

Exercices 1 à 7 du 9/10 : corrigés

Théorème de Bayes

Adrien Chiodo

Mercredi 14 octobre 2026

Les données réelles de ce fascicule viennent du jeu Hotel booking demand : N. Antonio, A. de Almeida, L. Nunes, « Hotel booking demand datasets », *Data in Brief*, 22 (2019), 41–49, doi:10.1016/j.dib.2018.11.126, licence CC BY 4.0. Deux hôtels portugais, 119 390 réservations, arrivées de juillet 2015 à août 2017. Une réservation y est dite « annulée » quand le client a annulé, ou qu'il ne s'est pas présenté sans prévenir.

EXERCICE 1 · AFFAIRES ET PETIT-DÉJEUNER

Dans un hôtel, 30 % des clients voyagent pour affaires. 80 % des clients d'affaires prennent le petit-déjeuner, contre 50 % des autres clients.

1. Dessinez l'arbre de probabilités.
2. Calculez la probabilité qu'un client pris au hasard prenne le petit-déjeuner.
3. Un client prend son petit-déjeuner. Quelle est la probabilité qu'il voyage pour affaires ?
4. Un client ne prend pas de petit-déjeuner. Même question.

Corrigé

1. Premier niveau : affaires 0,3 ; autre motif 0,7. Second niveau : sous « affaires », petit-déjeuner 0,8 et sans 0,2 ; sous « autre motif », 0,5 et 0,5.
2. $P(PD) = 0,3 \times 0,8 + 0,7 \times 0,5 = 0,24 + 0,35 = 0,59$.
3. $P(\text{affaires} | PD) = 0,24 / 0,59 \approx 0,407$.
4. $P(\text{affaires} | \text{sans PD}) = (0,3 \times 0,2) / 0,41 = 0,06 / 0,41 \approx 0,146$. Savoir qu'un client prend le petit-déjeuner fait passer la part des clients d'affaires de 30 % à 41 % ; savoir qu'il n'en prend pas la fait tomber à 15 %.

EXERCICE 2 · LE PASS ANNUEL

Parmi les visiteurs d'un parc de loisirs, 25 % ont un pass annuel. 60 % des détenteurs d'un pass viennent en semaine, contre 20 % des autres visiteurs.

1. Imaginez 2 000 visiteurs. Dessinez l'arbre des fréquences naturelles : des nombres de visiteurs à chaque nœud.
2. Combien de visiteurs viennent en semaine ? Combien d'entre eux ont un pass ?
3. Calculez $P(\text{pass} | \text{semaine})$ et $P(\text{pass} | \text{week-end})$.
4. Le parc veut proposer une offre de pass annuel à des visiteurs qui n'en ont pas encore. Vaut-il mieux la présenter en semaine ou le week-end ? Donnez deux arguments chiffrés.

Corrigé

1. 500 visiteurs avec un pass : 300 en semaine, 200 le week-end. 1 500 sans pass : 300 en semaine, 1 200 le week-end.
2. 600 en semaine, dont 300 avec un pass.
3. $P(\text{pass} \mid \text{semaine}) = 300 / 600 = 0,5$; $P(\text{pass} \mid \text{week-end}) = 200 / 1\,400 \approx 0,143$.
4. Le week-end : 86 % des visiteurs n'y ont pas de pass (1 200 sur 1 400), contre 50 % en semaine ; et ces visiteurs sans pass y sont quatre fois plus nombreux (1 200 contre 300). Un taux et un volume, comme dans la commande 1.

EXERCICE 3 · LA COMMANDE 1, PAR LA FORMULE

Source : Hotel booking demand, hôtel urbain. On ne garde de la commande 1 que trois nombres :

- $P(\text{agence en ligne}) = 0,488$: la part des réservations passées par une agence de voyage en ligne ;
 - $P(\text{annulée} \mid \text{agence en ligne}) = 0,374$;
 - $P(\text{annulée}) = 0,417$.
1. Avec la formule de Bayes, calculez $P(\text{agence en ligne} \mid \text{annulée})$.
 2. Comparez avec votre réponse à la question 3 de la commande 1.
 3. Les agences en ligne annulent moins que la moyenne, et pèsent pourtant 44 % des annulations. Expliquez-le à l'aide de la formule.

Corrigé

1. $P(\text{agence} \mid \text{annulée}) = 0,374 \times 0,488 / 0,417 \approx 0,438$ (0,4377 avec ces arrondis).
2. C'est le résultat de la commande, $14\,491 / 33\,102 \approx 0,438$: la formule retrouve ce que donnait le tableau.
3. $P(\text{agence} \mid \text{annulée}) = P(\text{agence}) \times P(\text{annulée} \mid \text{agence}) / P(\text{annulée}) = 0,488 \times (0,374 / 0,417) \approx 0,488 \times 0,897$. Le rapport est inférieur à 1 : elles annulent moins que la moyenne, ce qui réduit un peu leur part. Mais elles font près de la moitié des réservations : leur poids reste fort. La part d'un segment, son taux de base, compte autant que son taux d'annulation.

EXERCICE 4 · LES PUNAISES DE LIT

Un hôtel fait inspecter ses chambres par un détecteur. Une chambre sur 200 est infestée. Le détecteur signale 95 % des chambres infestées, mais aussi 3 % des chambres saines.

1. Imaginez 20 000 inspections. Combien de chambres infestées, et combien sont signalées ? Combien de chambres saines sont signalées ?
2. Une chambre est signalée. Quelle est la probabilité qu'elle soit infestée ?
3. La direction veut fermer chaque chambre signalée pour la traiter. Sur 100 chambres fermées, combien le seraient à tort, environ ?
4. *Pour aller plus loin.* Une chambre signalée est inspectée une seconde fois, et de nouveau signalée. On suppose les deux inspections indépendantes. Quelle est maintenant la probabilité qu'elle soit infestée ? Partez de la réponse à la question 2 comme nouveau taux de base.

Corrigé

1. 100 chambres infestées, dont 95 signalées ; 19 900 saines, dont 597 signalées ; 692 signalées en tout.
2. $P(\text{infestée} \mid \text{signalée}) = 95 / 692 \approx 0,137$.
3. Environ 86 sur 100 : $597 / 692 \approx 0,863$.
4. $0,95 \times 0,137 / (0,95 \times 0,137 + 0,03 \times 0,863) \approx 0,834$. Une seconde inspection indépendante change un signal faible en quasi-certitude. L'indépendance se discute : un détecteur qui se trompe sur une chambre (une odeur, un produit) risque de se tromper encore.

EXERCICE 5 · LES AVIS CINQ ÉTOILES

Sur une plateforme, 10 % des avis sont faux. 90 % des faux avis donnent cinq étoiles, contre 40 % des avis authentiques.

1. Quelle part des avis donnent cinq étoiles ?
2. Un avis donne cinq étoiles. Quelle est la probabilité qu'il soit faux ?
3. Un avis donne moins de cinq étoiles. Même question.
4. La plateforme envisage de masquer tous les avis cinq étoiles. Qu'en pensez-vous ?

Corrigé

1. $P(5 \text{ étoiles}) = 0,1 \times 0,9 + 0,9 \times 0,4 = 0,09 + 0,36 = 0,45$.
2. $P(\text{faux} \mid 5 \text{ étoiles}) = 0,09 / 0,45 = 0,2$.
3. $P(\text{faux} \mid \text{moins de 5}) = (0,1 \times 0,1) / 0,55 \approx 0,018$.
4. Elle masquerait 45 % des avis, dont 80 % sont authentiques (0,36 sur 0,45), pour écarter 90 % des faux. Cinq étoiles double le risque qu'un avis soit faux (de 10 % à 20 %) : c'est un indice, pas une preuve.

EXERCICE 6 · UN MODÈLE, DEUX HÔTELS

Un modèle de prédiction des annulations signale 75 % des réservations qui seront annulées ; il signale aussi 10 % de celles qui seront honorées. On l'installe dans deux hôtels. Dans l'hôtel A, 40 % des réservations sont annulées ; dans l'hôtel B, 10 %.

1. Imaginez 1 000 réservations de l'hôtel A. Combien d'alertes justes, combien de fausses alertes ? Calculez $P(\text{annulée} \mid \text{signalée})$.
2. Même question pour l'hôtel B.
3. Le fournisseur annonce : « 75 % des annulations détectées, seulement 10 % de fausses alertes. » Pourquoi ces deux chiffres ne suffisent-ils pas à dire si une alerte est fiable ?

Corrigé

1. 400 annulées, dont 300 signalées ; 600 honorées, dont 60 signalées. $P(\text{annulée} \mid \text{signalée}) = 300 / 360 \approx 0,833$.
2. 100 annulées, dont 75 signalées ; 900 honorées, dont 90 signalées. $75 / 165 \approx 0,455$.

3. La fiabilité d'une alerte, la précision du modèle, dépend du taux de base, propre à chaque hôtel. Le même modèle a raison cinq fois sur six dans l'hôtel A, moins d'une fois sur deux dans l'hôtel B. Vocabulaire : 75 % est le rappel, $P(\text{signalée} \mid \text{annulée})$; la précision est $P(\text{annulée} \mid \text{signalée})$.

EXERCICE 7 · VRAI OU FAUX

Justifiez chaque réponse en une phrase.

1. « 90 % des fraudes au paiement d'un hôtel viennent de cartes étrangères : 90 % des paiements par carte étrangère sont donc frauduleux. »
2. « Si $P(A \mid B) = P(B \mid A)$, avec des probabilités non nulles, alors $P(A) = P(B)$. »
3. « Un contrôle qui ne se trompe que dans 1 % des cas donne une alerte juste 99 fois sur 100. »
4. « Si un test signale aussi souvent les cas positifs que les cas négatifs, $P(B \mid A) = P(B \mid \bar{A})$, il n'apporte aucune information sur A. »

Corrigé

1. Faux. La phrase confond $P(\text{étrangère} \mid \text{fraude})$ et $P(\text{fraude} \mid \text{étrangère})$; la seconde dépend aussi de la rareté des fraudes et de la part des cartes étrangères.
2. Vrai. $P(A \mid B) \times P(B) = P(A \text{ et } B) = P(B \mid A) \times P(A)$: si les deux probabilités conditionnelles sont égales et non nulles, $P(A) = P(B)$.
3. Faux. Tout dépend du taux de base : avec une fraude sur 500, un contrôle presque aussi fiable ne donne qu'une alerte juste sur onze (séance 2).
4. Vrai. Alors $P(A \mid B) = P(A)$: A et B sont indépendants (séance 1), et observer B ne change rien à la probabilité de A.