L'épistémologie bayésienne

Christian Wüthrich

http://www.wuthrich.net/

BA2b Introduction à la philosophie des sciences

Remerciements: Pablo Carnino, Augustin Baas

Plan

Épistémologie/Théorie de la confirmation bayésienne

- un des développements les plus importants dans l'épistémologie du 20e siècle
- offre un modèle formel et mathématique rigoureux qui lie la croyance en certaines hypothèses et leur confirmation ou leur infirmation par des données
- le modèle est probabiliste: assigne des probabilités aux croyances
- idée générale: une donnée à l'appui E confirme une hypothèse H dans le cas où elle augmente la probabilité de H, i.e., $P(H|E) > P(H)$
- Les probabilités doivent être «mises-à-jour» de la façon prédite par le théorème de Bayes, de sorte que le degré de mis-à-jour de la croyance en une hypothèse soit la probabilité de l'hypothèse conditionnellement aux données.
- inclut un «test d'auto-défaite» pragmatique pour la rationalité épistémique (le meilleur après une justification basée sur la logique déductive)

$\Rightarrow$ les lois du calcul des probabilités comme contraintes sur des degrés rationnels de croyance (ou de confiance)

Andreï Nikolaievitch Kolmogorov (1903-1987)

- mathématicien russe, Université d'État de Moscou
- contributions en théories de la probabilité, topologie, logique intuitionniste, turbulence, mécanique classique, complexité algorithmique
- accomplissement principal: fondation axiomatique de la théorie de la probabilité
- «La théorie de la probabilité en tant que discipline mathématique peut et devrait être développée à partir d'axiomes exactement de la même façon que la géométrie et l'algèbre.»

(*Foundations of the Theory of Probability*, 1956, p.1)

Axiomes de Kolmogorov

Donnée: classe $\mathcal{S}$ de propositions $A, B, C, \ldots$

Introduire une fonction de probabilité sur $\mathcal{S}$ qui associe $\mathcal{S}$ à l'intervale clos $[0, 1]$ telle que les axiomes suivant soient valables:

Axiome (1: non-négativité)

$P(X) \geq 0$ *pour tout* X *dans* $\mathcal{S}$*; i.e. toutes les probabilités sont non-négatives.*

Axiome (2: unité de mesure)

$P(X) = 1$ *si* X *dans* $\mathcal{S}$ *est une tautologie; i.e. si* X *est une proposition qui est vraie dans tous les cas possibles, alors elle a la probabilité 1.*

Axiome (3: additivité)

Pout tout X *et* Y *dans* $\mathcal{S}$*, si* X *et* Y *sont des propositions mutuellement exclusives, alors* $P(X \vee Y) = P(X) + P(Y)$*.*

Probabilité conditionnelle

Définition (Probabilité conditionnelle)

Étant donné une fonction de probabilité $P(X)$ comme définie sur la diapositive précédente, la probabilité conditionnelle $P(S|T)$ de S étant donné T est définie comme

$$P(S|T) := \frac{P(S\&T)}{P(T)}.$$

Exemple: Jeu de cartes de jass

$$\begin{aligned} P(\heartsuit|\text{Rouge}) &= \frac{P(\heartsuit\&\text{Rouge})}{P(\text{Rouge})} = \frac{1/4}{1/2} = \frac{1}{2} \\ P(\text{Roi}|\heartsuit) &= \frac{P(\text{Roi}\&\heartsuit)}{P(\heartsuit)} = \frac{1/36}{1/4} = \frac{1}{9} \end{aligned}$$

Mise-à-jour bayésienne des croyances

Les bayésiens font (une version plus compliquée de) la présupposition épistémologique suivante:

Principe (Conditionnalisation)

À partir des probabilités initiales (antérieures) $P_i(H)$ de n'importe quel hypothèse H et ayant acquis de nouvelles données tel qu'on a observé les données E (c'est-à-dire $P(E) = 1$), la rationalité dicte que l'on mette-à-jour ses probabilités initiales afin d'obtenir ses probabilités finales (postérieures) pour «conditionner» sur E

$$P_i(H) \to P_f(H) = P_i(H|E).$$

Donc on devrait trouver un moyen de calculer $P_i(H|E)$ comme une fonction de $P_i(H)$.

Théorème du Révérend Thomas Bayes (1702-1761)

Pour toute proposition H et E, on a

$$P(H|E) = \frac{P(H) \cdot P(E|H)}{P(E)}$$
$$= \frac{P(H) \cdot P(E|H)}{P(E|H) \cdot P(H) + P(E|\neg H) \cdot P(\neg H)}$$

où

- $P(H)$: probabilité antérieure (*a priori*) de H
- $P(H|E)$: probabilité postérieure (*a posteriori*) de H sachant E (ou encore: de H sous condition E)
- $P(E|H)$: fonction de vraisemblance de H (pour un E connu)

Mise-à-jour bayésienne

- D'abord, déterminez la probabilité antérieure de H et les chances que E_1 soit observé étant donné H.
- Déterminez la probabilité d'observer E_1 indépendamment de H.
- Si E_1 est observé, calculez la probabilité postérieure $P(H|E_1)$ via le théorème de Bayes.
- Considérez cette probabilité postérieure comme votre nouvelle probabilité antérieure de H.
- Considérez la probabilité d'une nouvelle donnée à l'appui de E_2 et ses chances à la lumière de H.
- Si E_2 est observé, calculez la nouvelle probabilité postérieure de H via le théorème de Bayes.
- ...

Exemple 1: de quel bol vient le cookie?

Deux bols de cookies:

1. Bol_1 a 10 cookies au chocolat et 30 cookies nature
2. Bol_2 a 20 cookies au chocolat et 20 cookies nature

Question: Si l'on prend un cookie d'un bol au hazard, et qu'il est nature (E), quelle est la probabilité qu'il provienne du Bol_1 (H)?

a priori: $P(H) = P(\neg H) = 0.5$

vraisemblances: $P(E|H) = 0.75$ et $P(E|\neg H) = 0.5$

Utilisez le théorème de Bayes:

$$\begin{aligned} P(H|E) &= \frac{P(H) \cdot P(E|H)}{P(E|H) \cdot P(H) + P(E|\neg H) \cdot P(\neg H)} \\ &= \frac{0.5 \cdot 0.75}{0.5 \cdot 0.75 + 0.5 \cdot 0.5} = 0.6. \end{aligned}$$

Example 2: Vient-elle/il à la fête?

De: Peter Godfrey-Smith, *Theory and Reality: An Introduction to the Philosophy of Science*, p. 204.

- H: hypothèse qu'elle/il est à la fête
- E: donnée selon lesquelles son vélo est parquée dehors
- $P(H)$: probabilité initiale qu'elle/il soit à la fête (avant d'avoir vu le vélo); disons que c'est 0.5
- $P(E|H)$: chances que son vélo soit parquée dehors si elle/il est à la fête; supposez que c'est 0.8
- $P(E|\neg H)$: chances que son vélo soit parquée dehors si elle/il n'est pas à la fête; supposez que c'est seulement 0.1
- $P(H|E)$: probabilité qu'elle/il soit à la fête étant donné que son vélo est parquée dehors; utilisez le théorème de Bayes:

$$P(H|E) \;=\; \frac{0.5 \cdot 0.8}{0.5 \cdot 0.8 + 0.5 \cdot 0.1} \;=\; 0.89.$$

$\Rightarrow$ voir le vélo augmente la probabilité de H de 0.5 à 0.89.

Exemple 3: dans un tribunal

Un juré doit évaluer comment des données pèsent en faveur de ou contre la culpabilité de l'accusé:

- C: hypothèse que l'accusé est coupable
- D: données indiquant que l'ADN de l'accusé correspond à l'ADN trouvée sur les lieux du crime
- $P(D|C)$: chances de voir des données d'ADN correspondantes si l'accusé est coupable; dans les crimes capitals, typiquement très haute; ici présumée égale à 1
- $P(D|\neg C)$: chances de voir des données d'ADN correspondante si l'accusé n'est pas coupable; très basse, on présume 1 sur un million, ou 10^{-6}
- $P(C)$: probabilité initiale que l'accusé soit coupable (a priori) ; dépend largement des autres preuves, circonstances, etc. Deux cas de figure: soit (A) forte suspicion antérieure ($P(C) = 0.3$), ou (B) suspicion très basse ($P(C) = 10^{-6}$)
- $P(C|D)$: probabilité que l'accusé soit coupable si l'ADN correspond; c'est ce que l'on cherche à savoir!

Cas (A): $P(C) = 0.3$

$$\begin{aligned}
P(C|D) &= \frac{P(C)\cdot P(D|C)}{P(D|C)\cdot P(C) + P(D|\neg C)\cdot P(\neg C)} \\
&= \frac{0.3\cdot 1.0}{0.3\cdot 1.0 + 0.7\cdot 10^{-6}} \\
&= 0.99999766667
\end{aligned}$$

Cas (B): $P(C) = 10^{-6}$

$$\begin{aligned}
P(C|D) &= \frac{P(C)\cdot P(D|C)}{P(D|C)\cdot P(C) + P(D|\neg C)\cdot P(\neg C)} \\
&= \frac{10^{-6}\cdot 1.0}{10^{-6}\cdot 1.0 + (1-10^{-6})\cdot 10^{-6}} \\
&\approx 0.5
\end{aligned}$$

Example 4: La recherche du *Scorpion*

- Mai 1968: Le sous-marin nucléaire américain *Scorpion* n'arrive pas à destination à Norfolk, VA.
- La US Navy est convaincue que le vaisseau s'est échoué sur les côtes de l'Est, mais des recherches approfondies ne permettent pas de retrouver l'épave.
- John Craven, expert des profondeurs à la US Navy croyait qu'il était au sud-ouest de l'archipel portuguais des *Açores* selon une triangulation d'hydrophones controversée.
- attribution de ressources limitées (un navire) $\Rightarrow$ il les optimisât
- Craven travailla avec des mathématiciens pour optimiser la recherche, en utilisant la théorie bayésienne de la recherche.
- Octobre 1968: L'épave est retrouvée à 740 km au sud-ouest des Açores.

Théorie bayésienne de la recherche

1. mer divisée en carrés formant une grille
2. des commandants expérimentés de sous-marins sont interrogés, afin de formuler des hypothèses à propos de ce qui est arrivé au vaisseau
3. construction des lois de probabilité sur chaque carré correspondant à chaque hypothèse
4. construction de la loi de probabilité qu'on retrouve l'épave dans le carré X si elle est vraiment en X (fonction de la profondeur)
5. combinaison de toutes ces lois de probabilité (de 3 et 4) afin de produire une grille de l'ensemble des probabilités; elle donne la probabilité de trouver l'épave dans un carré si on fait des recherches dans ce carré (pour chaque carré)
6. construction d'un chemin de recherche partant du carré avec la plus haute probabilité qui passe ensuite par des zones de haute probabilité, puis de probabilité intermédiaire, avant d'arriver aux zones de basse probabilité
7. révision constant de la loi de probabilité d'ensemble à mesure que la recherche avance, i.e. si on a cherché dans un carré sans succès, alors la probabilité que l'épave y soit diminue grandement (quoique habituellement pas nulle), et la probabilité de la trouver ailleurs doit être augmentée; cette révision suit le théorème de Bayes

Exemple 5: Le V1@gra et les filtres bayésiens

- but: savoir les chances qu'un email soit du spam sur la base des mots qui y figurent
- idée de base: un filtre anti-spam bayésien
  - liste des mots dans les emails entrants,
  - assigne à chaque mot une probabilité qu'il apparaisse dans un email de spam (les mots mal épelés ont un score très élevé), et
  - emploie ces probabilités comme données entrantes dans la formule de Bayes afin de déterminer si un email est du spam ou non
- Tout d'abord il faut entraîner le filtre anti-spam en lui montrant des emails de spam et de non-spam (de plus en plus automatisé).
- Le filtre enregistre tous les mots des messages d'entraînement (incluant nom de host, adresse IP, tag HTML, etc.) dans une banque de données.

$\Rightarrow$ le filtre calcule la probabilité qu'un mot apparaisse dans un spam sur la base de sa fréquence dans la banque de données (la «spamicité» de chaque mot)

- Une spamicité de 0.5 est neutre, plus haute (moindre) signifie qu'il apparaît souvent dans des messages (non-)spam.
- Ensuite le filtre utilise la formule de Bayes afin de calculer la spamicité d'ensemble d'un email, sur la base de la spamicité de tous les mots qu'il contient.

$\Rightarrow$ le message est placé dans le filtre à spam si la spamicité est supérieure à 0.5

- Généralement: Les filtres à spam bayésiens sont très efficaces car (1) ils s'adaptent au circonstances individuelles (la banque de donnée est construite pour chaque usager), et (2) ils apprennent avec le temps et mettent à jour la banque de données

Exemple 6: Prédire le résultat des présidentielles aux USA

`Combien de votes aura Obama?` (en 2008; en anglais)

- Le lien ci-dessus mène à un blog où une simple méthode bayésienne est utilisée pour prédire le résultat de l'avant-dernière élection présidentielle des EU (le 31 Octobre 2008)
- Le blogger a prédit, sur la base des données de sondages de cnn.com et de 5000 élections simulées, que la probabilité que McCain gagne suffisamment de votes pour accéder à la Maison Blanche était de 0.0.
- Nate Silver a fait une prédiction similaire pour 2008. Voici sa prédiction pour l'élection présidentielle de 2012 basée sur les statistiques bayésiennes de la nuit d'avant (en anglais): `http://fivethirtyeight.blogs.nytimes.com/fivethirtyeights-2012-forecast/`
- Et par Drew Linzer à Votamatic (en anglais): `http://votamatic.org/final-result-obama-332-romney-206/` (plus sur la méthode de Linzer: `http://votamatic.org/how-it-works/`)

Échec en 2016...

Les méthodes bayésiennes ont échoué en 2016:

- Nate Silver et son équipe offrent des prédictions très détaillées (en anglais); voici leur dernières prédiction avant l'élection de 8 novembre 2016: `http://projects.fivethirtyeight.com/2016-election-forecast/?ex_cid=rrpromo`
- Drew Linzer s'est joint à l'équipe de «Daily Kos» (en anglais); voici sa dernière prédiction avant l'élection: `http://elections.dailykos.com/app/elections/2016`
- raisons de l'échec selon Silver: beaucoup des électeur-rice-s indécidé-e-s ont systématiquement voté pour Trump ⇒ crucial pour la perte de Clinton dans les états «de swing» du Mid-Ouest
- Il faut souligner que toutes les méthodes de prédiction des élections présidentielles de 2016 ont échoué pareillement (ou pire)...

Qu'est-ce que la probabilité?

(objective) hasard	vs.	(subjective) degré de croyance
$\downarrow$		$\downarrow$
Comment mesurer?		Comment mesurer?
$\downarrow$		$\downarrow$
fréquence des occurrences		attitude de pari

Les axiomes de la probabilité doivent s'appliquer aux deux!

Bayésiannisme subjectiviste

- «probabilité comme degré de croyance personnelle» (générée par le libre choix, la socialisation, l'évolution, etc.)
- La probabilité d'un événement est simplement la certitude avec laquelle un agent bayésien s'attend à ce que l'événement se produise.
- idée principale: «croyance rationnelle» devrait être comprise comme une généralisation de l'attitude de pari: étant donné une certaine quantité d'information et demandé une prédiction, à quelle cote est-ce qu'on parierait sur son prédiction?
- parier sur H à une cote de $x : 1$ signifie être prêt à risquer de perdre x CHF si H s'avère fausse, contre un gain de CHF 1 si H est vraie
- Si vos cotes subjectivement équitables pour un pari sur H sont $x : 1$, alors votre degré de croyance en H est $x/(x + 1)$.
- Il est bien sûr possible que le degré subjectif de croyance viole les axiomes, mais...

Le théorème du «Dutch book»

Si le degré de croyance subjective de quelqu'un viole un axiome de Kolmogorov, alors cette personne devrait accepter une combinaison de paris qui reviennent à un «Dutch book», i.e. la combinaison de paris qui, s'ils devaient être acceptés, garantissent à la personne une perte!

Exemple simple: votre degré de croyance que le prochain pile-ou-face tombera sur «face» (F) est de 0.55, et votre degré de croyance qu'il tombera sur «pile» (P) est 0.5 $\Rightarrow$ le bookmaker peut vous proposer un ensemble de paris qui garantit que vous perdrez 5 centimes sur chaque franc parié.

Table de décision sur une combinaison de paris [en francs]:

	$F(\neg P)$	$P(\neg F)$
pariez 0.55	1	0
pariez 0.50	0	1
payez 1.05 au total	gagnez 1	gagnez 1

La solution bayésienne au paradoxe «vleu» («noanc»)

- Supposez qu'on vous présente deux arguments inductifs à partir du même ensemble d'observations des corbeaux noirs, l'un argumentant que toutes les corbeaux sont noirs, l'autre qu'ils sont noancs.
- Pourquoi une des deux inductions est-elle meilleure que l'autre?
- réponse bayésienne standard: les deux sont OK, mais la plupart des gens assignerait une probabilité antérieure plus haute à l'hypothèse «noir» qu'à la «noanc»
- réaction: vrai, cela montre une différence, mais cela explique-t-il pourquoi l'hypothèse «noanc» à une probabilité antérieure plus basse?
- Le bayésiannisme n'offre aucune critique de la décision subjective d'assigner une probabilité antérieure plus haute à l'hypothèse «noanc», pour autant que les probabilités soient cohérentes de manière interne et mis-à-jour comme il faut.

(1) Problème des antérieures

- L'ensemble initial des probabilités antérieure peut être choisi librement (excepté pour 0 et 1).
- Mais comment peut-on critiquer une assignation étrange des probabilités antérieures, tant qu'elle suit les axiomes?
- réponse bayésienne: cela n'importe pas car l'ensemble initial des probabilités antérieures est évacué asymptotiquement (convergence, théorème de l'estimation stable)
- problème technique: conversement, on a aussi pour toute quantité de données, et toute mesure d'accord, un ensemble de probabilités antérieures telles que ces données ne vont pas finir par mettre les deux personnes d'accord (Henry Kyburg)
- Les conditions pour les théorèmes doivent être en place: par exemple, il doit y avoir un accord concernant les vraisemblances $P(E_i|H)$ et la pertinence de certaines données particulières.
- Les présuppositions des théorèmes ne s'appliquent même pas un petit peu dans des contextes réalistes scientifiquement.

(2) Explication et truismes méthodologiques (Glymour)

- La théorie de la confirmation devrait expliquer des truismes généraux méthodologiques ainsi que des jugements particuliers qui ont eu lieu dans l'histoire des sciences.
- Le bayésiannisme ne peut pas dire:
  - pourquoi une hypothèse est ad hoc,
  - pourquoi un ensemble de données est plus varié qu'un autre,
  - pourquoi on devrait préférer des données à l'appui plus variées,
  - pourquoi on devrait préférer les théories plus simples (e.g. problème de «curve-fitting» – ajuster la courbe)
- Généralement: Le bayésiannisme ne peut pas expliquer pourquoi on fait les projections qu'on fait et pourquoi des données diverses peuvent être pertinentes de manière différente (et pourquoi on pourrait être en désaccord sur l'attribution de cette pertinence).

(3) Problème de l'ancienne donnée (Glymour)

- Problème de l'ancienne donnée: D'anciennes données peuvent en fait confirmer une nouvelle théorie, mais d'après la cinématique bayésienne elles ne le peuvent pas.
- Supposez que E est connu avant que la théorie T ne soit introduite au temps t.
- Parce-que E est connu à t, $P_t(E) = 1$

⇒ la vraisemblance de T est aussi 1: $P_t(E|T) = 1$

$$P_t(T|E) = \frac{P_t(T) \cdot P_t(E|T)}{P_t(E)} = P_t(T)$$

⇒ La probabilité postérieure de T est la même que sa probabilité antérieure!

En guise de conclusion

Hájek and Hartmann, «Bayesian epistemology», *A Companion to Epistemology*, Blackwell, 2009.

Alan Hájek et Stephan Hartmann contrastent deux conceptions de l'épistémologie bayésienne:

> *Selon un point de vue, il ne peut pas y avoir d'épistémologie bayésienne: le bayésiannisme ne rend pas justice aux aspects essentiels de la connaissance et de la croyance, et de ce fait il ne peut fournir une vraie épistémologie. Selon un autre point de vue, le bayésiannisme devrait supplanter l'épistémologie traditionnelle: alors que cette dernière a été murée dans des débats sans fin sur le scepticisme et la gettierologie, le bayésiannisme offre à l'épistémologie un programme de recherche. Nous allons défendre une conception plus modérée: le bayésiannisme peut éclairer divers problèmes ancestraux de l'épistémologie, sans pour autant les régler tous; et bien que le bayésiannisme ouvre de nouvelles aires de recherche fascinantes, il ne met point fin aux préoccupations centrales de l'épistémologie traditionnelle. (93, trad.)*