

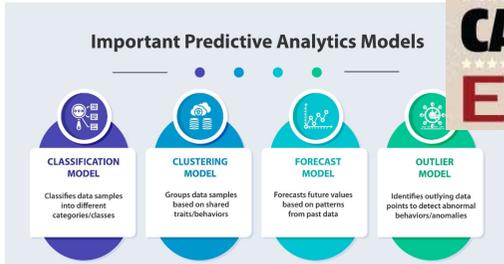
## Données et prévisions boursières



FOURTH EDITION  
**FACT  
FICTION  
AND  
FORECAST**  
Nelson Goodman

**BLOOMBERG**  
vs  
**FACTSET**  
vs  
**CAPITALIQ**  
vs  
**EIKON**

**Financial Modeling**  
[fə-'nan(t)-shəl 'mæd-lɪŋ]  
A forecast of a company's future financials based on assumptions and past performance.  
Investopedia



## Plan des diapositives

- Prévisions boursières ?
- Induction et prévision
- Validité logique de l'induction
- D'où vient le raisonnement inductif ?
- Encodage de l'information financière
- Multiplicité des règles de prévision
- La « nouvelle énigme de l'induction » : règles de prévision projetables
- Arbitrage statistique

## Prévisions boursières

5

## Prévoir le futur à partir du passé



La question est bien sûr la suivante : est-ce ces **relations entre données passées et futures** sont *fortuites* ou sont-elles *significatives* ?

6

## Prévoir le futur à partir du passé

- Fonction indicatrice de hausse du prix :
  - $I_t = 1$  si  $P_t - P_{t-1} > 0$
  - $I_t = 0$  si  $P_t - P_{t-1} \leq 0$
- Une chronique de hausses et de baisses est donc représentée par une suite binaire du type 1010001011
  - Ici, pour simplifier l'analyse, on ne prend pas en compte l'amplitude des hausses ou des baisses
  - Dans le cadre du HFT (High Frequency Trading), les prix ne varient pas plus que d'un échelon de cotation (tick)
- Supposons que les 10 dernières variations de prix à la hausse ou la baisse soient représentées par 1010001011
- Prévoir à partir de ces observations revient à compléter cette suite : 10100010111 ou 10100010110

7

## Prévoir le futur à partir du passé

- Prévoir à partir de ces observations revient à compléter cette suite : 10100010111 ou 10100010110
- Dans cet exemple, ce qui sert à prévoir (information disponible, inputs pour la prévision) se limite à la connaissance des hausses ou baisses du titre risqué pour les dix dernières transactions.
- Si l'on pense que les variations de prix d'un autre actif précèdent celui qui nous intéresse (phénomène lead-lag), on peut inclure ces variations dans l'ensemble des informations pertinentes
- 1010001011  
0100010110 la seconde ligne représente ici les variations de prix de l'actif lead

8

## Prévoir le futur à partir du passé

- Considérons la représentation binaire (en base 2) d'un nombre : 11010 ... (suite infinie de 1 et de 0)
  - *En finance, 1 (hit) si hausse des prix, 0 si baisse*
- « Machine learning » : prévoir à partir de données en repérant des « patterns » (motifs) répétitifs
  - *Exemple : les suites périodiques répètent un même motif.*
  - *La suite 1010101010 est une répétition du motif 10.*
    - La fréquence des 0 et des 1 est égale à  $\frac{1}{2}$  : **il y a autant de hausses que de baisses**
  - **Mais on peut gagner à tous les coups, en vendant après une hausse et en achetant après une baisse : stratégie « contrarian »**

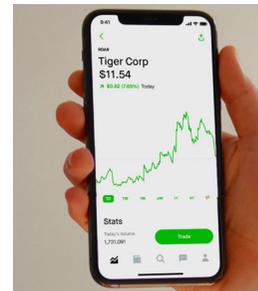
9

## Prévoir le futur à partir du passé

- L'information qui peut être utilisée pour la prévision peut être de nature financière et comptable, annonces de résultats, notes d'analystes financiers, ...
  - Voir la suite des transparents à propos du rôle des analystes financiers.
- L'information peut provenir de l'analyse textuelle - des tweets ou des discussions sur les forums en ligne, notamment reddit
  - <https://www.reddit.com/r/RobinHood/>
  - *RobinHood est un acteur majeur du brokerage en ligne à prix cassés*



10



1

## Utilisation d'informations textuelles

Social media



Natural language processing



Machine Learning



Sentiment analysis (état d'esprit des investisseurs)

Heart and Mind: Sentiment analysis in R



12

## Comment prévoir le futur à partir du passé

- Idée de base du machine learning
- Soit le motif (pattern, feature) 1010001011, on recherche une **fonction** qui dont l'entrée (input) est ce motif
  - Ici un nombre codé sur 10 bits
- Et la sortie (output) est soit 1 soit 0.
  - On a utilisé le terme « fonction » – l'entrée est une **variable**.
  - On aurait pu utiliser le terme **programme** ou **algorithme**.
- Dans notre exemple, il y a  $2^{10} = 1024$  entrées possibles
- Pour **calibrer** la fonction, on a une base d'apprentissage, constituée de motifs sur 11 jours.
- On va rechercher les motifs identiques (ou proches) de 1010001011

13

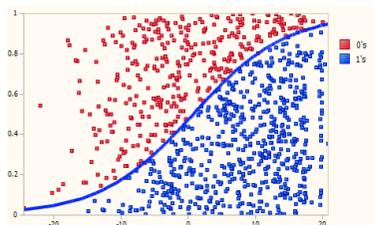
## Comment prévoir le futur à partir du passé

- Pour les motifs identiques à 1010001011, on va examiner si le successeur est un 1 ou si c'est un 0
- On peut aussi voir ce problème de **prévision**, comme une **classification** (voir infra)
  - Pour un motif donné, on veut savoir s'il est suivi d'une hausse ou d'une baisse
  - **Deux classes** : les antécédents d'une hausse et les antécédents d'une baisse
- Il s'agit d'**apprentissage supervisé** : on indique ce qui doit être prévu (hausse ou baisse) dans l'échantillon de données « historique » (base d'apprentissage)

14

## Comment prévoir le futur à partir du passé ?

- Classification binaire : chaque point est associé à un motif de variation des cours observé



- Si un motif est en bleu, il est suivi d'une hausse (attribution à la classe des hausses)
- En rouge, le motif est suivi d'une baisse (attribution à la classe des baisses)
- Ici tout fonctionne bien, la zone des points bleus et celle des points rouges sont bien dissociées : prévision parfaite
  - **Décision optimale** : achat si motif bleu, vente si motif rouge

15

## Comment prévoir le futur à partir du passé : difficultés théoriques et pratiques

- Cela a à voir notamment avec (la nature du problème considéré) le nombre d'observations dans la base d'apprentissage et le nombre possible d'entrées
- Exemple précédent : on se sert des dix dernières observations de hausse ou baisse pour la prévision, soit  $2^{10} = 1024 \approx 10^3$  antécédents (entrées) possibles
- Supposons que l'on ait  $10^5$  éléments dans la base d'apprentissage.
- On aura (en moyenne)  $10^5 / 10^3 = 100$  observations / cas.
- Pour un motif donné, si toutes ces 100 observations sont suivies d'une hausse (ou bien d'une baisse), on prédira une hausse (pour ce motif)

16

## Comment prévoir le futur à partir du passé ?

- Considérons à nouveau ce qui suit le motif 1010001011
  - Supposons qu'il y a 500 occurrences de ce motif dans la base d'apprentissage
  - Si elles sont suivies d'un 1 (hausse), on sera conduit à prédire qu'une nouvelle occurrence du motif, sera suivie d'une hausse.
  - Si elles sont toutes suivies d'un 0 (baisse), on prédira une baisse
    - Nous discuterons de la validité de ces inférences dites par **induction** ...
    - généralisation de la base d'apprentissage à l'ensemble des situations non encore observées
  - Induction probable : si 495 occurrences sur 500 sont suivies d'une hausse, on aura tendance à prévoir également une hausse :
    - $\frac{495}{500} = 99\%$  : fréquence conditionnelle d'apparition de 1 après 1010001011.
    - 99% est proche de 1
    - 1% : fréquence conditionnelle d'apparition de 0 après 1010001011.

17

## Comment prévoir le futur à partir du passé ?

- Considérons à nouveau ce qui suit le motif 1010001011
  - Il y a 500 occurrences de ce motif dans la base d'apprentissage
- Prédications imparfaites dans l'échantillon
  - Quelle prédiction faire si la fréquence conditionnelle d'apparition de 1, après le motif précédent est de 56% ?
    - Dans l'échantillon, le motif 1010001011 est plus souvent suivi d'une hausse 1 que d'une baisse.
    - Mais est-ce que l'écart 56% contre 44% est suffisant pour l'on décide d'acheter (anticipant une hausse) ?
  - Quel serait le seuil (en termes de fréquences) à partir duquel on devrait acheter ?
  - Est-ce qu'adopter la **règle majoritaire** (acheter dès que la fréquence conditionnelle est  $> 50\%$ ) est une stratégie pertinente ?

18

## Comment prévoir le futur à partir du passé ?

- **Distance de Hamming** entre deux motifs (voir exercices corrigés)
  - Hamming, Intro to The Art of Doing Science and Engineering: Learning to Learn (March 28, 1995)
    - <https://www.youtube.com/watch?v=AD4b-52jtos&list=PL2FF649D0C4407B30>
  - Considérons deux motifs de hausses et de baisses
    - 1010001011
    - 1111000011
- La distance de Hamming est ici de 3 : Nombre de cas (dates) où hausses et baisses ne coïncident pas
- Distance de Hamming nulle : deux motifs identiques
- Distance de Hamming faible : deux motifs proches.

19

## Comment prévoir le futur à partir du passé ? méthode des k – plus proches voisins (introduction)

- Le problème des petits échantillons
  - Supposons maintenant qu'il y ait 2 occurrences 1010001011 suivies de 1 et aucune suivie de 0, dans la base d'apprentissage
  - La fréquence conditionnelle d'apparition d'une hausse est de 100%, mais la robustesse de la prévision est faible
- Que se passe-t-il dans le cas précédent ou si un nouveau motif n'a pas été observé dans la base d'apprentissage ?
  - Une idée est de considérer des motifs proches, par exemple, des motifs dont la distance de Hamming avec le nouveau motif est inférieure ou égale à 1
    - C'est-à-dire des motifs qui ne diffèrent du motif qui apparaît que d'au plus pour une date.
  - L'hypothèse qui est faite ici est que des prémisses proches ont les mêmes conséquences (pas de disruption)

20

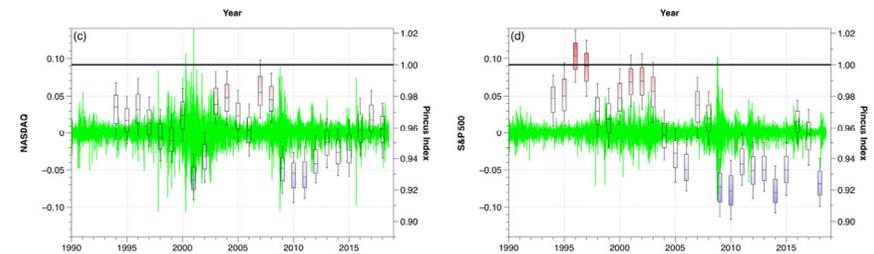
## Comment prévoir le futur à partir du passé ? méthode des $k$ – plus proches voisins (introduction)

- Le problème des petits échantillons (suite)
  - Quand la taille de l'échantillon (nombre de motifs observés) diminue, il y aura davantage de motifs qui n'auront que pas ou trop peu de voisins
  - Pour que l'on fasse une prédiction quelconque et donc que l'on puisse décider d'acheter ou de vendre.
- La taille de l'échantillon est à mettre en rapport avec la longueur des motifs
  - Supposons que l'on ait 1000 observations
  - Si motifs de longueur 3, en moyenne 125 observations par motif
  - Si motifs de longueur 10, en moyenne 1 observation par motif
  - Si motifs de longueur 15, en moyenne 30 motifs pour 1 observation
  - Nécessité de compresser l'information : encodage

21

## Comment prévoir le futur à partir du passé ? Évolution du degré de prévisibilité

- L'indice de Pincus a été considéré dans l'article suivant :
  - Delgado-Bonal (2019). Quantifying the randomness of the stock markets. Scientific reports, Nature.



- Une faible valeur de l'indice est associée à des répétitions plus fréquentes des motifs : marchés plus prévisibles
- Évolution de l'indice pour le Nasdaq et le S&P500 : graphique ba – chart échelle de droite
- On remarque des **fluctuations de la prévisibilité** (mesurée par l'indice de Pincus).

22

## prévisibilité : un peu de vocabulaire

what's the opposite of predictable?



unpredictable, unforeseen, unexpected, improbable, unlikely, capricious, freakish, unforeseeable, uncertain



Un peu de vocabulaire anglais : si l'on examine les termes de droite, on se rend compte qu'ils recouvrent des concepts différents. Par exemple, imprévisible, improbable, incertain et inattendu n'ont pas le même sens.

23

## prévisibilité : un peu de vocabulaire

what's the opposite of unpredictable?



predictable, certain, unchanging, stable, steady, reliable, unvarying, constant, sure, foreseeable



Même remarque que dans le transparent précédent. Par ailleurs, la forme littéraire permet de connoter de manière subtile les termes employés, mais n'a pas la précision du formalisme mathématique.

24

Induction is a process of inference; it proceeds from the known to the unknown.

John Stuart Mill

### Induction : Pourquoi est-ce important en finance ?

- L'induction est liée à l'**apprentissage** à partir des **données en vue de faire des prévisions**
  - Données : observations de cours boursiers, communications financières, informations textuelles.
  - Données massives (big data) associées au développement du trading à haute fréquence
- De passer de l'**échantillon** à la **population**.
  - Pour prévoir des cours boursiers futurs (market timing), de repérer des sociétés sous-évaluées (stock-picking).
  - **Évaluer des sociétés non-cotées par comparaison**
- Concevoir des algorithmes de trading
  - Stratégies boursières à partir de l'analyse des données
  - Apprentissage statistique (machine learning).



La finance utilise de plus en plus de données, notamment via l'émergence du trading à haute fréquence, la multiplicité des plates formes de négociation, l'utilisation de données comportementales (sentiment de marché), souvent issues des réseaux sociaux, l'analyse textuelle ... Ces données font l'objet de traitements automatisés, économétrie et statistiques « classiques », mais aussi apprentissage profond.

## Induction : Pourquoi est-ce important en finance ?

- La « logique inductive » sous-tend le problème de la prévision des cours boursiers, des ratios financiers.
- Deux approches pour les cours boursiers
  - *Les cours boursiers ont un certain degré de prévisibilité*
    - Et par conséquent, on peut envisager d'utiliser et construire certains algorithmes en vue de construire des stratégies boursières gagnantes
    - C'est la situation où la « logique inductive » trouve son application
  - *Les cours boursiers ne sont pas prévisibles*
    - Et par conséquent, aucune méthode inductive, aucun algorithme de trading, même utilisant des méthodes avancées de deep learning ne peut fonctionner de manière robuste
    - Problème (plus ou moins) caché de validation (back-testing) statistique des pseudo-stratégies gagnantes.

29

## Qu'est-ce que l'induction (par énumération) ?

- Induction de type (i) et de type (ii)
  - *Prémisse : tous les cygnes observés sont blancs*



- (i) : Le prochain cygne observé sera blanc
  - *Intéresse le spéculateur : en remplaçant blanc par hausse du marché et cygne par mouvement du marché boursier.*
  - *Inférence singulière ou par énumération simple ou prédictive*
- (ii) : tous les cygnes sont blancs
  - *Inférence généralisante ou par énumération ou nomologique : loi qui intéresse le théoricien.*

30

## Qu'est-ce que l'induction ?

- Soit un ensemble d'observations ou « faits empiriques »
  - *Exemple : motif (pattern) de trois hausses consécutives (codé en binaire 111) suivi d'une baisse (codée en binaire 0)*
  - *Supposons que toutes les observations dans l'ensemble donné (appelé échantillon en statistique) vérifient cette propriété.*
  - *L'induction énumérative ou de type (i) consiste à affirmer que la prochaine fois que l'observera 111, 0 (baisse des cours)*
    - L'induction (i) correspond à une prévision (utile au financier)
  - *L'induction de type (ii) consiste alors à affirmer que cette propriété est vraie pour tous les motifs, y compris ceux qui n'ont pas été observés, on n'observera jamais 1111*
    - Cad une hausse ne suivra jamais trois hausses consécutives
    - L'induction (ii) est nomologique : on établit une loi (utile au scientifique)

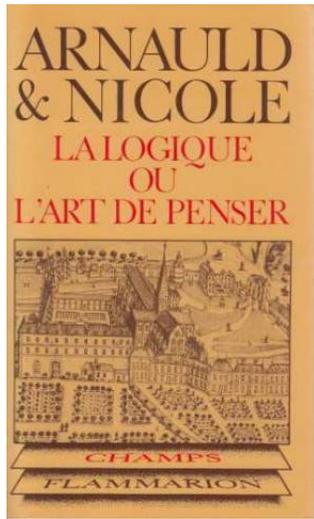
31

Induction  
=  
passage du particulier au général

"On appelle *induction* lorsque la recherche de plusieurs choses particulières nous mène à la connaissance d'une vérité générale. Ainsi, lorsqu'on a éprouvé sur beaucoup de mers que l'eau est salée, et sur beaucoup de rivières que l'eau en est douce, on conclut généralement que l'eau de la mer est salée et celle des rivières douce." [Arnauld & Nicole], p. 321

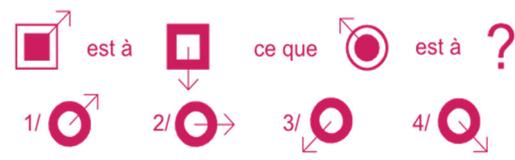
Antoine Arnauld - Pierre Nicole : la logique ou l'art de penser

32



*Qu'est-ce que l'induction ?*

- Dans sa forme la plus générale (ii), l'induction vise à établir des lois (ou régularités)
- Mais on est amené à procéder très souvent par **analogie** : dans l'exemple ci-dessous de test de QI, on ne dispose que d'une observation et il faut inférer une règle de prédiction

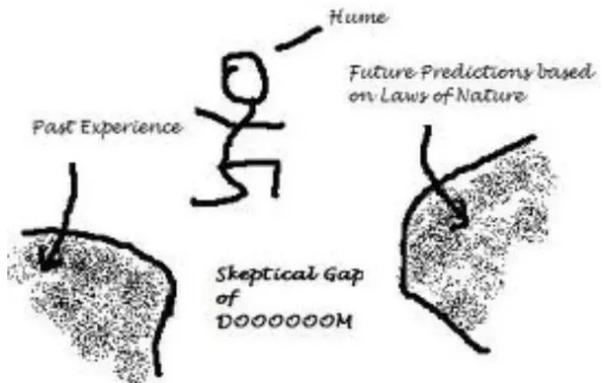


*Qu'est-ce que l'induction ?*

- Les raisonnements **analogiques** (par **association**) peuvent très souvent véhiculer des stéréotypes et leur **valeur prédictive** peut être **très douteuse** !
  - Exemple : types de personnalité de Kretschmer (1925)

| Type        | Physique                | Personnalité                      | Acteur                           |
|-------------|-------------------------|-----------------------------------|----------------------------------|
| Pycnique    | Petit et rond           | Expansif, gai, spontané, réaliste | Gérard Jugnot<br>Dany de Vito    |
| Leptosome   | Grand et mince          | Réservé, froid, rêveur            | Jean Rochefort<br>Clint Eastwood |
| Athlétique  | Large carrure et musclé | Impulsif, coléreux                | Lino Ventura<br>Harvey Keitel    |
| Dysplasique | Mal développé, anomalie | Asthénique, se sent inférieur     | N'a pas réussi dans le cinéma    |

- L'intérêt pédagogique de cet étrange tableau est de montrer le lien entre **prévision** et **classification**

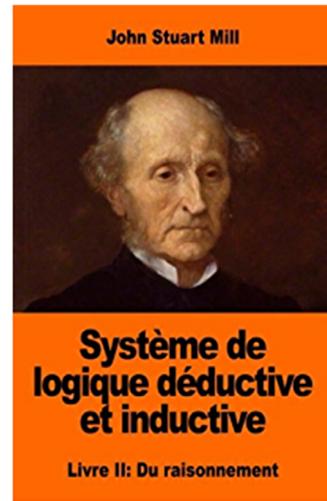


**The Inductive Leap**

**Saut qualitatif** non justifié logiquement entre données d'observations et données à prévoir

## Qu'est-ce que l'induction ?

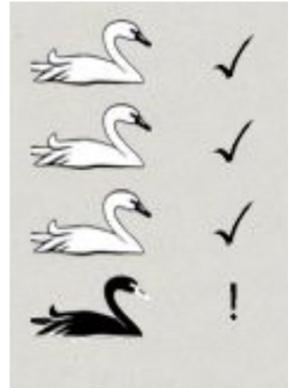
John Stuart Mill est l'un des premiers auteurs à avoir examiné l'utilisation de l'induction en vue de produire de nouvelles connaissances scientifiques



37

## Les limites des connaissances inductives sont connues (citation de Mill)

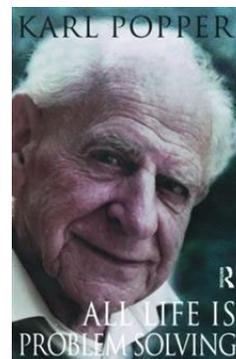
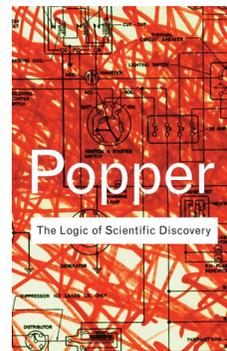
- Quelques unes, (...) crues bonnes pendant des siècles étaient, en réalité, mauvaises.
- L'induction que tous les cygnes sont blancs n'a pas pu être bonne puisque la conclusion a fini par se trouver fausse ».
- Cependant, l'expérience sur laquelle reposait la conclusion était de bon aloi.
- De temps immémorial, **le témoignage de tous les habitants du monde connu, concordante en un résultat commun, sans un seul exemple de déviation dans ce résultat, n'est pas toujours suffisante pour établir une conclusion générale.**



38

## Les limites des connaissances inductives sont connues (Karl Popper)

- “Now it is far from obvious, from a logical point of view, that we are justified in inferring universal statements from singular ones, no matter how numerous; for any conclusion drawn in this way may always turn out to be false:
- No matter how many instances of white swans we may have observed, this does not justify the conclusion that all swans are white.”
  - Popper, Karl, *The Logic of Scientific Discovery*



39

## Qu'est-ce que l'induction ?

- Approche empirique
  - *établir des régularités entre des observations (données, faits)*
- Loi (ou propriété)
  - *relation à caractère général entre les données*
  - *Données observées et non encore observées*
  - *Loi : passage de l'échantillon (statistique descriptive) à la population (statistique inférentielle)*
  - *Une loi portant sur des données non encore observées peut donc servir à faire des prédictions*

40

## Qu'est-ce que l'induction ?

- Induction ou inférence inductive : forme de raisonnement qui permet de passer du **particulier** au **général**
  - Le raisonnement est ampliatif (qui accroît, complète) : La conclusion contient davantage d'information que les prémisses
    - Prémisses : les données déjà observées
    - Conclusion : les propriétés des « données » non encore observées
  - Inférence factuelle ampliative : raisonnement par lequel, à partir de prémisses fondées sur des observations effectives (« factuelle »), on dépasse l'expérience finie pour atteindre une conclusion nouvelle (« ampliative »), sur ce qui n'a pas (encore) été expérimenté

41

## Déduction, induction, abduction

- Trois types de raisonnement
  - **Induction** : prévision des valeurs boursières
  - **Abduction** : recherche des déterminants (variables explicatives) des valeurs boursières (Pierce)
  - **Déduction** : relations logiques entre valeurs boursières : actions vs dette, analyse de la diversification des risques, relation entre risque et rentabilité
- Pour résumer, la **déduction**, qui repose sur des causes et des effets certains, aboutit à des **énoncés certains** ; l'**induction**, qui propose des causes certaines à des effets probables, aboutit à des **énoncés probables** ; et l'**abduction**, qui recherche des causes probables à des effets certains, aboutit à des **énoncés plausibles**.
  - Nicolas Chevassus-au-Louis

42

43

44

## Validité logique de l'induction

45

## Validité du raisonnement inductif

- Mill invoque le principe **d'uniformité des lois de la nature**
  - « Lorsque ces diverses uniformités sont constatées par une induction jugée suffisante, on les appelle communément des Lois de la Nature »
  - Pour justifier ce principe d'uniformité, Mill fait appel à un principe de causalité : toutes choses égales par ailleurs, **les mêmes causes produisent les mêmes effets**.
- Causes des mouvements des prix des actifs financiers ?  
Passages d'ordre (market impact) et **nouvelles** (non anticipées) **informations économiques et financières**
  - **Abduction** : recherche (difficile !) des **informations pertinentes** impactant les cours boursiers
    - Voir les transparents consacrés au biais de confirmation et au biais rétrospectif
- Mis à part les financiers formés par la physique, l'hypothèse d'uniformité rencontre peu de succès en finance.

46

## « Le » problème de l'induction (Hume)

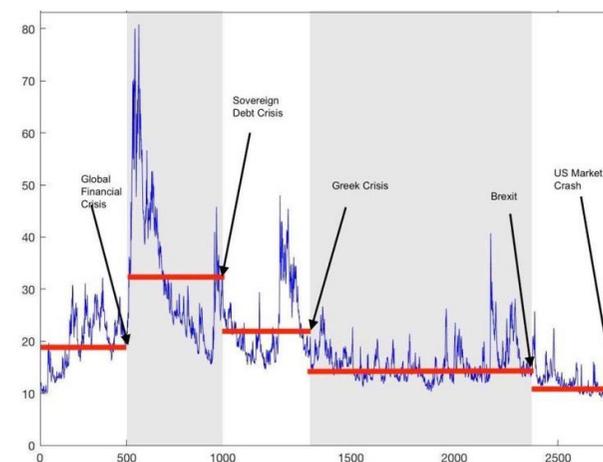
- Pour Nelson Goodman, ce qui ressort d'une **prétendue** uniformité de la nature s'apparente à une « **hypothèse non confirmée et peut-être même douteuse** » et il lui paraît « insensé » de justifier ainsi des prévisions
  - *Faits, fictions et prédiction* (page 78)
- Les économètres de la finance parlent de **stationnarité** (invariance des lois de probabilité des rentabilités) et de « structural breaks » (changement dans ces lois)



47

## Validité du raisonnement inductif

- Stationnarité ou « structural breaks » ? évolution du risque sur les actions

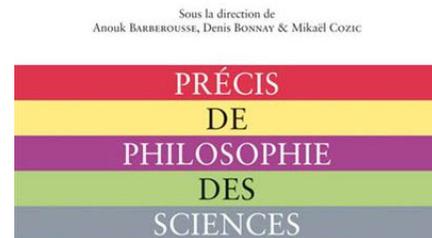


48

## Validité du raisonnement inductif

- Est-ce qu'une augmentation d'observations dans le sens attendu augmente la plausibilité de l'hypothèse ?
  - *Si l'on a vu beaucoup de cygnes blancs, est-ce qu'il est plus vraisemblable que le prochain cygne observé soit blanc ?*
  - *Ceci va dans le sens de ce que l'on appelle la théorie bayésienne de la confirmation (TBC)*

M. Cozic (2010), "Confirmation et induction" in Barberousse & al., eds, *Précis de Philosophie des Sciences*, vol.1, Paris: Vuibert



49

## Validité du raisonnement inductif : analogie, coïncidence ?

- Ian Fleming dans le roman Goldfinger, dont est issu le film homonyme

MR BOND, THEY HAVE A SAYING IN CHICAGO: 'ONCE IS HAPPENSTANCE. TWICE IS COINCIDENCE. THE THIRD TIME IT'S ENEMY ACTION!'

- Selon le proverbe français, jamais deux sans trois...Et pour Agatha Christie (Nemesis)

Any coincidence is worth noticing. You can throw it away later if it is only a coincidence.

50

## Validité du raisonnement inductif

- Il y a même des situations où **une seule expérience** nous permet une prévision fiable.



- Supposons que votre niveau d'anglais soit un peu limité : vous ne comprenez pas le contenu du panneau.
- Combien de fois allez-vous plonger votre main dans l'eau bouillante pour en conclure que ça brûle ?!

51

## Validité du raisonnement inductif

- Mais une seule expérience (ou expérience singulière) n'est pas forcément la meilleure manière de faire une prévision...

*Le chat qui s'assoit sur le couvercle d'un poêle brûlant [...] ne s'assiera plus jamais sur un poêle chaud et c'est très bien ainsi ; mais il ne s'assiera plus jamais non plus sur un poêle froid.*

– Mark Twain<sup>1</sup>

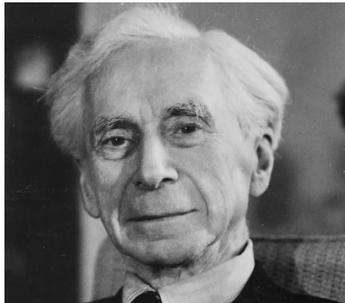
- On remarquera également qu'il suffit d'une seule expérience pour l'apprentissage, ce qui remet en cause la justification de l'induction par une « habitude » (Hume)

52

## Validité du raisonnement inductif

- L'abus du raisonnement inductif est problématique pour la prise de décision...
  - *"The man who has fed the chicken every day throughout its life at last wrings its neck instead, showing that **more refined views as to the uniformity of nature** would have been useful to the chicken."*

■ Bertrand Russell



The Demand For  
Certainty Is One  
Which Is Natural  
To Man, But Is  
Nevertheless An  
Intellectual Vice.

53

## Validité du raisonnement inductif

- Dans le même esprit que Russell, le philosophe (des sciences) Quine considère le contre-exemple suivant
- Chaque jour qui passe et où l'on constate que nous sommes vivants devrait augmenter notre croyance en l'immortalité ...



Willard V. O. Quine  
̄ Joseph S. Ullian  
LA FABRIQUE DES  
CROYANCES  
Une introduction au  
raisonnement scientifique

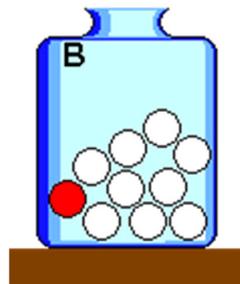


éditions  
markus haller

54

## Validité du raisonnement inductif

- Supposons que nous considérons une urne avec une boule rouge et des boules blanches
- Chaque tirage (sans remise) d'une boule blanche diminue et non pas augmente la probabilité de tirer une boule blanche au coup suivant



55

## Validité du raisonnement inductif

- Si l'on retient l'hypothèse d'efficacité des marchés sous sa forme initiale ...
  - *Les cours boursiers suivent des marches aléatoires*
- Une longue succession de hausses n'est en aucun cas un indicateur de hausse future
  - *Pas de tendances*
  - *Voir la partie du cours sur les biais cognitifs (biais de représentativité, gambler fallacy, hot hand fallacy)*
- Une hausse des cours supplémentaires

56

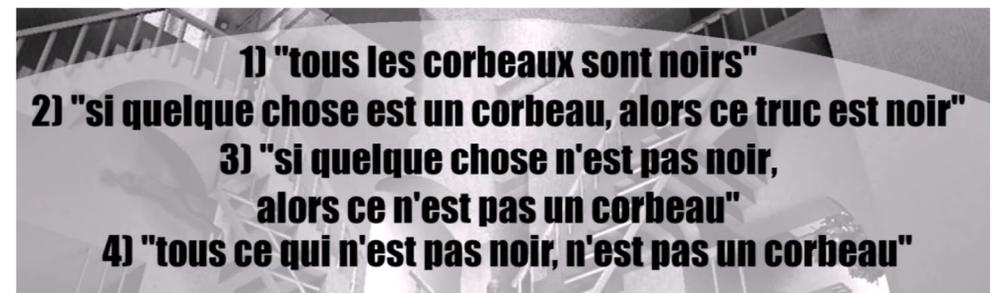
## Validité du raisonnement inductif

- Théorie bayésienne de la confirmation et paradoxe des corbeaux de Hempel



57

## Validité du raisonnement inductif



- Source : Maël Cavan
- La proposition 1) est une reformulation de 2)
- La proposition 3) est la contraposée de la proposition 2)
- Elle est logiquement équivalente à 2)
- Enfin 4) est une reformulation de 3)

58

## Validité du raisonnement inductif

- Le paradoxe ...

**Observer un parapluie jaune augmenterait la croyance envers le fait que tous les corbeaux sont noirs !**

- La solution ...

**Il y a beaucoup plus d'objets non noirs que de corbeaux donc observer un objet non noir qui n'est pas un corbeau augmente la crédence de manière beaucoup plus négligeable qu'observer un corbeau noir**

- Mais on ne connaît pas le nombre respectif d'objets non noirs et de corbeaux ...

59

## Validité du raisonnement inductif

- Pour la petite histoire, tout comme il y a des cygnes noirs, il y a des corbeaux blancs !
- L'induction ne fonctionne donc pas dans ce cas
  - *Même s'il est plus probable de rencontrer un corbeau noir*



60

## « Le » problème de l'induction (Hume)

### ■ Hume a remarqué que l'induction n'a pas de fondement logique.

- Justifier la méthode inductive par l'uniformité de lois de la nature ?
- Comment peut-on justifier le principe de l'uniformité des lois de la nature ?
- En répétant des expériences ...
- C'est-à-dire par la méthode inductive...
- **Raisonnement circulaire**
- On ne peut pas établir avec certitude une connaissance générale à partir de cas particuliers.



David Hume,  
philosophe écossais  
(1711-1776)

61

## « Le » problème de l'induction (Hume)

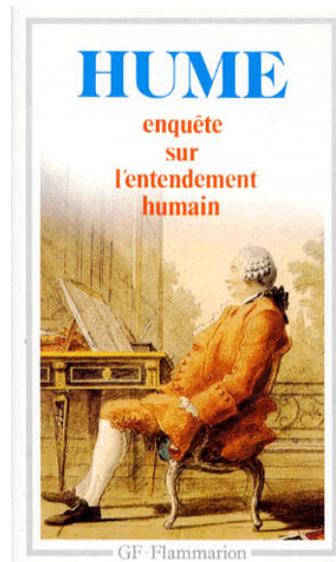
Dire qu'elle (la connaissance) vient de l'expérience, c'est une pétition de principe.

Car toutes les inférences tirées de l'expérience supposent, comme fondement, que le futur ressemblera au passé et que des pouvoirs semblables seront conjoints à de semblables qualités sensibles.

S'il y a quelque doute que le cours de la nature puisse changer et que le passé ne puisse être la règle pour l'avenir, toutes les expériences deviennent inutiles et ne peuvent engendrer d'inférence ou de conclusion.

Il est donc impossible qu'aucun argument tiré de l'expérience puisse prouver cette ressemblance du passé au futur, car tous les arguments se fondent sur la supposition de cette ressemblance.

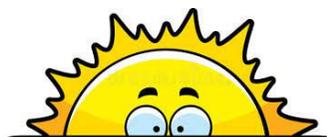
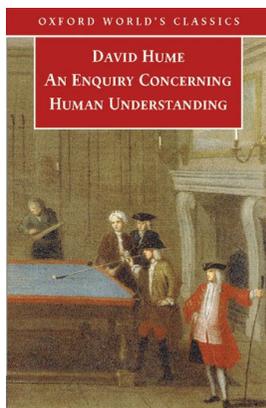
David Hume, 1748



62

## « Le » problème de l'induction (Hume)

- “That the sun will not rise tomorrow is no less intelligible a proposition, and implies no more contradiction than the affirmation, that it will rise. We should in vain, therefore, attempt to demonstrate its falsehood”.
- “Mr. Locke divides all arguments into demonstrative and probable. In this view, we must say, that **it is only probable (...) that the sun will rise tomorrow**”.



63

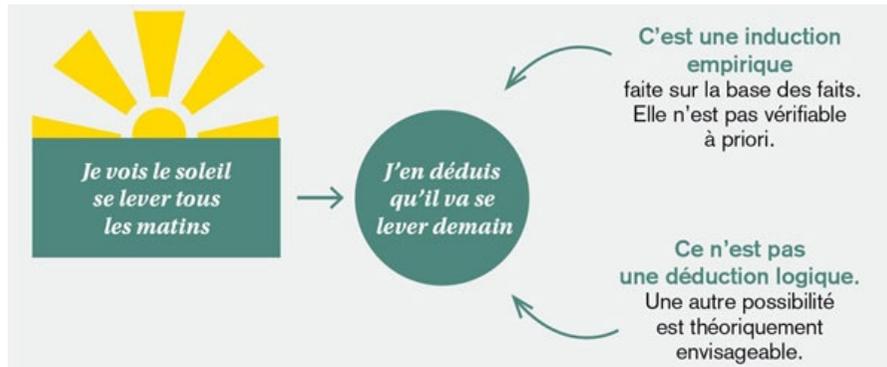
## « Le » problème de l'induction (Hume)

- Rappelons que la prévision des cours boursiers repose sur l'idée de répétition de « motifs » dans l'échantillon
  - Supposée persister dans le futur et permettre des gains boursiers.
- Hume (Traité de la Nature Humaine) ?
  - « Persuadez une bonne fois pour toutes les hommes de ce principe : **même après l'observation d'une fréquente ou constante conjonction d'objets, nous n'avons aucune raison de tirer aucune inférence au sujet d'aucun objet autre que ceux dont nous avons eu l'expérience** »
  - On a pu observer certaines régularités dans l'échantillon des observations, sans que celles-ci puissent s'étendre aux quantités non observées, cad à l'ensemble de la population.
  - Scepticisme quant à la possibilité de prévisions boursières.

64

## Notre connaissance du futur vient-elle de l'expérience ?

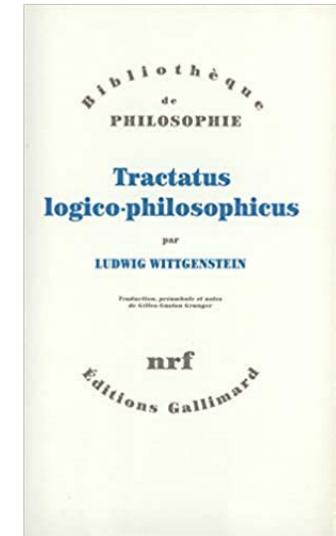
- Le problème de Hume : au cœur de la prévision boursière via les algorithmes de machine learning



- Limite à l'approche empirique (data driven), qui met en avant l'expérience comme base de la connaissance

65

« Que le soleil se lèvera demain est une hypothèse », citation extraite du Tractatus et clin d'œil à la problématique de l'induction



66

## Notre connaissance du futur vient-elle de l'expérience ?

- Le soleil se lèvera demain !
- Pour les Egyptiens de l'Antiquité, Rê voyage chaque jour à travers le ciel à bord de sa barque sacrée (parcours du Soleil), et chaque nuit au travers des mondes souterrains (la Douât).
- Chaque lever de soleil est une victoire remportée par Rê sur les « forces des ténèbres ».



67

## Notre connaissance du futur vient-elle de l'expérience ?

- Le soleil se lèvera demain (David Hume)
  - *Traité de la nature humaine* (1739-1740)
- Tous les jours et depuis toujours, le soleil se lève. Il en sera donc de même demain.
- Laskar (2013) : « Le Système solaire est-il stable ? »
  - *le mouvement des planètes dans le Système solaire est chaotique, ce qui interdit toute prédiction précise de leur mouvement au-delà de quelques dizaines de millions d'années. Les simulations montrent que les collisions planétaires ou les éjections sont possibles.*
  - <https://www.franceculture.fr/emissions/la-methode-scientifique/la-methode-scientifique-emission-du-mardi-23-juin-2020>

68

## Notre connaissance du futur vient-elle de l'expérience ?

### ■ La « provocation » de Laplace

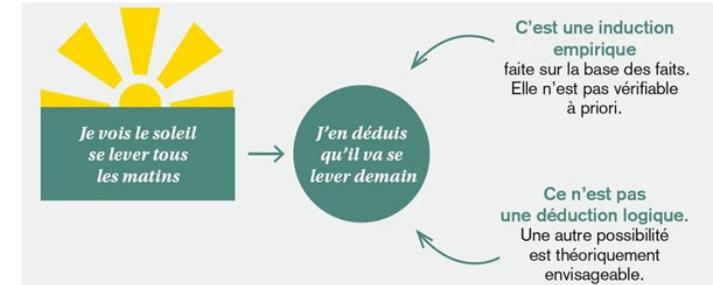
- Il suppose qu'on ne sait pas si le soleil se lèvera demain
- Il postule une distribution uniforme a priori sur la probabilité correspondante (mémoire, 1774) : principe d'indifférence
- Laplace suppose que l'on sait que le soleil se lève chaque matin depuis environ 5000 ans (l'invention de l'écriture)
  - Soit environ  $n = 1826200$  jours
- La probabilité que le soleil se lève demain est égale à  $\frac{n+1}{n+2}$  ...
  - (règle de succession)
  - Ou qu'il ne se lève pas demain à  $\frac{1}{n+2}$ , soit une chance sur deux millions...
    - A titre de comparaison, il y a 139 millions de combinaisons à l'EuroMillions
  - Emile Borel (Valeur pratique et philosophie des probabilités, 26-27), fait remarquer que l'hiver, au-delà du cercle polaire, le soleil ne se lève pas tous les jours ; d'où l'importance de bien spécifier une hypothèse ...

... tirée de l'événement observé. On trouve ainsi qu'un événement étant arrivé de suite un nombre quelconque de fois, la probabilité qu'il arrivera encore la fois suivante est égale à ce nombre augmenté de l'unité, divisé par le même nombre augmenté de deux unités. En faisant, par exemple, remonter la plus ancienne époque

69

## Notre connaissance du futur vient-elle de l'expérience ?

- C'est notre compréhension de la mécanique céleste qui nous permet de savoir que le soleil se lèvera demain



- Limite à l'approche « data driven », qui met en avant uniquement l'expérience comme base de la connaissance
- Pas de théorie dans un algorithme de machine learning

70

71

72

## D'où vient le raisonnement inductif ?

## D'où vient le raisonnement inductif ?

- Pourquoi est-il si couramment utilisé ?
- D'après Hume, c'est l'**habitude** qui guide notre pensée, pas la **logique**.
- Habitude selon Hume :
  - « transition coutumière de l'imagination d'un objet à celui qui l'accompagne habituellement ».
  - C'est une disposition habituelle de l'esprit humain
  - **Mécanisme associatif** et « psychologique »

73

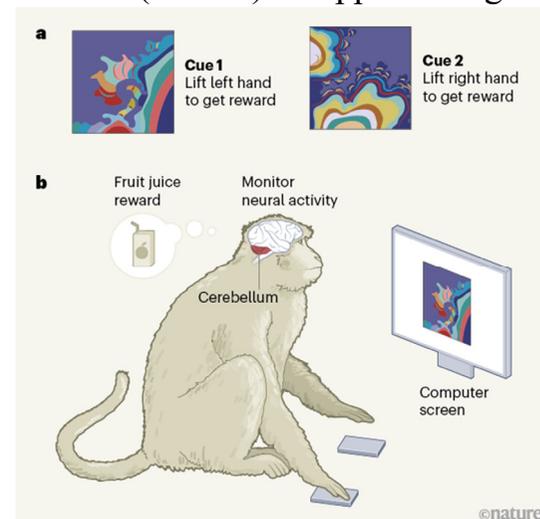
74

## D'où vient le raisonnement inductif ?

## D'où vient le raisonnement inductif ?

- Pour Hume, l'animal, comme l'homme, apprend à connaître les choses extérieures par l'**habitude**, c'est-à-dire par la perception répétée des mêmes successions de phénomènes.
- Mais le principe inductif n'est pas (uniquement) de l'ordre de la pensée ou du langage,
- Même si Daniel Kahneman a mis en évidence le problème de l'ancrage en psychologie cognitive.
  - *A quoi pensez-vous après le mot « ancre » ?*
  - *Normalement, bateau ; c'est le principe associatif*
  - *De manière plus étonnante, si on demande à quelqu'un de faire une prévision, elle sera plus élevée si elle est précédée de la vision d'un nombre élevé...*

- Bases biologiques de l'induction : *réflexe conditionné* (Pavlov) ou apprentissage



Reconnaissance d'image : Selon l'image qui apparaît à l'écran, le singe doit lever le bras droit ou le bras gauche à l'écran pour obtenir une récompense.

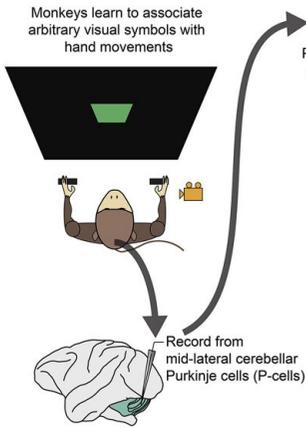
Le cervelet (cerebellum) est impliqué dans l'apprentissage par renforcement (reinforcement learning)

Ce que nous voudrions faire, c'est montrer un graphique des cours passés sur un écran Bloomberg et savoir si c'est un prédicteur de hausse ou de baisse des cours ...

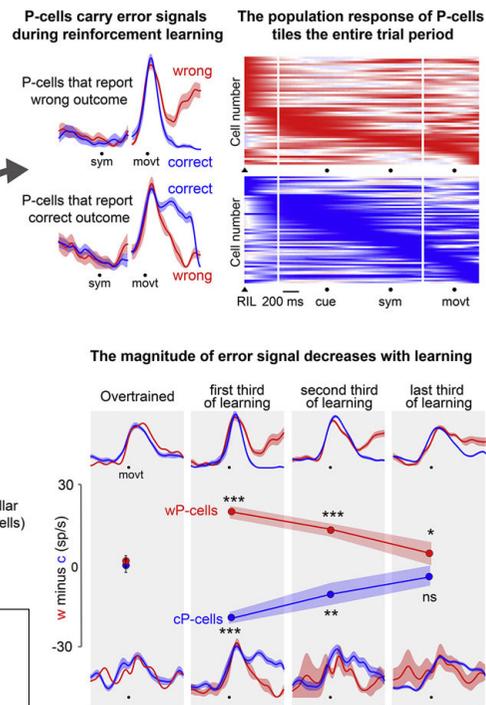
75

76

## Implications neurologiques de l'apprentissage



Sendhilnathan, Semework, Goldberg & Ipata (2020). Neural correlates of reinforcement learning in mid-lateral cerebellum. *Neuron*.



77

## D'où vient le raisonnement inductif ?

- On retrouve ces bases animales et biologiques de l'induction dans l'analyse du « conditionnement » de Pavlov
  - Pavlov. Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex. *Annals of neurosciences*
  - Pavlov s'intéresse aux « réflexes conditionnels »
  - Après un temps d'apprentissage, un stimuli donné ou information va provoquer une réponse comportementale donnée
  - De même qu'un enfant qui aurait par mégarde approché sa main d'une flamme sait instinctivement
- L'habitude de Hume, bien qu'elle puisse être vue sous un angle psychologique relève aussi de cette base animale

78

79

80

## Le problème de l'encodage des historiques de cours

81

## Quelles inductions (prévisions) faisons-nous ?

- L'habitude ou la propension de l'esprit à repérer des régularités par association n'épuise pas la question méthodologique
- **Reprenons Quine** : supposons que nous fassions la prévision (funèbre) de notre décès le 30 septembre 2023
  - *Nous sommes le 29 septembre 2023*
- Depuis le jour de notre naissance, chaque jour augmente la « crédence » de notre prévision.
  - *Si elle s'avère exacte, ce sera notre dernière prévision ...*
- Mais chaque jour augmente également la « crédence » entre notre immortalité
- Pourquoi devrions-nous induire notre décès demain ou jamais ?
  - *« Hume néglige le fait que certaines régularités établissent de telles régularités et d'autres pas » (Goodman)*

82

## Quelles inductions (prévisions) faisons-nous ?

- Revenons sur la prévision des cours boursiers à partir des motifs
- Quelle est la longueur pertinente des motifs ?
- Plus l'on considère des motifs de grande taille, moins on aura d'observations (voire pas d'observations)
- Mais chaque motif va apporter davantage d'information (être plus caractéristique d'une trajectoire spécifique)
- Quelle est la longueur pertinente d'un motif pour la prévision ?
  - *Par exemple, HHH peut prévoir H et BHH peut prévoir B*
  - *Auquel cas, HH ne permet pas de prévoir*
- Pour gérer la double contrainte précédente, on peut être tenté de considérer des résumés de l'information passée

83

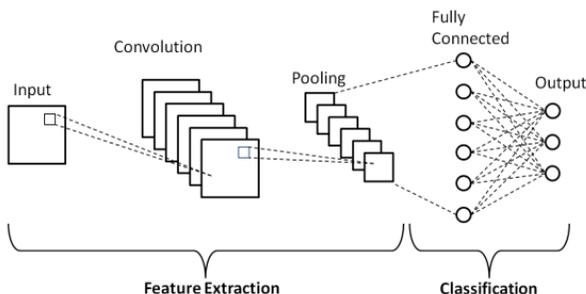
## Quelles inductions (prévisions) faisons-nous ?

- Exemple simple : momentum - tendance
  - *On considère les 9 derniers cours boursiers*
  - *Si le nombre de hausses est au moins de 5, le marché est haussier et l'on prédit une hausse (approche d'élan ou momentum)*
  - *Si le nombre de baisses est au moins de 5, le marché est baissier et l'on prédit une baisse (toujours l'approche momentum)*
- Exemple simple : reversal
  - *On peut s'intéresser à la longueur moyenne des séquences de hausse ou de baisse (runs) ou au nombre de cas où une hausse est suivie d'une baisse pour caractériser un marché cyclique.*
- En faisant cela, on va compresser l'information contenue dans les cours passés

84

## Prévisions boursières : encodage = angle mort de l'IA

- L'encodage consiste à compresser l'information disponible
- Avec le minimum de pertes
- Pour pouvoir la traiter par des méthodes de prévision (réseaux de neurones) ou de classification

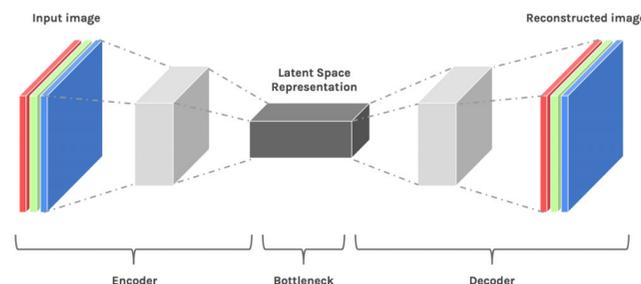


- L'encodage (feature extraction) permet une réduction de la dimensionalité des inputs

85

## Prévisions boursières : encodage = angle mort de l'IA

- Encodage : réduction de la dimensionalité des inputs



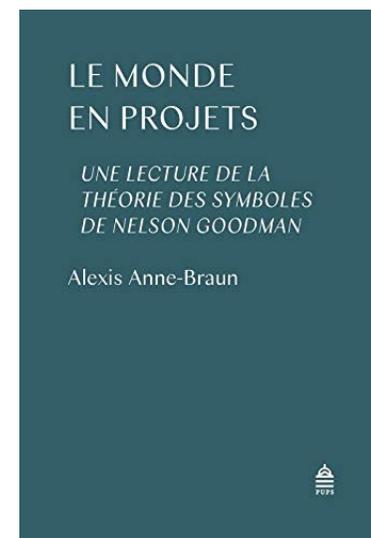
- Un encodage efficace permet non seulement de réduire la complexité
- Mais aussi de faciliter le traitement des données, une fois encodées : réduction des non-linéarités

86

## Quelles inductions (prévisions) faisons-nous ?

- Notre manière de caractériser l'histoire des cours par des mots, « haussier » (bull), « baissier » (bear), de tendance (momentum), cyclique (reversal) n'est pas neutre
- Dire que le marché est haussier, c'est dire qu'il a une disposition à la hausse.
- C'est donc déjà faire une prédiction
  - Pour Ian Hacking (le plus pur nominalisme), « les êtres humains classent et généralisent à la fois »
  - « Utiliser un nom pour une espèce, c'est vouloir réaliser des généralisations et former des anticipations concernant des individus de cette espèce ».
  - « Utiliser un prédicat, c'est lui faire jouer un rôle dans la formation des anticipations et l'établissement des généralisations »

87



88

## *Quelles inductions (prévisions) faisons-nous ?*

- Nous devons caractériser les représentations pertinentes des patterns boursiers
  - *Inputs de la prévision*
- Nous devons aussi nous intéresser à ce que l'on doit prévoir
  - *Supposons que le motif HHHH précède les motifs suivants : BHH, HBH.*
  - *Le motif HHHH ne peut pas être associé à une hausse ou bien à une baisse le jour suivant.*
  - *En revanche, on sait qu'il y aura deux hausses parmi les trois jours suivants, ce qui est suffisant pour gagner avec une stratégie buy and hold sur un horizon de trois jours.*

## Multiplicité des règles de prévision

93

## La « nouvelle énigme de l'induction » : Nelson Goodman



<https://www.youtube.com/watch?v=xZVVarMDso8>

94

## Comment former des prévisions boursières ?

- Nelson Goodman apporte un éclairage nouveau au « problème de l'induction »
  - Rôle fondamental du langage dans nos descriptions des phénomènes
- Ce qui est régulier dans un langage ne l'est pas dans une autre et vice versa
  - Comment compléter la suite 012345789 ?
  - 101112131415 en système décimal... ABCDEF en hexadécimal

95

## Comment former des prévisions boursières ?

- Induction : 12345 ... peut se prolonger par ... 6789
  - Récurrence  $x_{n+1} = x_n + 1$  ou  $x_{n+1} - x_n = 1$
  - On applique ici induction par énumération à la suite des accroissements
  - régularité généralisable ?
- La notion de régularité généralisables pose un problème d'ordre logique
  - Qu'induire à partir de 1235 ?
  - 1235 peut se prolonger ou par 81321
  - Récurrence  $x_{n+1} = x_n + x_{n-1}$  (suite de Fibonacci)
  - Ou par : 71113 (concaténation de la suite croissante des nombres premiers ou constante de Copeland-Erdős)
  - Ou par une suite croissante quelconque d'entiers ...

96

## Comment former des prévisions boursières ?

- Il y a donc plusieurs règles qui permettent d'extrapoler les données observées
  - Laquelle choisir ? (problème soulevé par Saul Kripke)
  - On retrouve la question de la multiplicité des inférences inductives
  - Ici, plusieurs règles (donc plusieurs théories prédictives) : sous-détermination des théories par l'expérience (Quine)
- Remarque : La notion de correspondances fortuites pose quant à elle des problèmes d'ordre psychologique.
  - Ce qui est fortuit n'a pas d'interprétation logique ou causale.
  - Ceci se heurte aux biais du système 1 (machine explicative de Kahneman) et biais rétrospectif de Bischof
  - Voir séances sur la finance comportementale

97

## Comment former des prévisions boursières ?

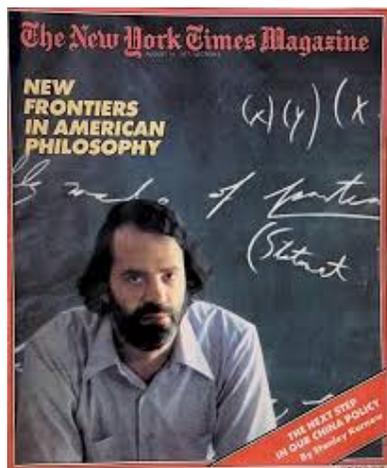
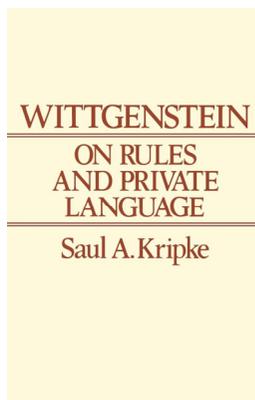
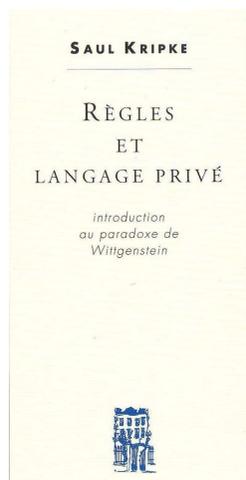
- Quelle règle choisir pour prolonger un historique de cours boursiers ?
- Le « paradoxe de Wittgenstein » selon Saul Kripke

« Supposons par exemple que «  $68 + 57$  » soit un calcul que je n'ai jamais effectué auparavant. Comme je n'ai pu effectuer par le passé (que ce soit en mon for intérieur ou en public) qu'un nombre fini de calculs, un tel exemple existe bien. [...]. J'effectue mon calcul et obtiens bien évidemment « 125 ». Me voilà persuadé, peut-être après une ultime vérification, que « 125 » est la bonne réponse. [...]. Supposons maintenant que je rencontre un étrange sceptique, qui entreprenne de saper la certitude de ma réponse en attaquant son sens « métalinguistique ». A l'en croire, d'après la façon dont j'utilisais le terme « plus » par le passé, j'aurais peut-être dû obtenir « 5 » en calculant «  $68 + 57$  » ! [Car] il se peut [...] que j'aie utilisé « plus » et « + » pour dénoter une fonction que j'appellerai « quus » et écrirai «  $\oplus$  ». Voici sa définition :

$$x \oplus y = x + y, \text{ si } x, y < 57, \\ = 5 \text{ autrement.}.$$

3

## Comment former des prévisions boursières ?



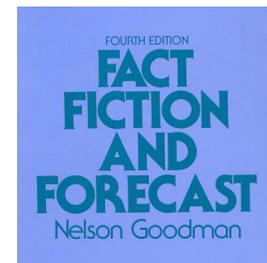
99

## Comment former des prévisions boursières ?

- Comme plusieurs régularités peuvent être trouvées dans les données, laquelle devrions-nous choisir ?

grueness does not. To say that valid predictions are those based on past regularities, without being able to say which regularities, is thus quite pointless. Regularities are where you find them, and you can find them anywhere. As we have seen, Hume's failure to recognize and deal with this problem has been shared even by his most recent successors.

- Citation de Nelson Goodman dans



100

## Comment former des prévisions boursières ?

- Autre exemple : Que dites-vous après 123 ?
  - (a) 123456 ...
    - Énumération de la liste des entiers naturels
    - $x_n = x_{n-1} + 1$
  - (b) 12357 ...
    - Début de la suite des nombres premiers
  - (c) 123123 ...
    - « Répéter 123 » ou bien  $x_n = x_{n-1}[3] + 1$
  - (d) 123☀️😊
    - Grandir avec Nathan <https://grandiravec.nathan.fr/activites-famille/jeux/regles-jeux/1-2-3-soleil.html>

101

## Comment former des prévisions boursières ?

- Observation de la série de nombres : 0010010101001
- Quelle est la prochaine valeur ?
  - *La plupart des individus (et des algorithmes programmés par des individus) vont chercher si cela peut être 0 ou 1*
- Univers des possibles donné par le contexte
  - *Si on joue à pile ou face*
    - et si on exclut que la pièce retombe sur la tranche
  - *Et que l'on code par 0, pile et 1 face, la valeur suivante est 0 ou 1*
  - *En l'absence de contexte, on peut prolonger 0010010101001 par 2, a, &, π, « je ne sais pas » selon que l'on considère tel ou tel ensemble alphanumérique.*

102

## Comment former des prévisions boursières ?

- Comment compléter la suite 0123456789 ?
  - *Comme c'est le début de la suite croissante des entiers naturels, on s'attend à la réponse 1011121314 ...*
- Codage en hexadécimal (base 16) des nombres entiers.
  - *Les « chiffres » en hexadécimal sont 0123456789ABCDEF*
  - *Codage standard en informatique car on passe facilement de la base  $2^4 = 16$  à la base 2 utilisée par les ordinateurs*
  - *L'hexadécimal a une écriture plus compacte*
    - Il est aussi utilisé dans le codage ASCII des caractères
    - 0000 correspond à 0, 0001 à 1, ..., 1110 correspond à E, 1111 correspond à F

103

## Comment former des prévisions boursières ?

- Codage en hexadécimal

|                                  |                |
|----------------------------------|----------------|
| Écriture décimale                | 1763           |
| Écriture binaire                 | 11011100011    |
| Écriture binaire en groupes de 4 | 0110 1110 0011 |
| Écriture hexadécimale            | 6E3            |

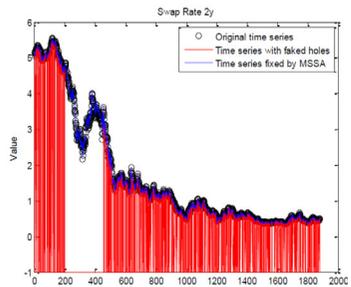
- *Pour un informaticien, dont la langue est l'hexadécimal, la suite 0123456789 se complète par ABCDEF101112131415 et non par 101112131415*
- *Celui qui égrène 0123456789 101112131415 a oublié ABCDEF*
  - Dans le langage hexadécimal : beaucoup plus utilisé que le décimal, car qui fait encore des multiplications à la main ?!
  - C'est comme si en langage décimal, on faisait suivre 0123456789 par 16171819

104



## Comment former des prévisions boursières ?

- Goodman établit qu'il n'y a aucune règle logique qui permette de choisir une généralisation d'une propriété dans un échantillon à la population, plutôt qu'une autre.
- Toutes les règles d'inférence inductive sont logiquement équivalentes.
- Reprenons le cas des données manquantes :



Une approche « raisonnable » et couramment utilisée est de recourir à une interpolation linéaire ou d'autres méthodes qui visent à avoir une trajectoire régulière (voir l'article, Goodman, Poincaré et la nouvelle énigme de l'induction). Mais dans un contexte de mesure des risques, une interpolation lisse va sous-estimer le risque. Il n'y a pas d'implantation « naturelle ».

## La nouvelle énigme de l'induction : les règles de prévision projetables

113

## Induction et langage



- Induction et ambiguïté du langage
- Le double sens de l'énoncé « le cygne est blanc »
  - « ce cygne est blanc » a un caractère singulier, blanc étant un prédicat (attribut du sujet) de ce cygne.
  - Il peut aussi énoncer une loi « tous les cygnes sont blancs », qui établit une relation permanente entre la blancheur et le cygne. On est passé de la constatation à la prédiction

114

## Induction : ambiguïté du langage et cygnes blancs

- Si l'induction, ou passage du particulier/spécifique au général, a des bases animales, ses caractéristiques humaines sont spécifiques
- Du fait de l'**utilisation du langage**.
- Nous allons travailler à partir d'un exemple simple.
- Considérons la proposition « le cygne est blanc ».



115

## Induction et langage

- Remarque : « cygne » est un concept qui dénote un objet particulier ou un ensemble délimité (ou pas) d'objets (l'espèce cygne).
  - A propos de la perception de l'objet cygne et de sa réalité ou non (approche nominaliste), nous suivons l'approche de W. James et de H. Putman, appelée « réalisme naturel »
- « est blanc » exprime une propriété du cygne : c'est un **prédicat nominal**
- La proposition « le cygne est blanc » attribue la propriété d'être blanc au cygne.
- « est » un verbe attributif ou substantif, « blanc » est un adjectif (attributif).

116

## Possibles et propositions réfutables

- Remarque : même si l'on n'a observé que des cygnes blancs, la proposition « le cygne est bleu » est logiquement et syntaxiquement valide.
  - Mais elle devient fausse dès que l'on a observé un cygne blanc.
- L'observation d'un cygne non-blanc rend la proposition « les cygnes sont blancs » fausse.
  - On dit que cette proposition est **réfutable**.
  - Remarque : toute proposition à caractère général est réfutable, ne peut donc avoir le caractère de loi universelle.
- Le langage permettant de construire des propositions « les cygnes sont blancs ou » « les cygnes sont bleus » permet de concevoir un ensemble de possibles.
  - La pluralité des possibles est le germe de l'aléatoire.

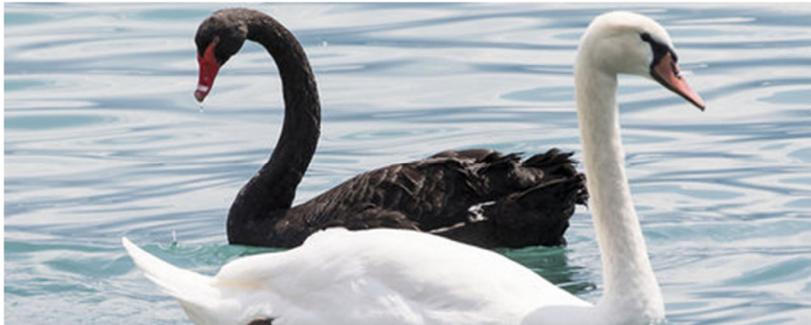
117

## Induction et langage

- La proposition « le cygne est blanc » peut avoir deux sens
  - Sens 1 : le cygne particulier (ce cygne) devant lequel je me trouve est blanc
  - Sens 2 : les cygnes en général (catégorie d'oiseaux) sont blancs.
- La première proposition est de nature descriptive.
- La seconde proposition a une portée générale.
- Selon le contexte, on saura (en principe !) si le locuteur utilise le sens 1 ou 2.
  - Mais l'énoncé est identique et l'on peut passer insensiblement du premier au second sens.
  - L'expérience ne peut être un moyen de preuve de la proposition générale : ce n'est pas parce que l'on a observé que des cygnes blancs que tous les cygnes sont blancs

118

## cygne noir et cygne blanc



- Pour mémoire, jusqu'au 18<sup>e</sup> siècle, on ignorait l'existence de cygnes noirs, « apparus » en Australie
- Un oiseau rare dans le pays, rare comme un **cygne noir**
  - « rara avis in terris nigroque simillima cygno »
  - Le cygne noir, expression qui remonte à l'auteur latin Juvénal (fin du 1<sup>er</sup> siècle) est un oxymore

119

## Induction : ambiguïté du langage et cygnes blancs

- Considérons la proposition « le cygne est blanc ».



- En tant que proposition logique, elle peut prendre deux valeurs de vérité : vrai ou faux.
- Ici, la proposition (ou énoncé d'observation) est vraie.
- La photographie dénote (renvoie à) un cygne qui est blanc.
- Il s'agit de ce cygne spécifique : cas particulier (singulier)

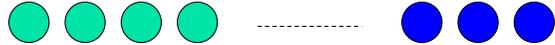
120

## Induction : ambiguïté du langage et cygnes blancs

- Considérons à nouveau la proposition « le cygne est blanc ».
- Le sujet de la proposition « le cygne » dénote-t-il le cygne particulier dont nous venons de voir la photo (à un instant donné) ou l'ensemble de tous les cygnes ?
  - La couleur blanche n'est plus seulement une caractéristique du cygne photographié, mais un **attribut** de l'espèce
- Dans le second cas, la proposition s'applique à l'ensemble de tous les cygnes, ceux que l'on a observés et ceux que l'on n'a pas encore observés.
- Ce n'est plus l'énoncé d'une propriété particulière, mais une loi qui a une portée générale.

121

## La « nouvelle énigme de l'induction »

- Le paradoxe Vleu de Goodman
  - Vleu : contraction de Vert et Bleu:
  - On considère un ensemble de boules placées en file indienne.  

  - Les premières sont de couleur verte, les dernières sont de couleur bleue.
  - Pour simplifier, on dira que toutes les boules sont vleues.
  - Considérons un expérimentateur qui observe d'abord la première boule, puis la seconde.
  - Il peut énoncer que les boules sont vertes. Il pourrait aussi dire que les boules sont vleues

122

## La « nouvelle énigme de l'induction » Le paradoxe Vleu de Goodman

- Vleu est bien défini
- L'expérimentateur (qui n'est pas encore arrivé aux boules bleues) peut énoncer deux hypothèses :
  - H0 : les boules sont vertes
  - H1 : les boules sont vleues
- Les deux hypothèses sont compatibles avec les observations
  - Chaque nouvelle observation vient aussi bien confirmer H1 que H0
- « est vleu » est un prédicat bizarre par rapport à « est vert » !
- Goodman examine pourquoi cela nous paraît bizarre.
- Renversons l'ordre dans lequel on examine les boules



- En allant de la droite vers la gauche, on peut dire que les boules sont bleues ou qu'elles sont « blertes »

123

## La « nouvelle énigme de l'induction » Le paradoxe Vleu de Goodman

- Analysons l'idée du « prédicat disjonctif »
- « est vert » ne serait-il pas disjonctif par rapport à « est vleu » ?
- En rupture avec l'idée d'uniformité de la nature souvent avancée pour justifier l'induction.
- En fait, si l'on considère comme couleurs primaires vleu et blert, on peut dire qu'une série de boules vertes est de couleur vleue jusqu'à un certain rang, puis blerte ensuite
- Dans ce système, c'est « est vert » qui est disruptif
- Donc, la question que l'on doit se poser est : **Pourquoi utilisons-nous vert et bleu comme couleurs primaires dans le langage ordinaire, plutôt que « vleu » et « blert » ?**
- Pour Goodman, « les racines de la validité inductive se trouvent dans notre façon d'utiliser le langage » (FFP 124).

124

## La « nouvelle énigme de l'induction » Le paradoxe Vleu de Goodman

- Ce qui est régulier dans un langage ne l'est pas dans un autre et vice-versa
- Goodman, contrairement à Hume, n'aborde pas le problème de l'induction comme une habitude psychologique contestable
  - Une « proposition reposante dans l'âme » selon le théologien franciscain Jean Duns Scot
- **C'est le concept même de régularité qui est remis en cause**
- **On utilise vert et bleu parce que ces termes sont ancrés (implantés) dans le français et qu'ils façonnent notre perception.**
- Goodman considère qu'une inférence inductive est **valide** si elle est couramment implantée
- Justification très différente de celle de Mill (uniformité de la nature)
- Approche de Goodman : « nominaliste » (en rapport avec les noms)

125

## Régularités et impermanence

- L'illusion de la permanence est entretenue par l'**ambiguïté du langage**
- Si on énonce « **le saphir est bleu** », on ne prend pas en compte le changement du bleu au violet selon la lumière qui éclaire le saphir
- « Bleu » n'est pas un attribut général et permanent du saphir.
  - « Ce saphir est bleu si on l'éclaire à la lumière du jour ».
  - Existence par conditions : caractère relatif



126

## La « nouvelle énigme de l'induction » Le paradoxe Vleu de Goodman

- Est-ce que les dénominations des couleurs sont uniquement affaire de langage ?
  - Connaissez-vous une langue qui utilise les couleurs vleues et blertes
  - Pouvons-nous apprendre à un singe le concept de couleur vleue ?
  - Goodman n'omet-il pas les bases animales de l'induction ?
- Mais les remarques de Goodman sont fondamentales pour une bonne méthodologie des prévisions boursières
  - Pas de langage naturellement pertinent de représentation (encodage) des données boursières et financières
  - Les données brutes (quantitatives, qualitatives, textuelles) sont « encodées », c'est-à-dire transformées en inputs (ratios financiers, ...) directement utilisables
  - Un réseau de neurones ne peut traiter qu'un nombre limité d'inputs

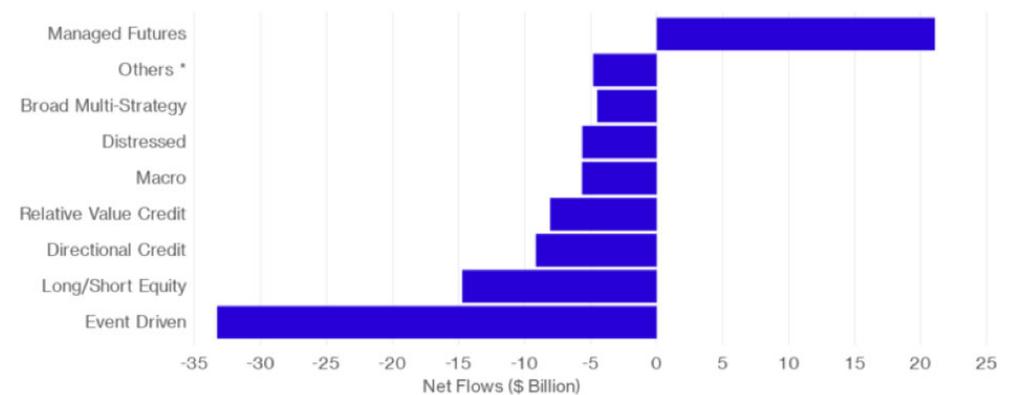
127

128

## Arbitrage statistique

129

## Computer-Driven Hedge Funds Rule



Source: eVestment

\*Includes MBS strategies, market neutral, convertible arbitrage

Bloomberg

En 2016, les hedge funds gérés selon des stratégies “systématiques” (rule driven) ont collecté 21 milliards de \$. Les hedge funds gérés de manière discrétionnaire ont subi des retraits de 60 milliards de \$.

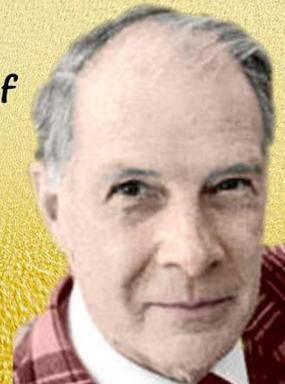
130

Peut-on investir avec succès à partir de règles « data-driven » ?

Mathematics is an interesting intellectual sport but it should not be allowed to stand in the way of obtaining sensible information about physical processes.

financial

— Richard Hamming



Richard Hamming, prix Turing 1968

131

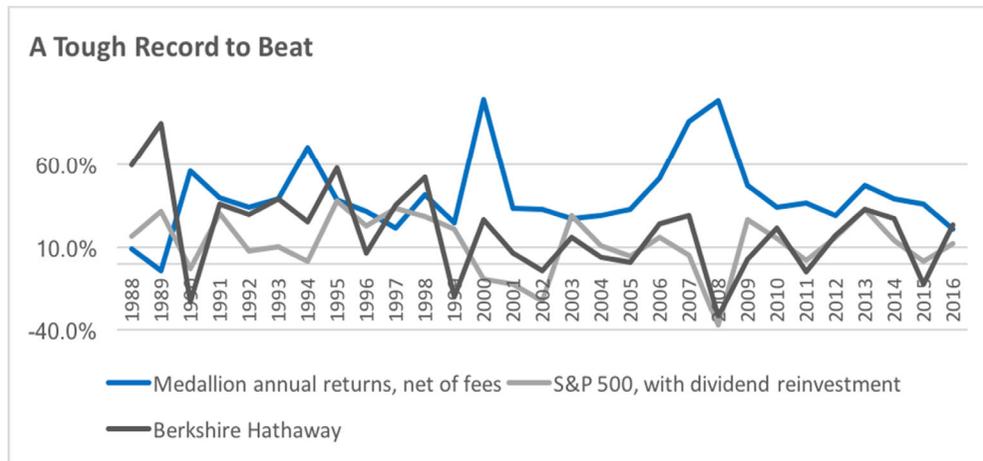


Jim Simons, mathématicien, patron de Renaissance Technologies : des stratégies d'investissement à partir d'analyses statistiques, de (très) bonnes données.

Une performance annuelle moyenne de 72% depuis le lancement du hedge fund, il y a 29 ans

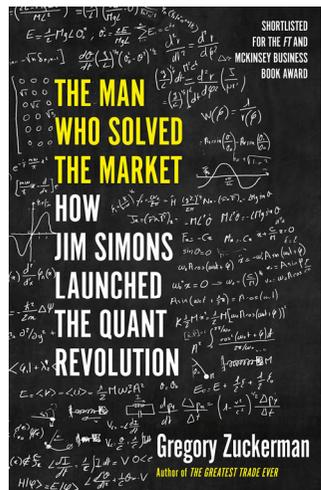
132

## Best in class: rentabilités du fonds Medallion, de Berkshire Hathaway et du S&P500



133

## Les bureaux de Renaissance Technologies à Stony Brook, près de l'Université

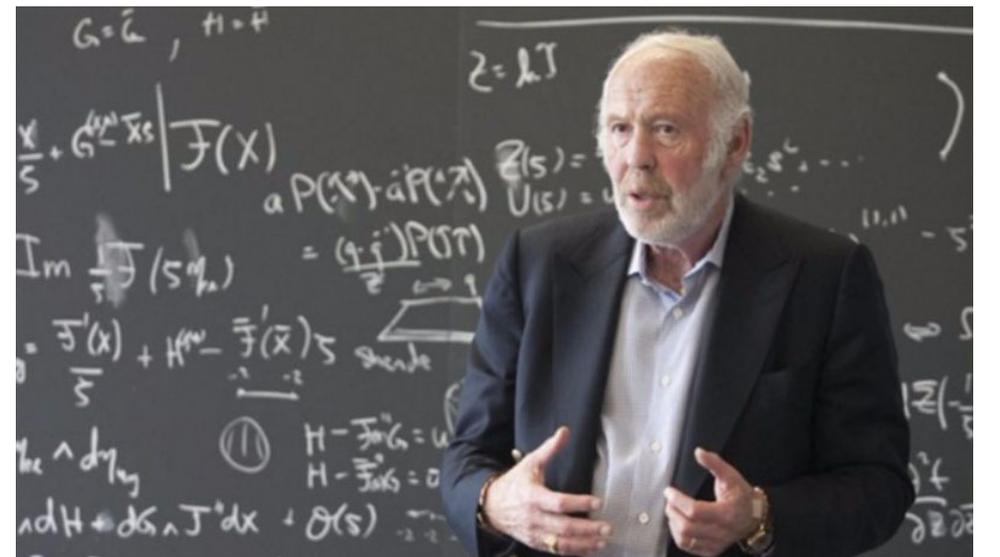


Entretien avec Jim Simons (sous-titré en français !)  
The mathematician who cracked Wall Street

<https://www.youtube.com/watch?v=U5kldtMJGc8>

In this fast-paced narrative, Zuckerman examines how Simons launched a quantitative revolution on Wall Street, and reveals the impact that Simons, the quiet billionaire king of the quants, has had on worlds well beyond finance.

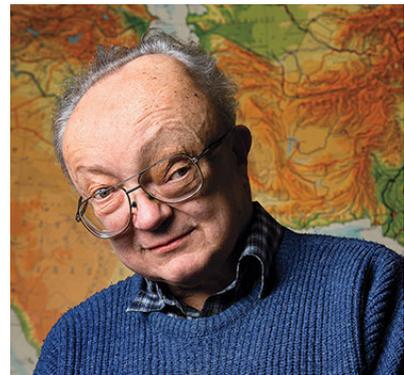
135



<https://awealthofcommonsense.com/2019/11/non-intuitive-lessons-from-the-man-who-solved-the-market/>

136

"LTCM's basic error was believing its models were truth, We never believed our models reflected reality — just some aspects of reality." Nick Patterson, RT partner.  
<https://awealthofcommonsense.com/2019/11/non-intuitive-lessons-from-the-man-who-solved-the-market/>



[https://en.wikipedia.org/wiki/Nick\\_Patterson\\_\(scientist\)](https://en.wikipedia.org/wiki/Nick_Patterson_(scientist))

"Some signals that make no intuitive sense do indeed work". Robert Mercer, ancient co-PDG de RT et personnage par ailleurs très controversé...



Questions pertinentes, relatives à la modélisation statistique et économique

## Long Term Capital Management LTCM

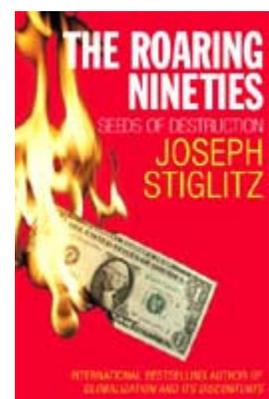


- Présentation du fonds
- Les protagonistes
- Les stratégies
  - Convergence trades
  - Paris sur le retour à la moyenne des spreads
- Communication financière
  - probabilités de gains / pertes
- 1998 : la chute
  - Divergences : faillite de l'État Russe
  - Sauvetage par la Fed



## Long Term Capital Management LTCM

- Création en 1994
- Plus grand hedge fund des années 1990
- Investissements sur les marchés jusqu'à 1280 milliards de \$
  - Surtout hors-bilan
  - Utilisation de produits dérivés
  - Très peu de fonds propres (fort effet de levier)
- Performances les premières années : +21%, +43%, +41%



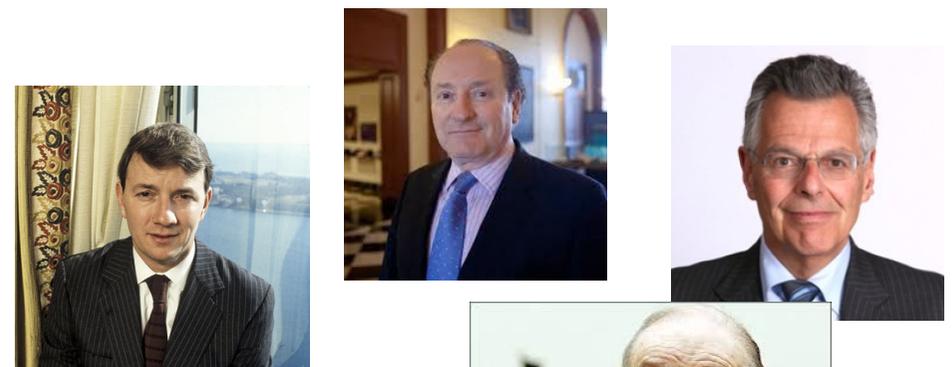
The Roaring Nineties  
 Seeds of Destruction  
 Joseph Stiglitz



## Long Term Capital Management LTCM



- Les principaux protagonistes



John Meriwether, Bob Merton, Myron Scholes, Alan Greenspan

## La stratégie de base du fonds



- Jouer sur les écarts anormaux entre les cours de deux titres proches (basis risk) et parier sur un retour à la moyenne
  - *Relative value, spread betting, pair trading, convergence trades*
  - *Hypothèse de stationnarité des spreads*
- Communication financière auprès des investisseurs utilisant largement le recours aux probabilités
  - *Probabilité de perdre 5% sur un horizon de placement d'un an = 12%*
  - *Probabilité de perdre 20% sur un horizon de placement d'un an = 2%*
  - *Approche « scientifique »...*

141

## Résultat des courses

Après un début en fanfare, faillite rapide et retentissante ayant menacé la stabilité du système financier

Organisation d'un sauvetage coûteux pour les grandes banques sous l'égide de la Fed

Au lieu de revenir vers leur niveau normal, les spreads avaient divergé

"A riveting account that reaches beyond the market landscape to say something universal about risk and triumph, about hubris and failure." —The New York Times

# WHEN GENIUS FAILED

WITH A NEW AFTERWORD ADDRESSING TODAY'S FINANCIAL CRISIS

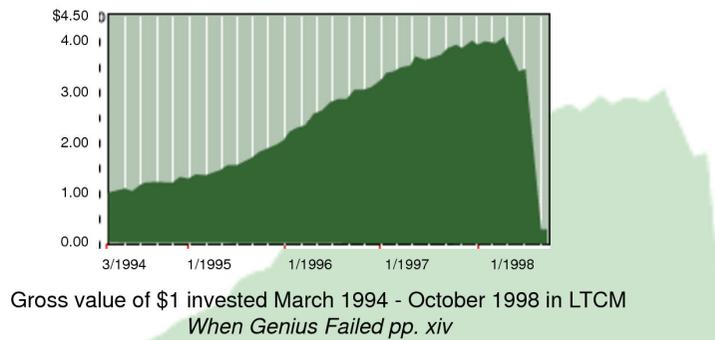
THE RISE AND FALL OF LONG-TERM CAPITAL MANAGEMENT

ROGER LOWENSTEIN



142

## Long-Term Capital Management

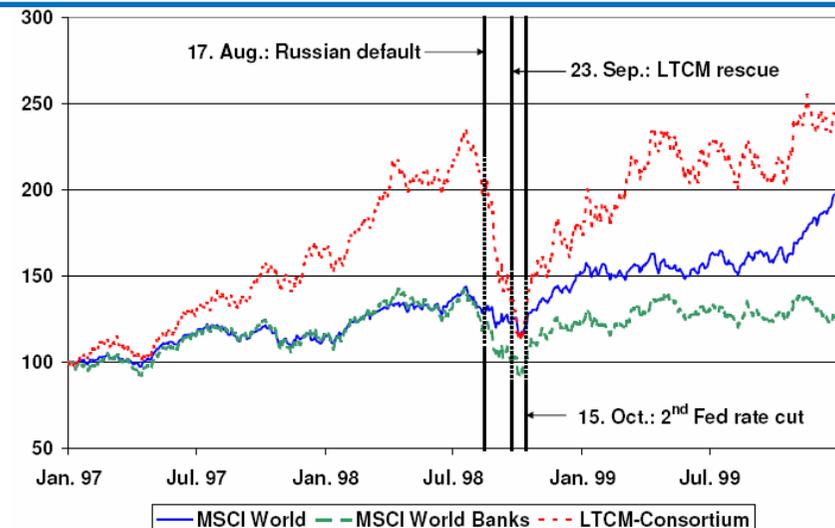


The company used complex mathematical models to take advantage of fixed income arbitrage deals (termed convergence trades) usually with U.S., Japanese, and European government bonds.

143

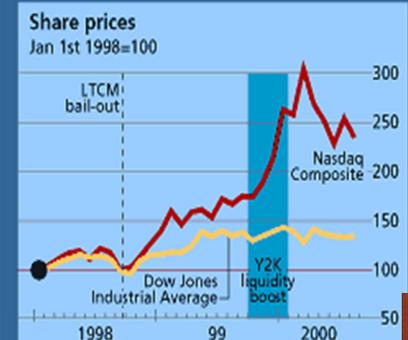
17 août 1998, défaut de la Russie, 23 Septembre bail-out de LTCM, 15 octobre nouvelle intervention de la Fed

Dans les « modèles statistiques », la baisse des prix des titres russes était anormale, et pourquoi le risque économique et financier, bien facile à appréhender



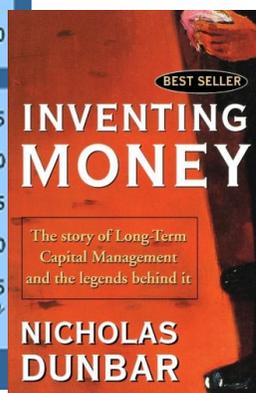
144

## The Fed effect



Source: Primark Datastream

**LTCM**  
The financial technology company

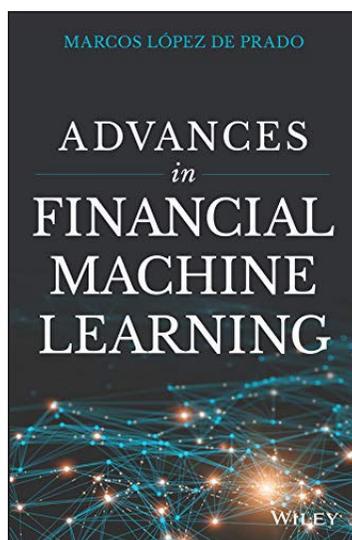


| 1998   | 2008   |
|--|--|
| <ul style="list-style-type: none"> <li>LTCM, JOHN MERIWETHER'S HEDGE FUND, COLLAPSES</li> <li>THE PIONEER OF QUANTITATIVE INVESTMENT LOST \$4.6BN IN 4 MONTHS</li> <li>TRIGGERED A FED-ORGANISED BAIL-OUT BY WALL STREET BANKS</li> <li>UNDONE BY MASSIVE LEVERAGE - UP TO \$50 FOR EVERY DOLLAR INVESTED</li> </ul> | <ul style="list-style-type: none"> <li>INVESTORS IN MERIWETHER'S JWM FUND TUMBLES 28PC</li> <li>REJECTS LTCM COMPARISONS - FUND 'HAS LESS RISK AND MORE LIQUIDITY'</li> <li>SOME INVESTORS WANT OUT DESPITE THE FUND'S REASSURANCES</li> <li>INVESTORS HAVE UNTIL MONDAY TO REQUEST WITHDRAWALS</li> </ul> |

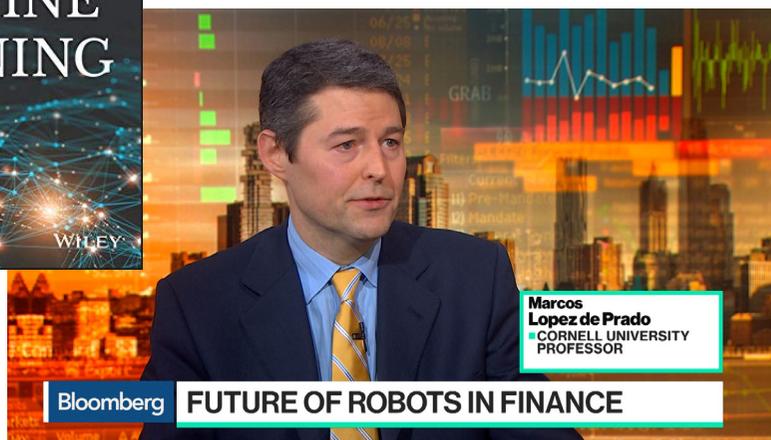
## Statistique et stratégies d'investissement

- Quel est le principe des stratégies systématiques / quantitatives d'investissement boursier ?
- Quelles en sont les limites ?
  - Critiques de la méthode inductive (Hume - Goodman)
  - Encodage des entrants (inputs)
  - Validation des stratégies
- Quels sont les problèmes liés à la validation des stratégies d'investissement (backtesting) ?
  - Spurious correlations : corrélations fictives
  - Validation ex-post (data snooping)
  - Données aberrantes (outliers)
  - Pas de méthode expérimentale (! au price impact)

146



Marcos Lopez de Prado, Professeur de « finance professionnelle » (financial practice) – Cornell & responsable de la recherche ADIA - Abu Dhabi Investment Authority



## Validation des stratégies : backtesting

- Principe d'un backtest : simulation d'une stratégie d'investissement sur un historique de données

## The Dangers of Backtesting

### 11.1 MOTIVATION

Backtesting is one of the most essential, and yet least understood, techniques in the quant arsenal. A common misunderstanding is to think of backtesting as a research tool. Researching and backtesting is like drinking and driving. Do not research under the influence of a backtest.

147

148

## Validation des stratégies : usual suspects, seven sins

1. **Survivorship bias:** Using as investment universe the current one, hence ignoring that some companies went bankrupt and securities were delisted along the way.
2. **Look-ahead bias:** Using information that was not public at the moment the simulated decision would have been made. Be certain about the timestamp for each data point. Take into account release dates, distribution delays, and backfill corrections.
3. **Storytelling:** Making up a story *ex-post* to justify some random pattern.
4. **Data mining and data snooping:** Training the model on the testing set.
5. **Transaction costs:** Simulating transaction costs is hard because the only way to be certain about that cost would have been to interact with the trading book (i.e., to do the actual trade).
6. **Outliers:** Basing a strategy on a few extreme outcomes that may never happen again as observed in the past.
7. **Shorting:** Taking a short position on cash products requires finding a lender. The cost of lending and the amount available is generally unknown, and depends on relations, inventory, relative demand, etc.

149

Seven sins of quant investing  
Survivorship bias  
Look-ahead bias  
Story telling  
Data mining and snooping  
Signal decay and turnover  
Outliers and data normalization  
Asymmetric pattern



Deutsche Bank  
Markets Research

Global

Quantitative Strategy  
Signal Processing

Seven Sins of Quantitative Investing

- Survivorship bias
- Look-ahead bias
- The sin of story telling
- Data mining and data snooping bias
- Signal decay and turnover
- Outliers – the story of spectacular successes and failures
- The asymmetric payoff pattern and shorting cost

150

151

152