

Philosophie des Probabilités

Claude Mayer

Emergence 23 mars 2015

Philosophie des Probabilités



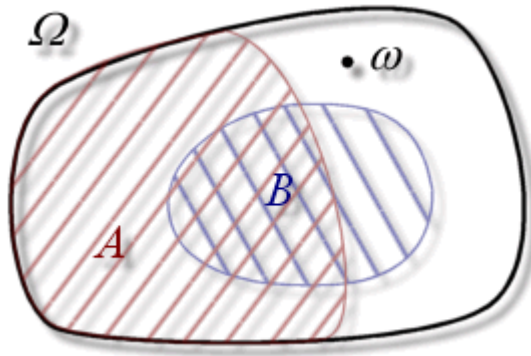
Qu'est-ce que la philo des probas ?

- La **théorie des probabilités** est une théorie axiomatique, abstraite, rigoureuse
 - Comme toute autre théorie axiomatique, elle n'exige pas d'être confrontée à une **expérimentation** quelconque
- **Objectif** : faire correspondre cette théorie à la **réalité du monde**
 - Correspondance avec le **monde physique** et les **sciences humaines**
 - Quelles sont les différentes **catégories d'incertitudes** ?
 - Comment interpréter la **valeur numérique** d'une probabilité ?
 - Que valent les probabilités **subjectives** ? Comment les **calculer** ?
 - **Applications** des probabilités
 - **Probabilités bayésiennes (recherche de causes)**
 - Comment **décider au mieux** dans un environnement **incertain** ?

Plan

- Rappels élémentaires sur les probabilités
- Les catégories d'incertitudes
- Interprétations quantitatives des probabilités
- Applications bayésienne : probabilité des causes
- Comment décider dans un univers incertain ?

Rappel des not(at)ions élémentaires



Univers, événement élémentaire, événements



$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

$$\omega = \{6\}$$

$$A = \{1, 2, 3, 4\}$$

$$B = \{3, 4, 5\}$$

Exemple

- Probabilité $\mathbb{P}(A)$ d'un événement A
 - $0 \leq \mathbb{P}(A) \leq 1$
- Axiome d'additivité (Kolmogorov)
 - Pour des événements $A_1, A_2, \dots$ *2 à 2 disjoints*, on a :
$$\mathbb{P}(A_1 \cup A_2 \cup \dots) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \dots$$
 [suite finie ou dénombrable]
- Probabilité conditionnelle
 - $\mathbb{P}(A|B)$: Probabilité de A *sachant* B : $\mathbb{P}(A|B) = \mathbb{P}(A \cap B) / \mathbb{P}(B)$
- Événements A et B indépendants si :
 - $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$ [On a alors $\mathbb{P}(A|B) = \mathbb{P}(A)$ et $\mathbb{P}(B|A) = \mathbb{P}(B)$]

Philosophie des probabilités

La nature des incertitudes

Interprétations quantitatives des
probabilités

La nature des incertitudes

- Différentes catégories de circonstances peuvent donner lieu à des **incertitudes**
 1. La simple *ignorance* de certains faits
 2. L'évolution des *systèmes complexes*
 3. L'évolution des systèmes *chaotiques* ou *instables*
 4. Les incertitudes *quantiques*
 5. Les *coïncidences* ou *circonstances fortuites*
 6. Exploitation *volontaire* du hasard

1. Ignorance partielle de paramètres ou d'événements **déjà déterminés**

(Incertitude sur *l'existant* ou le *passé*)

- *Pile ou Face* ? (*après* que la pièce soit retombée, mais *avant* de la regarder)
- Le *chat de Schrödinger*, juste *avant* d'ouvrir la porte (selon l'interprétation actuelle)
- *Imprécision de mesures* ou de *conditions initiales*
- *Ignorance totale* (Y a-t-il de la vie autour d'Arcturus ?)



2. Evolution des systèmes complexes

- La *complexité* est cause d'incertitude « épistémique »
 - Systèmes complexes *déterministes* et *non chaotiques* : *ignorance* due à une complexité *insaisissable* par l'esprit humain
 - Soit par *non-connaissance* des détails du système
 - Soit par *incapacité* d'en suivre ou d'en prédire l'évolution
 - **NB** : Pour certains systèmes, le nb d'états possibles > nb de protons dans l'univers
- Incertitudes *supplémentaires* liées à *certaines systèmes*
 - Systèmes complexes *non déterministes*
 - Systèmes complexes *instables* ou *chaotiques* (cf. planche suivante)
- Et pourtant... *prévisibilité* des systèmes complexes !
 - Il est quand même *possible*, dans de nombreux cas, de prévoir l'évolution *macroscopique* de systèmes complexes (*propriétés émergentes*)
 - Par *Connaissance*, par *Simulation*, ou par *Analogie* avec un système connu

3. Evolution de systèmes chaotiques ou instables

- Sensibilité « infinie » aux perturbations ou à l'imprécision des conditions initiales

- *Systèmes complexes instables*

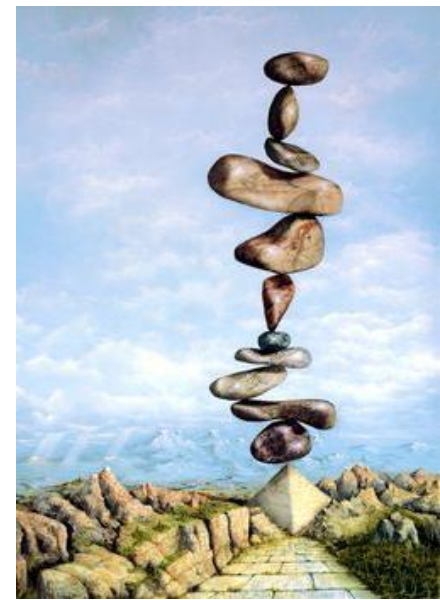
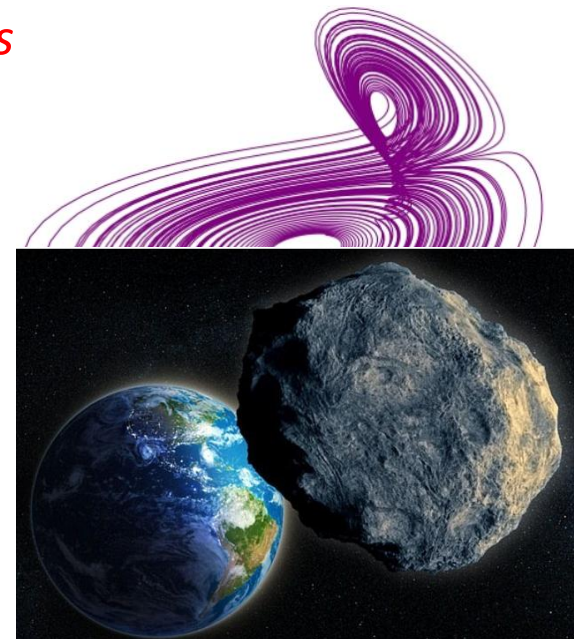
- Phénomènes d'*auto-amplification* à partir d'un *germe*
- « *Increasing returns* » en économie (Brian Arthur)
- Les *emballements collectifs*, les *modes*
- Les *crises* économiques ou boursières
- Une *décision* (oui/non) « sur le fil »

- *Prévisions météo* (chaotique)

- *Lancer de dé* (instable)

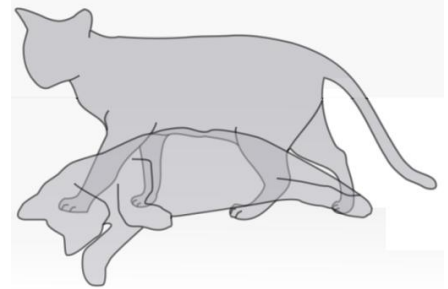


- Cet *astéroïde* va-t-il heurter la Terre ? (instable)



4. Incertitudes quantiques

- Phénomènes *quantiques* et leurs conséquences observables
 - *Mesures* effectuées par des appareils sensibles aux statistiques nucléaires
 - IRM, compteurs Geiger, scintigraphies...
 - Fiabilité des *microprocesseurs*
 - Diverses formes de *bruits de fond* (amplificateurs...)
 - *Images* photo *bruitées* en lumière faible (*grain*)
 - Le *chat de Schrödinger, avant* d'ouvrir la porte (*superposition, selon l'interprétation de Copenhague*)
 - Le *cerveau* et une *décision* « *sur le fil* »
 - Dans la mesure où le seuil d'activation neuronale serait sensible à des paramètres quantiques



5. Rencontres et Coïncidences

- **Rencontre** spatio-temporelle de plusieurs **processus indépendants**
 - Chaque processus peut être *aléatoire* ou *déterministe*
 - C'est *l'aléa temporel* (asynchronisme) et/ou *spatial* qui déclenche la coïncidence
 - On parle aussi d'événements « *fortuits* »
- Exemples très fréquents dans la **vie quotidienne**
 - *Rencontres* fortuites (parfois déterminantes...)
 - *Accidents, pannes, catastrophes...*
 - La *créativité* (rencontres sous un crâne)
 - Le *cerveau* et une *décision*
 - Une *prise de décision* peut être affectée par des circonstances *fortuites*, comme une *interruption* inopinée, ou même une *mouche*
- Certains processus mettent en jeu de **très nombreuses coïncidences**
 - *L'origine de la vie*
 - *L'évolution des espèces*



6. Exploitation délibérée du hasard

- La plupart des exemples précédents résultent de causes *naturelles* ou *spontanées*
- Mais il peut être judicieux de *provoquer délibérément* le hasard pour satisfaire certains *objectifs*
 - *Tirage au sort* d'un « volontaire » pour plus de *justice* (*courte-paille*,...)
 - *Tirage au sort* pour finaliser une *décision difficile*
 - *Jeux et paris* basés sur le hasard (*jeux de cartes*...)
 - *Créativité*, recherche de *solutions originales* en suscitant des *rapprochements aléatoires*
 - *Création artistique* par manipulation *contrôlée* du hasard
 - *Algorithmes* de tirages de *nombres (pseudo-)aléatoires* et leurs applications
 - *Simulations* par la méthode de *Monte-Carlo*
 - *Explorations optimales* comportant des pas aléatoires (*recherche d'optimums*)
 - Algorithmes de « *recuit simulé* » (*routage de circuits, organisation de réseaux*...)

Combinaisons de ces facteurs



- Exemple : Paul doit **prendre une décision**
 - Paul peut **ignorer** certains éléments importants du contexte (ex : il a oublié ce que lui a dit Jean)
 - Des perturbations externes imprévues (**coïncidences**) peuvent affecter son humeur, sa lucidité et sa décision 
 - Le cerveau de Paul est un **système complexe**, possiblement **chaotique** ou **instable** : de petites nuances au départ peuvent aboutir à des décisions opposées
 - Des **phénomènes quantiques** peuvent déclencher certaines activations neuronales, amplifiées ensuite par l'aspect chaotique du système (**chaos quantique**)
 - De guerre lasse, Paul tire la décision à **pile ou face...**

Résumé

Différentes catégories d'incertitudes

1. Ignorance de l'existant
2. Evolution des systèmes complexes
3. Evolution des systèmes chaotiques ou instables
4. Incertitudes quantiques
5. Rencontres et coïncidences spatio-temporelles
6. Recours délibéré au hasard
7. Combinaisons de ces facteurs

Incertitudes objectives vs. subjectives

- Incertitudes objectives
 - Ne dépendent pas de l'observateur
 - Vérifiables *expérimentalement*
 - *Systèmes instables ou chaotiques*
 - *Incertitudes quantiques*
- Incertitudes subjectives (épistémiques)
 - Dépendent de l'observateur pour un même phénomène
 - Traduisent *croyances, opinions, préférences, prédictions...*
 - *Ignorance totale ou partielle de l'existant*
 - *Systèmes complexes déterministes non chaotiques*
 - *Dépendent du contexte, des connaissances a priori*

Interprétation quantitative des probabilités

Sur quels fondements *calculer* ou
estimer une probabilité ?

Pourquoi faut-il une **interprétation** ?

- Une probabilité doit être **JUSTE** : elle doit s'accorder à la **réalité du monde** (dans un sens à définir)
- Une probabilité **FAUSSE** peut entraîner des erreurs et avoir parfois des **conséquences graves**. Ce type d'erreur est **fréquent** pour les probabilités subjectives
- Or la **théorie axiomatique** des probabilités est **abstraite** et indépendante de cette question... **philosophique**
- L'**interprétation quantitative** des probabilités a pour objectifs de **fonder** et/ou de **calculer** la valeur numérique des probabilités, quelles qu'en soient les causes

Principales interprétations

- Probabilités classiques
- Probabilités subjectives
 - Paris
 - « Dutch book »
- Propensions
- Interprétations fréquentistes
 - Cas fini
 - Cas infini
- Discussion
 - Défense de l'interprétation *fréquentiste infinie*

Probabilités classiques

- Laplace, Pascal, Bernoulli, Huygens, Leibniz... (XVII^{ème} et XVIII^{ème} siècles)
- Basées sur l'**hypothèse** de probabilités **équiréparties**
 - On suppose l'**équiprobabilité** ou la **symétrie** des cas élémentaires
 - **Probabilité d'un événement** $\stackrel{\text{def}}{=} (\text{Nombre de cas favorables}) / (\text{Nombre de cas possibles})$
 - En cas **d'ignorance totale** : une chance sur 2 (ou sur n selon le nombre de possibilités envisagées)
- **Problèmes**
 - Cette définition ne constitue pas un **fondement** de la notion de probabilité
 - Définir les **probabilités** à partir de **l'équiprobabilité a priori** revient à une définition **circulaire**...
 - L'équirépartition en cas d'**ignorance** peut conduire à des **absurdités**
 - L'équirépartition par **symétrie** peut être **illusoire**
 - Dés pipés, erreurs d'analyse...
- **En pratique**
 - Cette « définition » constitue un **moyen pratique et puissant** de **calculer** les probabilités lorsque la **symétrie** est avérée
 - Probabilités de distributions particulières au bridge, au poker... (*Analyse combinatoire*)
 - Modèles d'urnes, jeux de hasard, courte-paille...
 - **Rarement rencontré dans le monde naturel**

Probabilités subjectives

- Degré de croyance, de confiance, de préférence...
- Les estimations *a priori* sur la *vraisemblance* d'un événement sont fréquemment *subjectives*
- Ces probabilités peuvent être utilisées pour faire des *prévisions* et pour prendre des *décisions* (exemple des estimations *bayésiennes*)
- Il est *difficile de quantifier* une probabilité subjective
 - Influence des émotions, devinettes
 - Optimisme, pessimisme, « *wishful thinking* »
 - « Je ne sais pas : une chance sur deux... »
 - De nombreux *biais cognitifs* interfèrent avec une estimation subjective *rationnelle* de la probabilité*
 - Surestimation des petites probabilités
 - « Face est sorti 3 fois de suite : je mise sur Pile »

➤ *Comment faire ?*

Paris, Dutch book...

- Une manière « pratique » de quantifier une probabilité subjective
 - Quelqu'un désire *attribuer une probabilité subjective* p à un événement E
 - Pour cela, il *imagine accepter un pari* sur E où il gagne $k(1-p)$ contre kp
 - Il *définit* p comme *le* nombre lui permettant d'estimer que ce pari est *équilibré*
 - Cela signifie que sur un grand nombre de paris successifs, le joueur estime que ses gains auront tendance à équilibrer ses pertes (*espérance de gain nulle*)
 - NB : cette méthode ne *valide pas la valeur* d'une probabilité subjective, elle *aide* seulement à proposer un nombre « raisonné » *en accord avec la subjectivité*
- « Dutch book » : une *suite de paris* basés sur une probabilité subjective
 - *Théorème du Dutch Book* : 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 : si un joueur de pile ou face est persuadé qu'il existe une *mémoire* des tirages précédents (de sorte, croit-il, d'*équilibrer les fréquences*), alors il pourra accepter un pari perdant
 - *Remarque* : Ce théorème est basé sur la *loi des grands nombres*. Il exige la possibilité de *nombreux tirages*
 - A ce titre, il se ramène à la *conception fréquentiste* des probabilités (ci-après)

Les **propensions** (*propensities*)

- Concept soutenu à l'origine par Karl Popper (1957)
 - KP voulait rendre compte de probabilités physiques, en particulier **quantiques**
 - **D'où vient le fait** que la probabilité pour qu'un atome de radium se désintègre dans 1600 ans soit de $\frac{1}{2}$?
 - C'est parce qu'un atome de radium a la **propension** (naturelle) de se désintégrer en moyenne dans ce laps de temps
 - La probabilité est ainsi ramenée à une **causalité** qu'il est impossible (et inutile) d'expliquer d'une autre manière
 - Cette **définition** est **philosophique**. Elle n'exige pas **en principe** qu'on réalise des **expériences multiples**, même si la **vérification numérique** exige un grand nombre d'épreuves

Propensions (suite)

- Cette définition prétend s'étendre à de *nombreux types* de probabilités
 - Une pièce a 1 chance sur 2 de tomber sur Pile, car elle a la *propension*, lorsqu'elle est lancée, de tomber sur Pile autant que sur Face
 - C'est une propriété *de la pièce* (et de son processus de lancement), qui ne nécessite pas d'autre explication
- **Critiques** de cette interprétation
 - Cette terminologie *n'explique rien* : c'est comme la *vertu dormitive* de l'opium (une *propension* de l'opium à faire dormir...) (*convient aux probas quantiques*)
 - Cette « définition » ne s'applique qu'aux aléas *de nature causale*. Elle ne s'applique pas aux autres types d'aléas (*ignorance, coïncidences*)
 - En particulier, elle ne s'applique pas aux *probabilités bayésiennes* (*probabilités des causes*, voir plus loin)
 - L'objectif déclaré de se libérer du besoin de *tirages multiples* est *illusoire*
 - Il a fallu après Popper distinguer conceptuellement les propensions « *au long cours* » des propensions pour *tirage unique*

Interprétations fréquentistes

Préliminaire : rappel de la Loi 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 tirage

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

- Loi « forte » des grands nombres (*loi « probabiliste »*)

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

- NB : Cette loi est *démontrée*, pour toute *valeur de p*
 - Elle résulte des seuls *axiomes* abstraits de la théorie des probabilités
 - Elle est valide *pour toute probabilité p* satisfaisant à ces axiomes

Une confusion possible sur la LGN

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

- Cette formule se lit :
 - « Les fréquences convergent "*presque sûrement*" (= avec une probabilité de 1) vers p lorsque n tend vers l'infini »
- Mais certains comprennent :
 - « On peut être *presque sûr* que les fréquences convergent *vraiment* vers p lorsque n tend vers l'infini » : $\lim_{n \rightarrow \infty} (F_n) = p$ (limite *observée*)
- Or cette interprétation peut être *erronée*
 - Si un dé est *pipé*, la limite des fréquences d'une face *peut valoir 1/5* (par ex.)
 - Et pourtant la LGN *reste vérifiée pour $p = 1/6$* ... presque sûrement *selon $\mathbb{P}$*
 - *Problème* : la probabilité $\mathbb{P}$ est *définie à partir de p* . Si p est erroné, $\mathbb{P}$ l'est donc aussi (et même encore plus). La formule du haut *ne veut plus rien dire* si $\mathbb{P}$ est faux

➤ La loi des grands nombres semble être contredite par l'observation, car elle est *toujours démontrée*, même pour des *probabilités erronées*

Loi probabiliste vs. loi physique

- 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
- C'est la « loi physique » : $\lim_{n \rightarrow \infty} (F_n) = p$
- La loi physique 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
- C'est cette définition qui sera retenue pour *l'interprétation fréquentiste*

Interprétations fréquentistes

- Fréquences à tirage fini

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

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

- *Problème* : même à pile ou face, on trouve $p \neq 0,5$!
- *Définition instable* : le résultat dépend de n ; comment choisit-on n ?

- Fréquences hypothétiques à tirage infini

- On *prolonge* la définition précédente en faisant tendre n vers *l'infini*
- On envisage, par une *expérience de pensée*, d'effectuer une *infinité de tirages*, et d'en déduire :

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

- **NB** : Si l'on désire obtenir p *avec une précision donnée*, il est *suffisant* de considérer un nombre *fini* de tirages, assurant la précision demandée

Arguments contre le fréquentisme infini

- Certains auteurs* sont **opposés** au fréquentisme infini (ou hypothétique)
- Leurs **arguments** principaux sont :
 1. Ces tirages infinis successifs sont **impossibles** à réaliser. Certaines successions sont même **impensables**
 2. Cette définition est donc **impraticable** et par suite **dénuée de sens**
 3. **Rien** ne garantit que la suite des fréquences soit effectivement **convergente** à l'infini

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

Défense du fréquentisme infini

- *Aucune autre valeur* que p ainsi défini ne satisfera la loi physique des grands nombres (limite observée)
- En conséquence, *toute autre valeur* que p sera *erronée*
- Une définition par une *expérience de pensée* (ou « *contrefactuelle* ») est parfaitement admissible et même courante en philosophie ; par ex :
 - La *fragilité*
 - La *solubilité*
- Le caractère irréalisable de l'expérience de pensée n'est donc *pas un obstacle*
- De plus, on peut *approximer* p *autant qu'on veut* sans devoir recourir à un tirage *infini*
- *Aucune autre définition* ne peut donner un résultat correct (sauf si elle est *équivalente*)



Mise en pratique du fréquentisme infini

- *En pratique*, toute estimation (par exemple pour un événement non répétable) *devra s'approcher le plus possible* (en pensée) de cette conception
 - Même si on n'est pas capable d'appliquer en pratique cette définition, elle doit *servir de base au raisonnement*, afin de rechercher ou d'approcher la probabilité *juste*
 - Une approche alternative, plus « pratique » même si elle reste contrefactuelle, peut être proposée grâce à *l'ergodicité* :
 - Au lieu d'imaginer *plusieurs tirages successifs* de la même expérience, on peut considérer *un seul tirage*, appliqué à un grand nombre de phénomènes similaires
 - *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. Cela évite de faire « érupter » plusieurs fois le même volcan, ce qui est en effet *impensable*
 - Dans certains cas, on pourra réaliser un *modèle* et effectuer des *simulations*
 - L'*expérience de pensée* sera alors remplacée par une *simulation informatique*

Objection de la **non-convergence**

- La série des fréquences F_n est-elle **convergente** ?
 - Je ne suis pas encore parvenu à le **démontrer**, mais sans doute une démonstration existe-t-elle, au moins sous **certaines conditions de régularité** (?)
 - Dans de **toutes les applications** pratiques (Pile ou Face, etc.), la convergence est **observée**
 - En toute rigueur, il faudrait en attendant reformuler la définition du fréquentisme infini :
« [...] p est la limite, **si elle existe**, des fréquences observées ou imaginées »

Quelques remarques

Remarque sur les probabilités nulles

- On a tendance à *croire* qu'un événement de probabilité nulle *ne peut pas se produire* en pratique
- Or dans un univers où il existe une *infinité de possibilités*, *tous* les événements qui se produisent en fait avaient, *à l'avance*, une probabilité nulle ($1/\infty$) de se produire
- Exemples
 - *On pointe avec une aiguille* sur le segment $[0,1]$ A diagram showing a horizontal line segment from 0 to 1. A needle is shown pointing to a specific point on this segment. A pink squiggly line is drawn above the needle, representing the path of the needle's tip.
 - Quel que soit la coordonnée obtenue, elle avait à l'avance une *probabilité nulle* d'être précisément pointée
 - Il y a également une *probabilité nulle* pour qu'un M. Martin se gratte le coude gauche, dans un escalier à Niort, à minuit pile le 13 mai 2023
 - Et pourtant cela peut parfaitement arriver, comme une *infinité* d'autres événements tout aussi ordinaires

Interprétations des probabilités nulles

- 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*] (comme de pointer sur 2 si on est restreint au segment $[0,1]$)
- Mais ces deux interprétations fondamentalement différentes *ne se distinguent pas* en théorie des probabilités
 - Un événement de probabilité nulle est-il *impossible* ou bien *possible* et même *vraisemblable*, mais *imprévisible* à l'avance ?
- Cette ambiguïté peut entraîner des *confusions* lors de *l'interprétation physique* des probabilités nulles ou très petites
 - Ex : voir ci-après une objection à la « preuve » de *l'existence de Dieu*

0 1



Dépendance au temps des incertitudes

(temps écoulé entre la prévision et la constatation)

- Incertitudes **indépendantes** du temps
 - Probabilités d'événements déjà déterminés
 - Lancers de dés, Pile ou Face, tirages dans une urne...
- Incertitudes **croissantes** avec le temps
 - Processus évolutifs
 - Chaos
 - Systèmes complexes
 - Processus de diffusion
- Incertitudes **décroissantes** avec le temps (*certitudes à long terme*)
 - Pannes, usure, mortalité
 - Désintégration radioactive
- Incertitudes **portant sur le temps** lui-même
 - Coïncidences temporelles

Applications des probabilités

Probabilités bayésiennes

Décisions en univers incertain

Probabilités bayésiennes

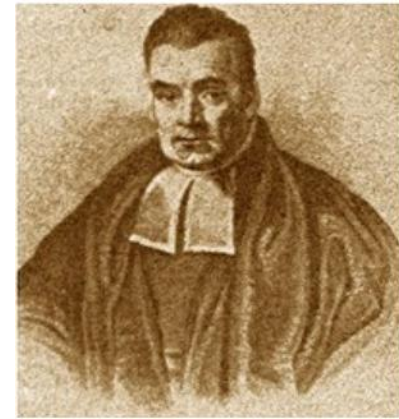
Application aux recherches des causes

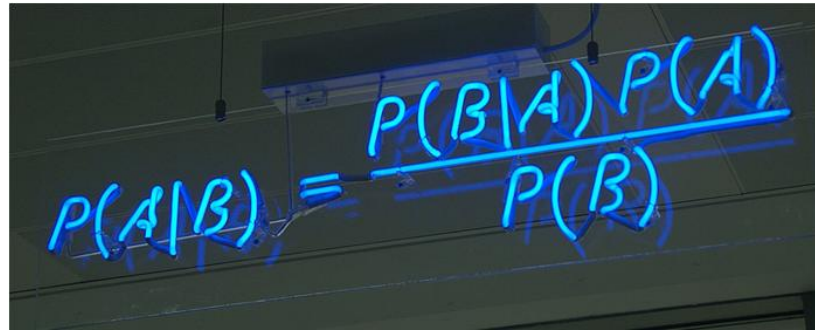
Une « preuve » bayésienne de
l'existence de Dieu

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 blackboard with the formula for Bayes' Theorem written in bright blue chalk. The formula is $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$. The text is slightly tilted to the right.

« Théorème » ou « formule » de Bayes

Démonstration

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

Application aux probabilités des causes

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

- 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 =
- C'est le cas en particulier pour déterminer la *probabilité des causes* d'un phénomène *observé*

Exemples de recherche de causes

- O : *observation* ou *mesures*
 1. Fumerolles, grondements et odeur de soufre sur un volcan
 2. Ensemble de symptômes et d'analyses sur un patient
 3. Un panneau routier lointain et peu lisible
- $H_1, \dots, H_n$: *causes hypothétiques* envisagées de O
 1. Le volcan prépare-t-il une éruption ? (*Oui / Non*)
 2. Quel est le bon diagnostic ? (*Hépatite B / Cirrhose / Indigestion*)
 3. Quelle ville indique ce panneau ? (*Lyon / Lens / Laon*)
- *Probabilité a priori* de chaque hypothèse H_k , en fonction du contexte, *indépendamment de l'observation O*
 1. Ce volcan est peu actif, une éruption est très peu probable
 2. Ce patient est alcoolique, une cirrhose serait vraisemblable
 3. Nous sommes dans l'Aisne : plutôt Laon que Lyon...

Probabilité bayésienne des causes

- Soit O une *observation* ou un ensemble de *mesures*
- On envisage pour O des causes possibles *hypothétiques* $H_1, \dots, H_n$
- On recherche la cause *la plus probable* parmi les H_k
- Dans ce contexte, la *formule de Bayes* s'écrit :

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

- $\mathbb{P}(O | H_k)$: Si l'hypothèse H_k est juste, c'est la probabilité pour qu'elle *rende bien compte* de O
 - $\mathbb{P}(H_k)$: C'est la *probabilité a priori* de l'hypothèse H_k c'est-à-dire sa vraisemblance *sans tenir compte* de O
 - $\mathbb{P}(O)$: Est parfois *ignorée* ou se *calcule* à partir des autres probabilités
- Cette formule sera appliquée à *chaque hypothèse* et utilisée pour déterminer *la cause la plus probable*, qui *maximise* $\mathbb{P}(H_k | O)$

Exemples d'applications

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 $\mathbb{P}(S|B) = 25\%$ *
- *A priori*, 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*
- *Aïe ! Que faire ?*
 - Il faudra *valoriser* les risques
Cf. Les décisions valorisées

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 a priori* $p = \mathbb{P}(D)$ pour qu'une telle entité puisse exister ; par exemple :
 - $p = 1$ (la *Foi*)
 - $p = 0,5$ (une attitude d'*ignorance ouverte*, « tolérante »)
 - $p = 10^{-1000}$ (un *très grand sceptisme*)
- 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

- $D = \{\textit{Dieu existe} : \text{créateur, tout-puissant, éternel}\}$
- $O = \{\textit{Observations} : \text{le monde, la vie, l'homme...}\}$
- Les observations étant ce qu'elles sont, quelle est la *probabilité* $\mathbb{P}(D|O)$?
 - On cherche à *démontrer* qu'elle vaut 1
- Pour cela on s'appuie sur la *formule de Bayes* :

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

NB : Le dénominateur $\mathbb{P}(O)$ est développé selon la formule des « probabilités totales »

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

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

- $\mathbb{P}(O|D) = d$: un nombre *proche de 1* (si Dieu existe, le monde est comme il est car !! l'a voulu ainsi)
- $\mathbb{P}(D) = p$: retenons $p = 10^{-1000}$ (*très grand scepticisme*)
- $\mathbb{P}(O|\bar{D}) = \varepsilon$: un nombre *infinitésimal* [$1/(10 \wedge 10^{123})$ selon *Penrose*]* (Pourquoi *ce* monde parmi une *infinité* de mondes possibles?)
- Comme $\varepsilon \ll d.p$, la *formule de Bayes* nous donne :

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

Commentaires

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

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

Objection 1

- 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 peut-être 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* par rapport à ε , pour tous ceux qui ne sont pas *résolument* athées

Objection 2 (*plus subtile ?*)

- Lorsque la probabilité d'un événement est *nulle* ou *infime*, on peut considérer cet événement, soit comme *impossible*, soit comme *possible* et même *vraisemblable*, mais *imprévisible*
- Le monde tel qu'il a évolué à partir du Big Bang, était *scientifiquement vraisemblable*, mais a priori *imprévisible*
- On peut à l'inverse considérer Dieu comme *invraisemblable*, car se situant résolument *en dehors de toute science*
- 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* : on traite de la même manière *vraisemblable* et *invraisemblable*
 - L'approche bayésienne perd peut-être sa validité dans ce cas extrême

Objection 3 : généralisations absurdes

- Le raisonnement présenté peut s'étendre à l'identique à toute théorie au fort *pouvoir explicatif apparent, aussi absurde soit-elle*
 - Intervention miraculeuse de *Saint-Antoine* pour avoir retrouvé 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*, même si elle est peu crédible, ce même raisonnement bayésien *conforte cette hypothèse* en lui attribuant une probabilité de 1 (elle est *prouvée* « *presque sûrement* »)



Extension : Philosophie bayésienne

- On évalue la probabilité d'une cause en *combinant deux facteurs* d'informations probabilistes indépendants
 - Un facteur lié à l'observation : la cause hypothétique évaluée est-elle une *cause possible* de l'observation ? (facteur de *causalité*)
 - Un facteur a priori : Quelle est le degré de *vraisemblance* de cette cause dans le contexte donné, *indépendamment de l'observation* ? (facteur de *vraisemblance*)
- Cette approche s'étend aux évaluations *qualitatives*
 - On combine ces deux types d'informations *sans faire de calculs*
 - On parle alors de *cerveau bayésien*
 - Exemple : *Reconnaissance intuitive de formes* (planche suivante)

Le cerveau bayésien (exemple)

A B C I 2 I 3 I 4

I 3

2 causes également possibles
2 contextes différents

Les **décisions valorisées** dans un environnement incertain

Gains ou Coûts aléatoires...
Quelle est la meilleure décision ?

Contexte de la **décision valorisée** (*modèle simplifié*)

- On fait face à une situation où peuvent survenir des **événements** aléatoires $E_1, E_2, \dots$ **incompatibles**, de probabilités $p_1, p_2, \dots$
- **Avant** de savoir quel événement se produira, on doit **choisir** entre **plusieurs décisions** $D^1, D^2, \dots$ dont les conséquences seront différentes selon celui des événements E_k qui surviendra
- On attribue des **valeurs** V_k^i (**gains** ou **coûts**) à ces décisions
 - V_k^i : **valeur** résultant de la décision D^i si l'événement E_k vient à se produire
 - Par convention $V_k^i > 0$ est un **gain**, $V_k^i < 0$ est un **coût**
 - Ces valeurs peuvent être **monétaires** ou **subjectives**
- Par exemple, E_1 ou E_2 : ce volcan va-t-il connaître une **éruption** ?
 - D^1 ou D^2 : le préfet va faire **évacuer la ville**, ou non
 - Les V_k^i mesurent les **conséquences** (gains ou coûts) d'une **bonne** ou d'une **mauvaise décision** (i.e. **catastrophe humanitaire** ou **évacuation inutile...**)
 - Cf. querelle Allègre – Tazieff sur la **Soufrière** (1975)



Critère de l'espérance de gain maximale



- *Espérance mathématique* : concept développé au XVII^{ème} siècle (Pascal, Huygens)
- *L'espérance de gain* de la décision D^i s'écrira :

$$G(D^i) = \sum_k p_k V_k^i$$

(V_k^i : valeur résultant de D^i si l'événement E_k , de probabilité p_k , se produit)

- La décision « *optimale* » selon ce critère est celle qui maximise *l'espérance de gain* :

$$D^i = \arg \max G(D^i)$$

- Ce critère est *universellement* considéré comme *le* critère de décision « *rationnel* »

Justification et limites du critère de l'espérance de gain maximale

- On *démontre* (LGN) que l'espérance mathématique de gain d'une décision est la *valeur limite* vers laquelle tendra le gain si on effectue un *grand nombre de décisions identiques* indépendantes successives
- La *justification théorique* de ce critère n'est donc valable que si on effectue un *grand nombre de décisions* successives
- On dit pourtant qu'un pari est *équitable* si chaque joueur a la *même espérance de gain, même si on ne parie qu'une fois*
- Ce critère est *limité*, voire parfois *absurde*, dans le cas d'une *décision unique* (cf. planches suivantes)



Le Paradoxe de St-Petersbourg

- Les origines du paradoxe

- Ce paradoxe est dû à Nicolas Bernoulli (1713)
- Il a été étudié par son cousin Daniel, en poste à St-Petersbourg
- Ce dernier a proposé pour le résoudre les *fonctions d'utilité* (cf. planches suivantes)

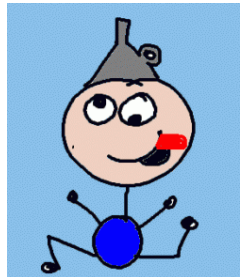
- Pierre et Paul jouent à Pile ou Face

- Pierre verse une certaine somme, qui reste acquise à Paul
- La partie commence, et s'arrête *dès que Face sort*
- Si Face sort au *$n^{\text{ième}}$ coup*, Paul verse *2^n ducats* à Pierre
- *Combien* Pierre accepte-t-il de verser à Paul pour avoir le *droit de participer* à ce jeu ?



Pourquoi est-ce un paradoxe ?

- Calcul de *l'espérance de gain* de Pierre (*hors versement initial*)
 - Si Face sort pour la première (et unique) fois au *$n^{\text{ième}}$ coup* :
 - La *probabilité* de cet événement est $(1/2^n)$ [*$n-1$ Pile, puis 1 Face*]
 - Le *gain* de cet événement est de 2^n ducats [*d'après la règle*]
 - L'*espérance de gain* correspondante est $(1/2^n) \times (2^n) = 1$ *ducat*
 - En *sommant sur toutes les valeurs de n* , l'espérance totale de Pierre est de :
 - $1 + 1 + 1 + 1 + \dots = \infty$ *ducats*
- L'espérance de gain est donc *infinie* : Pierre a *intérêt* à jouer à ce jeu pour *n'importe quelle mise initiale* (par exemple, il serait encore gagnant en misant *100 millions de ducats* !)
- Mais *personne de sensé* ne miserait ne serait-ce que *100 ducats* pour jouer à ce jeu !!
 - En effet, Pierre a déjà *15 chances sur 16* de n'empocher qu'*au plus 16 ducats* en contrepartie de son *énorme mise initiale*. Il n'arrivera *jamais* aux grosses sommes. Il n'est pas fou !!!
- *Faut-il être fou* pour faire confiance au critère de *l'espérance mathématique* ?
- C'est le *paradoxe de St-Petersbourg*, qui a fait couler beaucoup d'encre...



Les réponses « classiques »

- Daniel Bernouilli (1738)
 - Pour lui, le problème vient de la *sous-estimation* par Pierre de la *véritable valeur* de ses gains G (terminologie moderne : *aversion au risque*)
 - Bernouilli a donc créé le concept de *fonctions d'utilité* $U = f(G)$, et il propose de retenir le critère de l'*espérance d'utilité maximum*
 - Mais ce critère oblige à créer pour chaque cas une fonction d'utilité *ad hoc*. De plus, il faudrait une fonction d'utilité *absurde* pour *résoudre ce paradoxe*
 - *NB* : Cette remarque ne remet pas en cause l'intérêt, par ailleurs élevé, des fonctions d'utilité
- Emile Borel *
- Borel met d'abord en avant le caractère *irréaliste* du jeu (les joueurs devraient être *infiniment riches* et également *immortels*...). Mais, même en *modifiant les règles* pour les rendre plus réalistes, le paradoxe demeure...
- Borel constate alors qu'il n'y a qu'une *probabilité infime* pour que Pierre atteigne une grosse somme, et pour Borel, les événements de probabilité minuscule *n'arrivent jamais* *
- Ces réponses résultent du besoin implicite de *sauver le dogme* du critère d'*espérance maximale* comme *seul critère possible d'équité*. Du coup ces auteurs vont rechercher d'autres arguments...

* Emile Borel, *Valeur pratique et philosophie des probabilités*, Jacques Gabay (1939)

Mon point de vue

- Je suis *presque* d'accord avec la solution de Borel...
 - « Les événements de probabilité minuscule n'arrivent jamais »
- ...Mais *pas tout à fait d'accord* ; j'ajouterais :
 - « Les événements de probabilité minuscule n'arrivent qu'au bout d'un *très grand nombre de répétitions* »
- Or, le problème est que Pierre et Paul ne jouent qu'*une seule partie*
 - Pierre ne serait gagnant à long terme, en théorie, que s'ils pouvaient jouer un *nombre illimité de parties*, Pierre misant une grosse somme *avant chaque partie*, et chacune pouvant durer un *temps illimité*
 - Ces règles absurdes ne sont pas celles du jeu de St-Petersbourg
- Mon point de vue est donc le suivant
 - Borel essaye, lui aussi, de sauvegarder le critère d'*espérance maximale*
 - Mais il n'y a *pas de justification* pour appliquer ce critère à une décision ou une épreuve unique
 - Il faudra rechercher *d'autres critères de décision*

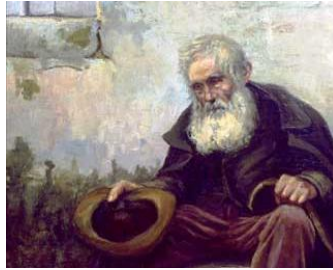
Un exemple d'aversion **rationnelle** au risque

- Le mendiant et le marchand
 - Un mendiant possède un billet de loterie lui permettant de gagner **20 000 €** avec une probabilité de $\frac{1}{2}$
 - *Son espérance de gain est de 10 000 €*
 - Un riche marchand lui propose de lui racheter son billet pour **9 000 € « cash »**
 - Le mendiant **accepte**
 - *Son espérance de gain passe à 9 000 €*
 - Une perte **irrationnelle** d'espérance ?
- Qu'auriez-vous fait à sa place ?



2 décisions opposées et... rationnelles

- Le mendiant



- Il a besoin d'argent ; 20 000 €, ou même 9 000 €, vont le *sortir de la misère* pendant un certain temps
- Risquer de ne *rien avoir du tout* serait *catastrophique*
- Il applique un adage de *bon sens* : « *mieux vaut tenir que courir* »



- Le riche marchand

- Il est prêt à perdre 9 000 €, ça ne va *pas le ruiner*
- En bon marchand, il serait content de *doubler sa mise*
- Et puis il a *bon cœur* (pourquoi pas ?), ça lui *plaît* de *dépanner* ce mendiant sympathique

- *Ils ont tous les deux raison !*

Renoncer au **dogme** de l'espérance maximale ?

- La plupart des auteurs considèrent comme *admis* que les *seules* décisions *rationnelles* sont basées sur la maximisation de l'*espérance de gain*
- Les personnes adoptant un autre comportement sont considérées comme effectuant des *choix irrationnels*
- Des « *mathématiciens-psychologues* » se sont penchés sur leurs cas et ont posé leurs *diagnostics*
 - *Aversion au risque* (considérée comme *irrationnelle*)
 - Travaux de Tversky et Kahneman* sur les *biais heuristiques* (également considérés comme *irrationnels*)
 - Fonctions d'utilité subjectives *ad hoc*, *distorsions* de probabilités...
- Or, pour des décisions *non répétables*, de tels comportements n'ont *rien de nécessairement irrationnel*
 - Il faut rechercher *d'autres critères* ou *d'autres approches*

* A. Tversky & D. Kahneman, *Judgment under uncertainty: Heuristics and biases* (Science, 1974)

Quels autres critères utiliser ?

- Rappel de l'*espérance de gain d'une décision* D^i : $G(D^i) = \sum_k p_k V_k^i$
 V_k^i est la valeur résultant de D^i si l'événement E_k , de probabilité p_k , se produit
- Au lieu de calculer les espérances de gain $G(D^i)$, on peut aussi considérer *séparément* les coefficients V_k^i (exemples planches suivantes)
 - Pour *maximiser son gain maximal possible*, on choisit la décision D^i ayant *le plus grand coefficient* V_k^i positif de toutes les décisions (*appât du gain*)
 - Pour *minimiser la perte maximale possible*, on choisit la décision D^i *minimisant le plus négatif* de ses *coefficient* V_k^i (*aversion au risque*)
- Ces critères maximax et minimax n'ont *rien d'irrationnel*, à condition cependant de tenir compte, au moins *qualitativement*, de la *probabilité* p_k associée à la valeur V_k^i ayant déclenché la décision
 - Si ce p_k a une valeur éloignée de 0, le critère est « raisonnable »
 - Si ce p_k est très proche de 0, il faut relativiser l'importance du critère
 - La valeur *précise* de p_k n'est pas importante, elle ne le serait que pour une série de *décisions répétées* (et c'est heureux car p_k est en général *inconnu*)

L'affaire des poisons (suite)

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



Approches non analytiques

- 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 (des experts, la population...)... cette approche combine, parfois *inconsciemment*, des *paramètres beaucoup plus nombreux et subtils* que l'approche analytique*
- Les décisions prises 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
 - *Une approche « rationnelle » peut souvent être trompeuse*, car on est *incapable* d'évaluer correctement les probabilités et les coûts impliqués ; on ignore aussi de *nombreux paramètres* qui restent *inconscients*. Cette approche n'a alors *de rationnelle que le nom...*

*Ap Dijksterhuis, "Think Different: The Merits of Unconscious Thought in Preference Development and Decision Making", *Journal of Personality and Social Psychology* (2004)

Merci de votre attention !

Compléments

Minimax et Principe de Précaution

- Dans ce cas, on ne considère que **deux décisions** D^1 , D^2 et **un seul événement** E
 - E est un événement **catastrophique** dont la probabilité p est en général **inconnue**
 1. **OGM** : risques pour la santé publique, les récoltes, domination mondiale de Monsanto
 2. **Gaz et pétrole de schiste** : pollution à l'extraction, pollution (CO_2) à l'utilisation
 3. **Centrales nucléaires** : catastrophe majeure, fuites radioactives, déchets
 4. **Réchauffement climatique** : catastrophes en série, disparitions d'îles, zones glaciaires, et pour certains : fin de la vie sur Terre
 - D^1 = agir préventivement contre E
 1. **Interdire** la consommation et la recherche sur les OGM
 2. **Interdire** la production et la recherche sur le gaz ou le pétrole de schiste
 3. **Démanteler** les centrales et favoriser les autres énergies (**éoliennes, centrales à charbon...**)
 4. **Révolutionner l'économie mondiale** pour limiter les émissions de CO_2
 - D^2 = agir dangereusement
 1. **Autoriser** certains OGM, encourager des producteurs français à faire concurrence à Monsanto
 2. **Encourager** recherches et expérimentations sur des techniques de production **propres**
 3. **Gérer** le parc nucléaire en fonction d'une politique globale et rationnelle à long terme (risques, coûts, sécurité...)
 4. **Affecter des milliards** à l'adaptation (digues, systèmes d'alerte et de prévention...) plutôt qu'à une politique ruineuse
- Le principe de précaution tient compte de **deux coûts** et **ignore la probabilité** p
 - **Coût de la catastrophe** (il est toujours extrêmement élevé)
 - **Coût de l'action préventive** (on devrait en tenir compte, en principe...)
- Le principe de précaution **ignore les gains éventuels** liés à la décision D^2
 1. Gain éventuel des **OGM** (épargne les insecticide, augmente les rendements, la qualité ou le goût)
 2. Gain éventuel de la production de **gaz de schiste** (indépendance énergétique de la France, économie)
 3. Gain d'**indépendance énergétique**, bénéfices de l'exportation de technologies de pointe, réduction des émissions de CO_2
 4. Bienfaits de **sauver des vies et des villes**
- Il s'agit d'une application du **critère minimax** : un cas, parfois extrême, d'**aversion au risque**

La théorie axiomatique des probabilités

Exposé élémentaire

Axiomes de Kolmogorov



- Les probabilités sont définies sur un ensemble Ω (*univers*) dont les éléments ω sont appelés des *événements élémentaires*
- Certains sous-ensembles de Ω sont dits *probabilisables* : ce sont les *événements* (élémentaires ou composés)
- Pour tout A probabilisable, on définit une *fonction réelle* $\mathbb{P}(A)$ dite *Probabilité de A* , respectant les axiomes suivants :
 - $\mathbb{P}(A) \geq 0$ et $\mathbb{P}(A) \leq 1$
 - $\mathbb{P}(\Omega) = 1$
 - Pour toute suite *finie* ou *dénombrable* $A_1, A_2, \dots$ composée d'événements *2 à 2 disjoints*, on a :

$$\mathbb{P}(A_1 \cup A_2 \cup \dots) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \dots$$

Probabilités conditionnelles

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

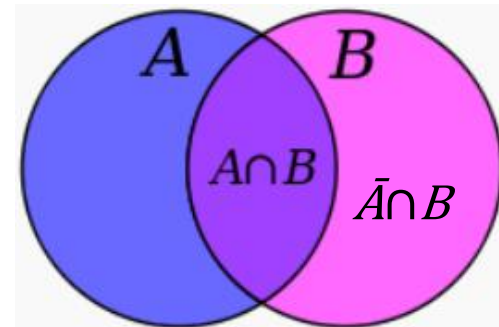
Se lit : « Probabilité de A sachant B », ou :

« Probabilité de A étant donné B », etc.

Pourquoi cette formule ?

- 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 B

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



Indépendance

- Indépendance de 2 événements A et B :

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

- Pourquoi cette formule représente-t-elle *l'indépendance* de ces deux événements ?

D'après la définition des probabilités conditionnelles, on peut déduire de cette formule que :

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

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

→ 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

Convergences à l'infini

Lois des grands nombres

Convergence vers la loi normale

Planche de Galton

Les lois des grands nombres



Loi 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 tirage

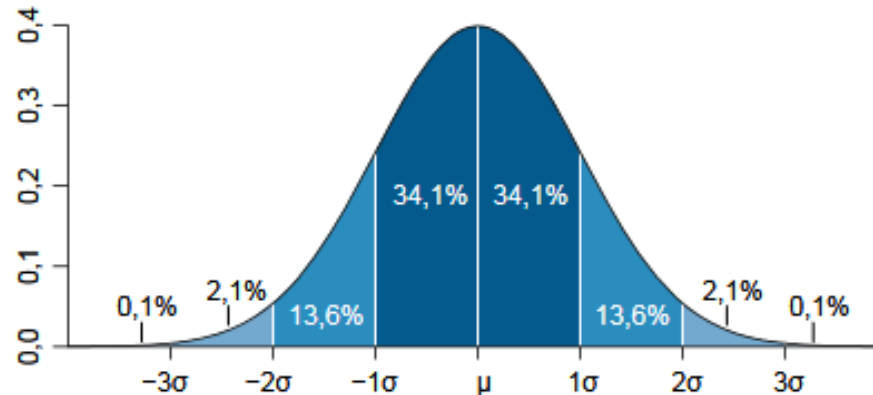
$$F_n(A) = \frac{(\text{nombre de réalisations de } A)}{n}$$

- Loi « forte » des grands nombres (*loi « probabiliste »*)
 - Si on effectue un nombre n croissant de tirages indépendant, la fréquence $F_n(A)$ va tendre vers p avec une probabilité de 1

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

(en jargon probabiliste, F_n va converger « *presque sûrement* » vers p)

La loi normale (*Laplace-Gauss*)



- La loi normale (courbe en cloche) est *omniprésente* en probabilités
- En effet, elle apparaît comme la distribution limite d'une *somme* de nombreuses variables aléatoires de *n'importe quelle distribution*, mais *identiques* et *indépendantes* (« *théorème central limite* »)
 - En particulier, une *mesure* entachée d'une *somme de petites erreurs indépendantes* se distribue selon une *loi normale* centrée sur la valeur moyenne de la mesure
 - Illustration : *Planche de Galton*

La planche de Galton

- Principe

Chaque bille est déviée par une *somme* de petits sauts aléatoires *indépendants*, vers la gauche ou la droite, avec une probabilité $\frac{1}{2}$

On se retrouve dans le deuxième cas d'application du *théorème central limite* (somme de petites erreurs indépendantes)

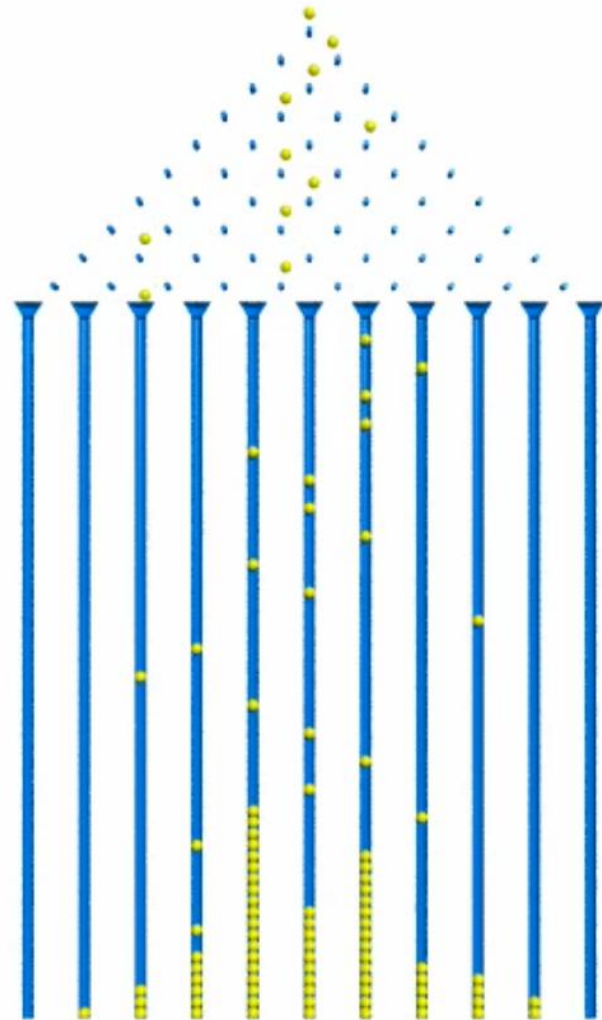


Planche de Galton

- Lancer l'animation

Illusion de la mémoire des tirages passés

- A la mi-temps de l'animation précédente, la pile centrale a « pris du retard »
- Par la suite, ce retard a été « rattrapé » et l'on retrouve à la fin une belle courbe en cloche
- Ce type de constatation est à l'origine d'une erreur très répandue : les probabilités auraient la *mémoire* des tirages passés, et auraient tendance à « *rattraper* » les *retards*, ou à corriger ce qui pourrait apparaître comme une « *injustice* »

