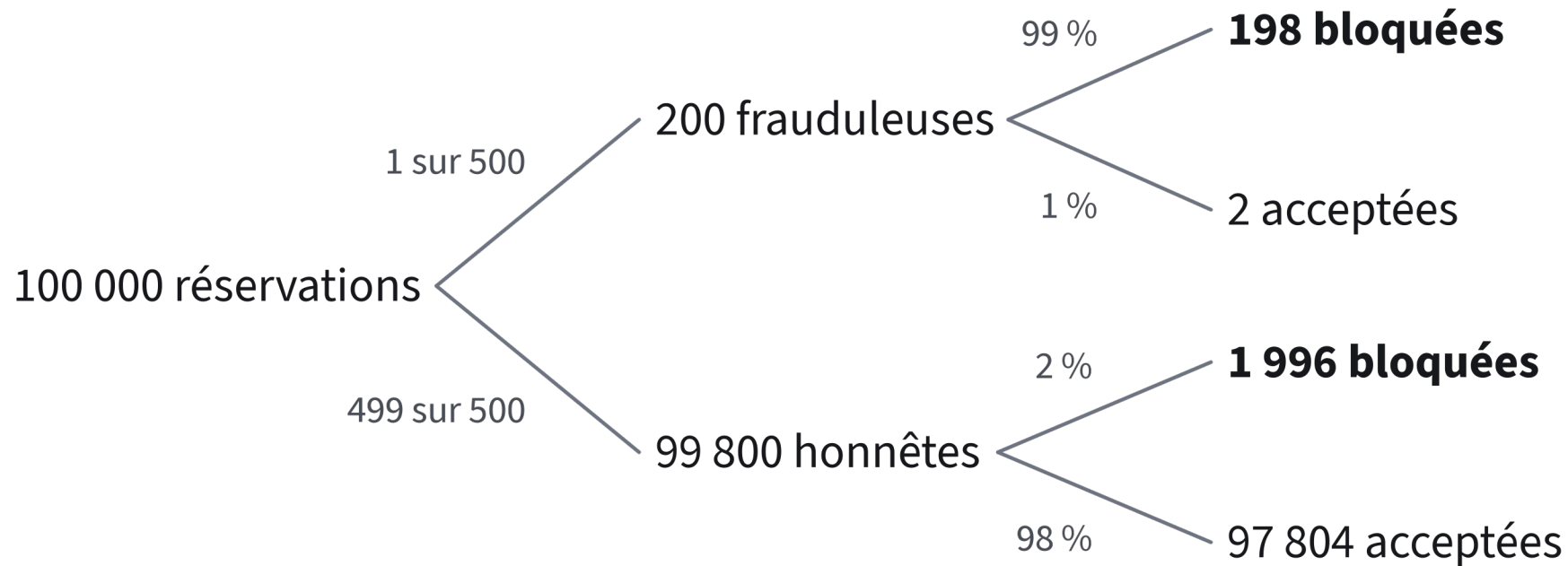
LE TAUX DE BASE

Un contrôle très fiable

- Une réservation sur 500 est frauduleuse.
- Le contrôle de paiement bloque 99 % des réservations frauduleuses.
- Il bloque aussi, à tort, 2 % des réservations honnêtes.
- Une réservation vient d'être bloquée. Est-elle frauduleuse ?

Chiffres fictifs.

Sur 100 000 réservations



Neuf alertes sur dix sont fausses

- 2 194 réservations bloquées, dont 198 frauduleuses : 9 %.
- La fraude est rare : 200 cas sur 100 000.
- Les honnêtes sont nombreuses : 2 % de 99 800, c'est 1 996.
- Les fausses alertes noient les vraies.

Avec la formule

- $P(\text{fraude}) = 0,002$; $P(\text{bloquée} \mid \text{fraude}) = 0,99$; $P(\text{bloquée} \mid \text{honnête}) = 0,02$.
- $P(\text{bloquée}) = 0,99 \times 0,002 + 0,02 \times 0,998 = 0,02194$.
- $P(\text{fraude} \mid \text{bloquée}) = 0,00198 / 0,02194 \approx 0,090$.

Le taux de base change tout

Part de réservations frauduleuses	P(fraude bloquée)
1 sur 500	0,09
1 sur 50	0,50
1 sur 10	0,85

Même contrôle dans les trois cas : 99 % des fraudes bloquées, 2 % de fausses alertes. Chiffres fictifs.



DEUX CAS MÉTIER

Prédire les annulations

- Un hôtel teste un modèle qui signale les réservations « à risque ».
- 80 % des réservations annulées avaient été signalées.
- 25 % des réservations honorées l'avaient été aussi.
- Taux de base : 37 % des réservations sont annulées.
- Une réservation est signalée : quelle chance qu'elle soit annulée ?

Taux de base : Hotel booking demand, Antonio et al., 2019, licence CC BY 4.0. Modèle et taux de signalement : fictifs.

Sur 10 000 réservations

- 3 700 annulées, dont 80 % signalées : 2 960.
- 6 300 honorées, dont 25 % signalées : 1 575.
- 4 535 signalées, dont 2 960 annulées : 65 %.
- Relancer les réservations signalées : une relance sur trois est inutile.

Les faux avis

- Sur une plateforme, 5 % des avis sont faux.
- Un filtre signale 90 % des faux avis.
- Il signale aussi 8 % des avis authentiques.
- Sur 1 000 avis : 45 faux signalés, 76 vrais signalés.
- $P(\text{faux} \mid \text{signalé}) = 45 / 121 \approx 0,37$.

Chiffres fictifs.

Exercices 4 et 6

- Fascicule, exercices 4 et 6.
- En binôme, 20 minutes.
- Les exercices 5 et 7 se font à la maison.
- Méthode conseillée : un arbre, ou un total imaginé qui tombe juste.

À retenir

- $P(A | B)$ et $P(B | A)$ répondent à deux questions différentes.
- Un arbre : multiplier le long d'un chemin, additionner les chemins.
- Les fréquences naturelles : un total imaginé, puis compter.
- Le taux de base, $P(A)$, pèse autant que la fiabilité du test.
- La formule de Bayes résume tout cela.

Cet après-midi

- Les lois de Bernoulli, binomiale et de Poisson.
- Espérance, variance, écart-type.
- Un notebook de simulation, dans votre codespace.
- Les quatre missions de l'agence, puis le tirage au sort.

Sources

- Hotel booking demand : N. Antonio, A. de Almeida, L. Nunes, *Data in Brief*, 22 (2019), 41–49, doi:10.1016/j.dib.2018.11.126. Licence CC BY 4.0.
- Sauf mention de cette source, les chiffres de cette séance sont fictifs.
- Supports et résumés des séances : ad-c.github.io/mth101