Le problème des données connues (*old evidence*)

Mikaël Cozic

IHPST (Paris I-ENS Ulm-CNRS)
GREGHEC (HEC-CNRS), Logiques de l'Agir (Besançon) & DEC (ENS Ulm)

PDI
1/II/2008

1. La théorie bayésienne de la confirmation

les théories de la confirmation

- un trait essentiel des sciences empiriques : l'évaluation des hypothèses/théories en fonction des données empiriques : des données favorables (resp. défavorables) à une hypothèse la *confirment* (resp. disconfirment)
- l'objectif des théories de la confirmation est de fournir "une approximation raisonnablement proche de la conception de la confirmation qui est implicite dans la procédure scientifique et dans la discussion méthodologique" (Hempel 1945)
- analogie avec la logique : codifier et systématiser nos intuitions et nos pratiques de raisonnement inductif (ampliatif)

le bayésianisme

- la théorie dominante est la théorie bayésienne de la confirmation (TBC). La TBC procède de l'épistémologie bayésienne :

1. **gradualisme** : une épistémologie adéquate doit considérer les *degrés* de croyance
2. **probabilisme** : les degrés de croyances d'un agent rationnel se laissent représenter par une fonction de probabilité (DB)
3. **conditionalisation** : les croyances d'un agent rationnel sont révisées par conditionalisation : la nouvelle probabilité P_{new} d'un agent qui apprend E et dont la probabilité initiale est P_{old} est
$P_{new}(H) = P_{old}(H|E) = P_{old}(H \wedge E)/P_{old}(E)$

la théorie bayésienne de la confirmation

- dans un cadre probabiliste également, Carnap (*Logical Foundations*, 1962) distingue deux notions de confirmation:

1. concept **absolu** de confirmation (*confirmation as firmness*): E a-confirme H ssi H est probable étant donné E: $P(H|E) > k$
2. concept **incrémental** de confirmation (*confirmation as increase in firmness*): E i-confirme H ssi E “augmente” la probabilité de H : $P(H|E) > P(H)$

- la TBC adopte le concept incrémental de confirmation. *Argument : apprendre E peut faire baisser la probabilité de H (i-disconfirmation) sans faire descendre cette probabilité sous le seuil k (a-confirmation).*

la théorie bayésienne de la confirmation

- Applications à la méthodologie des sciences empiriques :
  • P. Horwich, *Probability and Evidence*, 1982
  • C. Howson & P. Urbach, *Scientific Reasoning. The Bayesian Approach* (1989/2006)
  • J. Earman, *Bayes or Bust ?* (1992)
- Une partie du succès de la TBC provient du fait qu'elle propose de rendre compte de nombreux concepts de la méthodologie, de reconstruire des épisodes scientifiques et de "résoudre" des problèmes épistémologiques:
  • les vertus sans les vices de la théorie H-D
  • le paradoxe des corbeaux
  • la notion d'adhocité
  • la variété des données

théorème de Bayes

- le **théorème de Bayes** joue un rôle central dans ces applications:

$$P(H|E) = [P(E|H)/P(E)].P(H)$$

- La confirmation de H par E dépend de l'évolution de $P(H)$ qui dépend à son tour du "multiplicateur bayésien" (Strevens): $[P(E|H)/P(E)]$
- 2 conséquences immédiates pour la TBC:
  - *ceteris paribus*, plus E se laisse prédire par H, plus E confirme H (cas limites : $H \vDash E$ et $H \vDash \neg E$)
  - *ceteris paribus*, plus E est inattendu, plus E confirme H ("principe de surprise", Joyce)

mesures de confirmation

- le concept initial de confirmation est *qualitatif* ; on peut en donner une version quantitative basée sur une *mesure de confirmation*.
  $P(H|E) > P(H)$ ssi $d(H,E) = P(H|E) - P(H) > 0$
  $\Rightarrow$ prendre $d(.,.)$ comme mesure de la confirmation
- d'autres mesures de confirmation sont envisageables.
  $P(H|E) > P(H)$ ssi $P(H|E) > P(H|\neg E)$
  $\Rightarrow$ autre mesure, $s(H,E) = P(H|E) - P(H|\neg E)$
  Attention : quand $P(E) = 1$, $s(H,E)$ n'est pas nul (comme $d(H,E)$) mais *non défini*
- il existe encore d'autres mesures qui sont utilisées et la question des mérites respectifs est aujourd'hui largement débattue (voir Fitelson (thèse))

2. Le(s) problème(s) des données connues

Glymour contre la TBC

- Le PDC fait partie de la batterie d'objections que C.Glymour dans *Theory and Evidence* (1980) adresse à la TBC - chap III, "Why I Am Not a Bayesian" :

> *"Scientists commonly argue for their theories from evidence known long before the theories were introduced...The argument that Einstein gave in 1915 for his gravitational field equations was that they explained the anomalous advance of the perihelion of Mercury, established more than half a century earlier...Old evidence can in fact confirm new theory, but according to Bayesian kinematics it cannot."*

Einstein 1915

- Novembre 1915 : Einstein propose la Théorie Générale de la Relativité, notée H
  Le Verrier (1859) et Newcomb (1882) montrent qu'il y a un écart entre l'avance du périhélie de Mercure observée et celle prédite par la théorie newtonienne
- E est connu en 1915 donc $P_{1915}(E) = 1$
  $P_{1905}(H|E) = P_{1905}(H) \times [P_{1905}(E|H)/P_{1905}(E)] = P_t(H)$

  (le multiplicateur bayésien vaut nécessairement 1)
- Conclusion : la donnée E ne confirme (ni ne disconfirme) l'hypothèse H
- Problème : l'avance du périhélie de Mercure fut considéré comme l'argument principal en faveur de la TGR

le PDC et autres phénomènes apparentés

- le **PDC qualitatif**
  - si la donnée E est connue, elle ne confirme ni ne disconfirme H
  - si les données E et E' sont connues, alors elles ont la même relation confirmationnelle à H (y compris si $H \vDash E$ et $\neg H \vDash E'$)
  - si la donnée E est connue, alors elle a la même relation confirmationnelle à H et H' (y compris si $H \vDash E$ et $\neg H' \vDash E$)
- le **PDC quantitatif**
  - si $P(E) > 1 - \epsilon$, alors $-\epsilon < d(H,E) < \epsilon$

quel problème ?

- (presque) tout le monde reconnaît que le PDC est un problème très sérieux pour la TBC et que sa version standard est inadéquate :

  > *"It is now generally acknowledged that the simple Bayesian analysis of confirmation is inadequate, for a reason that Glymour (1980) drew attention to" (Maher 1996)*

- le problème est sérieux notamment parce qu'il concerne l'endroit où la TBC est censé être la plus performante, i.e. dans l'explication des relations entre théories et données
- un exemple comme celui d'Einstein 1905 est cependant complexe et combine différentes difficultés...

cacophonie de taxinomies, I

- Eells (1985, 1990) distingue
- Problem of **Old New Evidence** :
  • en t, l'agent apprend E qui augmente la probabilité de H :
  $P_t(H) = P_{t-1}(H|E) > P_{t-1}(H)$
  $(P_t(E) = 1 > P_{t-1}(E))$
  • pour tout $t' > t$, $P_{t'}(E) = 1$ donc $d_{t'}(H, E) = 0$
  Une donnée ne (dis)confirme qu'une seule fois ! - un carburant doxastique qui se consomme
- Problem of **Old Old Evidence** :
  • E est appris en $t' < t^*$ donc $\forall t \geq t'$, $P_t(E) = 1$
  • H est formulée en t^*
  • pour tout $t'' > t^*$, $P_{t''}(E) = 1$ donc $d_{t''}(H, E) = 0$

cacophonie de taxinomies, II

- Problem of **New Old Evidence** :
  - E est appris en $t' < t^*$ donc $\forall t \geq t'$, $P_t(E) = 1$
  - H est formulée en t^*
  - en t^*, $P_{t^*}(E) = 1$ donc $d_{t^*}(H, E) = 0$
- Christensen (1999) distingue le PDC **synchronique** du PDC **diachronique**
- PDC diachronique : il semble que l'on se sert de données connues pour augmenter sa confiance en une hypothèse
- PDC synchronique : un fort degré de croyance en E rend E incapable de fournir un support confirmationnel à une hypothèse H

cacophonie de taxinomies, III

- Joyce (1999) distingue
  - **Problem of New Hypothesis** : How should a rational agent assign an *initial* subjective probability to a newly formulated hypothesis in the light of the evidence she already has ?
  - **Problem of Logical Learning** : How should a less than ideally rational agent revise her *existing* subjective probability for a hypothesis when she discovers the logical fact that it *entails* some piece of evidence ?
  - **Problem of Evidential Relevance** : How can a rational agent ... make sense of evidential relationships between the hypothesis and evidence to which she assigns a subjective probability of 1 ?

- pour le reste de l'exposé, je vais essentiellement suivre la distinction de Christensen 1999 en utilisant la terminologie suivante, plus explicite :

- **problème de l'incrément** de confiance que peut conférer une donnée connue E en une hypothèse H
- **problème de la survie** confirmationnelle d'une donnée à son apprentissage

les prémisses

- prémisses principales :

(P1) probabilisme standard : l'état épistémique d'un agent rationnel en t se laisse caractériser par une distribution de probabilité $P_t(.)$ *qui obéit aux axiomes de Kolmogorov*

(P2) actualisme : les jugements confirmationnels d'un agent rationnel en t *surviennent* sur son état épistémique en t

(P3) différence : E confirme H en t ssi $P_t(H|E) > P_t(H)$ ssi $d_t(H, E) = P_t(H|E) - P_t(H) > 0$

(P4) omniscience logique

(P5) certitude : Si un agent rationnel sait que E en t, alors $P_t(E) = 1$

quelques propositions

(A1) **bayésianisme faillible** (rejette (P5))

(A2) **bayésianisme limité** (rejette (P4) et éventuellement (P1)) - Garber (1983), Jeffrey (1983), Eells (1990)

(A3) **bayésianisme historique** (rejette (P2) et modifie (P3)) - Garber (1983)

(A4) **bayésianisme contrefactuel** (rejette (P2) et modifie (P3)) - Howson (1984, 1991), Maher (1996)*

(A5) **bayésianisme re-mesuré** (rejette (P3) et éventuellement (P1)) - Christensen (1999), Joyce (1999)

3. Les principales solutions

- selon la prémisse (P5), $P(E) = 1$. Or, on peut contester qu'il soit possible, voire rationnel, d'attribuer une probabilité maximale (la même que celle attribuée aux vérités logiques) à des énoncés empiriques.
- si la probabilité de E est simplement très forte, alors il peut y avoir confirmation de H par E - au sens du critère incrémental de la TBC.
- solution au problème de la survie, mais pas au problème de l'incrément
- solution qui ne résiste pas à la version quantitative du PDC: si $P(E) > 1 - \epsilon$, alors $-\epsilon < d(H, E) < \epsilon$

données connues et omniscience logique

- le “bayésianisme limité” est une approche populaire défendue et élaborée principalement par Garber (1983) et Jeffrey (1983) - ce n’est pas la seule approche dévéloppée par Jeffrey.
- idée de départ : dans un épisode comme Einstein 1905, c’est le défaut (initial) d’omniscience logique qui explique pourquoi il y a incrément de confiance dans la TGR. Ce qu’Einstein apprend, ce n’est pas E mais $H \vDash E$
  Garber :
  “...in the case at hand, what *increases* S’s confidence in H is not E itself, but the *discovery* of some generally logical or mathematical relationship between H and E.”

PDC et OL

- Garber extrapole et considère que le PDC est un problème entièrement induit par l'OL:
  "The historical problem of old evidence, then, seems to be a consequence of the fact that the Bayesian framework is a theory of reasoning for a logically omniscient being."
- selon lui, il faut rétablir une symétrie entre l'apprentissage empirique et l'apprentissage logique

modélisation

- les probabilités sur des espaces non-standards permettent de rétablir cette symétrie et de représenter l'apprentissage logique
- apprendre que $H \vdash E$, c'est éliminer les états (impossibles) où H est satisfait mais où E ne l'est pas et renormaliser - voir Cozic (2007) pour les vertus de cette approche
- problème immédiat : si $P(E) = 1$, le modèle est impuissant
- remarque : dans ce modèle, apprendre que H implique E ne peut que faire baisser la probabilité de H ! - intuitivement, cela ne peut faire que diminuer la proportion des H-mondes
  moralité : je n'aurais pas dû m'intéresser au PDC

modèle de Garber

- langage propositionnel *L*
- langage enrichi $L*$ qui contient comme formules atomiques (i) les formules atomiques de *L* et (ii) $(\phi \vdash \psi)$ pour toutes formules (quelconques) ϕ, ψ de *L*

  contrainte extra-systématique (G) :

  $$\text{(G) } P(\phi \wedge (\phi \vdash \psi)) = P(\phi \wedge \psi \wedge (\phi \vdash \psi))$$

- remarques : (i) *G* est équivalent à $P(\phi \wedge \psi \wedge \neg(\phi \vdash \psi)) = 0$; (ii) condition suffisante : $[[\phi]] \cap [[\phi \vdash \psi]] \subseteq [[\psi]]$; (iii) conséquences :
  - $P_{H \vdash E}(H \wedge E) = P_{H \vdash E}(H)$
  - $P_{H \vdash E}(H) \leq P_{H \vdash E}(E)$

  (si $H \vDash E$, c'est déjà le cas !!)

- **Proposition (Non trivialité)** : Il existe P sur $L*$ tq pour toute formule $\phi \vdash \psi$ où ni ϕ ni ψ ne sont des tautologies, $0 < P(\phi \vdash \psi) < 1$
  La Proposition de Non-Trivialité ouvre la voie à une confirmation de H par l'apprentissage de $H \vdash E$
- **Proposition** : Pour toute formule $\phi \vdash \psi$ où (a) ψ n'est pas une antilogie, (b) non $\phi \vDash_L \neg\psi$ et (c) non $\psi \vDash_L \phi$ et pour tous $r, s \in (0, 1)$, il existe une infinité de distributions de probabilité sur $L*$ qui satisfont (G) et tq
  (A1) $P(\psi) = 1$
  (A2) $P(\phi \vdash \psi) = r$
  (A3) $P(\phi) = s$
  (C) $P(\phi|\phi \vdash \psi) > P(\phi)$

Jeffrey

- Question : quelles conditions peuvent conduire à une confirmation de H par E connue ?
- Réponse 1 : Jeffrey (1983) :

  • (A1)-(A3)
  • (G*) : $P(H \wedge (H \vdash \neg E)) = P(H \wedge (H \vdash \neg E) \wedge \neg E)$
  • (J1) $P((H \vdash E) \wedge (H \vdash \neg E)) = 0$ (confiance totale dans la cohérence de l'hypothèse)
  • (J2) $P(H|(H \vdash E) \vee (H \vdash \neg E)) \geq P(H)$ (apprendre décidabilité implique confirmation)

Earman

- Réponse 2 : Earman (1992) :
  - (A1)-(A3)
  - (G*) : $P(H \wedge (H \vdash \neg E)) = P(H \wedge (H \vdash \neg E) \wedge \neg E)$
  - (E) $P((H \vdash E) \vee (H \vdash \neg E)) = 1$ (certitude de la décidabilité)
  - (E') $P(H | H \vdash E) > P(H | \neg(H \vdash E) \wedge \neg(H \vdash \neg E))$

objections au bayésianisme limité

- objection gén .1 : le modèle de Garber est étrange !
- objection gén.2 : le bayésianisme limité explique au mieux comment $H \vDash E$ confirme H, pas comment E confirme H
- objection gén. 3 (Zynda 1995) : le bayésianisme limité ne couvre pas les cas où H n'implique pas E
- objection gén.4 (Earman 1992, Christensen 1999) : le bayésianisme limité ne traite que du problème de l'incrément pas du problème de la survie

bayésianisme historique

- Problem of **Old New Evidence** :
  • en t, l'agent apprend E ce qui augmente la probabilité de H :
  $P_t(H) = P_{t-1}(H|E) > P_{t-1}(H)$
  $(P_t(E) = 1 > P_{t-1}(E))$
  • pour tout $t' > t$, $P_{t'}(E) = 1$ donc $d_{t'}(H, E) = 0$
- Eells (1990) introduit un nouveau concept pour représenter le fait que E reste une "évidence" en faveur de H : en $t' > t$, le **support** conféré à H par E, $supp_{t'}(H, E)$ est la (dis)confirmation qui a eu lieu en t soit $d_t(H, E)$.

- le bayésianisme historique rejette la prémisse actualiste (P2) : la relation de support ne survient pas sur P_t
- trois choses à distinguer :
  • la relation de confirmation de H par E (en un t) : $d_t(H, E) > 0$
  • l'évément de confirmation de H par E (en un t) : E est appris en t et $d_{t-1}(H, E) > 0$
  • le support de H par E (en un t) : au moment où il a été appris, E a confirmé (au sens de l'événement) H

objections au bayésianisme historique

- objection de Christensen (1999) (et Maher 1996):
  H = il y a un cerf dans le bois de Vincennes ; E_1 = il y a des excréments de cerf dans le bois de Vincennes ; E_2 = il y a des ramures de cerf dans le bois de Vincennes
  à t_1, l'agent apprend E_1 et $P_{t_1}(H) = P_{t_0}(H|E_1) >> P_{t_0}(H)$
  à t_2, l'agent apprend E_2 et $P_{t_2}(H) = P_{t_1}(H|E_2) \approx_+ P_{t_1}(H)$
- intuitivement, E_1 et E_2 apportent toutes les deux un "support" important à H. Ici, on observe (i) dépendance du support à l'égard du déroulement de l'apprentissage et (ii) dissymétrie entre les deux données
- de manière générale, l'incrément produit par E sur H dépend du reste des croyances de l'agent

bayésianisme contrefactuel

- stratégie contrefactuelle (Howson 1984, 1991 , Maher 1996) : E confirme H ssi, si E n'avait pas été connu, alors $P(H|E) > P(H)$
  Howson (1991) : P est paramétré par les croyances d'arrière-plan K ; quand on juge de la confirmation de H par E i.e. de la différence $P(H|E) - P(H)$, il faut prendre comme croyances d'arrière-plan $K - \{E\}$.
  "When you ask yourself how much support E gives H, you are plausibly asking how much a knowledge of E would increase the credibility of H, which is the same thing as asking how much E boosts the credibility of H relative to what else you currently know. The 'what else' is just $K - \{E\}$." (Howson 1991)

objection au bayésianisme contrefactuel

- remarque : le bayésianisme peut être une solution au problème de la survie, pas à celui de l'incrément
- objection 1 : comment déterminer $P_c(.)$? (Chiara (1987) : "...if this sort of defence of Bayesianism is to work, one needs a reasonably clear procedure or rule for coming up with the required stock of background information.")
- objection 2 (Maher 1996) : il se peut que, bien que l'on juge que E confirme H, l'état épistémique contrefactuel où l'on ne sait pas que E est tel que $d(H, E) < 0$

objections, cont.

- exemple : M.Schreiber
  H : je suis un auteur de romans à succès
  E : j'écris des romans de qualité
  M.Schreiber peut juger que H confirme I mais se dire que s'il ne pensait pas qu'il est un auteur à succès, il ne prendrait pas la popularité comme gage de qualité littéraire
- réponse de Maher (1996) : il faut que l'état épistémique contrefactuel soit rationnel du point de vue actuel ; l'état épistémique contrefactuel de M.Schreiber n'est pas rationnel - biaisé par l'amertume
- objection 3 : quid de la version quantitative du PDC ?

bayésianisme re-mesuré

- solution récemment avancée par Christensen 1999 et Joyce 1999
- idée : dans le critère bayésien de confirmation et dans la mesure $d(H,E)$ sont pris en compte à la fois le "lien" entre H et E et la distance à parcourir pour que la probabilité de E parvienne à 1. Il faudrait *factoriser* ces deux paramètres de manière à ne prendre en compte que le premier. Christensen propose de contrôler la probabilité initiale de E en normalisant la différence $d(H,E)$: $s(H,E) = d(H,E)/P(\neg E) = P(H|E) - P(H|\neg E)$ Remarque : il n'est pas vrai que si $P(E|H) = P(E'|H)$, $s(H,E) = s(H,E')$

bayésianisme re-mesuré

- *une* forme d'indépendance par rapport à $P(E)$: $s(.,.)$ est stable par application de la **règle de Jeffrey** à $\{E, \neg E\}$ (la règle de Jeffrey satisfait en effet une propriété de rigidité : $P_{old}(.|E) = P_{new}(.|E)$)
- **Proposition** Si $P_{old}(E) > 0$ et $0 < P_{new}(E) < 1$, alors
  - $d_{P_{new}}(H, E) = d_{P_{old}}(H, E).[P_{new}(\neg E)/P_{new}(E)]$
  - $s_{P_{new}}(H, E) = s_{P_{old}}(H, E)$

- on peut voir l'apprentissage de E comme un processus continu "à la Jeffrey" sur l'intervalle $[0, 1]$. Dans ce cas,
  • $\lim_{t\to 1} s_{P_t}(H, E) = s(H, E)$
  • $\lim_{t\to 1} d_{P_t}(H, E) = 0$
- Problème : avec la conditionalisation, on ne peut pas traiter $s(.,.)$ pour $P^*(.) = P_1(.)$ puisque $P^*(\neg E) = 0$
- Proposition : recours aux mesures de Rényi-Popper

- une mesure de Réyni-Popper est une fonction $P(.|.)$ tq
  $P(.||\phi)$ est mesure de probabilité
  $P(.||\top) = P(.)$
  $P(\phi||\phi) = 1$
  $P(\chi||\phi \wedge \psi).P(\phi||\psi) = P(\chi \wedge \phi||\psi)$
- application à la mesure $s(.,.)$ quand $P^*(C) = 1$:
  $P^*(H||\neg E) = \lim_{t \to 1} P_t P_t(H|\neg E)$

objections

- objection (Earman 1992) :
  quand $P(E) = 1$ et $H \vDash E$, alors $P(H|\neg E) = 0$ et $P(H|E) = 1$ donc $s(H, E) = P(H)$
  cela implique que si E et E' sont deux données connues et qui sont impliquées par H, alors elles ont le même pouvoir confirmationnel vis-à-vis de H
- objection (Christensen 1999) :
  "...even measure s...does not succeed in matching our intuitive quantitative confirmation judgments. s is best seen not as a correct account of confirmation, but as a useful tool for understanding confirmation."

objections

- objection (Eells & Fitelson 2000) :
  $s(.,.)$ (contrairement à $d(.,.)$) ne respecte pas la propriété suivante
  Si $P(H|E) > P(H|E')$, alors $c(H,E) > c(H,E')$
  propriété intuitive et qui sert à la majorité des résolutions bayésiennes du paradoxe des corbeaux