


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

Probabilités Bayésiennes

Théorie

Applications

Philosophie

Plan

- **Rappel et notations**
 - *Axiomes des probabilités*
 - *Indépendance, probabilités conditionnelles*
- **Probabilités bayésiennes**
 - *Le théorème de Bayes*
 - *Application à la recherche des causes*
- **Exemples d'applications**
 - *Test de diagnostic médical*
 - *Quel poison ai-je avalé ?*
 - *Une « preuve » de l'existence de Dieu*
 - *Echo Sonar ambigu dans un sous-marin*
 - *Interprétation de mesures d'exploration pétrolière*
- **Le Cerveau Bayésien**
 - *Définition*
 - *Exemples*
 - *Implémentations cérébrales (discussion)*
- **Décision bayésienne valorisée**
 - *Principes, exemples, discussion*
- **Compléments : Philosophie des probabilités**
 - *Qu'est-ce qu'une probabilité ?*
 - *Nature de l'incertitude*
 - *Interprétations quantitatives des probabilités*
 - *Interprétation des petites probabilités*

Axiomes de Kolmogorov



Définition : *espace probabilisé* $(\Omega, \mathcal{F}, \mathbb{P})$

Ω : ensemble des « événements élémentaires » $\omega \in \Omega$

$\mathcal{F}$: ensemble non vide de *parties de Ω* formant une « tribu » (stable par union dénombrable et complémentation) (d'où $\Omega \in \mathcal{F}$ et $\emptyset \in \mathcal{F}$)

➤ Chaque $A \in \mathcal{F}$ est appelé un « *événement* »

$\mathbb{P}$: fonction de $\mathcal{F}$ dans $[0, 1]$ satisfaisant aux axiomes suivants (*Kolmogorov*) :

- $\mathbb{P}(\Omega) = 1$
- Pour toute suite *finie* ou *dénombrable* $(A_1, A_2, \dots)$ d'éléments de $\mathcal{F}$, *disjoints 2 à 2*, on a :

$$\mathbb{P}(A_1 \cup A_2 \cup \dots) = \sum_{i \geq 1} \mathbb{P}(A_i)$$

D'où $\mathbb{P}(\emptyset) = 0$; $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$ pour tout $A \in \mathcal{F}$

Indépendance

- Indépendance de 2 événements A et B :

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B) \quad \text{Pourquoi parle-t-on d'indépendance ?}$$

- Indépendance de N événements $A_1 \dots A_N$:

Pour toute partie $I = \{i_1, i_2, \dots, i_k\}$ de $\{1, \dots, N\}$, on a :

$$\mathbb{P}\left[\bigcap_{i \in I} A_i\right] = \prod_{i \in I} \mathbb{P}(A_i)$$

- NB** : Indépendance 2 à 2 $\nRightarrow$ indépendance pour N

- Exemple: on tire deux fois à Pile ou Face et on pose:
 - A_1 : Pile au 1^{er} lancer,
 - A_2 : Face au 2^{ème} lancer,
 - A_3 : On a le même côté aux deux lancers.

Probabilités conditionnelles

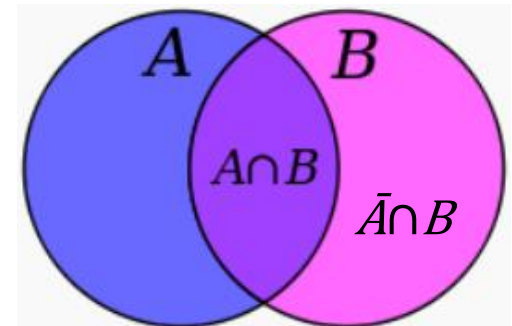
$$\mathbb{P}(A|B) \stackrel{\text{def}}{=} \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad (\text{si } \mathbb{P}(B) \neq 0)$$

Se lit : « Probabilité de A sachant B », ou :
« Probabilité de A étant donné B », etc.

Justification

- Si l'on connaît B , les seuls cas possibles sont $A \cap B$ et $\bar{A} \cap B$
- La probabilité $\mathbb{P}(A|B)$ est le ratio du cas favorable $A \cap B$ sur l'ensemble des cas possibles :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}[(A \cap B) \cup (\bar{A} \cap B)]} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$



Indépendance : justification

Pourquoi la formule $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$
représente-t-elle l'*indépendance* de A et de B ?

D'après les définitions des probabilités conditionnelles
et de l'indépendance, l'indépendance de A et B est équivalente à :

$$\mathbb{P}(A|B) = \mathbb{P}(A)$$

En effet,

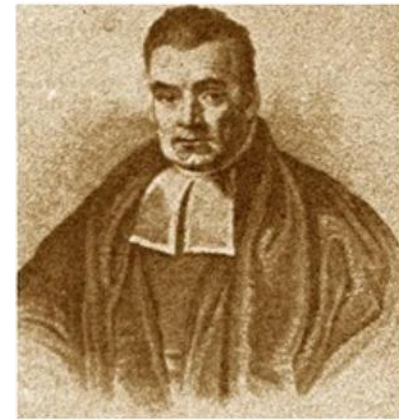
$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A)$$

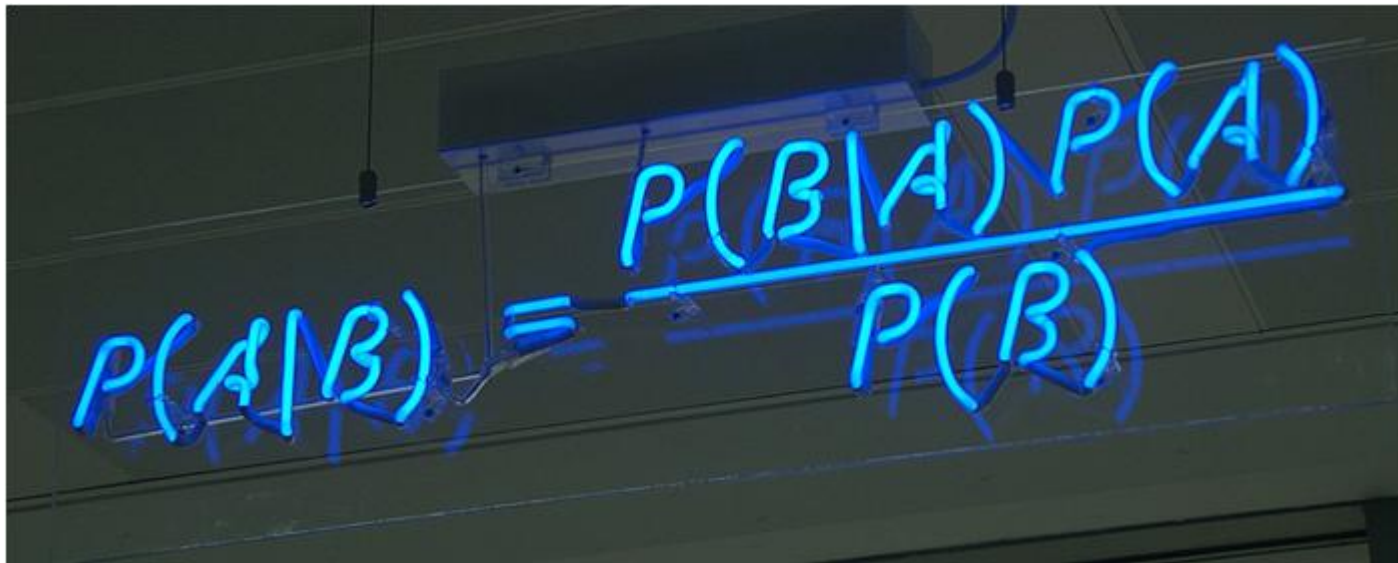
→ L'information apportée par la connaissance de B *ne modifie en rien la probabilité* de réaliser A : il y a bien *indépendance au sens commun*

Théorème de Bayes

Révérant Thomas Bayes (~1701-1761), Pasteur de l'Église presbytérienne et mathématicien britannique

« *Essai sur la manière de résoudre un problème dans la doctrine des risques* »



A photograph of a chalkboard with the formula for Bayes' Theorem written in blue chalk. The formula is $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$. The chalkboard is dark, and the blue chalk is clearly visible.

Démonstration

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \text{ et } \mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$$
$$\Rightarrow \frac{\mathbb{P}(A|B)}{\mathbb{P}(B|A)} = \frac{\mathbb{P}(A)}{\mathbb{P}(B)} \Leftrightarrow \text{Formule de Bayes}$$

Intérêt de cette formule

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

- Quel est l'utilité de cette formule ?
 - Elle permet de calculer $\mathbb{P}(A | B)$ lorsque les conditions du problème rendent *plus facile* le calcul des autres facteurs, à droite du signe =
- Quand se présente cette situation ?
 - Essentiellement pour calculer la *probabilité des causes possibles* A_k d'un *phénomène observable* B
 - *Observation* $\rightarrow$ *Cause* : aucune déduction naturelle $B \rightarrow A_k$
 - *Cause* $\rightarrow$ *Observation* : « modèle direct » de causalité $A_k \rightarrow B$

Recherche de causes

- Soit O une *observation* ou un ensemble de *mesures*
 - Fumerolles, grondements et odeur de soufre sur un volcan
 - Ensemble de mesures dans un puits de pétrole
 - Ensemble de symptômes et d'analyses sur un patient
 - Observation lointaine d'une silhouette
- Soient $H_1, \dots, H_n$ les causes *hypothétiques* envisagées de O
 - Le volcan prépare une éruption (ou non)
 - Y a-t-il du pétrole exploitable ?
 - De quel maladie souffre-t-il ? Quel traitement appliquer ?
 - Est-ce Paul ou Jean ?
- Pour chaque hypothèse H_k , le théorème de Bayes donne :

$$\mathbb{P}(H_k | O) = \frac{\mathbb{P}(O | H_k) \mathbb{P}(H_k)}{\mathbb{P}(O)}$$

Terminologie bayésienne usuelle

$$\mathbb{P}(H_k | O) = \frac{\mathbb{P}(O | H_k) \mathbb{P}(H_k)}{\mathbb{P}(O)}$$

- $\mathbb{P}(H_k | O)$: Probabilité *a posteriori* de la cause de l'observation (*posterior*) : résultat recherché
- $\mathbb{P}(H_k)$: Probabilité *a priori* de la cause H_k (*prior*), indépendamment de l'observation (proba souvent *subjective*)
- $\mathbb{P}(O | H_k)$: Probabilité de l'observation si sa cause est H_k
 - Résulte d'un *modèle causal* $H_k \rightarrow O$ (*modèle direct*)
- $\mathbb{P}(O)$: voir plus loin « Calcul du dénominateur »

Décision bayésienne (non pondérée)

$$\mathbb{P}(H_k|O) = \frac{\mathbb{P}(O|H_k)\mathbb{P}(H_k)}{\mathbb{P}(O)}$$

On retient l'hypothèse H_k qui *maximise* la probabilité $\mathbb{P}(H_k|O)$:

C'est l'hypothèse *la plus probable* (*maximum likelihood*)

- NB 1 : Il faut parfois tenir compte des *coûts* ou *bénéfices* des *décisions* associées à chaque choix*
- NB 2 : L'*estimation directe* de $\mathbb{P}(O)$ est en général *impossible* :
on applique alors :

$$\mathbb{P}(O) = \sum \mathbb{P}(O|H_k) \mathbb{P}(H_k)$$

Calcul du dénominateur $\mathbb{P}(O)$

$$\mathbb{P}(H_k | O) = \frac{\mathbb{P}(O | H_k) \mathbb{P}(H_k)}{\mathbb{P}(O)}$$

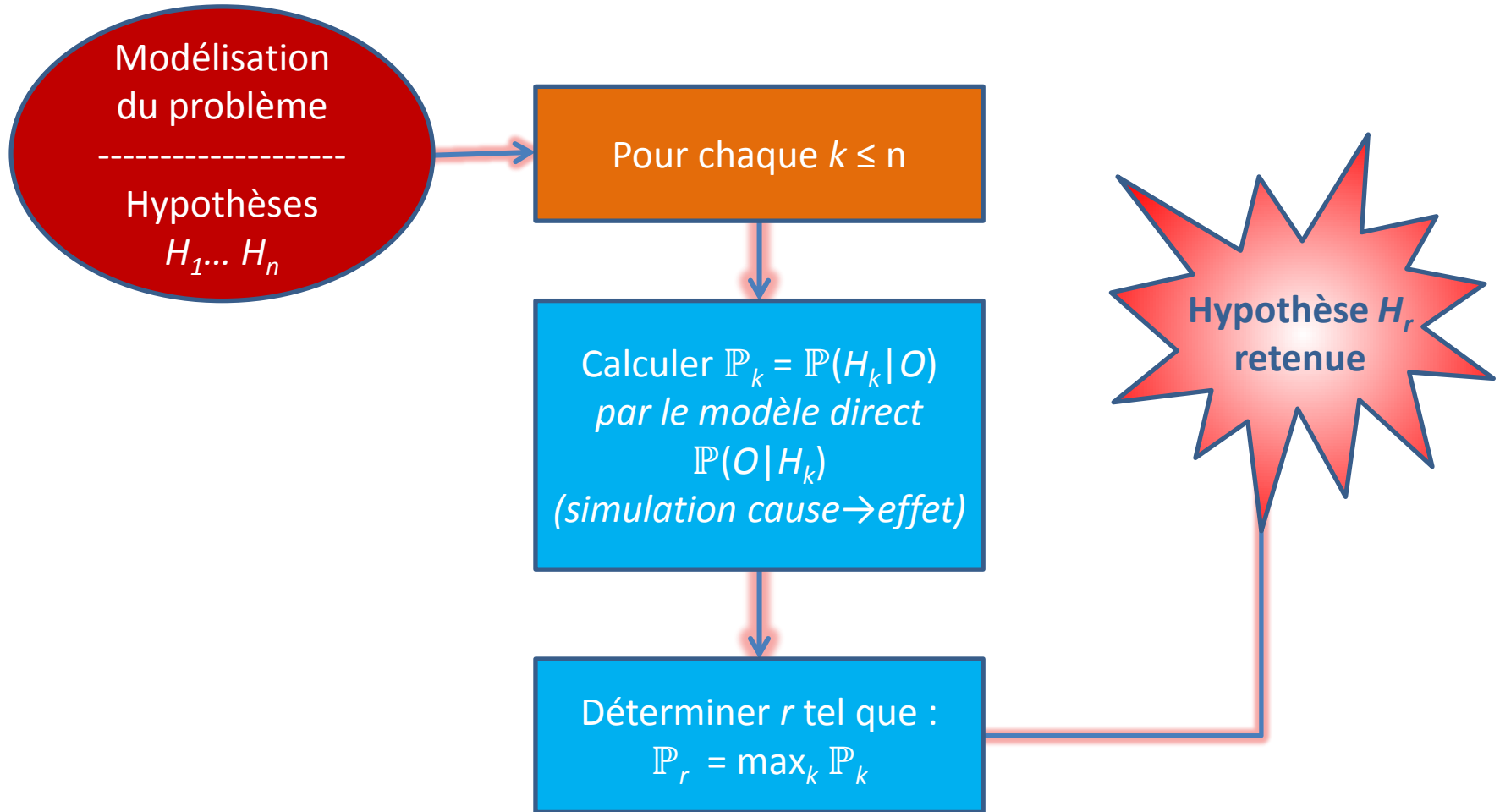
$$\mathbb{P}(O) = \sum \mathbb{P}(O | H_k) \mathbb{P}(H_k)$$

- Cette seconde formule (des « *probabilités totales* ») exige que les hypothèses H_k ($k \leq n$) constituent une *partition* de Ω :
 1. H_k disjoints 2 à 2 (les hypothèses sont clairement distinctes)
 2. Réunion des H_k = tout l'espace Ω
- Si la condition 2 n'est pas réalisée, on crée une *nouvelle hypothèse* H_{n+1} qui complète les hypothèses existantes
 - Mais il faut alors savoir évaluer $\mathbb{P}(H_{n+1})$ et $\mathbb{P}(O | H_{n+1})$
- Si le seul but est de *comparer les hypothèses* sans en calculer la probabilité, afin de *décider la plus probable*, cette comparaison *ne requiert pas* le calcul du dénominateur commun $\mathbb{P}(O)$

Approche bayésienne : *essais/erreurs, bottom-up...*

- **Processus essais-erreurs**
 - On propose différentes hypothèses, on les teste et on retient la meilleure (ici, la *plus probable*)
- **Bottom-up, systèmes inverses**
 - Au lieu de tenter l'impossible approche directe *Observation* → *Cause* (*top down*), on calcule plus facilement l'inverse *Cause* → *Observation* (*bottom-up*)
- **Modélisation, simulation**
 - Pour évaluer l'effet d'une *cause hypothétique* H , on a souvent besoin de *modéliser* et/ou de *simuler* le comportement du système afin d'obtenir l'*observation qui résulterait de H* et en déduire $\mathbb{P}(O|H)$

Organigramme de la décision bayésienne



Exemples d'applications

Test de diagnostic médical



- Pour un patient testé :
 - Soit P l'observation : le test est positif (si négatif : $\bar{P}$)
 - Soit M l'hypothèse : le sujet est malade (si non malade : $\bar{M}$)
- Caractéristiques *connues* et *fournies avec le test* :
 - *Sensibilité* (Se) : probabilité pour qu'un test réalisé sur une personne *malade* se révèle *positif* $Se = \mathbb{P}(P|M)$
 - *Spécificité* (Sp) : probabilité pour qu'un test réalisé sur une personne *saine* se révèle *négatif* $Sp = \mathbb{P}(\bar{P}|\bar{M})$
- Pour un sujet donné, *caractéristiques recherchées* :
 - *Valeur prédictive positive* (VPP) : probabilité pour que le patient, dont le test est positif, soit effectivement malade $VPP = \mathbb{P}(M|P)$
 - *Valeur prédictive négative* (VPN) : probabilité pour que le patient, dont le test est négatif, ne soit pas malade $VPN = \mathbb{P}(\bar{M}|\bar{P})$

Test de diagnostic (suite)

- Une donnée nécessaire au calcul bayésien :
 - *Prévalence* (p) : proportion observée de la maladie *dans la population considérée* (nombre de malades pour 1000 personnes de même catégorie – *par ex : fumeurs de moins de 50 ans*)

- La formule de Bayes nous donne pour $VPP = \mathbb{P}(M|P)$:

$$\mathbb{P}(M|P) = \frac{\mathbb{P}(P|M)\mathbb{P}(M)}{\mathbb{P}(P|M)\mathbb{P}(M) + \mathbb{P}(P|\bar{M})\mathbb{P}(\bar{M})}$$

- Soit, avec les notations précédentes :

$$VPP = \frac{Se.p}{Se.p + (1 - Sp)(1 - p)}$$

Se nsibilité
Sp écificité
 p révalence

- Remarques
 - Il est impossible de « *deviner* » directement VPP sans cette formule
 - Aucune des probabilités ci-dessus n'est *subjective*

Application numérique

$$VPP = \frac{Se.p}{Se.p + (1 - Sp)(1 - p)}$$

Se nsibilité
Sp écificité
p révalence

- Confusion répandue entre *Se* et *VPP*
 - *Problème posé à 160 gynécologues** :

On leur a demandé d'estimer la **probabilité pour qu'une femme soit atteinte d'un cancer du sein** en sachant qu'elle avait un **test de dépistage positif**. Il était précisé qu'elle appartenait à une population dans laquelle la **prévalence de ce cancer était de 1%**, et que le test de dépistage avait une **sensibilité et une spécificité de 90%**

– Réponses des gynécologues

- Seuls 1/5 ont donné la bonne réponse (< 10% **)
- Près des 2/3 ont répondu 90% (**confusion** de *VPP* avec *Se*)



* Gigerenzer et al., *Helping doctors and patients make sense of health statistics* (2007)

** La réponse exacte est $VPP = 8,3\%$ – Si l'on portait p à 10% $\rightarrow VPP = 50\%$

Quel poison ai-je avalé ?



A



B



- J'appelle le **centre anti-poison** avec des symptômes précis S
- Après enquête, le centre retient **deux poisons possibles**, A et B. Il précise que mes symptômes relèvent de A avec une probabilité $\mathbb{P}(S|A) = 75\%$, ou de B avec $\mathbb{P}(S|B) = 25\%$ *
- De mon côté, j'estime **à l'inverse** qu'il est peu probable que j'aie avalé A [$\mathbb{P}(A) = 30\%$] , et plus vraisemblable que j'aie consommé B [$\mathbb{P}(B) = 70\%$]
- Bayes nous donne : $\mathbb{P}(A|S) = 56\%$, $\mathbb{P}(B|S) = 44\%$
- Les deux traitements sont **incompatibles**. Une erreur de traitement pourrait entraîner de **graves complications**
- **Aie ! Que faire ?**
 - Il faudra **valoriser** les risques
Cf. *Théorie de la décision valorisée*

Echo sonar ambigu dans un sous-marin

On ne sait pas interpréter *directement* l'écho O (ambigu)



Hypothèses : $\{H_k\} = \{\text{sous-marin, baleine, bruit}\}$

Chaque hypothèse H_k a une *signature* sonar type S_k , que l'on sait calculer en y affectant une *tolérance*

On compare chaque signature à O (notion de *distance* $d(S_k, O)$)

On en déduit $\mathbb{P}(O|H_k)$

Si *une baleine est peu probable* dans ces eaux, on estime une faible valeur pour $\mathbb{P}(H_2)$; on estime de même $\mathbb{P}(H_1)$ et $\mathbb{P}(H_3)$

On applique le théorème de Bayes pour chaque hypothèse

On retient *la plus vraisemblable*, sauf si l'on doit tenir compte des *risques* associés à chaque choix*



* Cf. plus loin « Théorie de la décision valorisée »

Dieu existe-t-il ?



- Définissons Dieu (D) comme le **créateur du monde tel que nous le connaissons** : D est éternel, volontaire, tout-puissant, au dessus des lois (de la physique)
- Attribuons une **probabilité subjective p a priori** pour qu'une telle entité puisse exister ; par exemple :
 - $p = 1$ (la **Foi**)
 - $p = 0,5$ (une attitude d'**ignorance ouverte**, « tolérante »)
 - $p = 10^{-1000}$ (une **très forte incrédulité**, un rejet)
- Quelle que soit la valeur retenue pour p , je vais **démontrer** de manière bayésienne que **Dieu existe... presque sûrement** !

Une « preuve » de l'existence de Dieu - 1

$D = \{Il \text{ existe un Dieu (Eternel, Créateur, tout - puissant)}\}$

$O = \{Observations: le Monde, la Vie, l'Homme, le libre arbitre ... \}$

Les observations étant ce qu'elles sont, quelle est la probabilité de D ?



$$\mathbb{P}(D|O) = \frac{\mathbb{P}(O|D)\mathbb{P}(D)}{\mathbb{P}(O|D)\mathbb{P}(D) + \mathbb{P}(O|\bar{D})\mathbb{P}(\bar{D})}$$

$\mathbb{P}(O|D) = d$ (un nombre proche de 1, car Dieu a voulu le Monde tel qu'il est, et Il en avait le pouvoir. Si $d < 1$, c'est que Dieu n'a pas été précis)

$\mathbb{P}(D) = p$ (on retiendra l'hypothèse basse $p = 10^{-1000} = 1/10^{1000}$)

Estimation de $\mathbb{P}(O|\bar{D}) = \varepsilon$:

ε est la probabilité d'obtenir notre monde *naturellement*, depuis le Big Bang, sans intervention divine ni « *intelligent design* »

Cette probabilité est *très faible*, car une *quasi-infinité d'autres mondes* seraient également possibles. Penrose l'a estimée* à : $\varepsilon \sim 1/10^{10^{123}}$

Une « preuve » de l'existence de Dieu - 2

$$\mathbb{P}(D|O) = \frac{\mathbb{P}(O|D)\mathbb{P}(D)}{\mathbb{P}(O|D)\mathbb{P}(D) + \mathbb{P}(O|\bar{D})\mathbb{P}(\bar{D})}$$

$$\text{Avec : } \mathbb{P}(O|D) = d \sim 1$$

$$\mathbb{P}(D) = p \sim 1/10^{1000}$$

$$\mathbb{P}(O|\bar{D}) = \varepsilon \sim 1/10^{10^{123}}$$

Il s'en suit :

$$\mathbb{P}(D|O) = \frac{d.p}{d.p + \varepsilon(1-p)} \simeq 1$$

Car $\varepsilon \ll d.p$ selon les hypothèses retenues

Donc $\mathbb{P}(D|O) \simeq 1$: *Dieu existe presque sûrement*

➤ CQFD

Commentaire

$$\mathbb{P}(D|O) = \frac{d.p}{d.p + \varepsilon(1-p)} \simeq 1$$

- Comment peut-on contester (ou réfuter) cet argument ?
- Selon la formule ci-dessus et l'estimation des paramètres, le *seul moyen* est de poser a priori $p = 0$
- On ne peut réfuter cette « preuve » qu'en déclarant d'emblée qu'il est *strictement impossible* que Dieu existe
- Si l'on admet *la plus petite possibilité* (comme $p = 10^{-1000}$), la formule de Bayes la transforme en *quasi-certitude*

Objection 1

- **En résumé** : Dieu existe, ou non, selon que l'on croie ou non, à tort ou à raison, à sa *possibilité a priori*
- On pourrait contester cette « démonstration » en disant qu'elle *ne s'applique qu'au Dieu et au Monde particuliers* sur lesquels elle porte
 - Elle laisserait ainsi place à diverses *réfutations* si on voulait l'appliquer à des dieux et des mondes *définis autrement*
- **Réponse** : tant que le dieu considéré a une *volonté* et un *pouvoir créateur*, la démonstration s'applique, en changeant les probabilités ε , d et p mais en *conservant la relation $\varepsilon \ll d \cdot p$* , suffisante pour la démonstration :
 - En effet, ε est quasi-négligeable *par nature*, car tout événement possible parmi une quasi-infinité d'autres est *imprévisible a priori* tout en restant parfaitement *vraisemblable* (l'un d'eux se produira)
 - En revanche, d est *proche de 1* (Dieu a voulu ce monde), et p est « *petit* », mais *non négligeable*, pour les agnostiques non résolument athées



Objection 2 (*plus subtile ?*)

- Lorsque la probabilité d'un événement est *infime*, cet événement peut être considéré, soit comme *imprévisible*, soit comme *invraisemblable*
- De même, lorsque cette probabilité est *nulle*, on peut considérer cet événement, soit comme *imprévisible* (mais *possible* et même *vraisemblable*), soit comme *impossible*
 - Par exemple, si on choisit un nombre rationnel au hasard entre 0 et 1, il est *imprévisible* mais *possible* de choisir 0,84576, mais *impossible* de choisir 2. Or ces deux événements avaient a priori *la même probabilité* nulle
 - Ces deux interprétations sont pourtant *fondamentalement différentes*
- Le monde tel qu'il a évolué à partir du Big Bang, était *imprévisible* a priori, mais *pas invraisemblable*. On peut à l'inverse considérer Dieu comme *invraisemblable*. Les probabilités *ne distinguent pas* ces deux cas fondamentalement différents de probabilités infimes
 - La démonstration proposée *confond* donc *deux interprétations différentes* des probabilités dans la *même formule*. Ce mélange n'a pas de sens bien défini
 - Les concepts de probabilités sont *incapables de distinguer* ces deux acceptions des petites probabilités et peuvent donner lieu à des *paradoxes* si l'on tente de les interpréter en termes de degrés de *possibilité* ou de *vraisemblance*



Objection 3 : *généralisations absurdes*

- Ce type de raisonnement peut s'étendre à toute théorie au fort pouvoir explicatif, *aussi absurde soit-elle*
 - Intervention de Saint-Antoine pour retrouver un objet perdu
 - Souffrance ou maladie expliquée par un sorcier maléfique torturant une poupée vaudou
 - Dans tous ces cas, si l'on admet la cause absurde comme possible, l'observation s'en déduit avec la plus grande probabilité

Il y a subjectif... et subjectif

- Les probabilités *subjectives* évoquées ici sont un constituant important de la philosophie bayésienne
- Une probabilité *subjective* [comme $\mathbb{P}(D)$] peut s'interpréter de plusieurs manières principales :
 - Un *certitude intérieure* forte, une *révélation*... ou le simple « *pifomètre* »
 - Le résultat d'un *raisonnement argumenté*, explicite, riche, mais non totalement probant (sans quoi la probabilité ne serait plus subjective)
- La *confiance* qu'on peut accorder à une probabilité bayésienne *a posteriori* doit tenir compte de la *qualité de l'évaluation des probabilités a priori*

Exploration pétrolière

Principes du « logging »

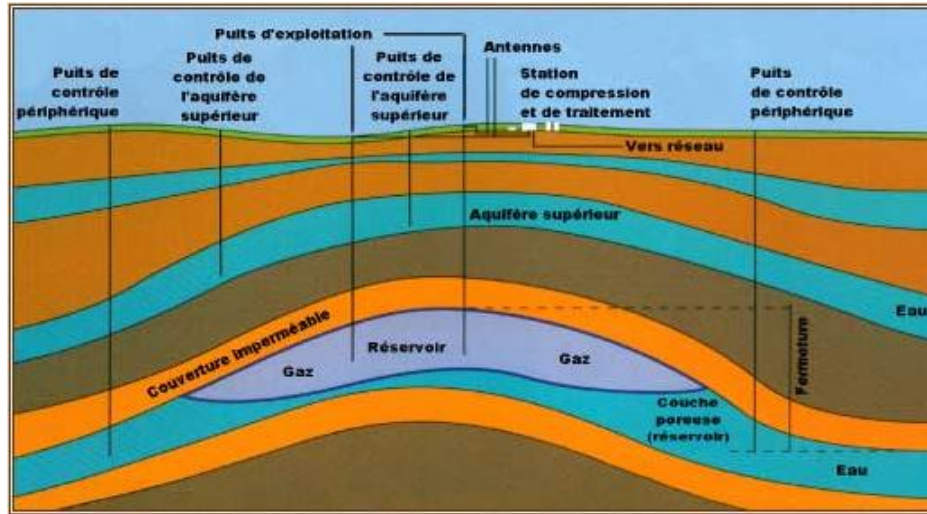
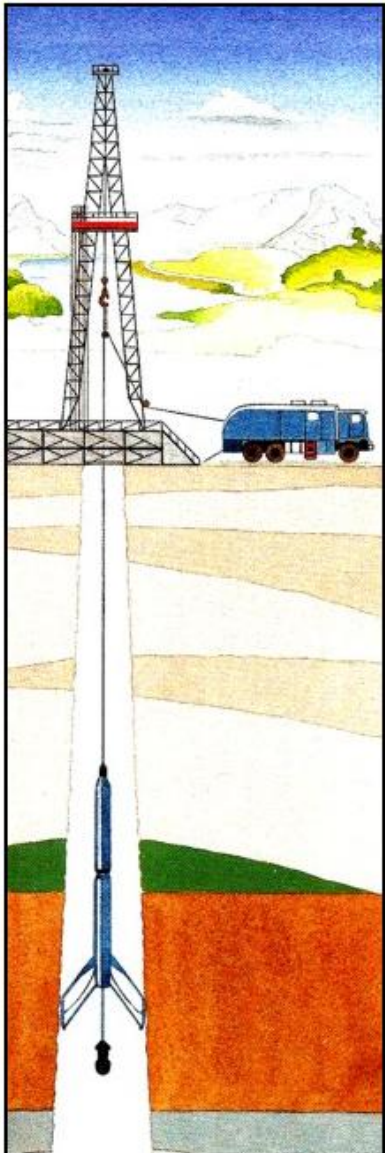
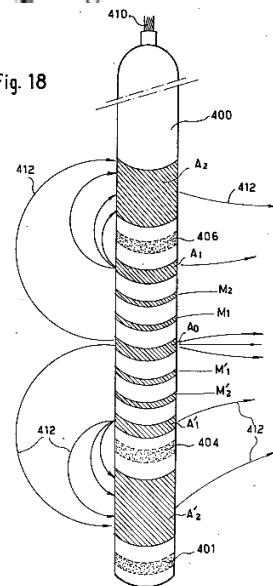


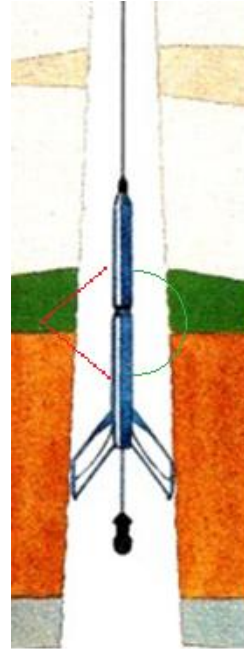
Fig. 18



Exploration pétrolière

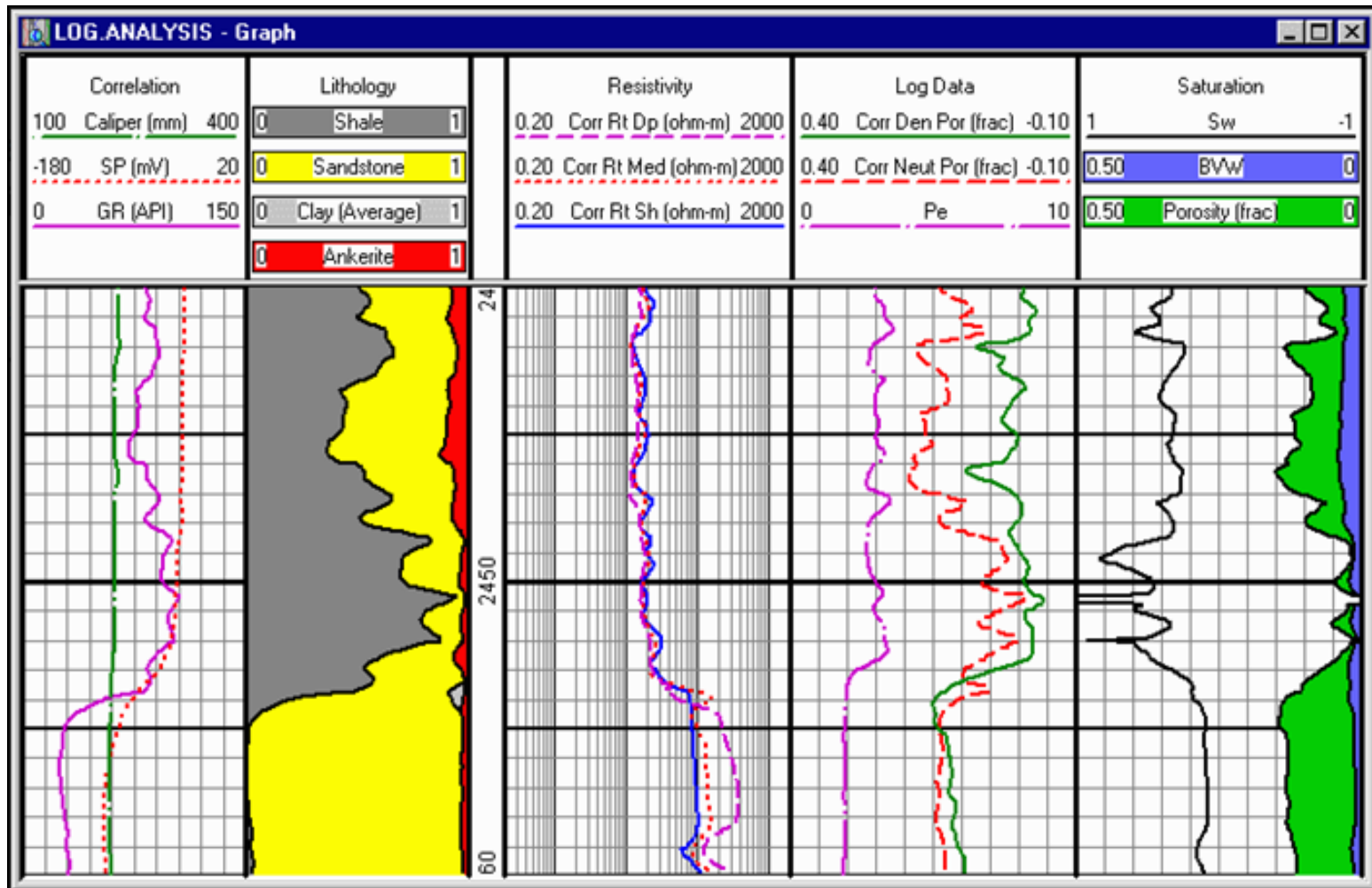
Principes du « logging »

- Mesures dans un puits d'exploration (« *logs* » ou « *diagraphies différées* ») par des « outils » de diverses technologies :
 - Réception de rayons gamma après avoir bombardé la formation
 - Par des *neutrons lents*
 - Par des *rayons gamma*
 - Vitesse et formes d'ondes sonores traversant la formation
 - Conductivité électrique
 - Potentiel spontané
 - Radioactivité naturelle
 - Résonance magnétique nucléaire (IRM)
- Interprétation : à chaque profondeur, on doit estimer :
 - La *composition solide* de la roche (lithologie) : minéraux, argiles...
 - La *porosité*
 - La *teneur des pores en fluides* : eau, huile, gaz



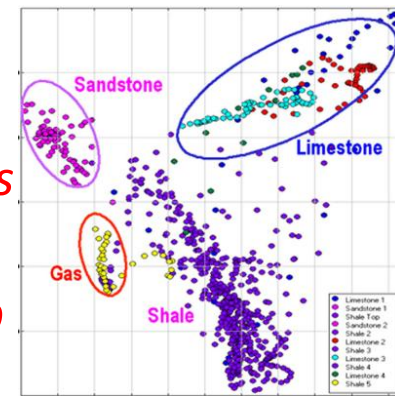
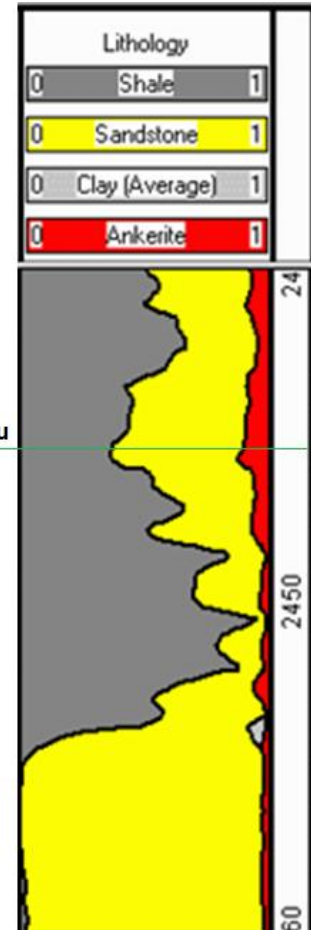
Une impression typique

Logs et résultats



Principes de l'interprétation

- *A chaque niveau de profondeur, on mesure* :
 - $\mathbf{a} = (a_1, \dots, a_n)$ (lectures des outils)
- *On recherche pour ce niveau* : 1 niveau
 - $\mathbf{x} = (x_1, \dots, x_k)$ ($k \leq n$)
 - ex : lithologie, porosité, teneur en huile...
- *Approche directe* (la seule utilisée avant 1976)
 - On recherche une *heuristique* $\mathbf{a} \rightarrow \mathbf{x}$
 - Mais il n'existe *aucune relation fonctionnelle* directe $\mathbf{x} = \mathbf{f}(\mathbf{a})$
 - On recourt à de nombreuses *abaques* et « *cross-plots* »
 - Cette approche est *rigide* et *non adaptée* à toute suite de logs
 - Elle n'est pas adaptée aux *lithologies complexes* (mica...)
 - Elle tient mal compte des *connaissances a priori*
 - *Elle n'utilise pas pleinement toutes les mesures disponibles*
 - Elle tient mal compte des *incertitudes de mesure*
 - Elle n'est pas adaptée aux *nouveaux outils issus de la R&D*

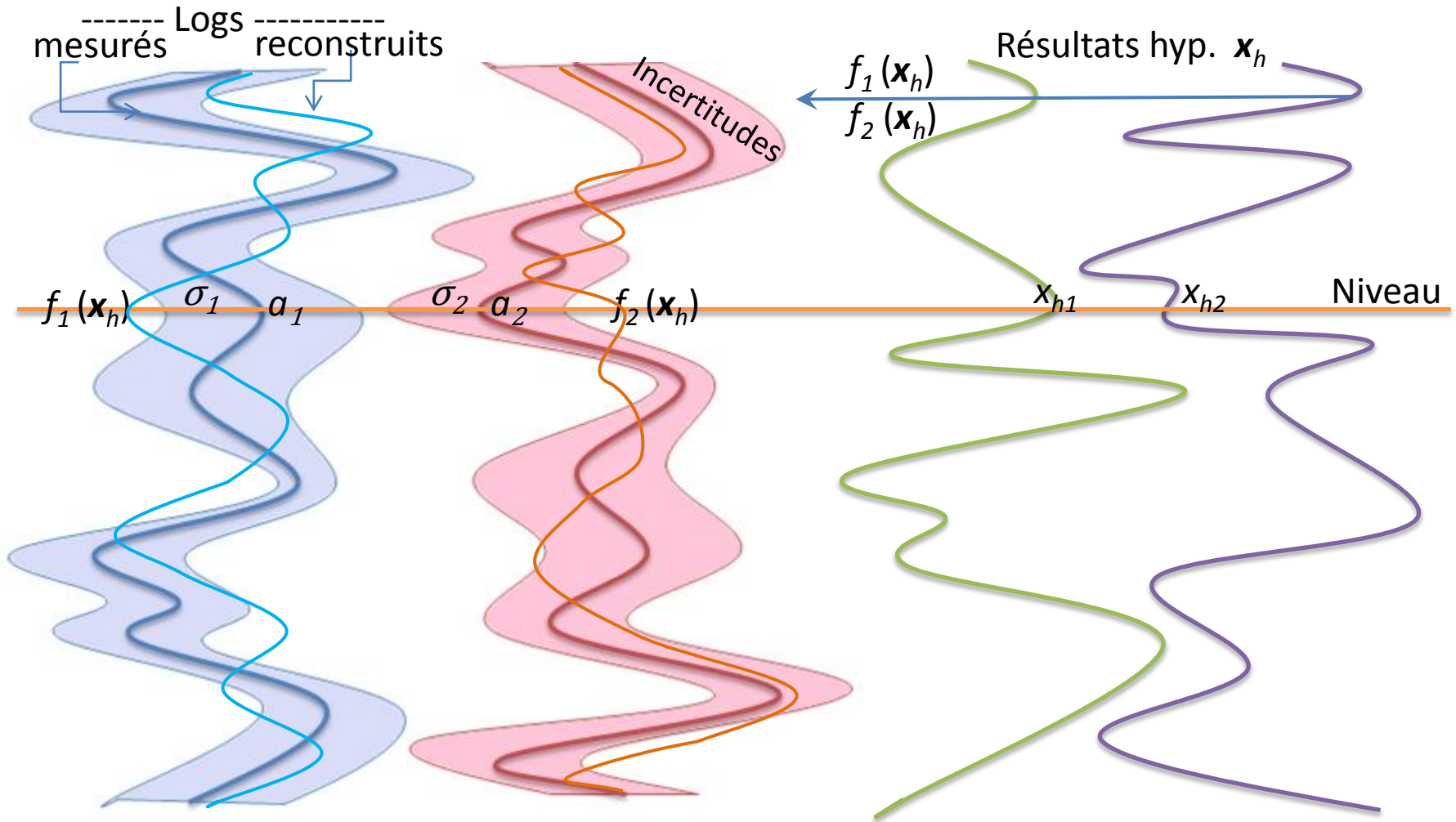


Principe de l'approche inverse

- Les *réponses des outils* en fonction des formations traversées sont connues:
 - Par la théorie (physique de l'outil)
 - Par des mesures de laboratoire (simulations)
 - Pour chaque mesure a_i , on peut écrire : $a_i = f_i(\mathbf{x}) + \varepsilon_i$, soit $\mathbf{a} = \mathbf{f}(\mathbf{x}) + \boldsymbol{\varepsilon}$
 - Où $\varepsilon_i(0, \sigma_i)$ est une v.a. d'erreur sur la mesure a_i
 - C'est le *modèle direct* cause $\rightarrow$ effet [*il n'existe pas* $\mathbf{f}'$ tel que $\mathbf{x} = \mathbf{f}'(\mathbf{a})$]
- Pour chaque *hypothèse de solution* $\mathbf{x}_h$, on calcule une *fonction d'écart* :
$$\Delta(\mathbf{x}_h) = \sum_i \{ [a_i - f_i(\mathbf{x}_h)]^2 / \sigma_i^2 \} + g(\mathbf{x}_h)^2$$
 - $f_i(\mathbf{x}_h)$ est la *lecture théorique* de l'outil i si la formation était décrite par $\mathbf{x}_h$
 - $g(\mathbf{x}_h)$ est une fonction de pénalité exprimant des *contraintes a priori* sur le résultat attendu, en fonction de l'environnement géologie exploré
 - Ex : connaissance de certains minéraux, de la nature huile/gaz, etc.

Schéma explicatif de $\Delta(\mathbf{x}_h)$

$$\Delta(\mathbf{x}_h) = \sum \{ [a_i - f_i(\mathbf{x}_h)]^2 / \sigma_i^2 \} + g(\mathbf{x}_h)^2$$



$$\Delta(\mathbf{x}_h) = \{ a_1 - f_1(\mathbf{x}_h) \}^2 / \sigma_1^2 + \{ a_2 - f_2(\mathbf{x}_h) \}^2 / \sigma_2^2 + g(\mathbf{x}_h)^2$$

Approche inverse bayésienne (« méthode globale »)

- Si on suppose les erreurs de mesure approximativement *gaussiennes* et *2 à 2 indépendantes*, on démontre que :
- La solution $\mathbf{x}$ qui maximise la *probabilité bayésienne* pour que $\mathbf{x}$ soit le *meilleur résultat* – compte tenu des *logs* et des *connaissances a priori* sur le terrain – est précisément celle qui minimise la fonction d'écart $\Delta(\mathbf{x})$

➤ On démontre que : $p(\mathbf{x}|\mathbf{a}) \propto \exp[-\lambda \Delta(\mathbf{x})]$

Avantages

par rapport à l'approche directe

- En minimisant $\Delta(\mathbf{x}_h) = \sum \{ [a_i - f_i(\mathbf{x}_h)]^2 / \sigma_i^2 \} + g(\mathbf{x}_h)^2$
 - On est *flexible* par rapport à la *suite d'outils* utilisée (*anciens* ou issus de la *recherche*)
 - Pour chaque outil i , il suffit d'établir sa *fonction de réponse* $f_i(\mathbf{x})$ et sa *sensibilité* σ_i aux erreurs de mesure
 - Chaque outil est introduit comme un *simple terme* dans la somme ci-dessus
 - On intègre ainsi *toutes les connaissances* sur la réponse des outils
 - On tient compte des *incertitudes sur les mesures*, dépendant elles-mêmes de l'environnement de mesure
 - On peut choisir les *composantes du vecteur de résultats* $\mathbf{x}$ en fonction du type de formation probable (*minéraux présents ou attendus*)
 - On prend en compte les *contraintes a priori* $g(\mathbf{x})$ (*environnement géologique*)
 - On obtient la *solution bayésienne la plus probable* (*maximum likelihood*)
 - On peut également estimer l'*incertitude sur le résultat* $\mathbf{x}_0$ et l'intégrer dans des estimations de réserves



Le cerveau bayésien

Définition

Applications

Hypothèses de fonctionnement

Définition du « cerveau bayésien »

- C'est la capacité pour le cerveau d'appliquer *implicitement* les *règles* ou les *principes* du calcul bayésien dans l'évaluation d'une situation ou la prise de décision
 - Les *règles* : le cerveau serait capable d'*effectuer les calculs* de la formule de Bayes en s'appuyant sur des probabilités *subjectives*
 - Les *principes* : le cerveau estimerait des probabilités ou des *préférences*, mais les combinerait de manière plus *qualitative* que quantitative en intégrant *plusieurs sources d'information indépendantes*
- Cette capacité est *largement admise* par les neurosciences actuelles, mais *l'aspect quantitatif reste discuté*
 - Il est notamment défendu par *Stanislas Dehaene* dans son cours au Collège de France 2011-2012*
 - Je présenterai aussi une *approche alternative*, plus qualitative

L'enfant bayésien

Citations de S. Dehaene (Collège de France)

- « Dès *huit mois de vie*, un bébé est capable d'anticiper le résultat approximatif d'un tirage aléatoire d'une urne. Plus surprenant encore, il est également capable d'inférer, en sens inverse, le contenu probable d'une urne après quelques tirages, ce qui constitue *le fondement même de l'inférence bayésienne* »
- « *Divers algorithmes* permettent d'approcher les règles de l'inférence bayésienne et pourraient être utilisés par l'enfant (T. D. Ullman, Goodman, & Tenenbaum, 2012) »

Exemples

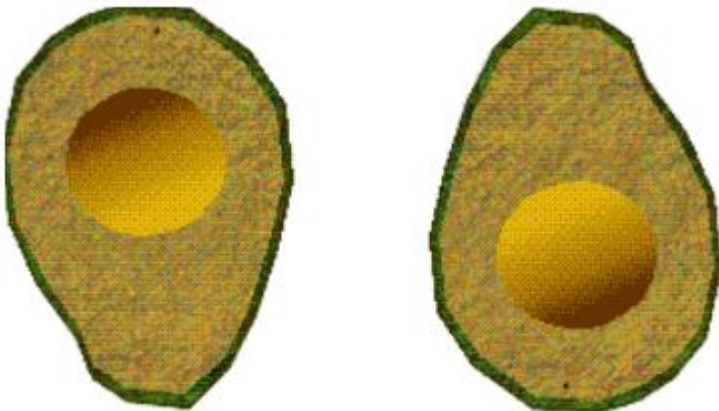
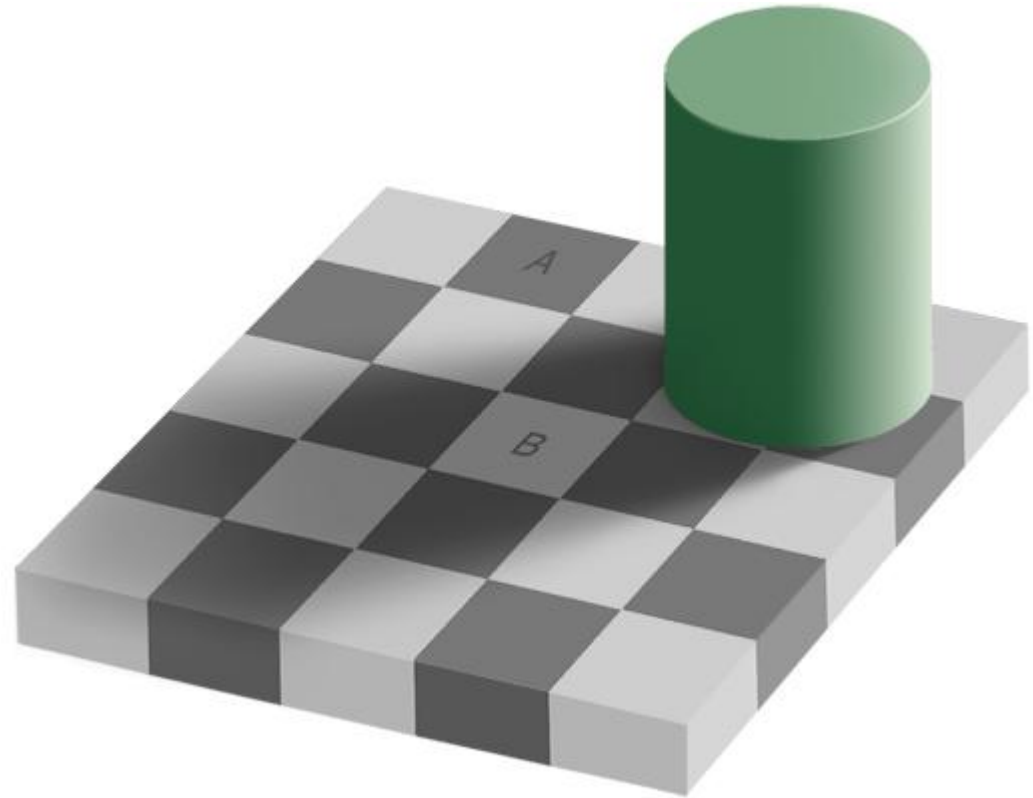
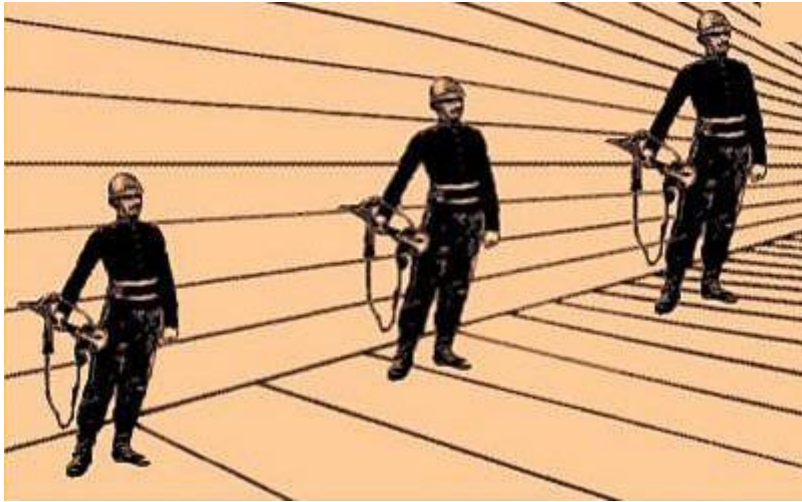
Interprétation *bayésienne* de scènes

ABC

121314

- On raisonne sur la figure de gauche G , puis on étendra à celle de droite
- Pour G , Il y a seulement **2 hypothèses** d'interprétation : $H_1 = \{ABC\}$ et $H_2 = \{A-13-C\}$
- Si l'on ne tient pas compte du sens mais seulement de la **forme** du tracé, le signe central **B** étant parfaitement ambigu, on a : $\mathbb{P}(G|H_1) = \mathbb{P}(G|H_2)$
- Si l'on ne tient compte que du sens proposé, la probabilité a priori de rencontrer une suite **homogène** de lettres ou de chiffres **consécutifs** est supérieure à celle de rencontrer une suite **hétérogène sans logique** : $\mathbb{P}(H_1) > \mathbb{P}(H_2)$
- On a donc, en appliquant la formule de Bayes :
 - $\mathbb{P}(H_1 | G) = \mathbb{P}(G|H_1) \cdot \mathbb{P}(H_1) / \mathbb{P}(G)$
 - $\mathbb{P}(H_2 | G) = \mathbb{P}(G|H_2) \cdot \mathbb{P}(H_2) / \mathbb{P}(G)$
- Pour comparer les deux hypothèses, il est inutile (heureusement) de calculer $\mathbb{P}(G)$
- On en déduit : $\mathbb{P}(\{ABC\} | G) > \mathbb{P}(\{A-13-C\} | G)$: c'est le **sens** qui a fait la différence
- Le cerveau bayésien choisit ici *inconsciemment* l'interprétation *la plus probable*

Autres illusions « bayésiennes »



Dans tous ces cas, le cerveau combine

deux sources d'information:

1. La perception brute (formes, couleurs...)
2. Le contexte, la vraisemblance du résultat

Reconnaissance d'un visage

- Visage parfaitement connu
 - Dans ce cas, il n'y a pas d'hypothèses à départager : le calcul n'est *pas bayésien*
- Visage mal connu : test d'hypothèses
 - Est-ce Paul ou bien Jean ?
 - Sans entrer dans les calculs, il faut tenir compte :
 - De l'*adéquation* du visage observé avec chacune des hypothèses
 - De la *vraisemblance* de voir Paul ou Jean à cet endroit (par exemple, nous sommes à Paris et Jean vit à Nice)

Cerveau bayésien : *conscient* ou non ?

- Le processus de test d'hypothèses est effectué :
 - Parfois *consciemment* (volontairement)
 - *Diagnostic médical* : test de plusieurs *hypothèses de maladies*, en comparant leurs symptômes théoriques avec les *symptômes observés*, en tenant compte de la *prévalence de chaque maladie* dans la population concernée (probabilité a priori)
 - *Recherche de pannes* : même processus
 - Parfois *inconsciemment* (cf. mon précédent exposé)
 - *Reconnaissance de formes*, de visages, de démarches, de mots...
 - *Interprétation de scènes* visuo-auditives, *illusions*
 - *Conscient ou inconscient ?* cela dépend de la *difficulté de la tâche*
 - Reconnaissance d'un visage : inconscient selon qu'il est familier ou non

Implémentations cérébrales

- Selon Dehaene *et al.*
- Par réseau neuronal



Le cerveau implémente-t-il la formule de Bayes ?

- Selon **Dehaene** et d'autres auteurs, le cerveau humain *implémente* une approximation de la formule de Bayes. Cela implique :
 - Qu'il existe des *réseaux de neurones* aptes à effectuer ce type de calculs
 - Que le cerveau soit capable d'*évaluer correctement les probabilités « subjectives »* des différents événements concernés
- Or, les travaux de Tversky et Kahneman **contredisent** cette seconde hypothèse *
 - Ils mettent en évidence de *nombreux « biais »* qui déforment les probabilités subjectives par rapport à une démarche rationnelle
 - Par exemple, la fonction de valeur subjective serait *concave pour les gains, convexe pour les pertes*

* Tversky A. & Kahneman D., *The Framing of Decisions and the Psychology of Choice*, Science (1981)

Autres difficultés pour le cerveau bayésien

- Même Stanislas Dehaene, partisan du cerveau bayésien, reconnaît **certaines difficultés** (*sujettes à de longues discussions*) :

[...] « Le cerveau, qui **n'est pas un ordinateur digital**, pourrait n'implémenter qu'un algorithme bayésien **imparfait**. Il se pourrait ainsi que la représentation des toutes petites probabilités, ou la multiplication de deux probabilités ou de deux distributions, posent des **difficultés particulières au cerveau** » *

- La **perception visuelle** est en principe un domaine privilégié d'application du cerveau bayésien; mais selon certaines études, le travail inconscient serait bien bayésien, mais **pas le résultat conscient (affiché)** :

[...] « Percevons-nous systématiquement la représentation **la plus plausible** ? **Pas nécessairement** [...] Les processus perceptifs **non conscients** manipuleraient l'ensemble de la distribution, tandis que l'accès à la **conscience** consisterait en un **tirage aléatoire d'un échantillon** de la distribution [...] L'échantillonnage serait une réponse à la **difficulté de réaliser des calculs Bayésiens** complets » *

Mes principales objections au cerveau « *strictement* » bayésien

- Le caractère « non-naturel » (pour le cerveau) de l'application d'une formule telle que celle de Bayes : *l'inconscient ne sait pas calculer numériquement avec précision* *
- *L'inutilité d'une telle précision*, alors que les *probabilités subjectives* prises en compte sont elles-mêmes largement *imprécises et/ou biaisées*
 - Pourquoi l'évolution aurait-elle sélectionné une aptitude *complexe et peu utile* ? (Ce ne serait pas *économique*...)
- Le *caractère plus naturel des solutions alternatives proposées*, par propagation et rétro-propagation d'activations neuronales en mémoire sémantique **

Une autre approche d'implémentation bayésienne

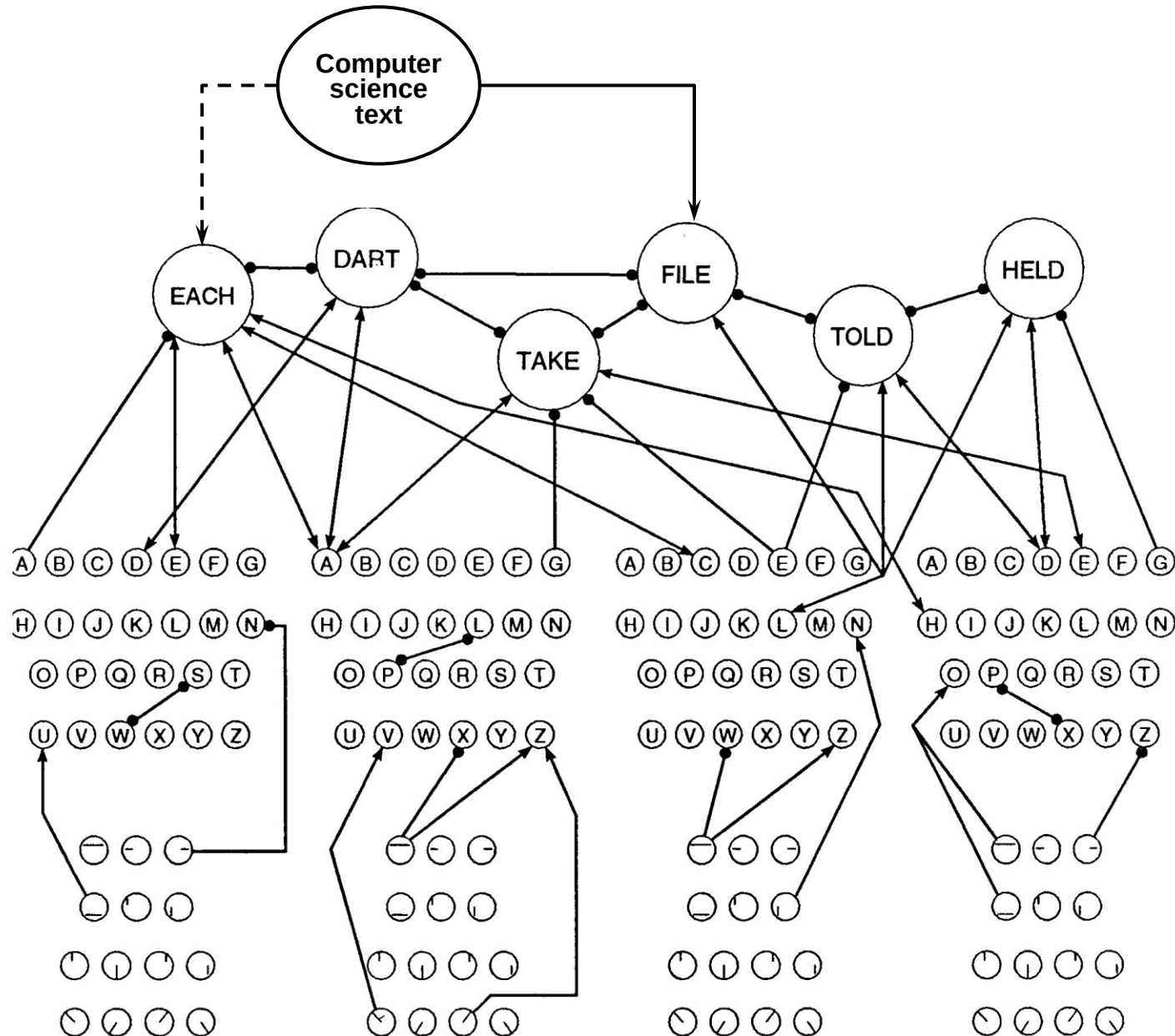
Reconnaissance de
mots peu lisibles

Reconnaissance de mots

- Contexte
 - On **cherche à identifier** un mot mal écrit ou peu distinct
 - Médicament sur ordonnance (le médecin écrit mal)
 - Un panneau routier (lointain ou sale)
 - Une enseigne (id°)
 - On **combine deux sources** d'informations
 - La **forme des lettres** (mal perçue dans les cas ci-dessus)
 - Le **contexte** : maladies possibles, ville attendue, catégorie du commerce...
- Il s'agit donc bien ici de **situations typiquement « bayésiennes »**
- Comment le cerveau procède-t-il ?
 - **Estimation** des probabilités et **calcul** de la formule de Bayes ?
 - **Propagation d'activations neuronales** en mémoire sémantique ?



Propagation mémoire pseudo-bayésienne*



Commentaires

- Ce réseau a été *simulé avec succès** pour illustrer la reconnaissance de mots de 4 lettres à partir d'*éléments graphiques élémentaires* composant les lettres, et comportant des signes mal reconnus
- Il montre les propagations d'activations *ascendantes et descendantes* dans la *hiérarchie mémoire* relationnelle (mémoire sémantique)
 - *Renforcement mutuel* d'une lettre mal reconnue et d'un mot comportant cette lettre au même emplacement
 - Le *renforcement ascendant* active l'*identification* de la forme cible (graphème ↗ lettre ↗ mot) [↗ : renforce, mais parfois ↘ : inhibe]
 - Le *renforcement descendant* met en œuvre le rôle du *contexte* (mot *possible*, mot *attendu* ... : contexte ↗ mot ↗ lettre)
- On retrouve les 2 facteurs d'une évaluation bayésienne : *forme* et *contexte*
- Ainsi *tous les éléments qualitatifs d'une évaluation bayésienne sont en place et fonctionnels* (et d'autres niveaux peuvent s'ajouter dans cette hiérarchie)

Pourquoi se baser sur la mémoire ?

- Parce que les *connaissances* utilisées pour l'estimation bayésienne s'appuient sur des *apprentissages préalables*
 - Savoir *identifier une lettre* à partir de ses composantes graphiques
 - Connaître *l'orthographe*
 - Connaître la *signification* des mots
 - Savoir *classer les mots* dans des *catégories* (noms de ville, termes informatiques...)
- Toutes ces connaissances étaient *déjà stockées en mémoire sémantique* avant l'épreuve de déchiffrage
 - Il est donc *incontournable* d'utiliser cette mémoire pour effectuer ce déchiffrage

Alors quel sens donner au « cerveau bayésien » ?

- Sans doute le terme « *bayésien* » doit-il être pris ici dans son *acception philosophique la plus large*
 - Recherche de *la cause la plus probable* d'un phénomène ou d'une observation
 - Evaluation et utilisation de *probabilités subjectives*
 - Il faudrait plutôt parler de *préférences*
 - Prise en compte de *deux ou plusieurs sources indépendantes* d'évaluation
 - La *reconnaissance des formes*, identification suite à une *observation* (ou interprétation d'autres données cérébrales)
 - La *vraisemblance* de la solution en fonction de *l'environnement et du contexte général*

Théorie de la décision valorisée

(exposé élémentaire)

- On se situe dans le **cadre bayésien** :
 - Les probabilités mentionnées sont issues d'un raisonnement bayésien (non reflété dans les notations ci-dessous)
- On a le choix entre **deux décisions** D^1 et D^2 , basées sur **deux événements hypothétiques** E_1 et E_2 de probabilités $\mathbb{P}(E_1)$ et $\mathbb{P}(E_2)$
- On attribue des **valeurs** V_k^i (**gains ou coûts**) aux décisions
 - V_k^i : valeur de la décision D^i si l'événement E_k se produit
 - Par convention $V_k^i > 0$ est un **gain**
- La **décision « optimale »** est celle qui maximise son **espérance de gain** $V(D^n)$ ($n = 1$ ou 2)
 - $V(D^n) = V_1^n \mathbb{P}(E_1) + V_2^n \mathbb{P}(E_2)$
- **NB** : cette théorie se généralise de manière évidente à **n décisions** et **p événements**

L'affaire des poisons (suite)

- Rappel : **je suis empoisonné** par l'un des poisons p_1 ou p_2
 - Après une **évaluation bayésienne**, on a estimé précédemment : $\mathbb{P}(p_1) = 56\%$ et $\mathbb{P}(p_2) = 44\%$
- On a le choix entre **3 décisions**
 - D^1 : **antidote de p_1** (*guérison* si p_1 , mais *interaction toxique* avec p_2)
 - D^2 : **antidote de p_2** (antidote *moins toxique*, mais inefficace si p_1)
 - D^3 : **se coucher et attendre** (*aucune toxicité, aucune guérison*)
- On évalue les conséquences de ces décisions :
 - D^1 : si c'est p_1 , $V^1_1 = +10$ (guérison)
 - D^1 : si c'est p_2 , $V^1_2 = -100$ (inefficacité + toxicité)
 - De même, $V^2_1 = -50$; $V^2_2 = +10$; $V^3_1 = -20$; $V^3_2 = -30$
- On calcule les valeurs « espérées » $V(D^n)$ ($n \leq 3$) :
 - $V(D^1) = -38.4$
 - $V(D^2) = -23.6$
 - $V(D^3) = -24.4$
- La **moins mauvaise décision** est D^2 (suivie de D^3 !)
- Décision « **minimax** » : D^3 (minimise le risque maximum)



Remarques

- Chaque **coefficient de valorisation** V_k^i doit tenir compte :
 - Du **gain attendu** si la décision est bonne
 - Gain objectif
 - Gain en termes de crédibilité, de carrière du décideur
 - De la **perte attendue** si la décision est mauvaise
 - Coût objectif
 - Perte en termes de crédibilité, de carrière du décideur
 - Du **coût d'exécution** de la décision (et du **bénéfice** pour certains...)
- Ces valeurs sont **subjectives** et parfois **impossibles** à déterminer **rationnellement**
 - Quel est **valeur de la vie** d'une personne ?
De **1000** personnes ?
 - Exemple : le préfet doit-il **évacuer un village** de 1000 habitants menacé par une peu probable **éruption volcanique** ???



Coulée de lave

Les dangers de l'illusion rationnelle

- Lorsque l'on se trouve devant une *décision risquée* avec un *fort enjeu* et dans un *univers incertain* (par ex. évacuer un village)
 1. On peut être tenté d'estimer des probabilités et des coûts, et d'appliquer une décision « rationnelle » (« optimale », ou minimax...) : c'est l'*approche analytique*
 2. On peut aussi appliquer une réflexion *globale*, de manière plus informelle : peser le pour et le contre, interroger son *intuition*, consulter... cette approche combine, *souvent inconsciemment*, des *paramètres beaucoup plus nombreux* que l'approche analytique (*propagations neuronales*)
- Les décisions retenues en pratique relèvent le plus souvent de la *seconde approche*
 - Ce n'est qu'*après-coup* que l'on peut inférer les probabilités et les coûts que l'on avait *implicitement pris en compte* et estimés de manière globale
 - *L'approche « rationnelle » peut souvent être trompeuse*, car on est *incapable* d'évaluer correctement les probabilités et coûts impliqués ; on ignore ainsi également de *nombreux paramètres*. Cette approche n'a alors *de rationnelle que le nom...*

Conclusion

Acceptions du terme « bayésien »

- Axé sur la *recherche des causes* d'événements ou d'observations
- Il existe *deux acceptions majeures* de ce terme
 - *Au sens strict*
 - Application de la *formule de Bayes*
 - Avec ou sans caractère *subjectif* des probabilités
 - *Au sens élargi* (le plus répandu)
 - Prise en compte de critères de préférence *subjectifs* pour prendre une décision tenant compte de *plusieurs sources* d'information *indépendantes*
 - Critères physiques ou logiques (pattern recognition)
 - Vraisemblance du résultat par rapport au contexte général
 - Apprentissage par *accumulation d'indices successifs*
- Mais attention aux *confusions* et aux *effets de mode*
 - *Cerveau bayésien* : vouloir *à la fois* appliquer la conception élargie, *non rigoureuse*, et tenter de la justifier par une *formule mathématique*
 - *Apprentissage* : les indices doivent être d'origines *indépendantes* et non des *raffinements successifs* d'une même source de connaissance

C'est tout pour aujourd'hui...

Compléments

Philosophie des probabilités

La nature de l'incertitude

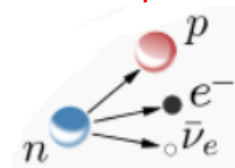
Interprétations quantitatives
des probabilités

Pourquoi une *philosophie des probabilités* ?

- La *théorie mathématique* des probabilité est largement développée et non contestée
- Mais elle ne traite pas de la question philosophique : « *qu'est-ce qu'une probabilité ?* »
 - Quelle est la *nature* d'un aléa ?
 - Quelle *interprétation* peut-on donner à la valeur quantitative d'une probabilité ?
- Certaines de ces questions restent ouvertes et discutées et sont pourtant fondamentales
 - Quel *sens* a la probabilité d'un *événement non répétable* et quelle *valeur* lui attribuer ?
 - Comment estimer et juger une *probabilité subjective* ?
- Ces questions intéressent les probabilités *bayésiennes*

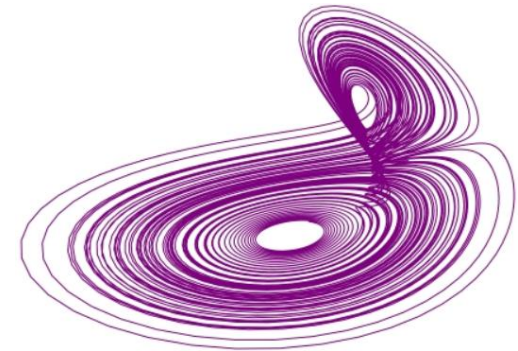
La nature de l'incertitude

- Soit un événement A inconnu passé ou futur
- 1^{ère} possibilité : **incertitude épistémique** (due à l'ignorance)
 - A est inconnu car on **manque de connaissance** sur un phénomène **déterminé** ou **déterministe**, mais on dispose de certaines informations utilisables
 - Pile ou Face ? (avant ou après le tirage)
 - Cette météorite va-t-elle tomber sur la terre ?
- 2^{ème} possibilité : **indéterminisme ontologique**
 - L'indétermination est due à la nature elle-même, il n'existe **aucune information possible** permettant de la réduire **au moment où l'on pose la question**
 - Physique quantique (→ *implications sur la terminologie Heisenberg*)
 - Emergence « forte » (ontologique)
- Cas litigieux ou difficiles à classer **Une troisième catégorie est-elle nécessaire ?**
 - Débat Einstein / Bohr sur la nature épistémique ou ontologique de la **mécanique quantique**
 - Le chat de Schrödinger (sa « superposition » est-elle épistémique ou ontologique ?)
 - Statut des systèmes déterministes **instables** ou **chaotiques** ? (Va-t-il pleuvoir mercredi ?)
 - **Rencontres fortuites imprévisibles** (Vais-je croiser Irène demain ? Aurai-je un accident ?)
 - Le fonctionnement du **cerveau** ? **L'émergence** (faible ou forte ?)

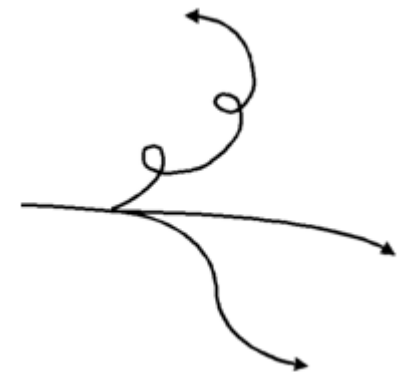


Une autre catégorie d'aléas

- Certains des exemples précédents relèvent d'une 3^{ème} catégorie : la *sensibilité « infinie » aux conditions initiales* (chaos, instabilités)
- Instabilités de *systèmes physiques*
 - Prévisions météo (systèmes chaotiques)
 - Eruptions, tornades, tsunamis...
 - Stabilité du système solaire
- Processus instables liés à la *vie* ou au *cerveau*
 - Evolution des espèces (le hasard et la nécessité)
 - Conséquences de *décisions imprévisibles* :
 - Accidents, rencontres fortuites, bourse, morsures de chiens...
- Ces cas doivent-ils être classés comme *épistémiques* ou comme *ontologiques* ?



Attracteur de Lorenz



Trajectoires divergentes

Interprétation quantitative des probabilités

Sur quels fondements *calculer* ou
estimer une probabilité ?

Principes généraux

- Les probabilités bayésiennes sont parfois nécessairement *subjectives*
- Comment *fonder* une estimation subjective ?
 - De nombreux *biais cognitifs* interfèrent avec une estimation subjective *rationnelle* de la probabilité
 - Emotions, erreurs logiques*, devinettes...
 - « Face est sorti 3 fois de suite : je mise sur Pile »
 - « Je ne sais pas : une chance sur deux... »
- *Conditions d'acceptabilité* d'une probabilité subjective
 - Respecter les axiomes finis de *Kolmogorov*
 - Cette contrainte est très forte pour des probabilités subjectives
 - Respecter l'*ordre des préférences* subjectives
 - $\{A \text{ préféré à } B\} \Leftrightarrow \{\mathbb{P}(A) > \mathbb{P}(B)\}$

Contraintes supplémentaires

- La notion qui va suivre n'appartient pas à la définition axiomatique des probabilités, ni aux théorèmes qui s'en déduisent (*comme la loi « probabiliste » des grands nombres*) :
- Une probabilité doit être JUSTE : elle doit s'accorder à la *réalité du monde* (dans un sens à définir)
- C'est tout l'objet des discussions qui vont suivre
 - Elle doit correspondre aux notions intuitives d'*ordre de préférence* pour des agents rationnels
 - Elle doit pouvoir être confirmée par des *tirages répétés à long terme* (lorsque ils sont réalisables)
 - Par exemple, attribuons à un dé une probabilité équirépartie de $1/6$ sur chaque face
 - Même si le dé est *pipé* vers le 1, cette probabilité de $1/6$ s'accorde avec tous les axiomes *et avec la loi des grands nombres**, mais elle est FAUSSE !
 - A long terme, la fréquence du 1 *ne va pas converger vers $1/6$*

* Cette loi est indépendante de la justesse de la probabilité

Discussion des principales interprétations des probabilités

Telles qu'elles sont débattues en
philosophie des sciences

Principales interprétations

- Probabilités classiques
- Probabilités subjectives
 - Pari
 - Théorème du « Dutch book »
- Interprétations fréquentistes
 - Cas fini
 - Cas infini : fréquence limite
- Propensions (« Propensities »)
- Lois de la physique, symétries
- Discussion

Probabilités classiques

- Laplace, Pascal, Bernoulli, Huygens, Leibniz...
 - (Nombre de cas favorables) / (nombre de cas possibles)
- Basées sur des probabilités *équiréparties*
 - En cas *d'ignorance totale* (une chance sur 2, ou sur n selon le nombre de possibilités envisagées)
 - En cas de *symétrie des possibilités élémentaires*
 - Pile ou face, jeux de dés, de cartes...
- Problèmes
 - L'équirépartition n'est pas vraiment *définie*
 - Définir la *probabilité* par *l'équiprobabilité* est en quelque sorte une définition *circulaire*...
 - L'équirépartition en cas d'*ignorance* conduit à des *absurdités*
 - L'équirépartition par *symétrie* peut n'être qu'*apparente*
 - Dés pipés, erreurs d'analyse...
- En pratique
 - Cette « définition » n'en est pas une, mais elle est un *moyen pratique et puissant* de calculer les probabilités lorsque la symétrie est avérée
 - Probabilités de distributions particulières au bridge, au poker...
 - Modèles d'urnes, jeux de hasard, courte-paille...
 - Rarement rencontré dans le monde naturel

Probabilités subjectives

- Degré de croyance, de confiance, de préférence...
- Avantages
 - Peut s'appliquer aux événements futurs *non répétables*
 - Prédications météorologiques, éruptions volcaniques...
 - N'exige pas de satisfaire les *axiomes d'additivité* de Kolmogorov
 - La *monotonie* est suffisante en fonction des préférences
 - Si A est préféré à B et B préféré à C, A doit être préféré à C
- Inconvénients
 - Peut *varier grandement* d'un individu à l'autre
 - Il est impossible de *départager deux opinions* (la loi « physique » des grands nombres ne peut être appliquée)
 - N'étant pas rigoureuse (quantitative), elle ne suffit pas pour *évaluer des coûts* ni pour élaborer des *décisions optimales* (comme l'évacuation « rationnelle » d'un village)
- En pratique
 - C'est la *seule possibilité* en cas de *répétitions limitées*, ou en l'absence de *critères objectifs* ou de *connaissances*

Paris, Dutch book...

- Une manière « **pratique** » d'évaluer une **probabilité subjective** :
 - **Définir** la probabilité p d'un événement en imaginant un pari, tel que :
 - On considère ce pari comme **équilibré** s'il est basé sur la probabilité p (= accepter de jouer à **p contre 1** sur cet événement)
- **Dutch book** : **suite de paris** basés sur une probabilité subjective
 - **Théorème** : si un joueur **se trompe** sur l'estimation d'une probabilité et base sa stratégie sur cette croyance, alors il existe toujours une **stratégie gagnante à long terme** contre ce joueur
 - Exemple : un joueur de pile ou face qui est persuadé qu'il existe une **mémoire** des tirages précédents (de sorte, croit-il, d'**équilibrer les fréquences**)
 - **Remarque** : Ce théorème est basé sur la **loi physique des grands nombres**. Il exige la possibilité de nombreux tirages
 - A ce titre, il se ramène à la **conception fréquentiste** des probabilités

Conceptions fréquentistes

Préliminaire : *les lois des grands nombres*

- Fréquence relative sur n tirages

- Soit A un événement de probabilité p . On effectue n tirages *indépendants* et on définit la *fréquence relative de A* sur ces n tirages

$$F_n(A) = \frac{(\text{nombre de réalisations de } A)}{n}$$

- Loi probabiliste des grands nombres

$$\mathbb{P} \left(\lim_{n \rightarrow \infty} F_n = p \right) = 1$$

- "Loi" physique des grands nombres

$$\lim_{n \rightarrow \infty} F_n = p$$

- Les deux lois diffèrent (*ne pas les confondre...*)

- La *loi physique* est *non démontrée* et *fausse* si la probabilité p est *erronée*
- La *loi probabiliste* est *démontrée* et *toujours vraie*, même si p est *erroné*
 - De fait la LGN *probabiliste* est une tautologie : $\mathbb{P}$ est défini à partir de p

Conceptions fréquentistes

- Fréquences à tirage fini

- On *définit* la probabilité p d'un événement E par la *fréquence* atteinte lors d'un « grand nombre » (ou un nombre « suffisamment élevé ») n de tirages indépendants :

$$p \stackrel{\text{def}}{=} F_n(E) \text{ (pour } n \text{ choisi « assez grand »)}$$

- *Problème* : même à pile ou face, on trouve $p \neq 0,5$!

- Fréquences virtuelles à tirage infini

- Même si c'est *impossible en pratique*, on définit p par une *expérience de pensée*, telle que :

$$p \stackrel{\text{def}}{=} \lim_{n \rightarrow \infty} [F_n(E)]$$

- Cette conception est attaquée par plusieurs auteurs* car elle est *inapplicable* ; et pourtant (*selon moi*)... →

* Alan Hájek, *Fifteen Arguments Against Hypothetical Frequentism*, Canberra (2009)

Défense du fréquentisme infini

- Seule la définition *fréquentiste infinie* p d'une probabilité satisfait à la *loi physique* des grands nombres
- *Aucun nombre* $p' \neq p$ ne pourra voir sa fréquence converger vers p'
 - Ceci, malgré le fait que la *loi probabiliste* des grands nombres sera satisfaite pour p' , comme pour toute autre valeur p'' , quelle qu'elle soit
- Même si cette définition est en général *inapplicable en pratique*, elle est *la seule* qui constitue une *définition philosophique cohérente* d'une probabilité « *juste* »
- *En pratique*, toute estimation (par exemple pour un événement non répétable) *devra s'approcher le plus possible* (en pensée) de cette conception
 - *Exemple* : probabilité d'*éruption prochaine d'un volcan* : selon les signes apparents et la géologie locale, il faut tenter d'imaginer, pour un grand nombre de *volcans hypothétiques* supposés identiques, la proportion de ceux qui connaîtront une éruption prochaine
 - Même si on n'est pas capable de faire cette estimation, elle *doit servir de base au raisonnement*, dans la mesure du possible

Autres définitions

« *contrefactuelles* »

- On distingue des propriétés *catégoriques* et des propriétés *dispositionnelles*
- Prenez un vase de chine de la période Ming
 - Son *poids*, sa *taille* sont des propriétés *catégoriques*
 - Sa *fragilité* est une propriété *dispositionnelle*
 - *Si* on le laisse tomber sur le sol, *alors* il se casse
 - Cette propriété a un sens *même si on ne fait pas l'expérience*
 - C'est une propriété *intrinsèque* du vase, mais sa *définition* est *conditionnelle*
- De même, la *probabilité* pour une pièce de tomber sur face est une propriété *dispositionnelle* de cette pièce
 - *Si* on la lance un grand nombre de fois, *alors* la fréquence relative de faces tendra vers $\frac{1}{2}$
 - Mais *on n'a pas besoin de faire l'expérience* pour que la pièce possède (ou non) cette propriété : elle est *intrinsèque mais conditionnelle*
 - Il en est de même pour la probabilité d'un événement *non répétable* (mais qui pourrait l'être *par la pensée* : expérience « contrefactuelle »)



Un autre problème d'interprétation

- Rappel d'une objection concernant l'entité D
- Il existe *deux interprétations* possibles pour un événement de probabilité nulle [*resp. négligeable*]
 - Soit l'événement est *imprévisible* mais *vraisemblable* (dans un univers infini)
 - Soit l'événement est *strictement impossible* [*resp. invraisemblable*]
- Ces deux interprétations fondamentalement différentes *ne se distinguent pas* dans la théorie des probabilités
 - En particulier, dans un univers infini, on est *certain* que la plupart des événements qui vont advenir seront de probabilité nulle : *les événements de probabilité nulle sont certains*
 - Inversement, dans les cas où la probabilité nulle signifie l'impossibilité, *les événements de probabilité nulle sont impossibles*
- Cela pourrait entraîner des *confusions* lors de *l'interprétation physique* des probabilités nulles ou négligeables

Les lois des grands nombres

0123456789

Lois des grands nombres

(énoncés fréquentiels)

- Ces lois s'appliquent à des *tirages répétés et indépendants* obéissant à une *même loi de probabilité*
- *Loi faible des grands nombres*

$$(\forall \varepsilon)(\forall \eta)(\exists N) : (n \geq N) \Rightarrow \mathbb{P}(|F_n - p| \geq \eta) \leq \varepsilon$$

- Sa démonstration s'appuie sur l'inégalité de *Bienaymé-Tchebychev*, valable pour toute variable aléatoire réelle :

$$\mathbb{P}(|X - \mu| \geq t\sigma) \leq (1/t^2) \text{ pour toute v.a. } X(\mu, \sigma)$$

- *Loi forte des grands nombres*

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} F_n = p\right) = 1$$

- *Une caractéristique des démonstrations*

- Les lois des grands nombres sont une *conséquence mécanique* des axiomes des probabilités, et *ne dépendent en rien de la validité physique* de la probabilité $\mathbb{P}$

Un paradoxe de la LGN

$$\mathbb{P} \left(\lim_{n \rightarrow \infty} F_n = p \right) = 1$$

- Cette formule se lit :
 - « Les fréquences convergent "*presque sûrement*" vers p lorsque n tend vers l'infini »
- Mais certains pourraient comprendre :
 - « On peut être *presque sûr* que les fréquences convergent *vraiment* vers p lorsque n tend vers l'infini »
- Or cette seconde interprétation est *fausse*
 - Supposons le dé *pipé* vers le $\{1\}$ *sans qu'on le sache*, avec :
 $p = 1/6$ (probabilité erronée, la « bonne » probabilité étant par ex. $1/5$)
On *observe* : $\lim_{n \rightarrow \infty} F_n(\{1\}) = 1/5 \neq p$ (*mesure physique*)
Mais on *démontre* que : $\mathbb{P}[(\lim_{n \rightarrow \infty} F_n(\{1\}) = p = 1/6)] = 1$ (**LGN**)

Comment expliquer ce paradoxe ?

La loi des grands nombres semble être contredite par l'observation !

On *observe* : $\lim_{n \rightarrow \infty} F_n(\{1\}) = 1/5 \neq p$ (*mesure physique*)

Mais on *démontre* que : $\mathbb{P}[(\lim_{n \rightarrow \infty} F_n(\{1\}) = p = 1/6)] = 1$ (*LGN*)

Posons $L = \lim_{n \rightarrow \infty} F_n(\{1\})$ (On *observe* $L = 1/5$)

On distingue la *fausse* (estimée) et la *vraie* probabilité de l'événement $\{1\}$: $p_f = 1/6$ et $p_v = 1/5$

On en déduit une *fausse* et une *vraie* probabilité applicables à la limite des fréquences : $\mathbb{P}_f$ et $\mathbb{P}_v$

$\mathbb{P}_f$ est *défini d'après* p_f tandis que $\mathbb{P}_v$ est *défini d'après* p_v
(selon la définition des probabilités sur un espace produit*)

LGN : $\mathbb{P}_f[L = 1/6] = 1$, tandis que : $\mathbb{P}_v[L = 1/5] = 1$

Comme $L = 1/5$, il s'en suit que : $\mathbb{P}_v[L = 1/6] = 0$

* $\mathbb{P}(A_1, A_2, \dots, A_N) \stackrel{\text{def}}{=} \prod_{i \leq N} \mathbb{P}(A_i)$ pour N tirages indépendants

Résumé

- Soit E un événement de probabilité p (*probabilité estimée*)
- La Loi « *probabiliste* » des grands nombres *n'assure pas* que la suite des $F_n(E)$ converge vers p , *même approximativement*
- Elle stipule seulement que la *probabilité* $\mathbb{P}$ pour qu'il en soit ainsi est *égale à 1*
- Mais *si p n'est pas juste, $\mathbb{P}$ ne l'est pas non plus*, car $\mathbb{P}$ est dérivé de p^*
 - Dé pipé
 - Probabilité *subjective*
 - *Evaluation difficile* de la probabilité réelle (prévision météo, etc.)
- Les fréquences ne convergent vers p *que si p est « juste »*

* Suivant la *définition* de la probabilité sur un *espace-produit*

Conclusion

- La loi probabiliste des grands nombres est une théorie *mathématique*, mais n'est pas une théorie *physique*. Dans ce sens, elle est *toujours vérifiée*
- Pour qu'une probabilité p soit « *juste* », il ne suffit pas qu'elle satisfasse la *loi probabiliste* des grands nombres (ce qui est *toujours le cas*), il faut que la suite fréquentielle *observée* converge *vraiment* vers p , ce qui n'est pas la même chose
- La loi physique des grands nombres assurerait la convergence *réelle* de la suite des fréquences vers p . Cette loi *n'est pas démontrée*. Elle n'est valide *que* si la probabilité p est « *juste* »
- Réciproquement, on pourrait *définir* p comme *le* nombre satisfaisant à la *loi physique* des grands nombres
- Cette observation sera utilisée lors de l'*interprétation des probabilités*, en faveur de l'*interprétation fréquentiste*