JOURNÉES TÉLÉCOM-UPS
2012

Conférences-débats

Enseigner les probabilités

en Grande École
...et bientôt en Classes Préparatoires ?

Stage LIESSE à Télécom ParisTech en collaboration avec l'UPS

9 et 10 mai 2012

http://perso.telecom-paristech.fr/~rioul/liesse.html

Journées organisées par Olivier Rioul, enseignant-chercheur à Télécom ParisTech

ups Union des Professeurs de Spéciales

Liesse

Télécom ParisTech
46 rue Barrault
75013 Paris
www. telecom-paristech.fr

TELECOM ParisTech

Table des matières

Sylvie Méléard — Aléatoire : les enjeux du cours de Probabilités en première année de l'Ecole Polytechnique **1**

Roger Mansuy — Des chaînes de Markov en classes préparatoires ? **19**

Laurent Decreusefond — L'enseignement des probabilités à Télécom ParisTech **45**

Jean-Marc Ginoux — Sur les conférences d'Henri Poincaré à l'Ecole Supérieure des Postes et Télégraphes (aujourd'hui Télécom ParisTech) **69**

Alain Maruani — Une promenade aléatoire au voisinage des probabilités **71**

Olivier Rioul — Les probabilités sans peine ? **99**

Gersende Fort & Eric Moulines — Chaînes de Markov finies et simulation **141**

Yves Guiard — Les probabilités comme un artisanat **165**

Synopsis

Enseigner les probabilités est un thème particulièrement important et d'actualité pour les professeurs de classes préparatoires scientifiques. En effet, la réforme actuelle du secondaire ayant déjà fait évoluer les contenus des enseignements en probabilités, on devine une révision des programmes en CPGE à la rentrée 2013, année qui produira les premiers bacheliers de la réforme des lycées. Les probabilités jusqu'alors absentes des CPGE devraient, avec une très forte... probabilité, apparaître à cette rentrée 2013 en Sup., en 2014 en Spé., et en 2015 pour les concours d'entrée dans les Grandes Ecoles. L'UPS a commencé à réfléchir au sujet depuis plus d'un an et la commission de rénovation des programmes va se réunir et donner les premières précisions sur les contenus courant 2012. Il apparaît donc qu'un nombre croissant de professeurs de CPGE sont demandeurs d'information et de formation en probabilités et statistiques.

Sylvie Méléard

Professeure et présidente du département de Mathématiques Appliquées de l'École Polytechnique

Aléatoire : les enjeux du cours de Probabilités en première année de l'Ecole Polytechnique

Nous présenterons l'enseignement du cours introductif de Probabilités donné en première année de l'École Polytechnique, pour les 500 polytechniciens réunis, issus de culture et formation mathématiques très différentes. Nous mettrons en évidence les spécificités de ce domaine :

– Le concept de hasard, très lié à des enjeux philosophiques ou religieux, a mis longtemps à émerger et ce domaine, bien qu'utilisant de nombreux outils d'analyse ou d'algèbre, a du mal à s'intégrer dans la culture mathématicienne ;
– Les développements probabilistes du 20ème siècle sont essentiels dans de nombreux domaines d'applications tels la physique statistique, les télécommunications, les mathématiques financières et la biologie. L'enseignement des probabilités à l'Ecole Polytechnique est porté par ces applications et par des ouvertures sur les statistiques ;
– La modélisation : les probabilités sont en lien étroit avec la vie quotidienne et il y a d'innombrables situations où le hasard intervient, de natures très différentes d'où la nécessité d'une abstraction mathématique pour donner un cadre général d'étude. Ce passage de la problématique appliquée au modèle est souvent difficile et a conduit à bien des paradoxes dans le passé ;
– La simulation est une méthode expérimentale essentielle pour comprendre ce qu'est la notion de loi ou de distribution. Chaque élève développe un projet numérique en Scilab lié à une modélisation probabiliste ;
– La nécessité d'outils d'analyse difficiles pour justifier complètement la

théorie probabiliste : la théorie de la mesure et l'intégrale de Lebesgue. Le cours de première année essaie de présenter les idées de la construction de l'espérance mathématique (intégrale de Lebesgue abstraite mais de masse 1), tout en évitant les écueils d'une trop grande et ennuyeuse difficulté théorique.

C'est ce grand écart entre l'apparente simplicité de certains problèmes probabilistes concrets et l'abstraction que nécessite leur résolution qui peut rendre le monde de l'aléatoire difficile ou inquiétant, mais c'est aussi ce qui en fait un domaine mathématique fascinant et palpitant.

Référence : Aléatoire, Éditions de l'École Polytechnique, 2010.

ALEATOIRE - Les enjeux du cours de Probabilités en première année de l'Ecole Polytechnique

Télécom ParisTech, 09 mai 2012

http://www.mathematiquesappliquees.polytechnique.edu/accueil/programmes/cycle-polytechnicien/annee-1/

Organisation du cours pour 500 élèves.

Un cours en amphithéatre, 22 groupes de petites classes en parallèle, et un projet numérique.

Le projet numérique compte pour un tiers de l'évaluation.

Spécificité de l'audience: un niveau de mathématiques très hétérogène.

- 400 élèves issus des concours français, des différentes filières de classes préparatoires.
- 100 élèves étrangers (concours ou dossier)
- quelques étudiants de l'Université.

Profils mathématiques très variés.

Cours de Probabilités: **Le premier contact des élèves avec les Mathématiques Appliquées**
Fondalement tourné vers les applications et la simulation.

Les probabilités

Une théorie mathématique pour quantifier le HASARD.

Fondamentale dans de nombreux domaines d'applications.
Les applications développées à l'Ecole dans le cadre de la troisième année ou dans les masters cohabilités avec l'UPMC ou Orsay:

- Biologie, Ecologie
- Finance, Assurance.
- Informatique et Réseaux de télécommunications
- Médecine, Imagerie médicale
- Physique (physique quantique, physique des particules)
- Traitement du signal, de la parole

Modélisation, Abstraction, Simulation

Nouveauté pour les élèves à découvrir les Mathématiques Appliquées:

- Modélisation
- Analyse théorique
- Simulation numérique.

Le cours de probabilités est construit sur cette idée. Les probabilités sont en lien étroit avec la vie quotidienne.

- Traduction d'une situation concrète en un problème mathématique abstrait: **La modélisation.**
- Innombrables situations où le hasard intervient, de natures très différentes: nécessité d'une **abstraction mathématique** pour donner un cadre général d'étude.

Le modèle probabiliste

Méthodes numériques

- Les probabilités sont beaucoup utilisées à des fins numériques (gros codes de calcul, modélisation de la méconnaissance, turbulence):

  Méthodes de Monte-Carlo

  - Efficaces en grande dimension (mathématiques financières, météorologie, aérospatial)
  - résultats très rapides (en temps réel)
  - Simulation de phénomènes irréguliers.
- Méthodes probabilistes de simulation: mise en place d'expérimentations fictives sur machine appelées **simulations**.

Les élèves apprennent scilab et développent un projet de modélisation qui va jusqu'à la simulation.

Exemples de projets de simulation

Percolation et processus d'invasion

Loi de Mendel et évolution d'une population sous neutralité

Points fixes et cycles de permutations aléatoires

Restauration d'un signal structuré

Simulation d'une couverture de réseaux

Du comportement de certains gaz à Google

Débruitage et théorie de l'information

La value at Risk

L'offre de cours en Probabilités et Statistiques à l'Ecole Polytechnique

Le cours de tronc commun (Année 1):

Va de la construction du modèle probabiliste au théorème de la limite centrale.
Quelques ouvertures aux dynamiques stochastiques.

Année 2: trois cours proposés.

Chaînes de Markov et martingales
Introduction aux méthodes statistiques
Modal: Simulation numérique aléatoire

Année 3 - Master 1

Un programme de Mathématiques Appliquées avec plusieurs cours sur les modèles probabilistes, liés aux parcours que l'on propose. Ces parcours sont associés à un M2 cohabilité avec d'autres établissements:

Mathématiques et sciences du vivant

Mathématique, vision, apprentissage

Probabilités et finances

Probabilités et statistiques.

Les cours de 3ème année liés à l'aléatoire:

Modèles stochastiques en finance

Apprentissage statistique et estimation non paramétrique

Communication networks, algorithms and probability

Modèles aléatoires en écologie et évolution

Stochastic simulation and Monte-Carlo methods

Processes and estimation

Le cours de tronc commun

7 cours de 1h30 + un amphi introductif avec un peu d'histoire.

7 PC (TD) de 2h

L'importance de l'histoire des Probabilités

Un défi de l'homme face au divin: la théorie des probabilités a mis beaucoup de temps à émerger.

- Très certainement d'origine arabe: az-zahr: le dé.
- En Inde, 4ème siècle, existence d'une science du dé et connaissance de ses rapports étroits avec une évaluation de type sondage. (Le Mahabharata)
- En Europe, aux 16-17èmes siècles, émergence d'une science du jeu de dé. Cardan, Kepler, Galilée.
  Théorie rigoureuse: Pascal.
- Résolution de controverses juridiques (Fermat, Leibniz).
- Autre impulsion motivée par des problèmes d'assurance (tables de mortalité et rentes viagères).
- Développement des Statistiques: outil puissant pour les organismes de décision...

Les développements mathématiques

Les probabilités sont un outil privilégié de modélisation des comportements humains, mais deviennent aussi un grand champ de développement des Mathématiques.

- 19ème siècle et début 20ème siècle: Essor des probabilités grâce aux méthodes d'analyse.
  - Calcul intégral et différentiel (Laplace, Gauss)
  - Théorie de la mesure (Borel, Lebesgue).
- A partir du 20ème siècle: Etude de phénomènes aléatoires qui évoluent au cours du temps. Processus de Markov
  - Problèmes de Physique statistique, Mouvement Brownien,
  - Problèmes de démographie. Processus de branchement, Processus de Poisson.
  - Théorie statistique, biométrie.

Aujourd'hui...

- Le modèle probabiliste actuel: Kolmogorov - 1933.
  Calcul stochastique: Itô, Doeblin, à partir de 1945.
- Les probabilités interviennent dans de nombreux développements mathématiques récents.
- L'Ecole Française de Probabilités est très active: Médaille Fields 2006, Wendelin Werner. (Probabilités et transformations conformes du plan).
- Les probabilités largement utilisées dans d'autres sciences (dures ou humaines), dans les entreprises et dans les banques.

Deux idées majeures incontournables

- Loi des grands nombres

  Hasard $=$ méconnaissance des paramètres intervenant dans une expérience.
  Les informations seront données par des répétitions de l'expérience.

- Conditionnement et indépendance
  La construction d'un modèle probabiliste repose sur l'information connue a priori sur l'expérience.
  Si l'information change, la probabilité de réalisation change.

  Si l'information donnée sur un phénomène n'a pas d'influence sur la probabilité de réalisation d'un autre phénomène, les phénomènes sont dits indépendants.
  L'hypothèse d'indépendance est fondamentale pour les calculs. **Notion liée à la probabilité.**

Une difficulté majeure: la notion de loi

Dans tout le cours, on adopte un point de vue fonctionnel: la notion de **variable aléatoire**.

Vocabulaire perturbant (et décalé) des probabilités.

Variable aléatoire = fonction de l'alea.

Intérêt d'une variable aléatoire: transporte l'espace de probabilité abstrait sur l'ensemble des valeurs de la variable aléatoire.

Loi de la variable aléatoire = mesure image de la probabilité abstraite par cette fonction.

Cela définit un nouvel espace de probabilité sur lequel on peut faire des calculs.

Simulations

Simulation: Expérimentation fictive sur machine d'un phénomène que l'on peut modélisé.

Elle permet de visualiser une expérience aléatoire, de calculer des quantités numériques et de vérifier certains résultats théoriques.

Chaque cours est émaillé de simulations et les élèves peuvent les retrouver sur le site internet et jouer avec.

Ils peuvent aussi en voir le code, en Scilab.

Marche aléatoire et mouvement brownien

Marche erratique sur la droite réelle: on avance ou l'on recule d'autant, avec même probabilité, à chaque pas de temps

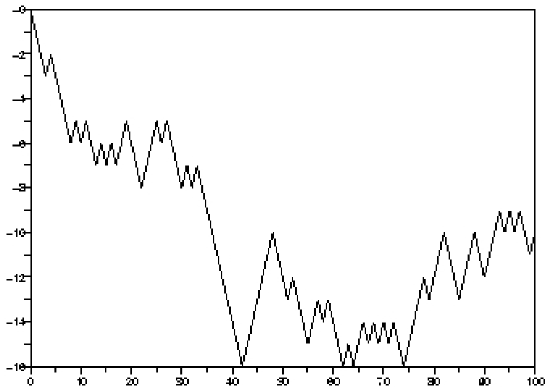


Figure: marche aléatoire en dimension 1

Vu d'une échelle plus macroscopique

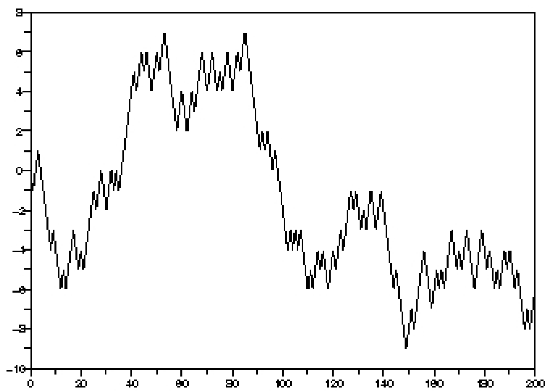


Figure: marche aléatoire en dimension 1

Encore plus...

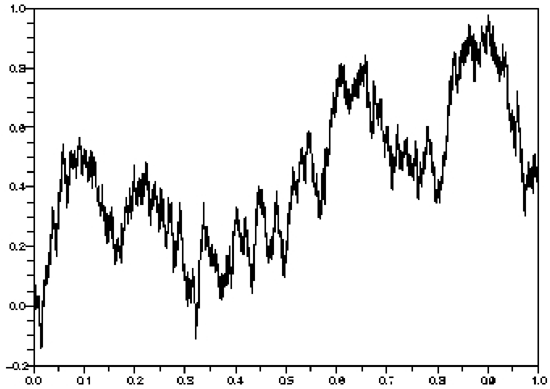
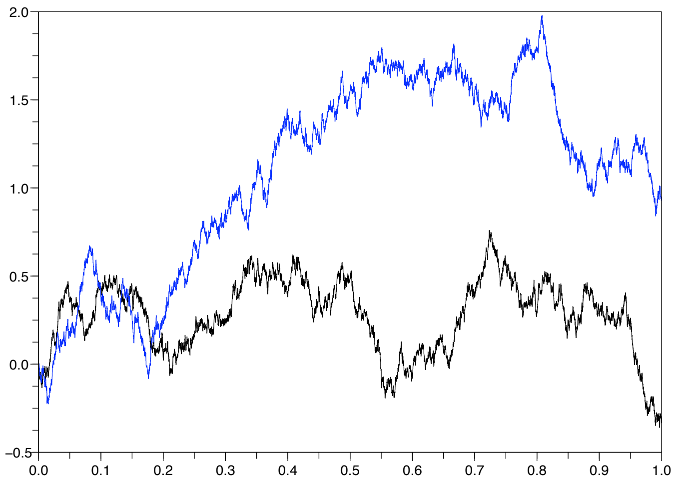


Figure: mouvement brownien en dimension 1

Trajectoires aléatoires

Réalisons plusieurs simulations:

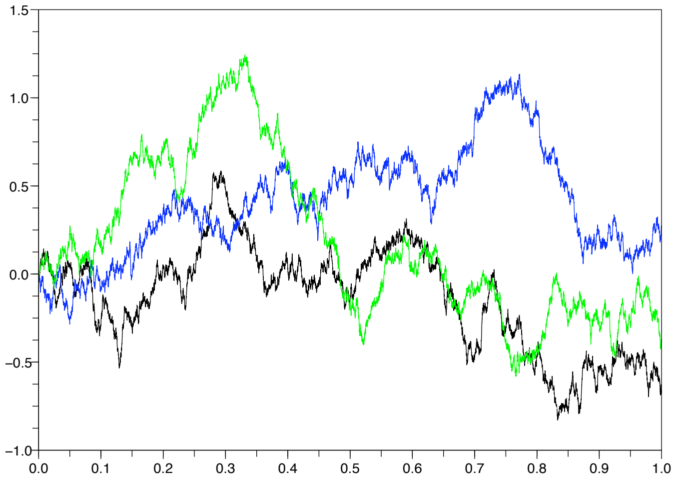
2



Trajectoires aléatoires

Réalisons plusieurs simulations:

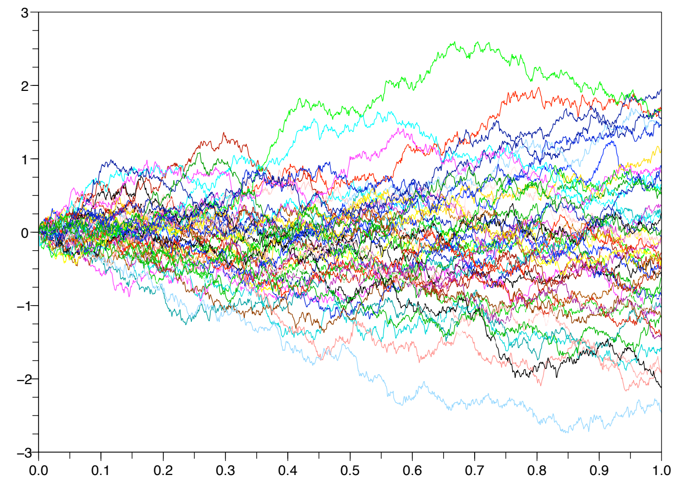
3



Trajectoires aléatoires

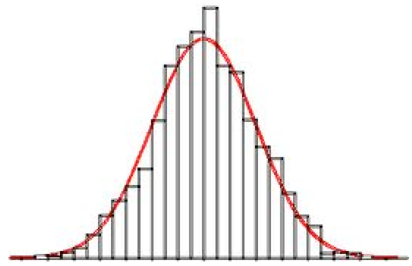
Réalisons plusieurs simulations:

50



La loi normale, loi de Gauss, courbe en cloche

Répartition des valeurs au temps $t = 1$ approchée par la courbe de Gauss: fonction de densité au temps 1.



Loi normale: une loi universelle

Les grands chapitres du cours

- Espace de Probabilité.
- Variables aléatoires sur un espace fini ou dénombrable.
- Variables aléatoires réelles.
- Vecteurs aléatoires, lois conditionnelles, calcul de loi.
- Convergences, Loi des grands nombres.
- Fonction caractéristique, vecteur gaussien, convergence en loi, théorème de la limite centrale.
- Introduction aux processus aléatoires

Espace de Probabilité

Expérience aléatoire et Espace d'états. Evénements aléatoires. **Notion d'information**.

Approche intuitive d'une probabilité: **limite de fréquences empiriques quand le nombre d'expériences tend vers l'infini.** Les premieres propriétés.

Exemples: les modèles d'urnes. Tirage avec ou sans remise, lois hypergéométrique et binomiale.

Définition générale d'une probabilité. (Cas du jeu de Pile ou Face).
Notion de tribu.
L'espace de probabilité $(\Omega, \mathcal{A}, P)$.
Probabilité sur un espace dénombrable.
Probabilités conditionnelles. Formule de Bayes, Exemples.
Indépendance. Expériences aléatoires indépendantes.

Théorème de Borel Cantelli

Variables aléatoires sur un espace fini ou dénombrable

Variable aléatoire et sa loi.

Probabilité sur un espace dénombrable.

Variable aléatoire discrète usuelles. Tout est défini à partir d'un jeu de Pile ou Face: lois de Bernoulli, géométrique, binomiale, convergence vers la loi de Poisson.

Le processus de Poisson: leur donner l'idée d'une dynamique aléatoire qui évolue au cours du temps.

Espérance, variance, fonction génératrice, calcul de moments.

Couple de variables aléatoires, cas discret. Lois conditionnelles, variables aléatoires indépendantes. Somme de variables aléatoires indépendantes.

Information et Entropie.

Variables aléatoires réelles

Variable aléatoire réelle et sa loi

Fonction de répartition. Caractérisation d'une probabilité sur $\mathbb{R}$. (admis)

Variable aléatoire à densité. Variable aléatoire uniforme.

Simulation d'une variable aléatoire (inversion de la fonction de répartition).

Variable aléatoire exponentielle, normale.

Construction de l'espérance (va étagées, va positives comme limites de va étagées, va quelconques).

$$\mathbb{E}(g(X)) = \int_\Omega g(X(\omega))\mathbb{P}(d\omega) = \int_\mathbb{R} g(x)\mathbb{P}_\mathbf{X}(dx).$$

Calcul d'espérance, de variance pour les va à densité.

Calcul de loi: si X a la densité f_X, quelle est la loi de $h(X)$?

Inégalités

Vecteurs aléatoires, lois conditionnelles, calcul de loi

Vecteurs aléatoires. Loi d'un vecteur aléatoire à densité.

Couple de variables aléatoires, covariance. Moments d'un vecteur aléatoire.

Lois conditionnelles, cas à densité. Vecteurs aléatoires indépendants. Calculs de lois de vecteurs aléatoires. (Théorème du changement de variable).

Méthode de simulation dite du rejet.

Convergences, loi des grands nombres

Justifier "les lois empiriques du hasard".

Différents types de convergence Proximité des variables aléatoires: CV en moyenne, en probabilité, presque-sure. Les liens entre ces convergences, exemples et contre-exemples. Théorème de CV dominée (admis).

Loi des grands nombres: loi faible, loi forte (prouvée pour des variables aléatoires de carrée intégrable).

Applications et commentaires: convergence des histogrammes vers la densité, théorème de Weierstrass et polynômes de Bernstein.

Méthode de Monte-Carlo: simuler $\int_A f(z)dz$.

Fonction caractéristique, vecteur gaussien, convergence en loi, théorème de la limite centrale

Fonction caractéristique (transformée de Fourier).

Calcul pour les variables aléatoires usuelles. Caractérisation des variables aléatoires indépendantes. Fonction caractéristique d'une somme de variables aléatoires indépendantes.

Fonction caractéristique et moments.

Vecteurs gaussiens et fonction caractéristique. Matrice de covariance pour un vecteur gaussien.

Convergence en loi, lien avec les autres convergences. Théorème de Paul Lévy.

Théorème de la limite centrale, caractère universel de la loi normale.
Applications (contrôle d'erreur, vitesse de CV dans Monte-Carlo, Statistiques - intervalles de confiance).

Introduction aux processus aléatoires

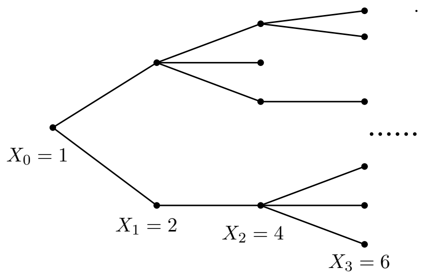
Suites récurrentes aléatoires.

Marche aléatoire.

Problème de la ruine d'un joueur.

Récurrence et transience de la marche aléatoire.

Processus de branchement.



processus de branchement

Files d'attente.

Roger Mansuy

Professeur de mathématiques en classe préparatoires MPSI au lycée Louis le Grand

Des chaînes de Markov en classes préparatoires ?

Quelle probabilités pour les CPGE ? Si tout le monde s'accorde sur l'obligation d'introduire le formalisme usuel pour des variables discrètes (ce qui serait un programme de première année), on reste plus partagé au sujet d'un programme de seconde année. Certains pensent à une théorie de l'intégration et aux calculs pour des variables continues comme dans les classes ECS et BCPST. Je propose ici de développer une autre piste de réflexions avec les chaines de Markov pour étudier un cas de dépendance tout en renforçant les liens entre les probabilités et le cours d'algèbre linéaire.

Des chaînes de Markov en prépa ?

Roger Mansuy

Introduction

Pourquoi étudier les chaînes de Markov en prépa ?

- un sujet très proche d'autres problématiques du cours (réduction en algèbre linéaire, automates ou graphes en informatique...)
- une grande source de sujets de TIPE
- un moyen de développer les compétences de modélisation
- un moyen de sortir les probas d'une vision « calculatoire » de premier cycle
- un accès à des mathématiques du XX-ième siècle
- une continuité avec le programme de la spécialité mathématiques en terminale

Extrait du bulletin officiel spécial n°8 du 13 octobre 2011

Exemples de problèmes

Marche aléatoire simple sur un graphe à deux ou trois sommets.

Marche aléatoire sur un tétraèdre ou sur un graphe à N sommets avec saut direct possible d'un sommet à un autre : à chaque instant, le mobile peut suivre les arêtes du graphe probabiliste ou aller directement sur n'importe quel sommet avec une probabilité constante p .

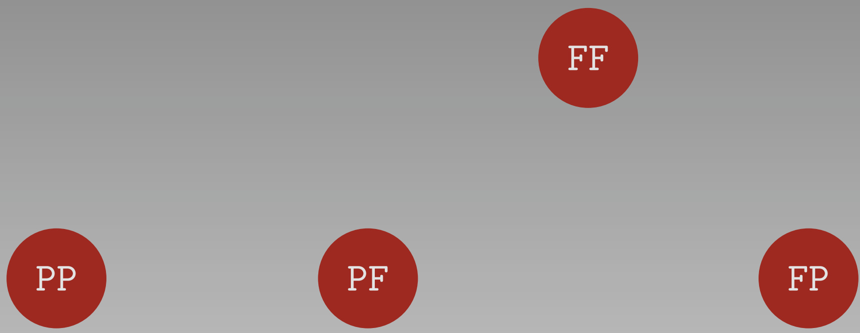
Etude du principe du calcul de la pertinence d'une page web.

Modèle de diffusion d'Ehrenfest : N particules sont réparties dans deux récipients ; à chaque instant, une particule choisie au hasard change de récipient.

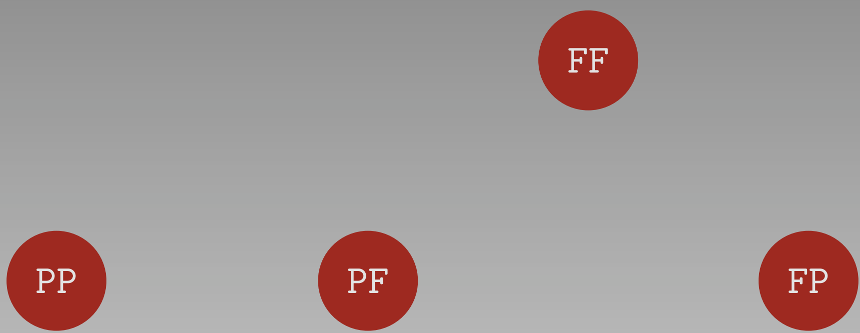
Considérons les lancers successifs d'une pièce équilibrée jusqu'à obtenir les résultats consécutifs FPP ou PPF.

- Si le motif FPP apparaît en premier, je suis gagnant.
- Si le motif PPF apparaît en premier, vous êtes gagnant.

Acceptez vous de jouer avec moi ?



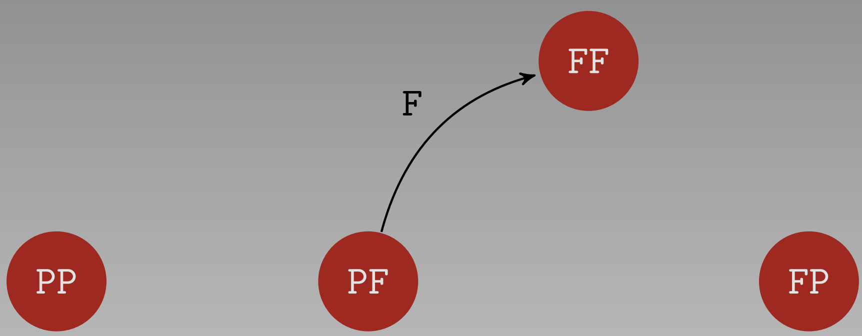
Définitions Matrice de transition Analyse à un pas Analyse asymptotique



PPF **gagne**

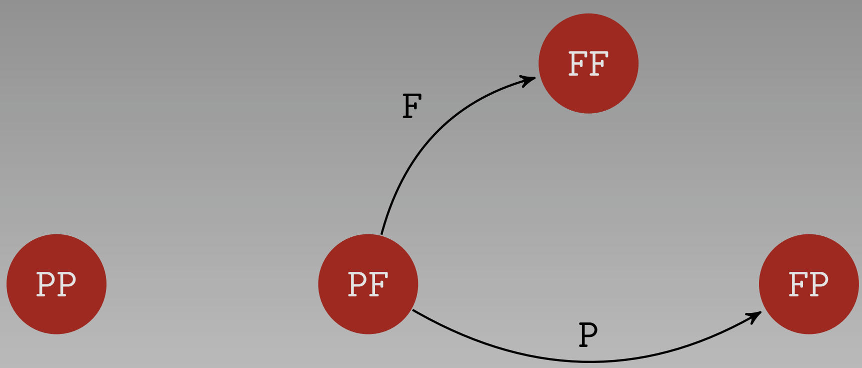
FPP **gagne**

Roger Mansuy Des chaînes de Markov en prépa?



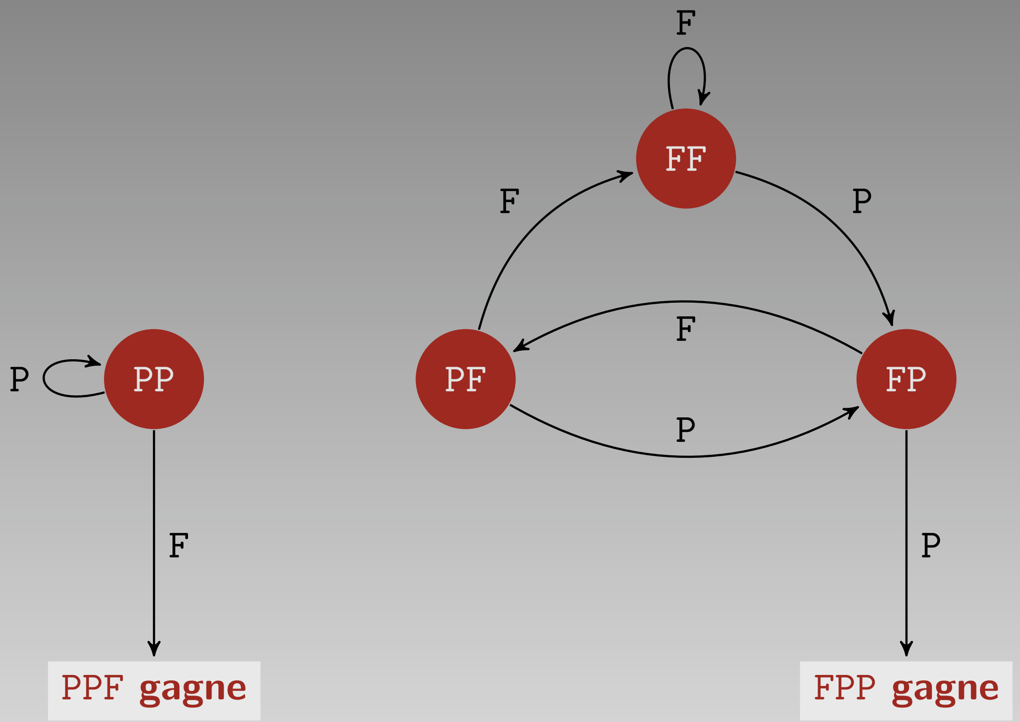
PPF **gagne**

FPP **gagne**



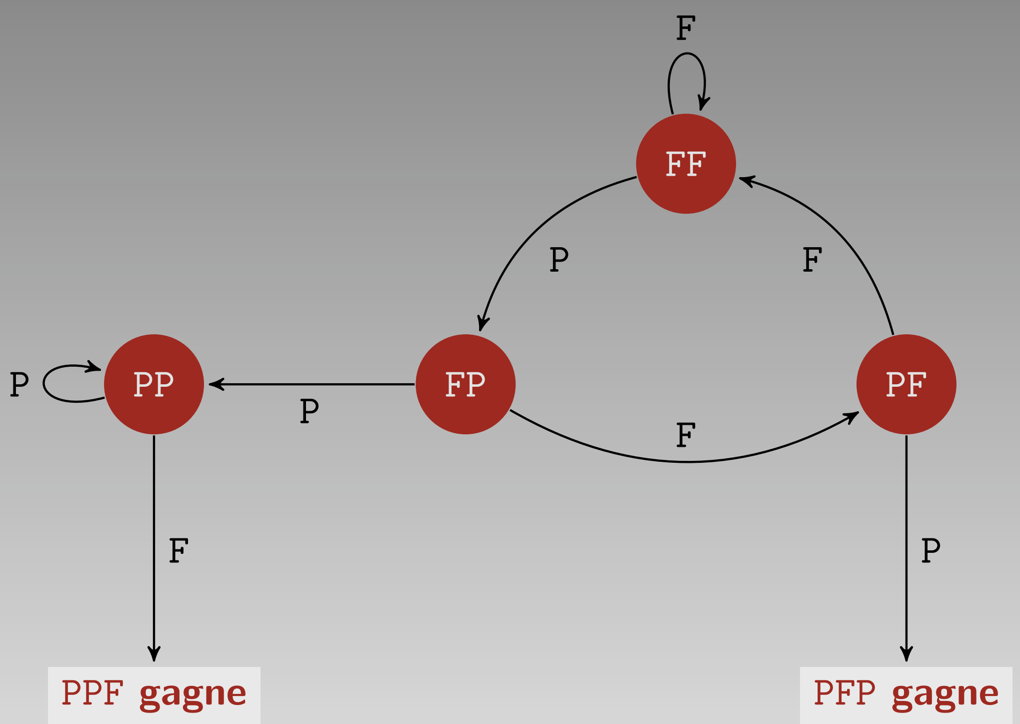
PPF **gagne**

FPP **gagne**



Définitions Matrice de transition Analyse à un pas Analyse asymptotique

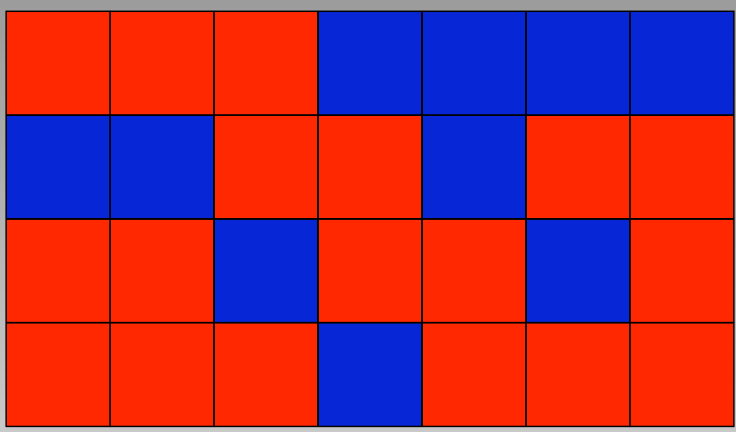
Avec les mots PFP et PPF, le graphe devient



Roger Mansuy Des chaînes de Markov en prépa?

On considère un « damier » dont les cases sont colorées en rouge ou en bleu. À chaque étape, une case est choisie uniformément et prend la couleur d'une de ses (au plus) quatre voisine. Quelle est la probabilité que le damier finisse entièrement rouge ?

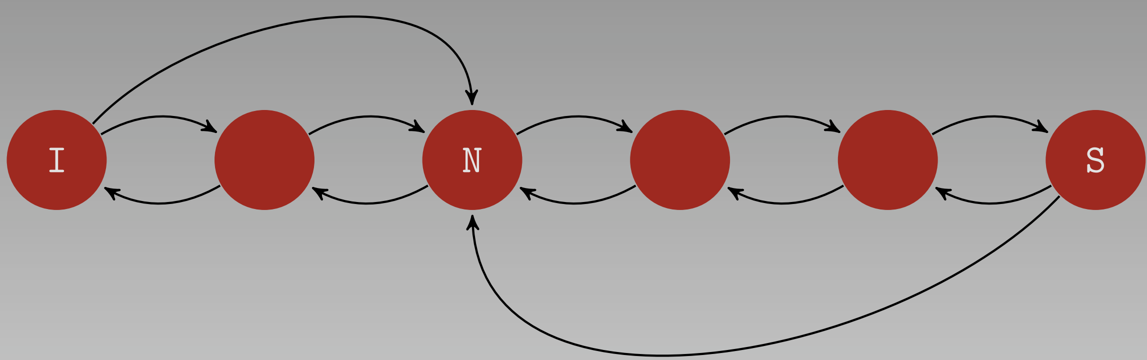
Cette fois le graphe compte 2^{mn} sommets qui correspondent à des configurations comme celle-ci :



Considérons le guichet d'une banque avec le fonctionnement suivant :

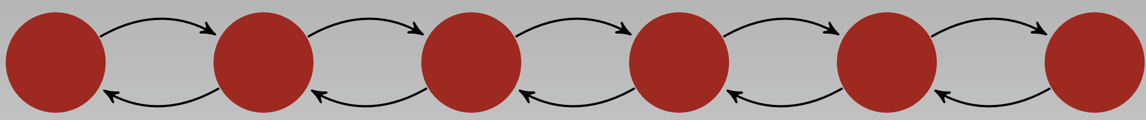
- chaque jour, le montant de liquidités en stock varie d'une unité (en positif ou en négatif) ;
- si le solde atteint le palier supérieur S, on effectue un transfert de fond pour se ramener à un solde normal N ;
- si le solde atteint le palier inférieur I, on effectue un transfert de fond pour se ramener à un solde normal N.

Existe-t-il une loi sur les entiers $[\![I, S]\!]$ stationnaire pour ce modèle ?



Simplifions cet exemple.

Considérons une marche aléatoire sur un segment d'entiers avec réflexion sur les bords.
Existe-t-il une loi stationnaire ?



Définitions Matrice de transition Analyse à un pas Analyse asymptotique

On considère le système de communication avec un centre qui émet vers l'extérieur (l'émetteur international de Hawaï) et des émetteurs satellites qui émettent vers le centre (sur chacune des îles de l'archipel). À des instants synchronisés, les émetteurs satellites envoient ou pas un message vers l'émetteur central.

1. Si deux (ou plus) messages arrivent en même temps, ils sont rejetés.
2. Tout message rejeté est automatiquement réémis vers l'émetteur central jusqu'à son envoi avec une probabilité $p \in]0,1[$.

Est ce que ce procédé de communication est stable ?

Roger Mansuy Des chaînes de Markov en prépa?

Programme

1. **Définitions, Exemples**
   Définition chaîne de Markov
   Matrice de transition
   Loi du n-ième terme
2. **Topologie de la matrice de transition**
   Relation d'équivalence de communication
   États absorbants
   Chaîne irréductible
3. **Analyse à un pas**
   Calcul de la probabilité d'absorbtion, du temps moyen avant absorbtion
4. **Analyse asymptotique**
   Loi stationnaire
   Période, chaîne apériodique
   Convergence vers la loi stationnaire

Définitions

Définition

Une chaîne de Markov est une suite de variables aléatoires $(X_n)_n$ d'états $E = \{e_1, \dots, e_N\}$ telle que
$\forall n \in \mathbb{N},\ \forall y, y_0, \dots, y_n \in E,$

$$\mathbb{P}(X_{n+1} = y | X_0 = y_0, \dots, X_n = y_n) = \mathbb{P}(X_{n+1} = y | X_n = y_n).$$

Elle est homogène si cette quantité ne dépend pas de n.

Une suite $(X_n)_n$ de variables indépendantes est une chaîne de Markov (à intérêt limité) puisque

$$\begin{aligned}\mathbb{P}(X_{n+1} = y | X_0 = y_0, \ldots, X_n = y_n) &= \mathbb{P}(X_{n+1} = y), \\ \mathbb{P}(X_{n+1} = y | X_n = y_n) &= \mathbb{P}(X_{n+1} = y).\end{aligned}$$

Soit $(Y_n)_n$ une suite de variables indépendantes identiquement distribuées. La suite

$$\left(X_n = \sum_{k=0}^{n} Y_k \right)_n$$

est une chaîne de Markov. En effet,
$\mathbb{P}(X_{n+1} = y | X_0 = y_0, \ldots, X_n = y_n)$

$$\begin{aligned} &= \mathbb{P}(X_n + Y_{n+1} = y | X_0 = y_0, \ldots, X_n = y_n) \\ &= \mathbb{P}(y_n + Y_{n+1} = y), \end{aligned}$$

$$\begin{aligned}\mathbb{P}(X_{n+1} = y | X_n = y_n) &= \mathbb{P}(X_n + Y_{n+1} = y | X_n = y_n) \\ &= \mathbb{P}(y_n + Y_{n+1} = y).\end{aligned}$$

Soit $(Y_n)_n$ une suite de variables indépendantes identiquement distribuées à valeurs dans F, X_0 une variable à valeur dans E indépendante de cette suite et $f : E \times F \to E$. La suite $(X_n)_n$ définie par la relation $X_{n+1} = f(X_n, Y_n)$ est une chaîne de Markov. En effet,

$\mathbb{P}(X_{n+1} = y | X_0 = y_0, \ldots, X_n = y_n)$

$$\begin{aligned} &= \mathbb{P}(f(y_n, Y_n) = y | X_0 = y_0, \ldots, X_n = y_n) \\ &= \mathbb{P}(f(y_n, Y_n) = y), \end{aligned}$$

$$\begin{aligned} \mathbb{P}(X_{n+1} = y | X_n = y_n) &= \mathbb{P}(f(y_n, Y_n) = y | X_n = y_n) \\ &= \mathbb{P}(f(y_n, Y_n) = y). \end{aligned}$$

Considérons une famille $(Y_{i,j})_{i,j}$ de variables aléatoires indépendantes, $X_0 \in \mathbb{N}$, et pour tout $n \in \mathbb{N}$,

$$X_{n+1} = \sum_{k=1}^{X_n} Y_{n+1,k}.$$

Alors, la suite $(X_n)_n$ est une chaîne de Markov homogène. En effet

$$\mathbb{P}(X_{n+1} = y | X_0 = y_0, \ldots, X_n = y_n) = \mathbb{P}\left(\sum_{k=1}^{y_n} Y_{n+1,k} = y \right).$$

Définition

La matrice de transition d'une chaîne de Markov homogène $(X_n)_n$ dont les états sont $e_1, \dots, e_N$ est la matrice stochastique d'ordre N dont le coefficient en position (i,j) est

$$\mathbb{P}(X_{n+1} = e_j | X_n = e_i).$$

Pour l'exemple du guichet de banque, la matrice de transition est

$$\begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}$$

Proposition

Soit $(X_n)_n$ une chaîne de Markov, d'états $E = \{e_1, \dots, e_N\}$, de matrice de transition A. Notons, pour tout $n \in \mathbb{N}$, la loi de X_n

$$\nu_n = \begin{pmatrix} \mathbb{P}(X_n = e_1), & \mathbb{P}(X_n = e_2), & \dots, & \mathbb{P}(X_n = e_N) \end{pmatrix}.$$

Alors, pour tout $n \in \mathbb{N}$, $\nu_{n+1} = \nu_n A$.

Démonstration

Il suffit d'appliquer la formule des probabilités totales :

$$\mathbb{P}(X_{n+1} = y) = \sum_{k=1}^{N} \mathbb{P}(X_{n+1} = y | X_n = e_k) \mathbb{P}(X_n = e_k).$$

∎

Proposition

Avec les mêmes notations, $\nu_n = \nu_0 A^n$.

Toujours avec la propriété de Markov.

Proposition

Soit $(X_n)_n$ une chaîne de Markov d'états $E = \{e_1, \dots, e_N\}$ de matrice de transition A, $f : E \to \mathbb{R}_+$. Alors

$$\mathbb{E}(f(X_{n+1})|X_0, \dots, X_n) = \sum_{j=1}^{N} [A]_{X_n,j} f(e_j).$$

« Topologie » de la matrice de transition

Définition

Soit $(X_n)_n$ une chaîne de Markov, d'états $E = \{e_1, \dots, e_N\}$.

- Un état $x \in E$ est absorbant si
$$\mathbb{P}(X_1 = x | X_0 = x) = 1.$$
- Une partie $A \subset E$ est fermée si
$$\mathbb{P}(X_1 \in A | X_0 \in A) = 1.$$

Pour le jeu de Penney, deux états sont absorbants (les états victorieux).
Pour l'automate cellulaire bicolore, les configurations monochromes sont absorbantes.

Définition

Soit $(X_n)_n$ une chaîne de Markov, d'états $E = \{e_1, \dots, e_N\}$.

- L'état e_j est accessible depuis l'état e_i s'il existe $p \in \mathbb{N}$ tel que
$$\mathbb{P}(X_p = e_j | X_0 = e_i) > 0.$$
- Les états e_i et e_j communiquent si e_j est accessible depuis e_i et e_i est accessible depuis e_j (relation d'équivalence).
- Une chaîne est irréductible si tous les états communiquent deux à deux (une seule classe d'équivalence pour la relation précédente).

La marche aléatoire réfléchie et le problème du guichet définissent des chaînes de Markov irréductibles.

Traduisons sur la matrice de transition ces définitions.

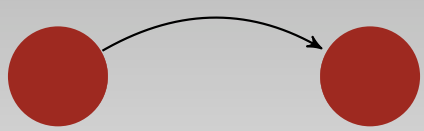
Proposition

Soit $(X_n)_n$ une chaîne de Markov, d'états $E = \{e_1, \dots, e_N\}$ et de matrice de transition A.
L'état e_j est accessible depuis e_i s'il existe $p \in \mathbb{N}$ tel que $[A^p]_{i,j} > 0$.

Démonstration

Il suffit de remarquer $[A^p]_{i,j} = P(X_p = e_j | X_0 = e_i)$. ∎

Attention, deux classes peuvent être reliée par un arc.



Pour le protocole de transmission Aloha, si on suppose que la loi μ d'émission de nouveaux messages à chaque instant vérifie

$$0 < \mu_0 < 1, \qquad 0 < \mu_1 < 1,$$

alors la chaîne associée est irréductible. En effet, la matrice de transition vérifie

$$[A]_{i,j} = \begin{cases} ip(1-p)^{i-1}\mu_0 & \text{si } j = i-1 \\ (1 - ip(1-p)^{i-1})\mu_0 + (1-p)^i\mu_1 & \text{si } j = i \\ (1-(1-p)^i)\mu_1 & \text{si } j = i+1 \\ \mu_{j-i} & \text{si } j \geq i+2 \end{cases}$$

Analyse à un pas

Idée générale

Pour calculer la probabilité d'absorbtion ou le temps moyen avant absorbtion, il suffit de se ramener à la résolution d'un système linéaire.

Analyse à un pas

Idée générale

Pour calculer la probabilité d'absorbtion ou le temps moyen avant absorbtion, il suffit de se ramener à la résolution d'un système linéaire.

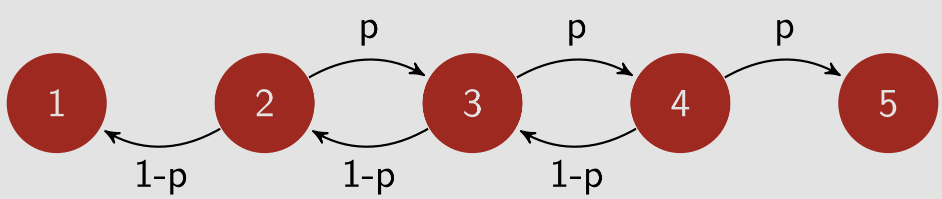
Considérons la fin d'un jeu au tennis entre deux joueurs (à partir de la situation `40A`).

Analyse à un pas

Idée générale

Pour calculer la probabilité d'absorbtion ou le temps moyen avant absorbtion, il suffit de se ramener à la résolution d'un système linéaire.

Considérons la fin d'un jeu au tennis entre deux joueurs (à partir de la situation `40A`). Le score suit alors la chaîne de Markov suivante



Notons p_i la probabilité que la partie se termine en l'état 1 depuis l'état i. Alors $p_1 = 1$, $p_5 = 0$ et pour $i \in [\![2, 4]\!]$,

$$p_i = pp_{i+1} + (1-p)p_{i-1}.$$

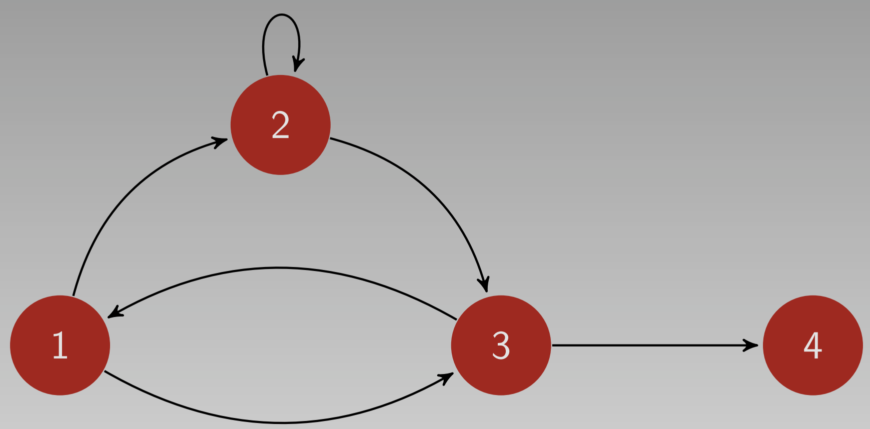
Notons p_i la probabilité que la partie se termine en l'état 1 depuis l'état i. Alors $p_1 = 1$, $p_5 = 0$ et pour $i \in [\![2, 4]\!]$,

$$p_i = pp_{i+1} + (1-p)p_{i-1}.$$

On résout le système et on trouve

$$\begin{aligned} p_2 &= \frac{(1-p)(1-p(1-p))}{1-2p(1-p)}, \\ p_3 &= \frac{(1-p)^2}{1-2p(1-p)}, \\ p_4 &= \frac{(1-p)^3}{1-2p(1-p)}. \end{aligned}$$

Reprenons maintenant le jeu de Penney avec les mots PPF et FPP conditionnellement à ce que les deux premiers lancers ne donnent pas PP.



En notant m_i le temps moyen avant absorbtion (par l'état 4) de la chaîne depuis l'état i. On trouve $m_4 = 0$ et,

$$\begin{aligned} m_1 &= \frac{1}{2}m_2 + \frac{1}{2}m_3 + 1 \\ m_2 &= \frac{1}{2}m_2 + \frac{1}{2}m_3 + 1 \\ m_3 &= \frac{1}{2}m_4 + \frac{1}{2}m_1 + 1 \end{aligned}$$

En notant m_i le temps moyen avant absorbtion (par l'état 4) de la chaîne depuis l'état i. On trouve $m_4 = 0$ et,

$$\begin{aligned} m_1 &= \frac{1}{2}m_2 + \frac{1}{2}m_3 + 1 \\ m_2 &= \frac{1}{2}m_2 + \frac{1}{2}m_3 + 1 \\ m_3 &= \frac{1}{2}m_4 + \frac{1}{2}m_1 + 1 \end{aligned}$$

Après résolution, on obtient

$$m_1 = 7, \ m_2 = 7, \ m_3 = 5.$$

Analyse asymptotique

Définition

Soit $(X_n)_n$ de matrice d'états E. Un état $y \in E$ est

- récurrent positif si le temps de retour en y partant de y est fini p.s. et d'espérance finie,
- récurrent nul si le temps de retour en y partant de y est fini p.s. mais que son espérance est infinie,
- transient sinon.

Proposition

Deux états qui communiquent sont de même nature.

Remarque

Les chaînes de Markov irréductibles finies sont toujours récurrentes positives.

Définition

Soit $(X_n)_n$ une chaîne de Markov. Une loi stationnaire μ est stationnaire si, lorsque $X_0 \sim \mu$, on a $X_1 \sim \mu$. Dans le cas fini, en identifiant μ à son vecteur ligne et en notant A la matrice de transition, cela équivaut à $\mu = \mu A$.

Remarque

Une loi stationnaire est donc un vecteur propre à gauche de la matrice de transition associé à 1, à coordonnées positives et de somme égale à 1.

Proposition

Une chaîne de Markov irréductible est récurrente positive si et seulement si elle admet une loi stationnaire.

Proposition

Soit A une matrice stochastique à coefficients strictement positifs. Alors il existe μ un vecteur ligne positif à coordonnées positives et de somme égale à 1 tel que $(A^n)_n$ converge vers M la matrice dont toutes les lignes sont μ.

Démonstration

Notons $\delta(y)$ l'amplitude d'un vecteur y, c'est-à-dire la différence entre sa plus grande coordonnée et sa plus petite et $d > 0$ la plus petite coordonnée de A. Alors, on vérifie que, pour tout vecteur y, $\delta(Ay) \leq (1-2d)\delta(y)$ et donc $(\delta(A^n y))_n$ est de limite nulle. Ceci entraîne que chaque colonne de $(A^n)_n$ converge vers une colonne constante à coordonnées positives : ceci définit le vecteur μ. Il est évident que μ est de somme 1 en considérant le vecteur y dont toutes les coordonnées sont égales à 1. ∎

Proposition

Avec les mêmes notations, $\mu A = \mu$ et s'il existe un vecteur v tel que $vA = v$, alors v est colinéaire à μ.

Démonstration

Le premier point résulte du passage à la limite dans la relation $A^{n+1} = A^n A$.
Pour le second remarquons que $v = vA$ entraîne $v = vM = a\mu$ en notant a le somme des coordonnées de v. ∎

Proposition

Avec les mêmes notations, pour toute loi ν, $(\nu A^n)_n$ converge vers μ.
Ainsi, pour une chaîne de Markov associée à A,

$$\lim \mathbb{P}(X_n = e_j) = \mu_j.$$

Les preuves suivantes se généralisent sans peine au cas où la matrice A admet un entier $r \in \mathbb{N}$ tel que A^r est à coefficients strictement positifs. Pour passer au cas général, on utilise le lemme suivant.

Lemme

Soit A stochastique irréductible. Alors la matrice $\widetilde{A} = \frac{1}{2}(I_n + A)$ admet un entier $r \in \mathbb{N}$ tel que $\widetilde{A}^r$ est à coefficients strictement positifs. De plus, A et $\widetilde{A}$ admettent les mêmes vecteurs propres à gauche associé à 1.

Remarque

Ainsi, toute matrice stochastique irréductible admet une loi invariante et tout autre vecteur propre à gauche associé à 1 est proportionnel à cette loi. Malheureusement, on n'a plus la convergence de $(A^n)_n$ comme on peut le voir avec

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Démonstration

- Soit $x \in \mathrm{Ker}\ (I_n - A + M)$. Alors $0 = \mu(I_n - A + M)x = (\mu - \mu + \mu)x = \mu x$. Mais alors, $Mx = 0$ et $x = Ax$ ce qui entraîne x constant d'après un calcul déjà effectué. La condition $\mu x = 0$ implique en conclusion $x = 0$. D'où $I_n - A + M$ est inversible.

∎

Démonstration

- Soit $x \in \mathrm{Ker}\ (I_n - A + M)$. Alors $0 = \mu(I_n - A + M)x = (\mu - \mu + \mu)x = \mu x$. Mais alors, $Mx = 0$ et $x = Ax$ ce qui entraîne x constant d'après un calcul déjà effectué. La condition $\mu x = 0$ implique en conclusion $x = 0$. D'où $I_n - A + M$ est inversible.
- Remarquons que $(I_n + A + \ldots + A^{n-1})(I_n - A + M) = I - A^n + nM$ et $M(I_n - A + M) = M$. Par conséquent,

$$\frac{1}{n}(I_n + A + \ldots + A^{n-1}) = \frac{1}{n}(I - A^n) + M \to M.$$

∎

Proposition

Soit $(X_n)_n$ une chaîne finie de loi invariante μ. Alors

$$\lim \frac{1}{n}\sum_{k=0}^{n-1}\mathbb{P}(X_k = e_j) = \mu_j.$$

Définitions Matrice de transition Analyse à un pas **Analyse asymptotique**

Définition

Soit $(X_n)_n$ une chaîne de Markov d'états $E = \{e_1, \dots, e_N\}$ et de matrice de transition A. La période de l'état e_i est

$$\mathrm{pgcd}\{k \geq 1,\ [A]_{i,i} \neq 0\}$$

avec la convention ∞ si l'ensemble est vide.

Proposition

Deux états qui communiquent ont la même période.

Par conséquent, on peut parler de la période d'une chaîne de Markov irréductible.

Roger Mansuy Des chaînes de Markov en prépa?

Démonstration

Soit A la matrice de transition, e_i et e_j deux états qui communiquent : il existe des entiers p et q tels que $[A^p]_{i,j} > 0$ et $[A^q]_{j,i} > 0$.
Alors, pour tout entier r tel que $[A^r]_{i,i} > 0$ et tout $n \in \mathbb{N}$,

$$[A^{p+q+rn}]_{i,i} \geq [A^p]_{i,j}[A^r]^n_{i,i}[A^q]_{j,i} > 0.$$

Ainsi, la période de e_i divise tous les entiers r tel que $[A^r]_{i,i} > 0$ donc la période de e_j.
Par symétrie, les deux périodes sont égales. ∎

Définitions Matrice de transition Analyse à un pas Analyse asymptotique

- La marche aléatoire réfléchie est de période 2 (en fait, on remarque que le graphe est bipartite).
- La chaîne associée au guichet de banque est de période 2 si $S - N = N - I[2] = 1[2]$, 1 sinon.

Définition

Une chaîne apériodique est une chaîne irréductible de période 1.

Roger Mansuy Des chaînes de Markov en prépa?

Proposition

Soit $(X_n)_n$ une chaîne de Markov apériodique d'états $E = \{e_1, \dots, e_N\}$ et de matrice de transition A. Alors, pour tout e_i, il existe un entier N tel que, pour tout $n \geq N$,

$$[A^n]_{i,i} > 0.$$

Proposition

Soit $(X_n)_n$ une chaîne de Markov irréductible d'états $E = \{e_1, \dots, e_N\}$ et de matrice de transition A. Il y a équivalence entre

- A est apériodique,
- il existe un entier N tel que, pour tout $n \geq N$ et tout (i, j)

$$[A^n]_{i,j} > 0.$$

Par conséquent, $(A^n)_n$ converge vers la matrice M déjà étudiée.

Définitions Matrice de transition Analyse à un pas **Analyse asymptotique**

Références

- Paolo Baldi, Laurent Mazliak, Pierre Priouret, Martingales et chaînes de Markov, Hermann, 2001.
- Pierre Brémaud, *Markov chains*, Springer, TAM 31, 1999.
- Arthur Engel, Processus Aléatoires pour les débutants, Cassini, 2011.
- Jean-François Le Gall, Intégration, Probabilités et Processus Aléatoires, polycopié FIMFA, 2006.
- E. Seneta, *Non-negative Matrices and Markov Chains*, Springer, 1980.
- Bernard Ycart, Modèles et algorithmes markoviens, Springer-SMAI, 2002.

Roger Mansuy Des chaînes de Markov en prépa?

Laurent Decreusefond

Professeur à Télécom ParisTech

L'enseignement des probabilités à Télécom ParisTech

Nous analysons les enjeux et les modes de fonctionnement des enseignements de probabilité au sein de Telecom ParisTech. Il s'agit moins de discourir sur l'organisation d'un cours (historique, auditoire, filières, équilibre amphis-TD, volume horaire ...) que de présenter les choix pédagogiques de contenu et de répondre aux questions suivantes :

- quels sont les pré-requis indispensables en analyse (intégration, mesure, Fourier,....) ?
- quels sont les choix pédagogiques retenus (approches, méthodes, impasses, raisons des choix, spécificité de l'Ecole) ?
- comment préparer les étudiants aux notions plus avancées (statistique, Markov, martingales, apprentissage....) ?
- quels sont les thèmes incontournables en raison de leurs nombreuses applications (en physique, mécanique quantique, communications, signal, finance....) ?
- quelles sont les évolutions prévisibles des cours (réformes, évolution de la recherche) ?

Une réflexion sera menée en rapport avec le programme actuel des CPGE toutes filières (MP, PSI, PC) et des bouts de calcul ou de démonstrations seront présentés au tableau à titre l'illustration.

L'enseignement des probabilités à Telecom Paristech

L. Decreusefond

TPT

1 Enjeux

2 Difficultés

3 Modèle de Cox-Ross-Rubinstein
- Modèle à 1 période
- Modèle à plusieurs périodes
- Espace de Rademacher et calcul de Malliavin

Le contexte

- Pas de résistance des matériaux, pas de mécanique, pas de thermodynamique, etc
- Mais, du traitement du signal et de l'image
- des communications numériques
- des réseaux et de l'informatique

Le cours de probabilités

- En tronc commun,
- 40 heures de cours/TD
- en petites classes

Compétences

- Formaliser mathématiquement un phénomène aléatoire
- Résoudre un problème formalisé en termes relevant de la théorie des probabilités

Document du groupe CTI

Ce qui n'est pas difficile (pédagogiquement)

- Dénombrements (ce n'est pas le cœur de l'enseignement)
- Les calculs de loi image par difféomorphisme
- Théorèmes abstraits admis (existence de la mesure de Lebesgue, classe monotone, ...) : existence d'un espace de probas qui rende compte d'une suite infinie de pile/face.

Hétérogénéité

- Des origines différentes : environ 150 étudiants,

  50% MP, 16% PC, 16% PSI, 2% TSI, 16% AST-L
- Des attendus différents : 15 % vont se lancer dans le parcours « Finances »

Difficultés

Qu'est-ce qui se caractérise par ...

- une fonction positive d'intégrale 1
- une fonction croissante, nulle en $-\infty$, qui vaut 1 en $+\infty$
- une fonction holomorphe sur le disque unité
- suites de nombres positifs de somme 1
- une transformée de Fourier ?

Une mesure

Indépendance et conditionnement

- La modélisation = variables aléatoires
- Les calculs = lois
- Calculs uni-dimensionnel : application des outils d'analyse (séries, intégrales)
- Somme de 2 v.a. indépendantes $\Longleftrightarrow$ convolution des 2 lois
- Dès que plus de 2 v.a. : représentation de la dépendance

Les probas en CPGE : une bonne idée ?

- Dans la continuité des enseignements de Terminale
- Un rafraîchissement des programmes
- Un autre mode de raisonnement
- Des mises en perspective du cours de maths : théorie des ensembles, séries génératrices, matrices
- Une réelle ouverture à la modélisation

Se limiter aux probas sur un espace dénombrable

Modèle binomial

Actif sûr

$$
\begin{tikzcd}
X_0 \arrow[r] & X_0(1+r)
\end{tikzcd}
$$

Actif risqué

$$
\begin{tikzcd}
 & S^u = S_0 u \\
S_0 \arrow[ur, "p"] \arrow[dr, "1-p"'] & \\
 & S^d = S_0 d
\end{tikzcd}
$$

Valeur du contrat :

$$
\begin{cases} V^u & \text{si } S_1 = S^u \\ V^d & \text{si } S_1 = S^d. \end{cases}
$$

Call européen

- $V^u = (S_0 u - K)^+$ et $V^d = (S_0 d - K)^+$.
- C'est la valeur d'une option d'achat (*call*) sur l'actif.

  Une option d'achat de prix d'exercice K est le droit d'acheter l'actif sous-jacent au prix K à l'instant T.

  - Si le prix réel $S_0 u$ est supérieur à K, le détenteur du call exerce son droit d'achat et revend aussitôt. Il gagne donc $S_0 u - K$.
  - Sinon, le détenteur du call ne fait rien et donc ne gagne, ni ne perd.

Question

3 paramètres

- S_0 prix initial
- K prix d'exercice
- T maturité

2 questions

- Prix du contrat
- Stratégie de couverture

Principe

- X_0 : prix de vente du contrat
- α_0 : nombre de parts de l'actif S achetées à l'instant 0.

Fortune finale

$$X_1 = \alpha_0 S_1 + (1+r)(X_0 - \alpha_0 S_0).$$

On veut que

$$X_1 = V_1.$$

Calculs

$S_1 = S_0 u$

$$V^u = \alpha_0 S_0 u + (1+r)(X_0 - \alpha_0 S_0).$$

$S_1 = S_0 d$

$$V^d = \alpha_0 S_0 d + (1+r)(X_0 - \alpha_0 S_0).$$

$$\iff \begin{cases} S_0(u-1-r)\alpha_0 & + (1+r)X_0 & = V^u \\ S_0(d-1-r)\alpha_0 & + (1+r)X_0 & = V^d \end{cases}$$

Résultats

$$\begin{aligned} \alpha_0 &= \frac{V^u - V^d}{S_0(u-d)} \\ &= \frac{V^u - V^d}{S^u - S^d}. \end{aligned}$$

$$\begin{aligned} X_0 &= \frac{1}{1+r}\left(\frac{1+r-d}{u-d}\,V^u + \frac{u-1-r}{u-d}\,V^d\right) \\ &= \frac{1}{1+r}\,\mathbf{E}_p\left[V_1\right]. \end{aligned}$$

Commentaires

- Le prix et la stratégie de couverture ne dépendent pas de la probabilité *a priori*.
- Seuls deux paramètres comptent : u et d.
- Risque uniquement lié au modèle.
- Comment estimer u et d ?
- On ne peut pas envisager un modèle où l'actif peut prendre plus de 2 valeurs.

Modèle à 3 périodes

- $S_0 = 4$; $r = 0,4$
- $u = 2$, $d = 1/2$ donc $p = 1/2$.
- Contrat = call européen maturité 3, prix d'exercice 6.

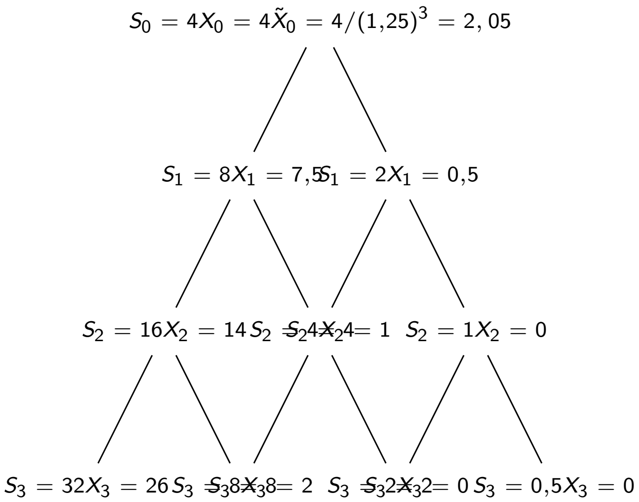


FIGURE: Evolution à 3 pas

Formules

Prix

$$\text{Prix} = \frac{1}{(1+r)^N} \mathbf{E}_p \left[V_N \right].$$

Stratégie de couverture

$$\alpha_k = \frac{\mathbf{E}_{1/2} \left[D_k V_N \,|\, \mathcal{F}_k \right]}{D_k S_k}.$$

Méthode

- Remonter l'arbre : trop long !
- Être malin !
- On suppose $r = 0$ (juste pour simplifier la présentation)

Objectif

- Il *suffit* de trouver $\gamma_0, \cdots, \gamma_N$ tels que

$$V_N = \mathbf{E}_{1/2}\left[V_N\right] X_0 + \sum_{j=0}^{N-1} \gamma_j (S_{j+1} - S_j).$$

Espace de Rademacher

- Aléa contenu dans $\Omega = \{-1, 1\}^N$
- $\{-1, 1\}^{\mathbf{N}}$=espace de rademacher
- $U_j(\omega) = \omega_j$
- Sous $\mathbf{P}_p$, les $(U_j, j \geq 1)$ sont iid de loi

$$\mathbf{P}_p(U_j = 1) = p = 1 - \mathbf{P}_p(U_j = -1)$$

Prix

- Si $U_j = 1$ alors $S_j = S_{j-1}u$
- Si $U_j = -1$ alors $S_J = S_{j-1}d$
- par conséquent

$$S_k = \left(\frac{u+d}{2}\right)^k \prod_{j=1}^{k}(1 + \frac{u-d}{u+d}\ U_j)$$

Objectif

- Il *suffit* de trouver $\gamma_0, \cdots, \gamma_N$ tels que

$$V_N = \mathbf{E}_{1/2}\left[V_N\right] U_0 + \sum_{j=0}^{N-1} \gamma_j (S_{j+1} - S_j).$$

Opérateur de différence

$$D_k F(U_1, \cdots, U_N) = \frac{1}{2}(F(U_k^+) - F(U_k^-)) = \mathbf{E}_{1/2}\left[U_k F \mid U_l,\ l \neq k\right]$$

où

$$\begin{aligned} U_k^+ &= (U_1, \cdots, U_{k-1}, \quad 1,\ U_{k+1}, \cdots) \\ U_k^- &= (U_1, \cdots, U_{k-1}, -1,\ U_{k+1}, \cdots) \end{aligned}$$

Exemple

$$D_k(1 + p_k U_k) = \frac{1}{2}\left(1 + p_k - (1 - p_k)\right) = p_k$$

Formule de Clark

Théorème

$$F = \mathbf{E}_{1/2}\left[F\right] + \sum_{k=1}^{N} \beta_k U_k.$$

$$\beta_k = \mathbf{E}_{1/2}\left[D_k F \,\middle|\, U_1, \cdots, U_{k-1}\right]$$

Définition

Les fonctions cylindriques sont les fonctions de la forme

$$F = \prod_{j=1}^{N} (1 + \rho_j U_j),$$

ρ_j déterministe.

L'espace vectoriel engendré par les fonctions cylindriques est dense dans L^2.

Preuve

- $\mathcal{F}_N = \sigma(U_1, \cdots, U_N)$, $F \in L^2 = L^2 - \lim_{N\to\infty} \mathbf{E}_{1/2}[F \mid \mathcal{F}_N]$
- $\mathbf{E}_{1/2}[F \mid \mathcal{F}_N] = F_N(U_1, \cdots, U_N)$
- $N = 1$

$$\begin{aligned} 0 &= \mathbf{E}_{1/2}[F_1(U_1)(1+\rho_1 U_1)] \ \forall \rho_1 \\ &= \frac{1}{2}[\rho_1 \ (F_1(1) - F_1(-1)) + F_1(1) + F_1(-1)] \end{aligned}$$

- Donc

$$F_1(1) - F_1(-1) = 0 \text{ et } F_1(1) + F_1(-1) = 0$$

donc $F_1 \equiv 0$

N quelconque

- $F_N^{\pm}(U_1, \cdots, U_{N-1}) = F_N(U_1, \cdots, U_{N-1}, \pm 1)$
- En notant $Z_N = \prod_{j=1}^N (1+\rho_j U_j)$, pour tout ρ_N on a

$$\begin{aligned} 0 = \mathbf{E}_{1/2}\left[F_N \prod_{j=1}^N (1+\rho_j U_j)\right] &= \rho_N \mathbf{E}_{1/2}\left[(F_N^+ - F_N^-) Z_{N-1}\right] \\ &\quad + \mathbf{E}_{1/2}\left[(F_N^+ + F_N^-) Z_{N-1}\right] \end{aligned}$$

- On conclut par récurrence

Itô-Clark

Pour $F \in L^2$, on a

$$F = \mathbf{E}_{1/2}\left[F\right] + \sum_k \mathbf{E}_{1/2}\left[D_k F \mid U_1, \cdots, U_{k-1}\right] U_k. \tag{1}$$

Preuve

- Par récurrence $Z_k = \prod_{j\le k}(1+\rho_j U_j)$. $Z_k = 1 + \sum_{j=1}^k Z_{j-1}\rho_j U_j$
- Vrai pour $k = 0$ et $k = 1$.
- Si c'est vrai au rang k, on a

$$\begin{aligned} Z_{k+1} = Z_k(1+\rho_{k+1}U_{k+1}) = (1 + \sum_{j=1}^k Z_{j-1}\rho_j U_j) + Z_k\rho_{k+1}U_{k+1} \\ = 1 + \sum_{j=1}^{k+1} Z_{j-1}\rho_j U_j. \end{aligned}$$

- Donc c'est vrai pour $F \in \mathfrak{E}$.
- Il suffit de montrer que l'application

$$\begin{aligned}\Theta \,:\, \mathfrak{E} \subset L^2 &\longrightarrow L^2(\Omega;\ l^2(\mathbf{N}))\\ F &\longmapsto (k \mapsto \mathbf{E}_{1/2}\left[D_k F \,|\, \mathcal{F}_{k-1}\right])\end{aligned}$$

est prolongeable par continuité sur L^2.

- Or pour $F \in \mathfrak{E}$,

$$F = \mathbf{E}_{1/2}\left[F\right] + \sum_k \mathbf{E}_{1/2}\left[D_k F \,|\, \mathcal{F}_{k-1}\right] U_k$$

donc

$$\begin{aligned}\mathbf{E}_{1/2}\left[(F - \mathbf{E}_{1/2}\left[F\right])^2\right] &= \mathbf{E}_{1/2}\left[\sum_k \mathbf{E}_{1/2}\left[D_k F \,|\, \mathcal{F}_{k-1}\right]^2\right]\\ &= \|\Theta F\|^2_{L^2(\Omega;\, l^2(\mathbf{N}))}.\end{aligned}$$

Conséquence

$$\begin{aligned}S_j &= S_{j-1}(1 + \frac{u-d}{u+d} U_j)\\ S_j - S_{j-1} &= S_{j-1}\frac{u-d}{u+d} U_j\\ D_j S_j &= \frac{1}{2} S_{j-1}\left[1 + \frac{u-d}{u+d} - (1 - \frac{u-d}{u+d})\right]\\ &= S_{j-1}\frac{u-d}{u+d}\end{aligned}$$

$$S_j - S_{j-1} = D_j S_j \, U_j$$

Application à V_N

$$V_N = U_0 + \sum_{j=1}^{N} \mathbf{E}_{1/2}\left[D_j V_N \,|\, \mathcal{F}_j\right] U_j$$
$$= U_0 + \sum_{j=1}^{N} \frac{\mathbf{E}_{1/2}\left[D_j V_N \,|\, \mathcal{F}_j\right]}{D_j S_j} \quad (S_j - S_{j-1})$$

Références

- On trouvera l'étude complète de l'espace de Rademacher à travers les chaos dans

N. Privault, *Stochastic analysis in discrete and continuous settings with normal martingales*, volume 1982 of *Lecture Notes in Mathematics*, Springer, 2009.

Compétences minimales en probabilités pour des ingénieurs en fin de formation

Conventions :

En adoptant une taxonomie simplifiée, on distingue ci-après quatre niveaux de savoirs :

- Initiation = I
- Savoir-faire = SF
- Maîtrise = M
- Appropriation, développement = A

Au niveau Initiation, l'élève sait résoudre une question par la méthode imposée explicitement ou implicitement par l'énoncé.

Au niveau Savoir-Faire, l'élève sait résoudre une question en choisissant la méthode la plus adaptée.

Au niveau Maîtrise, les questions d'évaluation sont plus ouvertes et les réponses nécessitent l'emploi de plusieurs étapes élémentaires.

Si le niveau A relève plus d'une formation doctorale, nous avons indiqué son existence pour montrer que le niveau M n'est pas l'ultime degré de connaissance mais bien un niveau de maîtrise.

Pour chaque ensemble de connaissances, nous avons indiqué un niveau minimal qui nous semble devoir être celui que doivent atteindre tous les élèves ingénieurs et un niveau idéal pour ceux dont la formation nécessite de plus solides bases en probabilités. Aux quatre niveaux précédents, nous avons donc ajouté le stade NV, pour non-vu.
Lorsqu'une seule indication est donnée dans la colonne afférente, c'est que le niveau d'exigence est le même pour tout le monde.

Outils méthodologiques généraux :

- On demande que l'élève ait acquis avant ces enseignements le niveau N2 en calcul de séries numériques et d'intégrales unidimensionnelles.

Compétence n°1 : formaliser mathématiquement un phénomène aléatoire

Dans tous les domaines d'activité d'un ingénieur, les probabilités constituent un outil essentiel de la représentation formelle du réel. Elles permettent de formaliser les parties inconnues ou mal connues d'un phénomène, quelle que soit sa nature. Par ailleurs, les probabilités ne sauraient être disjointes des statistiques (voir fiche correspondante), qui constituent le moyen par excellence d'analyser quantitativement le monde réel.

Dans le développement la compétence n°1, on s'attachera à développer la capacité à passer d'une situation concrète énoncée en langage naturel à une description formelle développée dans le langage des probabilités.

Enfin, il va de soi que cette aptitude à la modélisation n'est en rien spécifique aux probabilités. Cet aspect doit être développé tout le long de la formation d'ingénieur et dans tous les autres enseignements qui s'y prêtent.

<table>
<tr><th>Compétences</th><th>Commentaires/limites</th><th>Connaissances</th><th>D°. Ex.</th></tr>
<tr><td>Construire le cadre mathématique pour la description d'un phénomène aléatoire en termes d'événements et de variables aléatoires.</td><td>Cadre formel des probabilités. On introduira les concepts dans le cadre des espaces d'événements discrets. On insistera ensuite sur l'apport de la formalisation abstraite pour les situations « mixtes » (discret-continu).</td><td>Espace probabilisé,
variables aléatoires, probabilité, lois.

Être capable d'écrire formellement la question à résoudre.</td><td>SF</td></tr>
<tr><td rowspan="3">Savoir choisir la structure de dépendance pertinente associée à un phénomène aléatoire.</td><td>Première approche de la représentation de la dépendance.
Pas de processus aléatoire.</td><td>Indépendance, probabilité conditionnelle,
coefficient de corrélation.</td><td>SF</td></tr>
<tr><td>Deuxième approche de la représentation de la dépendance.</td><td>Loi conditionnelle, espérance conditionnelle</td><td>NV à SF</td></tr>
<tr><td>Introduction de la notion de temps dans la modélisation.
Augmentation de l'espace d'état pour garantir le caractère markovien du modèle.</td><td>Chaînes de Markov</td><td>NV à SF</td></tr>
</table>

Compétence n°2 : résoudre un problème formalisé en termes relevant de la théorie des probabilités

Une fois le problème réel formalisé, il reste à le résoudre, c'est-à-dire lui trouver une solution qui permette des évaluations quantitatives. Il importe de remarquer que la connaissance des outils théoriques conditionne la richesse des modèles que le futur ingénieur sera susceptible d'invoquer au cours de sa carrière. On ne saurait donc faire l'économie d'une solide formation théorique.

Compétences	**Commentaires/limites**	**Connaissances**	**D° Ex.**
Calculer Créé le 09/04/08 09:56des probabilités d'événements élémentaires dans le cadre des probabilités discrètes	Puisque le but est de se familiariser avec les méthodes et les lois usuelles, cette partie ne sera pas plus que les autres, prétexte à des exercices savants.	Méthodes usuelles de calculs de probabilité d'événements -Dénombrements usuels : combinaison, arrangements, etc. -Formule d'inclusion/exclusion -Utilisation de l'indépendance et probabilité conditionnelle -Fonction génératrice	M
		Variables aléatoires discrètes -Loi, loi jointe, loi marginale -Espérance, variance, moments -Indépendance -Fonction génératrice	M
Calculer des probabilités d'événements élémentaires dans le cadre des probabilités sur continuum	Puisque le but est de se familiariser avec les méthodes et les lois usuelles, cette partie ne sera pas plus que les autres, prétexte à des exercices savants.	Variables aléatoires réelles uni-dimensionnelles -Fonction de répartition -Densité -Fonction caractéristique	M
		Variables aléatoires vectorielles - Loi jointe, loi marginale -Indépendance -Loi image	SF à M

Savoir mener à bien une résolution analytique d'un problème formalisé.	Insister sur la notion de loi et les différentes caractérisations d'un tel objet de manière à unifier les présentations des probabilités discrètes et continues.	-Définition générale et propriétés élémentaires d'une mesure -Thm de Kolmogorov -Intégration par rapport à une mesure quelconque : définition et principaux théorèmes. -Lien entre indépendance et mesure produit	NV à SF
Calculs approchés de probabilités	Être capable de calculer une moyenne sans nécessairement connaître les lois sous-jacentes. Le calcul des moments est une méthode qui permet d'avoir simplement des résultats partiels. À comparer aux calculs de lois et à rapprocher des théorèmes limites.	Calculs des moments de variables aléatoires : -Additivité de l'espérance -Additivité de la variance dans le cas indépendant -Utilisation d'outils analytiques -Inégalités de Bienaymé-Tchebychev, de Markov.	SF à M
	Pas de difficultés théoriques soulevées sur ces théorèmes.	-Loi des grands nombres -Théorème de la limite centrée	SF
	On mettra en évidence les difficultés numériques.	Construire et programmer une simulation d'un modèle aléatoire. Analyser la précision des résultats.	I à SF
Pouvoir représenter précisément les dépendances entre variables aléatoires.	Savoir calculer des lois conditionnelles, lois a priori et a posteriori pour les statistiques	-Loi conditionnelle -Espérance conditionnelle -Cas gaussien	NV à SF
	Calculer (analytiquement ou numériquement) la loi stationnaire et les principales caractéristiques d'une chaîne de Markov.	Chaînes de Markov à espaces d'états discrets -Propriétés structurelles des chaînes de Markov -Loi stationnaire Savoir classer les états.	NV à SF

En marge de ce stage LIESSE, une conférence sur l'histoire des sciences en soirée, à l'occasion du centenaire de la disparition d'Henri Poincaré.

Jean-Marc Ginoux

Maître de conférences, IUT de Toulon

Sur les conférences d'Henri Poincaré à l'Ecole Supérieure des Postes et Télégraphes (aujourd'hui Télécom ParisTech)

Le 4 juillet 1902, Henri Poincaré est nommé professeur d'Electricité Théorique à l'Ecole Supérieure des Postes et Télégraphes qui se trouve alors rue de Grenelle. En 1901, le nouveau directeur Edouard Estaunié propose de compléter les cours par des conférences. Henri Poincaré inaugure le cycle et effectue ainsi en mai-juin 1904, 1906, 1908, 1910 et 1912 différents exposés sur les sujets de son choix. En ce début de XXème siècle, la Télégraphie Sans Fil (T.S.F.) en est à ses balbutiements et l'un des moyens employé pour transmettre un signal est le dispositif inventé par William Du Bois Duddell et qu'il a nommé "arc chantant". Basé sur le principe de l'excitateur (ou éclateur) de Hertz il présente l'avantage considérable d'entretenir les oscillations produites au lieu de les amortir ouvrant ainsi de belles perspectives aux procédés de télécommunication.

Les ingénieurs et expérimentateurs sont à cette époque confrontés à deux problèmes : amorcer les oscillations et les entretenir. Alors que le premier a été pratiquement résolu du point de vue théorique comme du point de vue pratique, le second semble résister à l'analyse mathématique. C'est précisément ce sujet qu'a choisi Poincaré pour sa conférence de 1908 dans laquelle il va mettre en application le concept de "cycle limite" (introduit vingt ans auparavant dans un contexte purement mathématique) pour établir l'existence d'un régime stable d'ondes entretenues. Il va ainsi démontrer que lorsque solution de l'équation différentielle caractérisant les oscillations entretenues par un "arc chantant" a la forme d'un "cycle limite", c'est-à-dire, d'une courbe fermée vers laquelle converge asymptotiquement toutes les autres courbes, le système présente un régime stable d'oscillations entretenues.

La découverte de ces conférences "oubliées" de Poincaré bouleverse quelque peu l'historiographie qui considérait jusqu'alors que c'était le mathématicien russe Aleksandr' Andronov qui avait le premier obtenu ce résultat, c'est-à-dire, qui avait réalisé cette correspondance entre science et technique en 1929. Cet exposé a pour but de présenter d'une manière pédagogique et accessible au plus grand nombre le contenu de cette conférence "oubliée" de 1908 ainsi qu'une partie des travaux de Poincaré en T.S.F.

Voir les conférences d'Henri Poincaré sur la T.S.F.

Alain Maruani

Président du département de première année à l'Ecole des Ponts ParisTech, professeur émérite à Télécom ParisTech

Une promenade aléatoire au voisinage des probabilités

Après une brève mention dans les lycées et avant un retour en force dans les grandes Écoles, les probabilités apparaissent de manière subreptice dans ce qui se nommait jadis SUP et SPE : La notion de densité de probabilité est dans le filigrane de ce que l'on dit de la théorie cinétique du gaz parfait, ou des orbitales atomiques et il n'est guère de hors programme plus contourné que l'interprétation statistique de l'entropie.

Il semble que les probabilités frappent en ce moment à la porte des programmes et y demandent quelque légitimité, en déclenchant des débats d'une intensité imméritée entre les gardiens d'une rigueur sans concession et les tenants d'une intuition non écrantée par le formalisme.

C'est dans ce cadre que j'indiquerai quelques rencontres que, en tant que praticien (et sous un timbre ou un autre), j'ai eu l'occasion de faire avec les probabilités ; je me restreindrai à la partie qui me semble pertinente pour une éventuelle présence de cette discipline dans les programmes et, à l'intérieur de ce cadre, de ce qui me semble le plus propice à nourrir le dialogue entre les mathématiques et les sciences physiques.

De ce point de vue, probabilités discrète et continue, méthodes bayesiennes et principe Maxent ne manqueront pas d'être, entre autres, mentionnés.

Une promenade aléatoire
en probabilités

Alain Maruani
École des Ponts ParisTech
Télécom ParisTech
alain.maruani@enpc.fr
alain.maruani@telecom-paristech.fr

Contenu

Quelques rencontres, au fil des années : regards rétrospectifs

Quelques points de vue pour les CPGE : regards prospectifs

Une question associée : Statut de la Mécanique Quantique ?

Au fil des années 1

Première affirmation

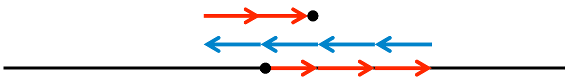
Probabilité = fréquence statistique, à travers des problèmes de dénombrement

Exemple : Probabilité qu'un paquet de quatre cartes pris dans un jeu dont on a retiré les trèfles contienne au moins un as rouge ?

Premiers développements

- Formule asymptotique de Wallis : $n! = n\ln n - n + O(\ln n)$
- Marche aléatoire, durée Δt, amplitude a : *la solution avant l'équation !*

$$P(n, N) \text{ pour } N \longrightarrow \infty \text{ , } n \longrightarrow \infty \text{ et } n/N \ll 1.$$



$$t = N\Delta t$$

$$x = Na \qquad P(x,t) = \frac{1}{\sqrt{\pi D t}} \exp\left(-\frac{x^2}{4Dt}\right) \qquad \left(\partial_t = D\,\partial_{x^2}\right)$$

$$D = \frac{a^2}{t}$$

Au fil des années 2

En CPGE

Facteur de Boltzmann : $FB = \exp\left(-\frac{E}{kT}\right)$

Distribution de Maxwell $DM = \exp\left(-\frac{1}{2}\frac{mv^2}{kT}\right)$; équipartition, viriel.

Entropie et température : $\mathrm{d}S = \frac{\delta Q}{T}$

La température est l'inverse du facteur intégrant de la quantité de chaleur, la forme de Pfaff associée est l'entropie.

Un peu plus tard

L'entropie :

$$S = -k'\ln\Omega = -k'\sum_n p_n \ln p_n$$
$$= -k'\int p(u)\ln\left[p(u)\right]\mathrm{d}u$$

Les statistiques quantiques $f_{FD/BE} = \dfrac{1}{1 \pm \exp\left(\dfrac{E-\mu}{kT}\right)}$

Au fil des années 3

Encore plus tard (pêle-mêle)

La classification
La mesure
Les problèmes mal posés (radar, traitement d'image)
Des lois nouvelles (Weibull, …)
Des techniques nouvelles (Transformée de Mellin, …)
Des concepts nouveaux, ou rajeunis (la pseudo inverse) (réseaux de neurones, recuit simulé, …)

Je n'ai pas cherché les probabilités, elles se sont insérées naturellement dans monpaysage.

Les probabilités sont probablement le seul enseignement commun à toutes les GE.

- ***C'est donc une excellente raison de l'introduire en CPGE.***
- ***C'est donc une excellente raison de ne pas l'introduire en CPGE.***

Un but de cet exposé

- Ne pas se préoccuper du contenu précis d'un éventuel programme (les dés sont-ils jetés ?)

- Proposer la présence d'une idée (sans se préoccuper de la technique qui pourrait l'accompagner) : méthodes bayésiennes.

- Rechercher des adhérences entre probabilités (au sens noble du terme : mesure, tribus boréliennes, …) et autres éléments de cours de CPGE (versions dégradées) :

Physique
Chimie
Sciences industrielles (MP !)
Informatique

(Retour sur) La Mesure 1 : généralités

Signe de confiance dans l'intelligibilité de la Nature.

- Fidélité, justesse et précision.
- Résultats : nombres, courbe, affirmation d'existence (Neptune) ou jugement (théories).

Transducteurs, traitement de données

- Conversion d'une grandeur en une autre.

  Il est souhaitable que le principe de fonctionnement de l'appareil soit différent du principe régissant le système étudié.

- Erreurs, inconsistances et problèmes mal posés sont le lot quotidien de la mesure. Théorie de Gauss.

  Variabilités de l'objet, du système de mesure et de l'observateur, en supposant ces éléments séparables. On cherche un représentant moyen. Dans une classe, la moyenne représente la performance de l'élève prototype. Invocation de la loi des Grands Nombres.

- Inférence et théorie de la confirmation :

  Peut-on quantifier la manière dont la découverte de Neptune a conforté la théorie newtonienne ?
  Que vaut la performance de ce télépathe ?

(Retour sur) La Mesure 2 : questions

Inférence et théorie de la confirmation

- Comment modifier nos conclusions au fur et à mesure que les données s'accumulent ?
- Peut-on quantifier la confiance mise en un modèle ?
- Pourquoi des expériences indépendantes sont-elles censées mieux conforter un modèle que la répétition inlassable de la même expérience ?

Duhem : On ne saurait déduire les lois de Newton de l'observation du mouvement des planètes.
Newton : Deux révolutions du même objet ne sont jamais identiques !

Les degrés de croyance ne sont ni logiques ni objectifs, ils relèvent des lois de probabilité et du pari.

Hypothèse

Le crédit accordé à une hypothèse après l'acquisition de données expérimentales (probabilité ***a posteriori***) est proportionnelle au produit de :

- la vraisemblance de cette hypothèse
- la probabilité d'obtenir ces résultats expérimentaux, en tenant l'hypothèse pour vraie (probabilité ***a priori***)

Principe d'inférence (Bayes) 1

Une probabilité n'est pas le passage à la limite d'une fréquence, mais la traduction numérique d'un état de connaissance.

$$\begin{aligned} P(H \text{ et } e) &= P(H|e) \times P(e) \\ &= P(e|H) \times P(H) \end{aligned}$$

Ces règles résultent de

$$\text{Addition : } P(A \cup B \ |C) = P(A|C) + P(B \ |C) - P(A \cap B|C)$$

$$\text{Multiplication : } P(A \cap B \ ) = P(A|B) \times P(B) = P(B|A) \times P(A)$$

La connaissance des causes possibles d'une observation permet de remonter aux causes.

Exemple : Vrai (V) ou Faux (F) $P(V \ |E) = 0{,}9$ $P(F|\overline{E}) = 0{,}95$

Théorème de Bayes : $P(E \ |V) = \dfrac{P(V \ |E)}{P(V)} \times P(E)$. Probabilité a priori.

Principe d'inférence (Bayes) 2

Modéliser l'attente en début de processus plutôt que de se fixer une méthode puis traiter les données.

- Possibilité de diminuer le poids des aprioris au fur et à mesure de l'acquisition des données.

Quoi utiliser ?

- Informations abondantes ou peu onéreuses : statistiques classiques.
- Autrement : méthodes bayesiennes.

Les résultats sont asymptotiquement identiques.

Le test bayésien utilisé pour déterminer la plausibilité d'une distribution par rapport à des observations est asymptotiquement convergent avec celui des statistiques classiques (grand nombre d'observations).

Principe d'ignorance maximale. *Principe d'entropie maximale.*

Principe d'inférence (Bayes) 3 : exemples

La boîte A contient30 biscuits au chocolat et 10 biscuits ordinaires.
La boîte B contient 20 biscuits au chocolat et 20 biscuits ordinaires.

Le biscuit choisi au hasard dans une boîte au hasard est au chocolat.
De quelle boîte a-t-il le plus de chances d'être issu, et avec quelle probabilité ?

Intuitivement : de la boîte A.

H_X = la proposition « le gâteau vient de la boîte X ». C = « le biscuit est au chocolat ».

$$P(H_A) = P(H_B) = 0{,}5.$$
$$P(C \mid H_A) = 30/(30+10) = 0{,}75$$
$$P(C \mid H_B) = 20/(20+20) = 0{,}5$$

Après le choix, cette probabilité est $P(H_A \mid C) = P(C \mid H_A) \times P(H_A) / P(C) = 0{,}75 \times 0{,}5/(50/80) = 0{,}6$.

Le gain de vraisemblance (1, 76 dB) ne dépend pas de la probabilité a priori des boîtes.

Autre exemple

Des résultats de double diffraction électronique contredisant la théorie relativiste de Dirac n'ont pas entraîné une remise en cause de cette équation : cette dernière se révèle robuste par rapport à une différence entre le résultat obtenu et le résultat attendu.

Information et surprise. Quantification de l'information

La quantité d'information est une fonction

Décroissante de la probabilité p (croissante de la surprise)

Additive

$$I_m = k \log_b(1/p_m)$$

$$I = \Sigma p_m I_m$$

C'est le manque d'information avant toute observation.
I est la seule quantité satisfaisant à des exigences de bon sens (Kinchine : jeu équiprobable, p rationnel, p réel)

L'incertitude devient une grandeur mesurable.

Que dire lorsque l'on ne sait rien ?

$$\textit{Entropie}\ S = -\Sigma p_m \log(p_m)$$

Principe d'ignorance maximale : $p_m = p = 1/W$

S = −*I* est maximale (convexité)

$$I = \Sigma p_m I_m = k\,\Sigma(1/W)\log(1/W)$$

$$S = \mathrm{k.}\log(W)$$

Maxent : loi 1

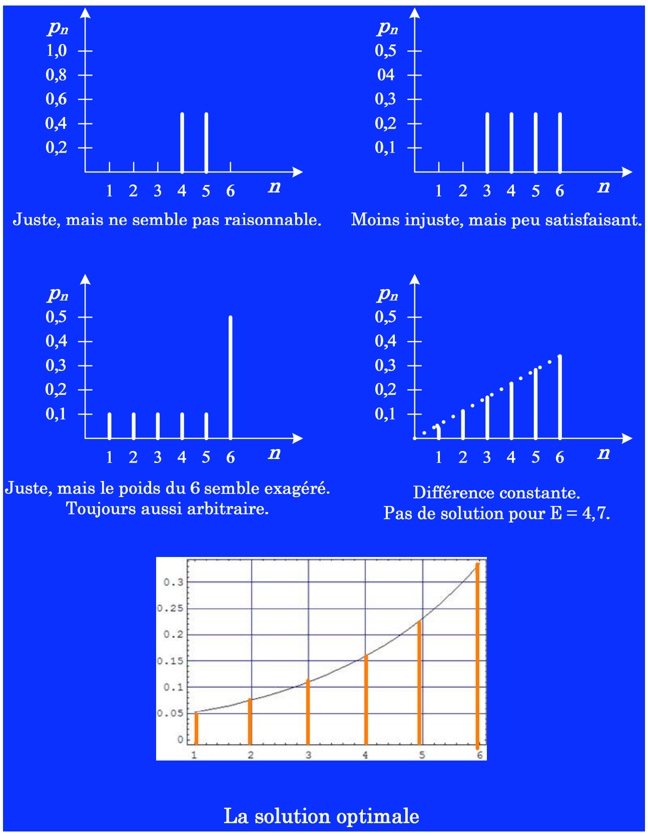
1) Dé (ou dés)

$E = 4{,}5$ ($E_{\text{loyal}} = 3{,}5$)

Maximiser I, sous la contrainte

$E_{\text{triche}} = 4{,}5$

$S_{maxent} = 1{,}61$
$S_{uniforme} = 1{,}79$



Maxent : loi 2 : le facteur de Boltzmann

Énergie individuelle ε, valeur moyenne E *imposée*.

Quelle est la loi de probabilité ?

Entropie maximale pour

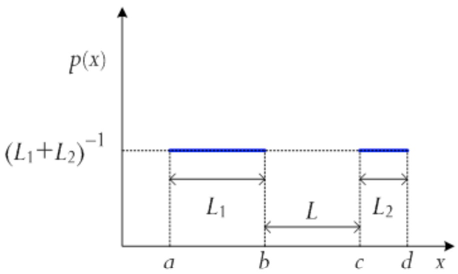
$$E = \sum_n p_n \varepsilon_n$$

$$S = -\sum_n p_n \ln p_n$$

$$\frac{\partial (S - \lambda E)}{\partial p_n} = -\ln p_n - 1 - \lambda \varepsilon_n = 0$$

$$p_n = A \exp(-\lambda \varepsilon_n) \qquad \begin{cases} \sum_n p_n = 1 \\ \sum_n p_n \varepsilon_n = E \end{cases} \Rightarrow A\left(= \frac{1}{Z}\right) \text{ et } \lambda \left(= \frac{1}{kT}\right)$$

Incertitude ou dispersion ?



$$\Delta x = \mathrm{F}(a,b,c,d) = \Phi(L_1, L_2, L)$$

$$S = \int p(x) \ln[p(x)] \mathrm{d}x = \ln(L_1 + L_2)$$

Quelques avatars de l'entropie

Thermodynamique : Clausius 1865

$$\mathrm{d}S = \frac{\delta Q}{T} \qquad S(2) - S(1) = \int_{\text{Réversible}} \frac{\delta Q}{T}$$

Mécanique statistique : Boltzmann 1877

$$S = k \ln W$$

Mécanique quantique : von Neumann 1927

$$S = -k\,\mathrm{Tr}(\rho \ln \rho)$$

Et après : Kullback, Kolmogorov, …

Vers l'information : exorcismes

Démon de Maxwell : 1871

Réfutation Szilard : 1929

Hartley : 1928, information et surprise ($i = -\ln p$)

Shannon : H est la moyenne de l'information sur tous les messages

Cox, *Jaynes*, Kullback : L'entropie mesure ce que l'on ignore d'un système connu uniquement par sa loi de distribution.

Grandeurs déduites

Entropie mutuelle :

$$H(x,y) = -\sum_{i,j} p_{ij} \ln p_{ij}$$

Entropie conditionnelle :

$$H(x,y) = H(x) + H(y|x) = H(y) + H(x|y)$$

Information mutuelle (redondance)

$$\begin{aligned} I(x,y) &= H(x) + H(y) - H(x,y) \\ &= H(x) - H(x|y) \\ &\le H(x) \end{aligned}$$

Un exemple

***X* = le temps qu'il fait** **$\quad$ *Y* = le temps prédit**

Pro		X (Il fait ...)		
		Beau	Pas beau	
Y (Il fera ...)	Beau	5/8	1/16	11/16
	Pas beau	3/16	1/8	5/16
		13/16	3/16	

Pr (erreur)	4/16
H(X)	0,696
H(Y)	0,896

Paresseux		X (Il fait ...)		
		Beau	Pas beau	
Y (Il fera ...)	Beau	13/16	3/16	1
	Pas beau	0	0	0
		13/16	3/16	

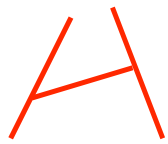
Pr (erreur)	3/16
H(X)	0,696
H(Y)	0

Amateur		X (Il fait ...)		
		Beau	Pas beau	
Y (Il fera ...)	Beau	403/512	93/512	31/32
	Pas beau	13/512	3/512	1/32
		13/16	3/16	

Pr (erreur)	0,207
H(X)	0,696
H(Y)	0,200

H(X|Amat.)	0,695
H(X|Pro.)	0,605

Qu’est ceci ?



- Statistiquement : A ?
- Structurellement : H ?

Principes de décisions

- Espace des formes Ω (tous les objets possibles **a**, *a*, *a*, a, a, …).
- Espace de représentations R (mesures, forme découpée en segments, graphes).
- Espace des classes J $(=\{0,1, \ldots, 9, ?\})$
- Fonction de classement idéale C : $\Omega \rightarrow J$
- *Fonction de classement d :* $R \rightarrow J$

Les méthodes statistiques

- **On connaît :**
  - La loi du vecteur d'état **x,** $p(\mathbf{x})$
  - Le nombre de classes, c, notées ω_i
  - La probabilité *a priori* d'apparition de chaque classe $P(\omega_i)$
  - La probabilité conditionnelle $p(\mathbf{x} \mid \omega_i)$

- **On cherche à :**
  - évaluer les probabilités a posteriori $p(\omega_i \mid \mathbf{x})$

Les grandes familles

- Statistiques
  - Classification bayesienne
    - Cas gaussien
  - Estimation des lois
    - Estimation paramétrique
    - Estimation non paramétrique
- Méthodes fondées sur une partition
  - Cas bayesien
    - Bayesien gaussien
  - Discrimination paramétrique non bayesienne
    - Cas linéaire : perceptron
- Segmentation de l'espace : k-ppv
- Approche neuromimétique

Estimation des lois (cadre bayésien)

- Approche paramétrique : estimer les paramètres d'une loi affirmée.
- Estimation sans modélisation

L'inférence bayésienne

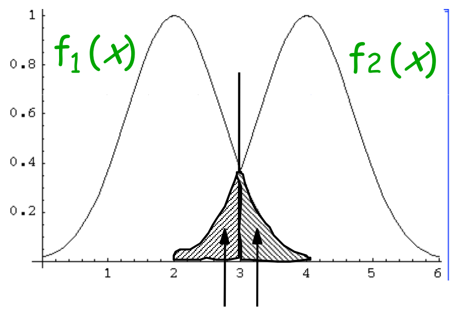
- L'étiquette
- La probabilité d'erreur

$$P(\omega_i|\mathbf{x}) = \frac{P(\omega_i) \times p(\mathbf{x}|\omega_i)}{p(\mathbf{x})}$$

$$R(d) = \int \left[1 - P(d(\mathbf{x}))\right] p(\mathbf{x}) \mathrm{d}\,\mathbf{x}$$

Le but

- Min. erreur moyenne de classification.
  - Place la séparatrice au point d'intersection des gaussiennes
    - dans la zone hachurée de gauche, on attribuerait « 1 » à un objet « 2 ».
- Non paramétrique
  - les formes peuvent ne pas être gaussiennes.
- Minimiser l'aire hachurée



Image

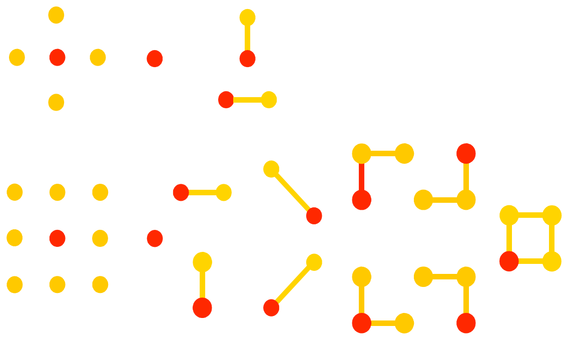
Information véhiculée > donnée des niveaux de gris en chaque site : zones, contours, structures définis par les contrastes, textures …
Le niveau de gris en un site est en relation avec les pixels voisins.

Les interactions locales entre niveaux de gris voisins définissent les différentes régions de l'image.
Elles permettent d'utiliser le formalisme markovien en restauration, segmentation, analyse …

Principe : définir des énergies locales entre groupes de sites reflétant les interactions entre niveaux de gris.
L'énergie globale est reliée à la probabilité d'apparition de l'image

Image

L'image : ensemble fini S de sites s_i correspondant aux pixels.
S $\in \mathbb{Z}^2$ est un réseau discret fini, muni de la notion de voisinage V.
Site $\rightarrow$ descripteur de son état (niveau de gris, …) à valeurs dans E.
Clique : ensemble de sites voisins.



$$U = \sum_{c \in C} U_c \quad U_s = \sum_{c \in C / s \in c} U_c$$

Image : modélisation probabiliste

Site $s \rightarrow$ variable aléatoire Xs, valeurs x_s dans E.
Le champ aléatoire $X = (X_s, X_t, ..)$ est à valeurs dans $\Omega = \mathrm{E}^{|S|}$.

L'image est une réalisation x de X : P($X = x$), est une sorte de vraisemblance de l'image.

- L'hypothèse markovienne permet d'évaluer le lien statistique entre un niveau de gris et le reste de l'image.

Configuration de l'image sans le site s : $x^s = (x_t)_{t \neq s}$

X est un champ de Markov ssi la probabilité conditionnelle locale en un site n'est fonction que de la configuration du voisinage du site considéré

$$P\left(X_s = x_s \middle| x^s\right) = P\left(X_s = x_s \middle| x_t, t \in V_s\right)$$

Mesure de Gibbs

Pour l'énergie $U : \Omega \rightarrow \mathbb{R}$: la probabilité P sur Ω

$$P(X=x)=\frac{1}{Z}\exp\left[-U(x)\right]=\frac{1}{Z}\exp\left[-\underbrace{\sum_{c\in C}U_c(x)}_{\text{Somme d'énergies locales}}\right]$$

Image binaire 512×512 : card(Ω) = 2^{262114}

Champ de Gibbs de potentiel associé au système de voisinage V : c'est le champ aléatoire X dont la probabilité est une mesure de Gibbs associée au système de voisinage V.

Théorème de Hammersley-Clifford

- S dénombrable,
- Système de voisinage V borné,
- Espace des états discret

X est un champ de Markov relativement à V et P($X = x$) $\forall\ x \in \Omega$ ssi X est un champ de Gibbs de potentiel associé à V.

Exemple de la 4-connexité

$$U(x) = \sum_{c=(s)\in C_1} U_c(x_s) + \sum_{c=(s,t)\in C_2} U_c(x_s, x_t)$$

Après quelques calculs (non triviaux), le résultat, remarquablement simple, ne fait intervenir que les potentiels des cliques contenant le site s :

$$\text{Prob. cond. loc.} : P\left(X_s = x_s \middle| X^s = x^s\right) = \frac{\exp\left[-\sum_{c\in C/s\notin c} U_c(x) - U_s(x_s|V_s)\right]}{\sum_{x_s\in E}\exp\left[-\sum_{c\in C/s\notin c} U_c(x) - U_s(x_s|V_s)\right]}$$

$$= \frac{\exp\left[-U_s(x_s|V_s)\right]}{\sum_{x_s\in E}\exp\left[-U_s(x_s|V_s)\right]}$$

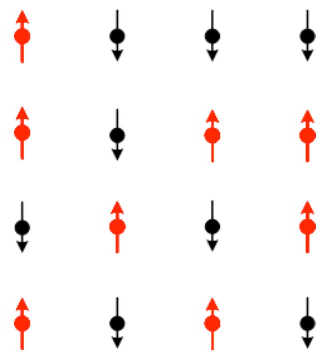
Pas possible d'accéder à la probabilité d'une configuration *x* donnée (à cause de la constante de normalisation), mais possible de calculer en chaque site la probabilité conditionnelle locale.

Application : Modèle d'Ising 2D, par une approche markovienne

- Réseau carré de spins interagissant

$$U = \sum_{(rs)} U_{rs}$$

$$U_{rs} = \begin{cases} -J\sigma_r\sigma_s & (\sigma_i = \pm 1) \quad \text{Ising-Heisenberg} \\ -K\delta_K(x_r, x_s) & (x_i = 0,1) \quad \text{Potts} \quad (K = 2J) \end{cases}$$



- Prototype des champs de Markov aléatoires :

Probabilité spin en $s = x_s$ conditionnellement au réseau, S_s

=

Probabilité spin en $s = x_s$ conditionnellement au voisinage, N_s

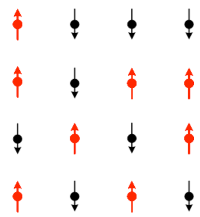
Hammersley-Clifford + Bayes

- Réseau carré de spins interagissant

$$\text{Prob. cond. loc.} : P\left(x_s \middle| S_s\right) = P\left(x_s \middle| N_s\right)$$
$$= \frac{1}{\underbrace{\sum_{\{x_s\}} \exp\left[-\frac{U\left(x_s|N_s\right)}{kT}\right]}_{=Z_{N_s}}} \exp\left[-\frac{U\left(x_s|N_s\right)}{kT}\right]$$

- Règle de Bayes

$$\begin{aligned} p &= P\left(x_s = 1\right) \\ &= \sum_{\{N_s\}} P\left(x_s = 1 \middle| N_s\right) P\left(N_s\right) \\ &= \sum_{\{N_s\}} P\left(N_s\right) \frac{1}{Z_{N_s}} \exp\left[-\frac{U\left(x_s|N_s\right)}{kT}\right] \end{aligned}$$

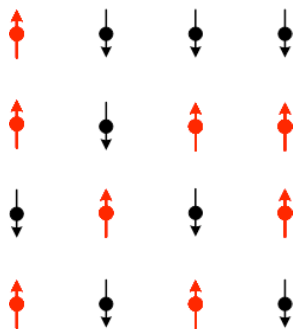


Équation résolvante

- Notation : $\frac{K}{kT} \mapsto \kappa$

- **Les spins de N_s ne sont pas corrélés**
- Probabilité invariante par translation

$$p = \frac{\binom{4}{0}\exp(4\kappa)}{1+\exp(4\kappa)} p^4 + \frac{\binom{4}{1}\exp(3\kappa)}{\exp(\kappa)+\exp(3\kappa)} p^3(1-p)$$
$$+ \frac{\binom{4}{2}\exp(2\kappa)}{\exp(2\kappa)+\exp(2\kappa)} p^2(1-p)^2$$
$$+ \frac{\binom{4}{3}\exp(\kappa)}{\exp(\kappa)+\exp(3\kappa)} p(1-p)^3 + \frac{\binom{4}{4}}{1+\exp(4\kappa)}(1-p)^4$$
$$= \varphi(p)$$



Solution (invariances)

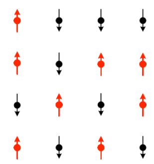
$$p = \frac{\exp(4\kappa)}{1+\exp(4\kappa)}p^4 + \frac{4\exp(2\kappa)}{1+\exp(2\kappa)}p^3(1-p) + 3p^2(1-p)^2 + \frac{4}{1+\exp(2\kappa)}p(1-p)^3 + \frac{1}{1+\exp(4\kappa)}(1-p)^4$$
$$= \varphi(p)$$

$$p = \frac{1}{2} \quad \forall\kappa$$

Centre de symétrie $\left(\frac{1}{2},\frac{1}{2}\right)$

$$\{x_s \leftrightarrow 1-x_s\} \quad \{p \leftrightarrow 1-p\} \quad \{U_{rs} \leftrightarrow U_{sr}\}$$

$$[a] \quad p = \varphi(p) \Leftrightarrow [b] \quad 1-p = \varphi(1-p)$$



$$[a]+[b] \Rightarrow \varphi(p)+\varphi(1-p) = 1 : d°\varphi \text{ impair } (=3) \ !$$
$$[a]-[b] \Rightarrow \mu = 2p-1 = \varphi(p)+\varphi(1-p)$$

$$\Rightarrow \begin{cases} \left[p^2+(1-p)^2\right]\tanh(2\kappa)+4p(1-p)\tanh\kappa = 1 \\ p = \dfrac{1}{2} \end{cases}$$

Analyses (équation de degré 2 !)

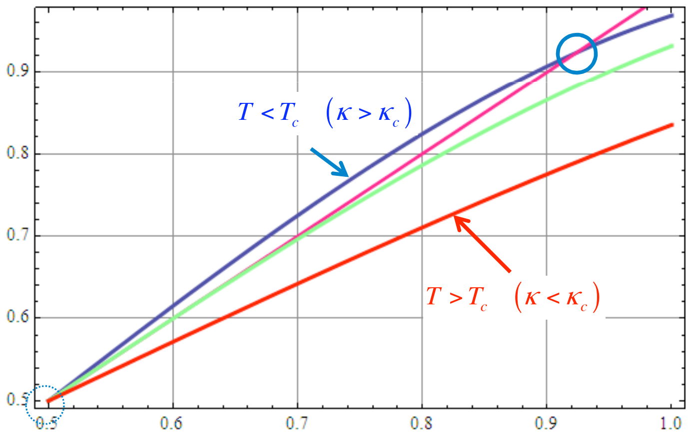
$$\left[p^2+(1-p)^2\right]\tanh(2\kappa)+4p(1-p)\tanh\kappa = 1$$

- Constante critique : ½ racine double.

$$\frac{1}{2}\tanh(2\kappa)+\tanh\kappa = 1$$

$$\kappa_c = 0{,}653$$
$$(\kappa_c)_{\text{Onsager}} = \ln\left(1+\sqrt{2}\right) \approx 0{,}882$$
$$(\kappa_c)_{\text{CM}} = 0{,}5$$

$$\frac{T_c}{J} = 3{,}063$$
$$\left(\frac{T_c}{J}\right)_{\text{Onsager}} = 2{,}269$$
$$\left(\frac{T_c}{J}\right)_{CM} = 4{,}0$$



Magnétisation spontanée

Constante critique : ½ racine double.

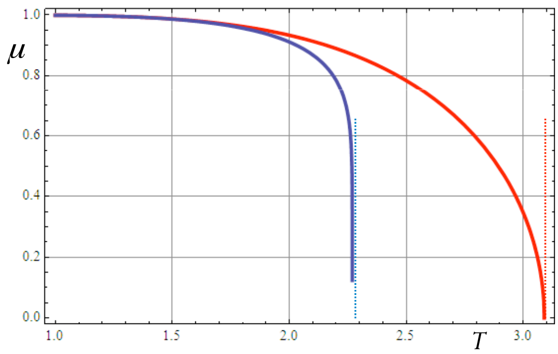
$$\left[p^2+(1-p)^2\right]\tanh(2\kappa)+4p(1-p)\tanh\kappa=1$$

$$\frac{1}{2}(1+\mu^2)\tanh(2\kappa)+(1-\mu^2)\tanh\kappa=1$$

$$x=\exp(-\kappa)$$

$$\mu=\sqrt{1+\frac{4x^4(1+x^2)}{(1-x^2)^3}}=1-2x^4-8x^6-20x^8+O(x^{10})$$

$$(\mu)_{\text{Yang}}=\left[\frac{(1+x^2)\sqrt{1-6x^2+x^4}}{(1-x^2)^2}\right]^{\frac{1}{4}}=1-2x^4-8x^6-34x^8+O(x^{10})$$



Du côté de la Mécanique quantique

- Le sens de la mesure (une projection)
- Une théorie linéaire. **Ce n'est pas une approximation.**
- Une fonction d'onde réellement complexe.
- Des corrélations bizarres.
- De belles inégalités (Weyl, Bell, …)
- Des applications immédiates (calcul quantique, cryptographie quantique)
- Les techniques mathématiques sont familières. Difficultés :
  - quelles notions garder (espace et temps)
  - et quelles notions redéfinir (à peu près tout).

Densités de probabilité

- Univocité de la définition (forme bilinéaire définie positive, …).
- Addition cohérente de tous les processus possibles.
- Pas de précédent en Physique non quantique. Image.
- Les fonctions d'onde sont, aussi bien, des vecteurs d'un espace normé.

Notation de Dirac

ket $|f\rangle$ représente la fonction d'onde.

bra $\langle g|$ représente son adjoint hermitique.

Bracket $\langle f|g\rangle$ représente le produit scalaire et $\langle f|g\rangle = \left(\langle g|f\rangle\right)^*$

$$f(x) = \sum_n c_n \psi_n(x)$$

$$|f\rangle = \sum_n c_n |\psi_n\rangle \qquad c_n = \langle \psi_n | f\rangle$$

$$|f\rangle = \sum_n |\psi_n\rangle\langle\psi_n|f\rangle = \underbrace{\left(\sum_n |\psi_n\rangle\langle\psi_n|\right)}_{=\mathbf{Id}}|f\rangle$$

Produit tensoriel $|f\rangle \otimes |g\rangle \mapsto |f\rangle|g\rangle$

Pas de clonage !

$$|f\rangle \rightarrow |f\rangle|f\rangle$$

$$|g\rangle \rightarrow |g\rangle|g\rangle$$

$$\left(|f\rangle+|g\rangle\right) \rightarrow \left(|f\rangle+|g\rangle\right)\left(|f\rangle+|g\rangle\right) \quad (\text{définition})$$

$$\neq |f\rangle|f\rangle+|g\rangle|g\rangle \qquad (\text{linéarité})$$

Intrigante intrication …

Si $\{S_1+S_2\}$ décrit par produit tensoriel entre un état de S_1 et un état de S_2 : état dit *séparable* ou *factorisable* (S_1 non altéré par mesures sur S_2).

$$|\Psi_{1+2}\rangle = |\psi_1\rangle \otimes |\psi_2\rangle = |\psi_1\rangle|\psi_2\rangle$$

$$|\Psi_{\text{sep}}\rangle = \frac{1}{\sqrt{2}}(|+\rangle_1|-\rangle_2 - |-\rangle_1|-\rangle_2) = \frac{1}{\sqrt{2}}(|+\rangle_1 - |-\rangle_1)\otimes|-\rangle_2$$

État intriqué : *non séparable* $\quad |\Psi_{\text{int}}\rangle = \frac{1}{\sqrt{2}}(|+\rangle_1|-\rangle_2 - |-\rangle_1|+\rangle_2)$

Impossible de séparer conceptuellement les systèmes, de parler de « l'état de S_1 » : seul le système global $\{S_1 + S_2\}$ a un état défini.
Corrélation parfaite des mesures réalisées sur S_1 avec les mesures réalisées sur S_2. Si S_1 donne « + » (50% des cas), le système total est projeté dans l'état $|+\rangle_1 \otimes |-\rangle_2$. La mesure de S_2 donnera « – » avec certitude.

$$\left[\begin{aligned} {}_1\langle+|\Psi_{\text{int}}\rangle &= \frac{1}{\sqrt{2}}\left(\underbrace{{}_1\langle+|+\rangle_1}_{=1}\otimes|-\rangle_2 - \underbrace{{}_1\langle+|-\rangle_1}_{=0}\otimes|+\rangle_2\right) \\ &= \frac{1}{\sqrt{2}}|-\rangle_2 \end{aligned}\right]$$

Inégalité triviale, calcul exact

- Paire de particules:

$$P(0,\overline{90}) = P(0,45,\overline{90}) + P(0,\overline{45},\overline{90})$$

$$P(0,\overline{45}) + P(45,\overline{90}) \geq P(0,\overline{90}) \qquad P(0,45,\overline{90}) \leq P(45,\overline{90}) \qquad P(0,\overline{45},\overline{90}) \leq P(0,\overline{45})$$

$$P(\uparrow\uparrow) = P(\downarrow\downarrow) = \frac{1}{2}\sin^2\left(\frac{a-b}{2}\right) \qquad P(\uparrow\downarrow) = P(\downarrow\uparrow) = \frac{1}{2}\cos^2\left(\frac{a-b}{2}\right)$$

$$\mathbf{a.b} = \begin{cases} \uparrow\uparrow = \downarrow\downarrow = 1 \\ \uparrow\downarrow = \downarrow\uparrow = -1 \end{cases}$$

$$E(a,b) = \langle \mathbf{a.b} \rangle = \sin^2\left(\frac{a-b}{2}\right) - \cos^2\left(\frac{a-b}{2}\right) = -\cos(a-b)$$

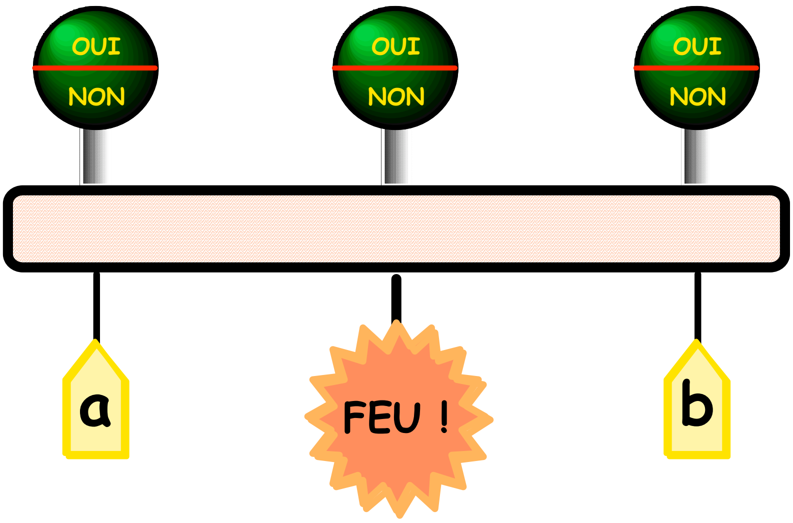
- Conflit ouvert

$$\underbrace{\frac{1}{2}\sin^2\left(\frac{\pi}{8}\right) + \frac{1}{2}\sin^2\left(\frac{\pi}{8}\right)}_{0{,}1464} \geq \underbrace{\frac{1}{2}\sin^2\left(\frac{\pi}{4}\right)}_{0{,}2500}$$

Hypothèses

- Aucune hypothèse sur l'appareillage.
- Aucune mention du déterminisme.
- Aucune référence particulaire (spin ...).
- Aucune référence à la mécanique quantique.

Rien que des corrélations.



- $t = t_1$: Feu ! puis vérification (oui central).
- $t = t_1 + T - \delta$: injection de a et de b.
- $t = t_1 + T$: réponse des autres indicateurs.
- $c\,\delta << L$.

Théorème

- Résultats de A et B corrélés.
- Corrélations inexplicables localement ; il faut une explication.
- Si on peut identifier plusieurs facteurs causaux, les fluctuations résiduelles sont indépendantes.

$$P(A,B|a,b) \neq P_1(A|a) \times P_2(B|b)$$

$$P(A,B|a,b,\lambda) = P_1(A|a,\lambda) \times P_2(B|b,\lambda)$$

$$P(A,B|a,b) = \int P(A,B|a,b,\lambda) f(\lambda) \mathrm{d}\lambda$$

$$I(a,b,a',b') = E(a,b) + E(a,b') + E(a',b) - E(a',b')$$

$$|I(a,b,a',b')| \leq 2$$

Cas particulier

$$I(a,b,a',b') = E(a,b) + E(a,b') + E(a',b) - E(a',b')$$
$$\left|I(a,b,a',b')\right| \leq 2$$

$a = 0$, $a' = \pi/2$, $b = \pi/4$ et $b' = -\pi/4$

$$I(a,b,a',b') = -3\cos\left(\frac{\pi}{4}\right) + \cos\left(\frac{3\pi}{4}\right) = -2\sqrt{2}$$
$$\left|I(a,b,a',b')\right| \approx 2{,}828$$
$$2{,}828 > 2$$

$$E(a,b) = P_{a,b}(\text{oui},\text{oui}) + P_{a,b}(\text{non},\text{non}) - P_{a,b}(\text{oui},\text{non}) - P_{a,b}(\text{non},\text{oui})$$

$$P(A,B\,|\,a,b,\lambda) = P_1(A\,|\,a,\lambda) \times P_2(B\,|\,b,\lambda) \qquad P(A,B\,|\,a,b) = \int P(A,B\,|\,a,b,\lambda)\, f(\lambda)\,\mathrm{d}\lambda$$

$$E(a,b) = \int \overline{A}(a,\lambda)\overline{B}(b,\lambda) f(\lambda)\,\mathrm{d}\lambda \qquad \begin{cases} \overline{A}(a,\lambda) = P_1(\text{oui}\,|\,a,\lambda) - P_1(\text{non}\,|\,a,\lambda) \\ \overline{B}(a,\lambda) = P_2(\text{oui}\,|\,b,\lambda) - P_2(\text{non}\,|\,b,\lambda) \end{cases}$$

$$E(a,b) \pm E(a,b') = \int \overline{A}(a,\lambda)\left[\overline{B}(b,\lambda) \pm \overline{B}(b',\lambda)\right] f(\lambda)\,\mathrm{d}\lambda$$

$$0 \leq P_1 \leq 1 \Rightarrow \left|\overline{A}(a,\lambda)\right| \leq 1 \qquad 0 \leq P_2 \leq 1 \Rightarrow \left|\overline{B}(b,\lambda)\right| \leq 1$$

$$\left|E(a,b) \pm E(a,b')\right| \leq \int \underbrace{\left[\overline{B}(b,\lambda) \pm \overline{B}(b',\lambda)\right]}_{\leq 1} f(\lambda)\,\mathrm{d}\lambda \qquad \left|E(a',b) \pm E(a',b')\right| \leq \int \underbrace{\left[\overline{B}(b,\lambda) \mp \overline{B}(b',\lambda)\right]}_{\leq 1} f(\lambda)\,\mathrm{d}\lambda$$

$$\int f(\lambda)\,\mathrm{d}\lambda = 1 \Rightarrow \left|E(a,b) \pm E(a,b')\right| + \left|E(a',b) \pm E(a',b')\right| \leq 2$$

Relations entropiques d'inégalité ?

- Bach : pas de staccato sur les basses : $\Delta f\,\Delta t \geq c$

- Gabor (1940) : durée de signal et largeur spectrale

$$G(\omega)=\frac{1}{\sqrt{2\pi}}\int g(t)\exp(-i\omega t)\,\mathrm{d}t \qquad g(t)=\frac{1}{\sqrt{2\pi}}\int G(\omega)\exp(i\omega t)\,\mathrm{d}t$$

$$\langle t\rangle=\frac{\int t\,\bar{g}(t)g(t)\,\mathrm{d}t}{\int \bar{g}(t)g(t)\,\mathrm{d}t} \qquad \langle \omega\rangle=\frac{\int \omega\,\bar{G}(\omega)G(\omega)\,\mathrm{d}\omega}{\int \bar{G}(\omega)G(\omega)\,\mathrm{d}\omega}$$

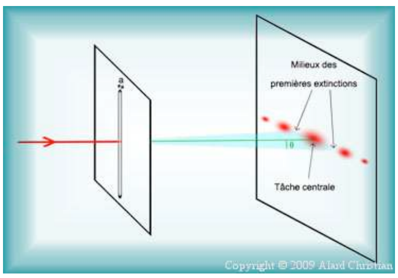
$$\Delta\omega\,\Delta t \geq 1$$

$$\psi(x,t)\mapsto p(x,t)=\left|\psi(x,t)\right|^2$$
$$\varphi(p,t)\mapsto p(x,t)=\left|\varphi(x,t)\right|^2$$

Relations entropiques d'inégalité ?

$$-\frac{a}{2}\leq x\leq\frac{a}{2} \qquad \psi(x)=\frac{1}{\sqrt{a}} \qquad \varphi(p)=\sqrt{\frac{a}{2\pi}}\,\frac{\sin\left(\frac{ap}{2}\right)}{\frac{ap}{2}} \qquad \Delta p_y \geq \frac{2\pi\hbar}{\Delta x}=\frac{2\pi\hbar}{a} \qquad \theta\approx\frac{p_y}{p_x}=\frac{2\pi\hbar}{a\left(\hbar k_y\right)}=\frac{\lambda}{a}$$

$$\Delta x=\frac{a}{2\sqrt{3}} \qquad \Delta p=\infty$$

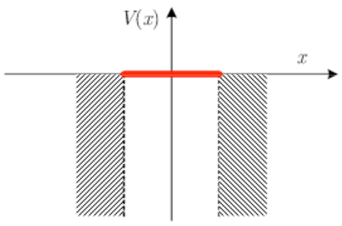


$$S_p=-\int_{-\infty}^{\infty}\frac{a}{2\pi}\left[\frac{\sin(ap/2)}{ap/2}\right]^2\ln\left\{\frac{a}{2\pi}\left[\frac{\sin(ap/2)}{ap/2}\right]^2\right\}\mathrm{d}\,p=\frac{J}{\pi}+\ln\left(\frac{\pi}{2a}\right)$$

$$S_x=-\int_{-\frac{a}{2}}^{\frac{a}{2}}\frac{1}{a}\ln\left(\frac{1}{a}\right)\mathrm{d}\,x=\ln a$$

$$S_p+S_x=C$$

Relations entropiques d'inégalité ?



$-a \leq x \leq a$

$$u_n^+ = \frac{1}{\sqrt{a}} \cos\left[\pi\left(n-\frac{1}{2}\right)\frac{x}{a}\right] \qquad E_n = \frac{\pi^2\hbar^2}{2ma^2}\left(n-\frac{1}{2}\right)^2$$

$$(\Delta x)^2 = \frac{a^2}{3}\left[1-\frac{3}{2\pi^2\left(n-\frac{1}{2}\right)^2}\right] \qquad (\Delta p)^2 = \frac{\pi^2\hbar^2}{a^2}\left(n-\frac{1}{2}\right)^2$$

$$\Delta x \Delta p = F(n)$$

$$S_p + S_x = Cte$$

Ubiquité de l'entropie ?

Thermodynamique
Information
Quantique

Olivier Rioul

Maître de conférences à Télécom ParisTech, professeur chargé de cours à l'Ecole Polytechnique

Les probabilités sans peine ?

Parlons de ce qui fâche : les Probabilités forment un domaine mathématique flou et hasardeux par essence, qui fourmille de paradoxes et de calculs contre-intuitifs. Elles nécessitent en préalable une formation solide en théorie de la mesure et de l'intégration de Lebesgue : la simple notion de variable aléatoire fait appel à celle de fonction mesurable sur un espace probabilisé muni d'une tribu — que dire alors des concepts plus avancés de stationnarité et d'ergodicité, si indispensables à l'ingénieur ? Elles demandent, pour être bien faites, une très grande expérience et un langage spécifique, décorrélés des mathématiques ordinaires. Réduites à la modélisation statistique du type vu en Terminale, elles sont très pauvres ; enseignées pour elle-mêmes, elles deviennent trop abstraites. Et l'on peut légitimement craindre leur inflation probable aux concours d'entrée aux Grandes Écoles.

Pour chacune de ces idées reçues, je tenterai quelques possibles remèdes ou pistes de solutions.

Les Probabilités Sans Peine?

Olivier RIOUL

Télécom ParisTech
Institut Telecom
CNRS LTCI
olivier.rioul@telecom-paristech.fr

Journées Télécom-UPS 2012
Paris, France
10 mai 2012

Les Probabilités Sans Peine?
Olivier RIOUL

Domaine hasardeux ?
Lieu de paradoxes ?
Trop difficile ?
Monde à part ?
Ennuyeuses / trop abstraites ?
Inflation aux concours ?
Conclusion

Parlons de ce qui fâche :

1. Les Probabilités : un domaine hasardeux ?
2. Les Probabilités : un lieu de paradoxes ?
3. Les Probabilités : une théorie trop difficile ?
4. Les Probabilités : un monde à part ?
5. Les Probabilités : ennuyeuses et trop abstraites ?
6. Les Probabilités : probable inflation aux concours ?
7. En guise de conclusion

Le hasard en mathématiques

Pas de culture mathématique du hasard

- à l'encontre du déterminisme des sciences (Laplace)
- incompétence à décrire complètement notre environnement
- lois des grands nombres, configurations moyennes
- mathématiques appliquées (c-à-d impures)

Distinguer deux problèmes

- logique : axiomes, théorèmes (basé sur l'analyse)
- physique : pertinence, sensations

Exercice

Coïncidence d'anniversaires

Je parie qu'au moins deux d'entre vous sont nés le même jour (il n'y a pas de jumeaux). Ai-je raison ?

Oui :

- Effectif N
- Interprétation linéaire (fausse) : $N > 365/2$
- $365 \cdot 364 \cdot (365 - N + 1)$ arrangements de N dates distinctes :

$$\mathbb{P}(\text{tort}) = \prod_{k=1}^{N} (1 - k/365) \approx \exp\left(-\frac{N^2}{2 \times 365}\right)$$

qui est < 0.0001 pour $N = 84$.

Exercice

Vous êtes candidat au jeu du “Big Deal”

La voiture est derrière l’un des 3 rideaux. Vous choisissez au hasard un rideau ; le présentateur (qui sait où se trouve la voiture) faire ouvrir un autre rideau derrière lequel se trouve une chèvre, et vous offre la possibilité de changer d’avis.

Je change d’avis :

- l’argument d’uniformité ne s’applique pas :
- Problème d’inférence baysienne : hypothèse H « la voiture est derrière le premier rideau » + événement E « la voiture n’est pas derrière au autre rideau »
- $\mathbb{P}(H) = \frac{1}{3}$ $\quad \mathbb{P}(E|H) = \mathbb{P}(E|H^c) = \frac{1}{2}$ donc $\mathbb{P}(H|E) = \mathbb{P}(H)$ inchangé $= \frac{1}{3}$.

Exercice

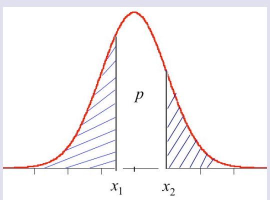
Les Probabilités Sans Peine?
Olivier Rioul
Domaine hasardeux ?
Lieu de paradoxes ?
Contre-Intuitif
Langagier
Trop difficile ?
Monde à part ?
Ennuyeuses / trop abstraites ?
Inflation aux concours ?
Conclusion

Vous êtes (encore) candidat au jeu TV

On présente deux enveloppes contenant des sommes d’argent, vous en choisissez une et prenez connaissance de son contenu. Vous avez la possibilité de garder celui-ci ou de choisir l’autre enveloppe.

Je prends ma décision sur un tirage :

- Soit x la somme connue, y une réalisation normale (par ex).
- Si $x \leqslant y$, on choisit l’autre enveloppe.
- $\mathbb{P}(\text{raison}) = \frac{1-p}{2} + p > 0.5$.



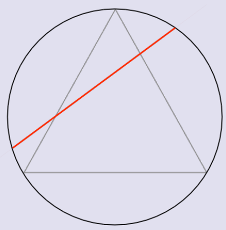
Exercice

On choisit au hasard une corde d'un cercle

Quelle est la probabilité pour qu'elle soit de longueur supérieure au côté du triangle équilatéral inscrit ?

Tout dépend du « choisi au hasard »

- extrémités au hasard : $\mathbb{P} = 1/3$
- direction + hauteur au hasard : $\mathbb{P} = 1/2$.
- milieu au hasard : $\mathbb{P} = 1/4$.
- longueur au hasard : $\mathbb{P} = 1 - \sqrt{3}/2$.



Exercice

Les Probabilités Sans Peine?
Olivier Rioul
Domaine hasardeux ?
Lieu de paradoxes ?
Contre-Intuitif
Langagier
Trop difficile ?
Monde à part ?
Ennuyeuses / trop abstraites ?
Inflation aux concours ?
Conclusion

Votre voisin a deux enfants.

Vous sonnez à sa porte. Une fille ouvre (E). Quelle est la probabilité que l'autre enfant soit un garçon (H) ?

Sans autre information :

- $\mathbb{P}(H) = 2/3$ (sachant qu'il y a une fille).
- $\mathbb{P}(E|H) = 1/2$ (équiprobabilité), $\mathbb{P}(E|H^c) = 1$

$$\frac{\mathbb{P}(H|E)}{\mathbb{P}(H^c|E)} = \frac{\mathbb{P}(H)}{\mathbb{P}(H^c)} \cdot \frac{\mathbb{P}(E|H)}{\mathbb{P}(E|H^c)} = 2 \cdot \frac{1}{2} = 1$$

d'où $\mathbb{P}(H|E) = 0.5$.

Les choix possibles

Questions

- se farcir les *tribus* et autres *σ-algèbres* ?
- abandonner les lois continues ? (tout est discret)
- et l'hypothèse normale ? (TCL)

Quel est le but ?

- bénéficier (rapidement) des outils
- besoins pratiques de l'ingénieur
- sans technicité excessive
- sans abandonner la rigueur

Quel est ce mystérieux Ω ?

Les Probabilités Sans Peine?

Olivier Rioul

Domaine hasardeux ?
Lieu de paradoxes ?
Trop difficile ?
Mystérieux Ω
Axer sur les « v.a. »
Discret & continu
Distributions
Tout le reste
Monde à part ?
Ennuyeuses / trop abstraites ?
Inflation aux concours ?
Conclusion

L'univers de tous les possibles

- Théorie élémentaire des probabilités (ensembliste) :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \qquad (= \mathbb{P}(A) \quad \text{si } A \perp\!\!\!\perp B)$$

  $\implies$ critiquable.
- Borel-Cantelli ($\limsup A_n$) $\implies$ délicat.
- Choix judicieux de $(\Omega, \mathcal{A}, \mathbb{P})$ $\implies$ ad-hoc.
- Variables aléatoires $X_n(\omega)$, $\omega \in \Omega$ $\implies$ mystère !

$\implies$ *raisonner sur les quantités aléatoires plutôt que sur les événements.*

Tout axer sur les « v.a. »

Qu'est qu'une variable aléatoire

- une quantité (X) et non une fonction ($\omega \mapsto X(\omega)$)
- caractérisée (définie !) par les probabilités $\mathbb{P}(X \in A)$
- "toujours" numérique : $X \in \mathbb{R}$ ou $X \in \mathbb{R}^n$ (quitte à transcoder)

Harald Cramér, Mathematical Methods of Statistics, Princeton Univ. Press, 1946 (réimpression 1999).

Les cas discret & continu

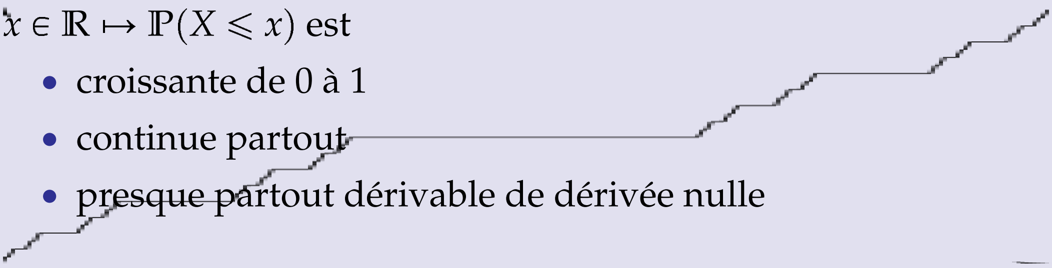
v.a. la plus générale (Lebesgue-Radon-Nikodym) :

- Partie discrète (masses ponctuelles au plus dénombrable)
- Partie (absolument) continue (définie par une densité)
- Partie diffuse (dégénérée)

Escalier du Diable

$x \in \mathbb{R} \mapsto \mathbb{P}(X \leqslant x)$ est

- croissante de 0 à 1
- continue partout
- presque partout dérivable de dérivée nulle



Quelles applications ?

Les distributions de probabilité

Formalisme unifié

- Tout est discret ou continu (ou un mélange)
- Notation sommatoire

$$\mathbb{P}(X \in A) = \sum\!\!\!\!\!\!\int_{x \in A} p(x)$$

où $p(x)$ est soit une probabilité, soit une densité.

- Comparer

$$\mathbb{E}(X) = \sum\!\!\!\!\!\!\int x\, p(x)$$

à

$$\mathbb{E}(X) = \int_{\Omega} X(\omega)\, \mathrm{d}\mathbb{P}(\omega)$$

Les théorèmes savoureux

TELECOM ParisTech
Les Probabilités Sans Peine?
Olivier Rioul
Domaine hasardeux ?
Lieu de paradoxes ?
Trop difficile ?
Mystérieux Ω
Axer sur les « v.a. »
Discret & continu
Distributions
Tout le reste
Monde à part ?
Ennuyeuses / trop abstraites ?
Inflation aux concours ?
Conclusion

On peut (on doit ?) faire rapidement

- loi (faible/forte) des grands nombres
- théorème central limite
- lois du tout ou rien (Borel-Cantelli, Kolmogorov)
- processus aléatoires : stationnarité, ergodicité

On peut ne pas (on ne pourra pas) faire rapidement

- l'espace de Hilbert L^2
- les mouvements browniens
- les martingales
- calcul différentiel stochastique
- intégrale de Itô, etc.

Monde exotique ?

Langage exotique ?

- *expérience, événement, probabilité, espérance, moments, lois...*
- à comparer à : *corps, anneau, idéal, trace, base, noyau...*
- proximité avec le langage humain, les applications

Discipline exotique ?

- omniprésente pour la modélisation – applis innombrables
- plus utile pour les ingénieurs que pour les commerciaux (prépa HEC)

$\Longrightarrow$ son absence en CPGE était incompréhensible pour tous.

Enseignement exotique ?

$\Longrightarrow$ isolé par rapport au reste des mathématiques ?

Weierstraß par Bernstein

Théorème

Toute fonction continue sur un segment est limite uniforme d'une suite de polynômes.

Démonstration

Soit f continue sur $[0,1]$, X une v.a. binomiale (p,n) et $B_n(p) = \mathbb{E}\Big(f\big(\frac{X}{n}\big)\Big)$, polynôme en p. Si $\varepsilon > 0$:

$$\begin{aligned}|B_n(p) - f(p)| &\leqslant \mathbb{E}\Big(\Big|f\big(\tfrac{X}{n}\big) - f(p)\Big| 1_{\left|\frac{X}{n}-p\right|<\varepsilon}\Big) + \mathbb{E}\Big(\Big|f\big(\tfrac{X}{n}\big) - f(p)\Big| 1_{\left|\frac{X}{n}-p\right|\geqslant\varepsilon}\Big) \\ &\leqslant \max_{\left|\frac{k}{n}-p\right|<\varepsilon} \Big|f\big(\tfrac{k}{n}\big) - f(p)\Big| + 2\|f\|_\infty \,\mathbb{P}\Big(\Big|\frac{X}{n} - p\Big| \geqslant \varepsilon\Big)\end{aligned}$$

On conclut avec la continuité uniforme de f et l'*inégalité de Bienaymé-Chebyshev* $\square$

Les Probabilités Sans Peine?
Olivier Rioul
Domaine hasardeux ?
Lieu de paradoxes ?
Trop difficile ?
Monde à part ?
En Analyse
En Algèbre
En Arithmétique
En Géométrie
Ennuyeuses / trop abstraites ?
Inflation aux concours ?
Conclusion

Stirling par le TCL

Théorème

$\frac{n^{n+1/2}}{n!}e^{-n} \xrightarrow{n\to\infty} \frac{1}{\sqrt{2\pi}}.$

Démonstration

Soit $(X_n)_n$ une suite i.i.d. de v.a. Poisson(1). La v.a. $S_n = X_1 + X_2 + \cdots + X_n$ suit une loi de Poisson(n), et par le *théorème central limite*, $\frac{S_n-n}{\sqrt{n}} \xrightarrow{\text{loi}} Y \sim \mathcal{N}(0,1)$. Comme $x \mapsto x^-$ est continue,

$$\mathbb{E}\left\{\left(\frac{S_n-n}{\sqrt{n}}\right)^-\right\} \xrightarrow{n\to\infty} \mathbb{E}(Y^-) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0} ye^{-y^2/2}\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}}$$

Mais dans $\mathbb{E}\left\{\left(\frac{S_n-n}{\sqrt{n}}\right)^-\right\} = \sum_{k=0}^{n}\left(\frac{n-k}{\sqrt{n}}\right)\frac{n^k}{k!}e^{-n}$, tous les termes s'éliminent en cascade sauf $\dfrac{n^{n+1/2}}{n!}e^{-n}$. □

Les Probabilités Sans Peine?

Olivier Rioul

Domaine hasardeux ?

Lieu de paradoxes ?

Trop difficile ?

Monde à part ?

En Analyse

En Algèbre

En Arithmétique

En Géométrie

Ennuyeuses / trop abstraites ?

Inflation aux concours ?

Conclusion

Sperner

TELECOM ParisTech

Théorème

On peut trouver au plus $\max_k \binom{n}{k} = \binom{n}{\lfloor n/2 \rfloor}$ parties de $\{1,2,\ldots,n\}$ dont aucune n'est contenue dans aucune autre.

Démonstration

Soit $\mathcal{A}$ une telle famille de parties, la chaîne aléatoire : $\mathcal{C}_\sigma = \{\varnothing, \{\sigma_1\}, \{\sigma_1,\sigma_2\}, \ldots, \{1,2,\ldots,n\}\}$ et $X = |\mathcal{A} \cap \mathcal{C}_\sigma|$ ($= 0$ ou 1).

$$1 \geqslant \mathbb{E}(X) = \mathbb{E}\Big(\sum_{A\in\mathcal{A}} 1_{A\in\mathcal{C}_\sigma}\Big) = \sum_{A\in\mathcal{A}} \mathbb{E}(1_{A\in\mathcal{C}_\sigma}) = \sum_{A\in\mathcal{A}} \mathbb{P}(A \in \mathcal{C}_\sigma)$$
$$= \sum_{A\in\mathcal{A}} \frac{1}{\binom{n}{|A|}} \geqslant \frac{|\mathcal{A}|}{\max_k \binom{n}{k}}.$$

□

Les Probabilités Sans Peine?

Olivier Rioul

Domaine hasardeux ?

Lieu de paradoxes ?

Trop difficile ?

Monde à part ?

En Analyse

En Algèbre

En Arithmétique

En Géométrie

Ennuyeuses / trop abstraites ?

Inflation aux concours ?

Conclusion

Ky Fan (concavité du log det)

Théorème

$\forall A, B \in S_n^{++}(\mathbb{R})$ et $\forall \lambda > 0, \mu > 0$ tels que $\lambda + \mu = 1$,

$$\det(\lambda A + \mu B) \geqslant (\det A)^{\lambda} (\det B)^{\mu}.$$

Démonstration (1988)

Soit $X_0 \sim \mathcal{N}(0, A)$ et $X_1 \sim \mathcal{N}(0, B)$, Θ une v.a. binaire $\perp\!\!\!\perp (X_0, X_1)$ de loi de Bernoulli (λ, μ). La matrice de covariance de $Y = X_\Theta$ est $= \lambda A + \mu B$. Inégalités sur les entropies différentielles :

$$\lambda h(X_0) + \mu h(X_1) = h(Y|\Theta) \leqslant h(Y) \leqslant h(Z)$$

où $Z \sim \mathcal{N}(0, \lambda A + \mu B)$. On conclut en calculant l'entropie de la loi $\sim \mathcal{N}(0, C)$; c'est $= -\int p \ln p = \frac{n}{2} \ln(2\pi e) + \frac{1}{2} \ln \det C$. $\square$

Hardy-Ramanujan

Théorème

$\forall \varepsilon > 0$,

$$\frac{1}{n}\left|\Big\{N \leqslant n\ ;\ (1-\varepsilon)\ln\ln N < \omega(N) < (1+\varepsilon)\ln\ln N\Big\}\right| \xrightarrow{n\to\infty} 1.$$

Démonstration

Soit $N = \prod_p p^{X_p}$ une v.a. entière uniforme sur $\{1, 2, \ldots, n\}$. Alors $\omega(N) = \sum_{p \leqslant n} X_p^*$ où $X_p^* = \min(X_p, 1)$ suit une loi de Bernoulli de paramètre $\frac{1}{n}\lfloor\frac{n}{p}\rfloor$. Mais $E(\omega(N)) = \sum_p \frac{1}{n}\lfloor\frac{n}{p}\rfloor = \ln\ln n + O(1)$, et un calcul un peu long montre que $\mathrm{Var}(\omega(N)) = O(\ln\ln n)$. On conclut avec l'*inégalité de Bienaymé-Chebyshev* :
$\mathbb{P}\left(\left|\frac{\omega(N) - E(\omega(N))}{\ln\ln n}\right| \geqslant \varepsilon\right) \leqslant \frac{\mathrm{Var}(\omega(N))}{(\ln\ln n)^2\varepsilon^2} \xrightarrow{n\to\infty} 0.$ $\square$

Cette preuve est reprise par Hardy... en effaçant les traces !

Erdös

Théorème

Tout ensemble $A = \{a_1, a_2, \dots, a_n\}$ de n entiers non nuls contient $> n/3$ entiers $\{a_{i_k}\}_k$ tels que $a_{i_k} + a_{i_l} \neq a_{i_m}$ pour tous k, l, m.

Démonstration

Soit $p = 3k + 2$ premier, plus grand que tous les $2|a_i|$, X un v.a. uniforme à valeurs dans $\mathbb{Z}_p^*$. Alors $X_i = a_i \cdot X \bmod p$ suit aussi une loi uniforme. Soit $\mathbf{B} \subset A$ constitué des entiers a_i tels que $X_i \in \{k+1, k+2, \dots, 2k+1\}$. Sa taille moyenne est $\mathbb{E}(|\mathbf{B}|) = \sum_{i=1}^n \mathbb{P}(k < X_i <= 2k+1) = \sum_{i=1}^n \frac{k+1}{p-1} = \frac{n(k+1)}{3k+1} > \frac{n}{3}$ donc il existe une valeur $X = x$ donnant un $|B| > n/3$. Les éléments $a_i \in B$ sont t.q. $a_i x \bmod p \in \{k+1, k+2, \dots, 2k+1\}$. Si $a_{i_k} + a_{i_l} = a_{i_m}$, en multipliant par x modulo p on trouverait une impossibilité. $\square$

Gram

Les Probabilités Sans Peine?
Olivier RIOUL
Domaine hasardeux ?
Lieu de paradoxes ?
Trop difficile ?
Monde à part ?
En Analyse
En Algèbre
En Arithmétique
En Géométrie
Ennuyeuses / trop abstraites ?
Inflation aux concours ?
Conclusion

Théorème de Gram

Soit α_i $(0 \leqslant i \leqslant n)$ la somme des angles solides intérieurs des i-faces d'un polyhèdre convexe en n dimensions. Alors

$$\sum_{i=0}^{n} (-1)^i \alpha_i = 0.$$

Démonstration (1994)

Choisir la direction d'un hyperplan H au hasard et considérer le projeté orthogonal du polyhèdre P sur H...

Zubkov, 1979

Théorème

Soit $C = [0,1]^n$ l'hypercube unité de $\mathbb{R}^n$, H_t l'hyperplan d'équation $x_1 + x_2 + \cdots + x_n = t$, et S le simplexe défini par l'enveloppe convexe des $n+1$ points 0 et $e_1, 2e_2, \ldots ne_n$. Alors $\mathrm{vol}(C \cap H_t) = \mathrm{vol}(S \cap H_t)$.

Démonstration

Soit $X_1, X_2, \ldots, X_n$ des v.a. i.i.d. exponentielles de paramètre $\lambda > 0$, réordonnés par ordre croissant. Ecrire de deux manières $\mathbb{P}(\forall i, X_i \leqslant 1)$ et conclure par injectivité de la transformée de Laplace...

Probabilités ennuyeuses ?

Premiers pas en probabilité. . .

Combinatoire

- les boules ... dans les urnes, les cartes à jouer, etc.
- raisonnements de pure dénombrement

Statistique

- les écarts-type et inter-quartiles sur tableur
- pure statistique descriptive

Théorie abstraite

- mesure
- probabilités pour elles-mêmes

Equilibre théorie / applications

Modélisation

- simulations en `scilab`
- T.I.P.E.

Théorie enseignée

- doit être utile aux sciences physiques
- formatrice (apprentissage de la logique, du raisonnement, de la rigueur)

Concours : présents & futurs

Les Probabilités Sans Peine?
Olivier Rioul

Domaine hasardeux ?
Lieu de paradoxes ?
Trop difficile ?
Monde à part ?
Ennuyeuses / trop abstraites ?
Inflation aux concours ?
Conclusion

Probabilités déjà présentes, mais cachées : (au CCMP)

PC 2011	un théorème ergodique
PC/PSI 2010	théorème central limite
MP 2009	problème des moments et loi log-normale
PSI 2009	matrices aléatoires
PC/PSI 2007, MP 2006	matrices stochastiques
PC/PSI 2004	séries génératrices de v.a. de Poisson

À l'avenir

- une introduction explicite (moins d'hypocrisie. . .)
- gagner du temps
- enrichir la palette du candidat

Conclusion

Domaine hasardeux ?

Lieu de paradoxes ?

Trop difficile ?

Monde à part ?

Ennuyeuses / trop abstraites ?

Inflation aux concours ?

Conclusion

Il faudra bien s'adapter…(bon gré mal gré)

Formation des professeurs

- livres pédagogiques
  (niveau intermédiaire entre Bac et Master)
- formations LIESSE
- ?

Les probabilités sans peine ?

Olivier RIOUL

10 mai 2012

Parlons de ce qui fâche : les probabilités forment un domaine mathématique flou et hasardeux par essence, qui fourmille de paradoxes et de calculs contre-intuitifs. Elles nécessitent en préalable une formation solide en théorie de la mesure et de l'intégration de Lebesgue : la simple notion de variable aléatoire fait appel à celle de fonction mesurable sur un espace probabilisé muni d'une tribu — que dire alors des concepts plus avancés de stationnarité et d'ergodicité, si indispensables à l'ingénieur ? Elles demandent, pour être bien faites, une très grande expérience et un langage spécifique, décorrélés des mathématiques ordinaires. Réduites à la modélisation statistique du type vu en Terminale, elles sont très pauvres ; enseignées pour elle-mêmes, elles deviennent trop abstraites. Et l'on peut légitimement craindre leur inflation probable aux concours d'entrée aux Grandes Écoles.

Pour chacune de ces idées reçues, je tenterai quelques possibles démentis, remèdes ou pistes de solutions.

Table des matières

1 Les probabilités : un domaine hasardeux ? 3

2 Les probabilités : lieu de paradoxes ? 4

2.1 Des résultats corrects mais contre-intuitifs 4
2.2 Paradoxes de langage 6

3 Les probabilités : une théorie trop difficile ? 8

3.1 Se débarrasser du Ω 9
3.2 Tout axer sur les « v.a. » 10
3.3 Isoler les cas discret et continu 11
3.4 Tout axer sur les distributions 12
3.5 Tout le reste 13

4 Les probabilités : un monde à part ? 13

4.1 Deux exemples en analyse 14
4.2 Deux exemples en algèbre 15
4.3 Deux exemples en arithmétique 17
4.4 Deux exemples en géométrie 19

5 Les probabilités : ennuyeuses et trop abstraites ? 21

6 Les probabilités : probable inflation aux concours ? 22

7 En guise de conclusion 23

A Demandez le programme 24

A.1 Première S (septembre 2011) 24
A.2 Terminale S (septembre 2012) 24
A.3 Maths Sup. (septembre 2013) 25
A.4 Maths Spé. (septembre 2014) 25
A.5 Programme typique en Grande École 25

1 Les probabilités : un domaine hasardeux ?

Les probabilités traitant du hasard, on peut craindre un domaine hasardeux par excellence.

Il faut d'abord comprendre pourquoi cela peut faire peur à l'étudiant comme à l'enseignant. Le hasard va à l'encontre de cette idée de belle rigueur qui caractérise *la Mathématique*. Il y a en effet une horreur de l'esprit humain pour le désordre, le flou, l'incertain de notre environnement. Cela s'est reflété dans le passé par le besoin de certitude dans les sciences, allant jusqu'au Déterminisme, conception selon laquelle toute la Nature, notre propre existence comprise, est entièrement déterminée, dans le passé ou dans l'avenir, par des lois immuables. Décrire scientifiquement le *hasard* est donc un gageure : tel le marquis de Laplace, nous pouvons croire qu'il ne reflète que notre incompétence à décrire précisément tous les facteurs qui contribuent à l'état de notre environnement, dans un lieu ℓ et à un instant t donnés. Le hasard serait donc insaisissable et échapperait à notre connaissance.

Mais voilà que depuis Pascal et Fermat, on découvre – ou on invente... – que le hasard est lui-même soumis aux règles mathématiques, en particulier si l'on considère des configurations *moyennes*. C'est ce qu'on explique généralement par la *loi des grands nombres*. Pour décrire le hasard, il faut alors probablement distinguer deux problèmes... :

1°) Le premier problème est d'ordre logique : comment faire pour inventer une définition opérationnelle, la *probabilité*, qui obéira aux axiomes que l'on voudra bien poser et aux théorèmes que l'on en déduira ? La théorie mathématique ne commence vraiment qu'une fois tous les événements considérés affectés de nombres réels du segment $[0,1]$ appelés « probabilités ». La rigueur reprend alors tous ses droits.

2°) Le deuxième problème est plus physique : comment cette affectation de probabilités aux événements a lieu dans la réalité ? Les événements physiques obéissent-ils aux axiomes mathématiques ? Le désordre et ses fluctuations naturelles sont-elles bien expliquées par la théorie ? Comment alors se définit le vrai hasard, celui dont nous avons la sensation et l'intuition ? Pour toutes ces questions, on est bien obligé de constater une certaine impuissance à trouver des réponses définitives. En tout cas, on sort du domaine mathématique. [1]

Si plonger dans le monde de l'incertain peut inquiéter voire horrifier, c'est une sensation humaine qui ne semble pas provenir des mathématiques elles-

1. Je n'ai évidemment pas la prétention de tout traiter et je ne m'attacherai ici qu'à la description mathématique des Probabilités. J'invite l'auditeur de cette conférence désireux d'explorer d'autres axes à se tourner vers les exposés d'Alain MARUANI pour la physique et d'Yves GUIARD pour les sciences humaines.

même. Pour prévenir la peur que l'on pourrait ressentir, démystifions-les : ce seront toujours des mathématiques ordinaires [2] et elles n'ont en elles-même rien d'hasardeux. Après tout, les mathématiques des probabilités sont conçues comme les autres : il y aura toujours des axiomes, des déductions logiques et des théorèmes. L'interprétation physique des modèles mathématiques du hasard est, par contre, sujette à un autre débat.

2 Les probabilités : lieu de paradoxes ?

On a eu souvent coutume de mettre en exergue des paradoxes liés au calcul de probabilités, en particulier dans les journaux et ouvrages de vulgarisation. Voici ci-dessous des exemples assez classiques – avec les solutions ! – qu'il m'a été donné d'enseigner [3].

2.1 Des résultats corrects mais contre-intuitifs

Exercice (von Mises, 1939)**.** *Le professeur parie qu'au moins deux de ses élèves sont nés le même jour (il n'y a pas de jumeaux dans sa classe). À partir de quel effectif a-t-il raison de faire ce pari ?*

Réponse. En supposant pour simplifier qu'il y a 365 jours par an et un effectif $N \leqslant 365$, une interprétation linéaire conduirait à dire qu'il faudrait une classe très surchargé de plus de 180 élèves. Mais le nombre d'arrangements de n dates distinctes parmi 365 est $= 365 \cdot 364 \cdot (365 - N + 1)$ et la probabilité d'une coïncidence est donc $P = 1 - \prod_{k=1}^{N}(1 - k/365)$ soit environ $1 - \exp(-\frac{N^2}{2\times 365})$ qui est $> 1/2$ dès que $N \geqslant 23$, chiffre confirmé par un calcul exact. □

Cet exercice peut être traité en Première ou en Terminale (c'était déjà le cas lorsque j'étais élève). La difficulté n'est pas particulière aux Probabilités, mais aux mathématiques en général : il s'agit de bien identifier les données du problème (en l'occurrence l'univers des possibles).

—

Exercice (Selvin, 1975)**.** *Vous êtes candidat à un jeu télévisé ; la voiture est derrière l'un des trois rideaux. Vous choisissez au hasard un rideau ; le présentateur (qui sait où se trouve la voiture) faire ouvrir un autre rideau derrière lequel se trouve une chèvre, et vous offre la possibilité de changer d'avis : que faites vous ?*

2. J'ai déjà rencontré un élève de Terminale qui avait peur de rentrer en classes de mathématiques « spéciales »…

3. Je suppose une connaissance de base de la théorie (disons jusqu'au niveau Licence).

Réponse. Beaucoup de gens soutiennent *mordicus* qu'il reste un chance sur deux, donc il n'y a pas d'intérêt particulier à changer. Mais cela suppose en fait que le probabilité initiale (1/3) change par le conditionnement (l'action du présentateur). Or justement, ici ce n'est pas le cas : en effet, si on utilise la méthode d'« inférence baysienne » : l'hypothèse est H : « la voiture est derrière le premier rideau », $\mathbb{P}(H) = 1/3$. L'événement supplémentaire est E : « le présentateur montre que la voiture ne se trouve pas derrière un autre rideau », on a toujours $\mathbb{P}(E|H) = \mathbb{P}(E|H^c) = \frac{1}{2}$, et donc les « côtes » sont inchangées :

$$\frac{\mathbb{P}(H|E)}{\mathbb{P}(H^c|E)} = \frac{\mathbb{P}(H)}{\mathbb{P}(H^c)} \cdot \frac{\mathbb{P}(E|H)}{\mathbb{P}(E|H^c)} = \frac{1}{2} \cdot 1 = \frac{1}{2}$$

Par conséquent $\mathbb{P}(H|E)$ est toujours $= 1/3$, donc la voiture à deux chances sur trois d'être derrière le troisième rideau, et il est préférable de changer d'avis. □

La nature contre-intuitive du résultat provient ici de l'utilisation des probabilités conditionnelles (inférence baysienne), thème souvent délicat pour les étudiants.

—

Exercice. *On attend le bus dans un trafic perturbé où les arrivées des bus suivent un processus de Poisson, avec en moyenne un bus toutes les T minutes. Trouver le temps moyen d'attente du bus.*

Réponse. La réponse naïve est $T/2$ car il y a un bus toutes les T minutes. Ce serait vrai si les arrivées des bus étaient uniformément réparties, car l'attente suivrait un loi uniforme entre 0 et T. Mais le processus étant poissonnien, les intervalles entre les arrivées consécutives de bus suivent des v.a. i.i.d. exponentielles de paramètre $\lambda = T$ (moyenne de chaque intervalle). La loi exponentielle étant sans mémoire, sachant qu'on arrive à un certain instant t, l'attente moyenne restante suit encore une loi exponentielle de même paramètre $\lambda = T$. Le temps moyen d'attente du bus est donc T, le double de ce qu'on obtient par un raisonnement naïf. □

Ce très joli résultat peut être vérifié expérimentalement (et pour les plus malchanceux d'entre nous, quotidiennement). Il n'est contre-intuitif qu'au premier abord.

—

Exercice. *Vous êtes (encore) candidat à un jeu télévisé ; on vous présente deux enveloppes contenant des sommes d'argent, vous en choisissez une et prenez connaissance de son contenu. Vous avez la possibilité de garder celui-ci ou de choisir l'autre enveloppe. Que faites vous ?*

Réponse. [4] Là encore, on a tendance à répondre qu'il n'y a rien de particulier à

4. Merci à Aslan TCHAMKERTEN de m'avoir signalé cet exercice.

faire. Mais contrairement aux apparences, il existe une méthode (probabiliste !) pour faire son choix de sorte d'avoir raison avec une probabilité $> 1/2$. Notons $x_1 < x_2$ les deux montants inconnus du candidat, X le montant observé (v.a. binaire avec $\mathbb{P}(X = x_1) = \mathbb{P}(X = x_2) = 1/2$). Sans même aucune connaissance de la façon dont ont été choisies les deux montants, on tire au hasard un nombre réel $Y = y$ selon une loi telle que la probabilité de se trouver dans un intervalle de longueur > 0 est toujours > 0 (par exemple, une v.a. Y gaussienne, indépendante de X). Si le montant observé est $\geqslant y$, on le garde ; s'il est $< y$, on prend l'autre enveloppe (de montant X'). La probabilité d'avoir raison est

$$\mathbb{P}(X1_{X\geqslant Y} + X'1_{X<Y} = x_2) = \mathbb{P}(X = x_1)\mathbb{P}(x_1 < Y) + \mathbb{P}(X = x_2)\mathbb{P}(x_2 \geqslant Y) = \frac{1+p}{2} > \frac{1}{2}$$

où on a noté $p = \mathbb{P}(x_1 < Y \leqslant x_2) > 0$. L'espérance de gain qui en résulte est

$$\mathbb{E}(X1_{X\geqslant Y} + X'1_{X<Y} = x_2) = x_1\frac{1-p}{2} + x_2\frac{1+p}{2} = \frac{x_1+x_2}{2} + (x_2 - x_1)\frac{p}{2} > \frac{x_1+x_2}{2} \qquad \square$$

Exercice. *Même jeu télévisé, mais on sait à l'avance qu'un des montant de l'enveloppe est égal au double de l'autre. Quitte ou moitié/double ?*

Réponse. Disons $x_1 = s$ et $x_2 = 2s$. Un raisonnement fallacieux consiste à dire que sachant $X = x$, l'espérance de l'autre montant $\frac{2x+x/2}{2} > x$ et qu'il faut donc toujours choisir l'autre enveloppe.... C'est faux car sachant $X = x$, X' devient une v.a. déterministe $= 2x$ si $x = s$ et $= x/2$ si $x = 2s$, et $\mathbb{E}(X'|X = x)$ a la même valeur ($= 2x$ si $x = s$ et $= x/2$ si $x = 2s$). On ne peut pas moyenner ces deux valeurs. Une réponse possible est d'utiliser la méthode ci-dessus, qui assure une espérance de gain $> 3s/2$ [5]. $\square$

Il est clair que dans ces deux exercices, la difficulté est également renforcée par le langage utilisé, en particulier la question vague “qui tue” : « que faites-vous ? ». On ne pourrait pas imaginer ce genre de question à l'écrit d'un concours, difficilement à l'oral.

2.2 Paradoxes de langage

L'interprétation du langage (utilisé dans l'énoncé) est souvent déterminant pour la résolution. Quelques exemples :

5. Pour les deux variantes du jeu des enveloppes, on peut faire l'hypothèse que les deux montants (ou la somme s) ont été choisis selon une loi de probabilité connue (ou avec une contrainte, par exemple de montant minimum). Dans ce cas on peut affiner la méthode exposée pour maximiser l'espérance du gain, en moyenne sur cette loi.

Exercice (Bertrand, 1888)**.** *Dans un cercle donné, quelle est la probabilité P qu'une corde du cercle choisie au hasard soit de longueur supérieure au côté d'un triangle équilatéral inscrit ?*

Réponse. Tout dépend de ce qu'on entend par « choisi au hasard ». Si on choisit une direction au hasard à l'aide d'un rayon du cercle et un point au hasard sur ce rayon pour définir la corde perpendiculaire en ce point à ce rayon, un côté du triangle équilatéral inscrit sera obtenu lorsque le point choisi et au milieu du rayon, donc $P = \frac{1}{2}$. Si ce sont les deux extrémités de la corde qui sont choisis au hasard, il s'agit de la probabilité pour qu'une extrémité ayant été choisie, considérant le triangle équilatéral inscrit de sommet cette extrémité, l'autre extrémité de la corde se trouve entre les deux autres sommets du triangle, d'où $P = \frac{1}{3}$. Si on choisit un point au hasard dans le cercle définissant une corde dont ce point est le milieu, la corde sera plus petite que le côté d'un triangle équilatéral inscrit si le point est choisi à l'extérieur du cercle inscrit à ce triangle, qui est de rayon moitié du rayon du cercle initial ; donc $P = \frac{1}{4}$. Si on choisit au hasard la longueur de cette corde entre 0 et deux fois le rayon, sachant que le côté d'un triangle équilatéral inscrit est $\sqrt{3}$ fois le rayon, on trouve $P = 1 - \frac{\sqrt{3}}{2}$. □

Cet exemple montre qu'il y a parfois aucune hypothèse a priori préférable à une autre et qu'il *faut* préciser la façon dont un choix au hasard est fait.

Exercice (Gardner, 1954)**.** *Votre voisin vous a dit qu'il avait deux enfants. Lorsque vous sonnez à sa porte, une fille ouvre. Quelle est la probabilité que l'autre enfant soit un garçon ?*

Réponse. Il manque ici beaucoup de données : Tout d'abord, il est entendu par défaut que les naissances sont indépendantes et qu'il y a une chance sur deux d'avoir une fille (ou un garçon). Bref, sans connaissance supplémentaire, on fait l'hypothèse d'équiprobabilité et d'indépendance (même si ce n'est pas forcément tout à fait vrai en pratique). Ce peut être une règle par défaut universelle pour les exercices.

Maintenant, si (disons) l'aînée est une fille, il y a une chance sur deux pour que le deuxième soit un garçon (les naissances sont supposées indépendantes). Mais si un des deux enfants est une fille, il y a deux chances sur trois pour que l'autre soit un garçon (les possibilités sont (G,F) (F,G), (F,F)). Disons que par défaut si on ne dit rien, c'est qu'on est dans le deuxième cas (on n'a pas spécifié l'ordre de naissance).

Il manque néanmoins une information sur la probabilité de qui répond à la porte. Disons par exemple (par défaut) que l'un ou l'autre des enfants, avec la même probabilité 1/2, répond à la porte. On applique donc la méthode d'« inférence baysienne » : l'hypothèse est H : « l'autre enfant est un garçon » (sachant qu'il y a une fille) et donc $\mathbb{P}(H) = 2/3$. Mais il y a un événement supplémentaire E : « une fille ouvre » tel que $\mathbb{P}(E|H) = 1/2$, $\mathbb{P}(E|H^c) = 1$ et on cherche à détermine $\mathbb{P}(H|E)$: alors

$$\frac{\mathbb{P}(H|E)}{\mathbb{P}(H^c|E)} = \frac{\mathbb{P}(H)}{\mathbb{P}(H^c)} \cdot \frac{\mathbb{P}(E|H)}{\mathbb{P}(E|H^c)} = 2 \cdot \frac{1}{2} = 1$$

d'où $\mathbb{P}(H|E) = \frac{1}{2}$. □

Ainsi, dans le cas d'hypothèses reflétant la plus grande méconnaissance des données, il se trouve que c'est la réponse naturelle qui est la bonne : « une chance sur deux, voyons ! ». On voit que le problème est une question d'interprétation mathématique du langage commun, il faut parfois préciser les hypothèses faites ou si c'est possible, adopter un principe de méconnaissance maximale par défaut.

D'autres énoncés de « paradoxe » restent encore sujets à polémiques et controverses aujourd'hui (paradoxe de la Belle au bois dormant, paradoxe de l'Apocalypse, etc.), faute de s'être mis d'accord pour clarifier leur interprétation mathématique.

Ces contreverses n'ont pas pour origine les mathématiques elles-même, seulement de l'*interprétation* qui est faite de l'énoncé. Comme ailleurs en sciences, il s'agit de faire preuve de la plus grande vigilance en essayant de poser des énoncés univoques. D'ailleurs, ne pourrait-on pas dire logiquement que s'il y avait réellement paradoxe de contenu en théorie des Probabilités, celle-ci n'étant que purement mathématique, il impliquerait paradoxe de *toutes* les mathématiques ?

3 Les probabilités : une théorie trop difficile ?

Depuis Kolmogorov et son approche axiomatique de la Théorie des Probabilités, il est bien reconnu que les outils mathématiques ainsi donnés aux concepts fondamentaux des probabilités sont puissants et efficaces. C'est pourquoi la plupart des livres et manuels sur ce sujet, dès le niveau de la Licence, obligent à se familiariser d'abord avec la théorie de la *mesure* et de l'*intégration* avant d'aborder le calcul de probabilités proprement dit [6].

Le passage obligé par la théorie de la mesure et de l'intégration constitue évidemment une difficulté initiale importante, qui s'avère rebutante pour de

6. La théorie de la mesure la plus adaptée aux probabilités n'est d'ailleurs pas nécessairement celle préconisée pas les mathématiciens « purs » (restreinte aux mesures de Radon sur les espaces topologiques localement compacts) mais plutôt celle de la mesure "abstraite" (avec le théorème d'unicité de Carathéodory) qui permet ensuite d'appliquer les résultats à des processus aléatoires généraux.

nombreux étudiants. L'étudiant (comme l'enseignant, d'ailleurs) doit alors se farcir ces « tribus » barbares, ces « clans » redoutés, ces « σ-algèbres » hermétiques, ces espaces boréliens, mesurables, probabilisables, probabilisés.... On est assez loin des préoccupations quotidiennes de l'ingénieur, et la théorie de la mesure est souvent perçue – par les étudiants comme les enseignants ! – comme ennuyeuse et trop longue.

À tel point qu'on se pose sérieusement la question si, pour être explicite sans perdre du temps en classes préparatoires, on ne va pas simplement abandonner les probabilités continues et se restreindre aux probabilités discrètes (tout ensemble discret est mesurable...). Ce serait dommage, car il est déjà prévu d'enseigner les densités de probabilité (gaussiennes, exponentielles, uniformes) en Terminale, en vue notamment d'aborder l'approximation gaussienne (pour ne pas dire le *théorème central limite*). Ces considérations constituent tout de même, comme son nom l'indique, un théorème central [7] pour tout ingénieur qui se respecte.

Car vue d'une École d'ingénieurs, quel est le but de l'enseignement des Probabilités ? Il s'agit de rapidement bénéficier des outils de calcul de probabilités pour des besoins pratiques (sans pour cela nécessairement abandonner la saveur et l'intérêt d'une étude suffisamment rigoureuse des concepts). Dans les besoins pratiques, il y a naturellement l'hypothèse normale et les lois continues. Par ailleurs, l'exposé systématique de la théorie de la mesure et de l'intégration n'est peut-être pas un préliminaire indispensable pour commencer à « faire des probabilités » : les résultats utiles liés à la théorie de la mesure peuvent-ils être démontrés au fur et à mesure des besoins sans technicité excessive ?

Voici résumées un certain nombre de pistes de solutions explorées pendant plus de dix ans dans mes cours, et exposées dans mon livre [4]. Si elles ne conduisent pas toujours à une rigueur parfaite (ce point devrait être améliorable), elles tendent en pratique à atteindre l'objectif premier : être rapidement capable de « faire des probabilités ».

3.1 Se débarrasser du Ω

Dans de nombreux manuels, on se familiarise avec les probabilités en se référant à un univers de réalisations possibles souvent noté Ω, et en manipulant à l'aide de diagrammes ensemblistes des probabilités des ensembles d'une tribu $\mathscr{A}$ de parties de Ω. C'est ainsi qu'on définit rapidement les probabilités conditionnelles par la formule $\mathbb{P}(A|B) = \mathbb{P}(A\cap B)/\mathbb{P}(B)$, ainsi que la notion d'événements indépendants : on dit que « $A \perp\!\!\!\perp B$ » si $\mathbb{P}(A\cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Cela donne

7. Le terme *central limit theorem* (dû à G. Pólya) désigne à l'origine un théorème central de la limite, et non comme on le voit trop souvent un « théorème de la limite centrale » (d'ailleurs, la limite en question n'est pas spécialement centrale, mais normale).

lieu à de multiples exercices plus ou moins subtiles de probabilités « élémentaires » dont la résolution passe par des choix plus ou moins judicieux de l'espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$.

Mieux (ou pire) : en admettant ou en construisant un espace probabilisé produit (infini) permettant de modéliser la même expérience répétée une infinité dénombrable de fois, on peut très rapidement énoncer l'indépendance d'une infinité d'événements et démontrer les lemmes de Borel-Cantelli. C'est ainsi qu'on prouve, par exemple, que tout événement probable, même très peu probable, se réalisera presque sûrement une infinité de fois.

Bien que ce pan du cours de probabilités a le mérite de mettre en perspective des éléments de théorie des ensembles, il est critiquable : Borel-Cantelli est déjà d'une compréhension délicate [8] mais surtout, certaines notions – même les plus simples – apparaissent un peu artificielles : pourquoi tel choix de Ω serait-il plus naturel que tel autre pour résoudre tel ou tel exercice ? que déduire du fait que tirer un as ou un pique d'un jeu de cartes sont deux événements indépendants ?

Je soutiens qu'il est possible, et même salutaire à un niveau relativement simple, de se passer du fameux Ω. En effet, il est souvent plus intuitif et plus commode d'analyser un problème de calcul de probabilités en identifiant d'abord les *quantités* variables aléatoires, puis en raisonnant sur leurs distributions de probabilité de ces variables aléatoires [9]. Comme on va le voir maintenant, nul besoin de Ω pour cela.

3.2 Tout axer sur les « v.a. »

On définit classiquement, selon l'approche de la théorie de la mesure, une *variable aléatoire* (v.a.) X comme une application mesurable de Ω vers un espace probabilisable (typiquement $\mathbb{R}$ dans le cas d'une v.a. réelle). On écrit donc $X(\omega)$ où $\omega \in \Omega$. Cependant, au regard des applications pratiques, le rôle du ω dans l'espace probabilisé Ω ne correspond à rien de bien palpable et reste donc très mystérieux. Par exemple, supposons que l'on modélise des échantillons d'un bruit thermique par une suite de v.a. normales corrélées : à quoi correspond ω ?

Ce qui est important en pratique, c'est que l'utilisation d'une variable aléatoire X est entièrement déterminée par la donnée des probabilités $\mathbb{P}\{X \in A\}$, et l'utilisation de la lettre X ne constituera finalement qu'une commodité de notation. Cela suffit largement pour « faire des probabilités », tout du moins au niveau souhaité (jusqu'à la loi des grands nombres et même les processus

8. La notion de réalisation une infinité de fois presque sûrement fait appel à celle de limite supérieure d'ensembles.

9. On peut toujours, en exercice, retrouver les définitions élémentaires (ensemblistes) à l'aide de variables aléatoires définies par des fonctions indicatrices.

aléatoires). Cette approche, qui est celle suivie par Cramér [2] dans son ouvrage de référence, est plus simple et plus directe que celle utilisant le mystérieux ω.

Une variable (ou vecteur) aléatoire X sera, dans cette approche, *défini(e)* par la donnée des *probabilités* $\mathbb{P}\{X \in A\}$ de sous-ensembles A de $\mathscr{X}$ appelés *événements*. Le fait qu'un résultat satisfait à une propriété donnée correspond à la réalisation d'un certain événement.

3.3 Isoler les cas discret et continu

Si une variable aléatoire ne prend pas, à proprement parler, une ou plusieurs valeurs numériques (par exemple un résultat qualitatif comme pile ou face, ou la couleur du ciel), on peut généralement adopter un codage numérique (comme 0 pour pile et 1 pour face, ou le code RGB d'une couleur). On se ramène ainsi au cas de variables aléatoires *réelles* (en abrégé v.a.r., dans $\mathbb{R}$) ou de vecteurs aléatoires réels [10] (v.a.r. dans $\mathbb{R}^n$). Bien que l'on élimine ainsi le cas d'un espace d'arrivée fonctionnel, cela couvre presque tous les cas donnant lieu à des modélisations pratiques, ce qui est bien suffisant pour démarrer. De la sorte, les mesures considérées sont concrètes (non abstraites) selon la terminologie traditionnelle, on peut même se restreindre à considérer des événements boréliens (la seule tribu à considérer est celle engendrée par les intervalles/pavés) [11]..

Une variable (ou vecteur) aléatoire peut être *discrète* (à valeurs dans un ensemble fini ou infini dénombrable) ou *continue* (à valeurs dans un continuum de valeurs comme un intervalle ou un domaine de l'espace). Il y a aussi d'autres possibilités : variables « dégénérées » ou « mixtes » (discrètes/continues). Dans un premier temps, il est facile de dégager la partie *discrète* d'une v.a. quelconque (ensemble au plus dénombrable de masses ponctuelles dites de Dirac). Il est plus difficile, mais possible (en admettant le théorème de Radon-Nikodym) d'isoler sa partie (absolument) continue (c'est-à-dire définie par une densité). Il reste alors une partie « dégénérée », c'est-à-dire entièrement concentré sur un volume de mesure nulle.

Il est difficile d'imaginer un seul exemple d'une v.a.r. dégénérée utile dans la pratique des mathématiques appliquées. Sa fonction de répartition $P(X \leqslant x)$ serait :

10. Le cas d'une v.a. X *complexe* est un cas particulier d'un couple aléatoire $(\operatorname{Re} X, \operatorname{Im} X) \in \mathbb{R}^2$ où on identifie $\mathbb{C}$ à $\mathbb{R}^2$.

11. Notons au passage que R. Solovay a montré qu'on ne peut obtenir aucun exemple explicite d'ensemble non mesurable sans faire intervenir l'axiome du choix général, et qu'on peut donc parfaitement imposer que tout ensemble et toute fonction est mesurable (au sens de Lebesgue) en restant compatible avec l'axiome du choix dépendant et donc toutes les mathématiques utiles pour les applications. La notion même de mesurabilité n'a alors plus d'intérêt et peut disparaître de l'enseignement.

1° croissante de 0 à 1, comme toute fonction de répartition ;

2° continue partout, car il n'y a pas de masse ponctuelle ;

3° presque partout dérivable de dérivée nulle, puisque concentrée sur un ensemble de mesure nulle (et donc en ce sens presque partout constante) !

On peut donc bien parler de cas pathologique [12]. En résumé, on peut fort bien, sans trop de perte de généralité, se limiter exclusivement aux v.a. discrètes ou continues.

3.4 Tout axer sur les distributions

On peut d'ailleurs adopter un formalisme unifié pour étudier en même temps les propriétés des cas discret et continu, ce qui permet de rendre les calculs généraux plus agréables. Ce formalisme unifié utilise la notion centrale de *distribution de probabilité* $p(x)$: par exemple, la probabilité que $X \in A$ s'écrit sous la forme :

$$\mathbb{P}\{X \in A\} = \sum\!\!\!\!\!\!\!\int_{x \in A} p(x)$$

où la sommation peut être discrète (pour une variable X discrète où $p(x) = \mathbb{P}(X = x)$) ou continue (sommation intégrale pour une variable X continue) ; dans ce dernier cas, $p(x)$ désigne une densité de probabilité. Cette formule est d'un grand intérêt pratique lorsqu'on veut traiter les cas discret et continu ensemble sans devoir tout réécrire deux fois.

Chaque fois que c'est possible, les notions importantes (comme les changements de variable, l'indépendance et le conditionnement) peuvent être exprimées directement sur les distributions de probabilité, plutôt que de faire le détour – comme c'est souvent l'usage – par les fonctions de répartition ou les probabilités d'ensembles. Le formalisme obtenu est plus direct et plus simple.

Par exemple, si l'on tenait à présenter les v.a. comme des applications mesurables définies sur le mystérieux Ω, l'espérance est une intégrale (de Lebesgue-Stieltjes, par exemple) par rapport à la mesure de la variable ω :

$$\mathbb{E}(X) = \int X(\omega)\, dP(\omega).$$

Ici, avec le formalisme basé sur la distribution $p(x)$, l'espérance $\mathbb{E}(X)$ est définie par la formule utile pour les calculs pratiques :

$$\mathbb{E}(X) = \sum\!\!\!\!\!\!\!\int_{x \in \mathscr{X}} x \cdot p(x)$$

12. Un exemple classique est appelé *escalier du diable*. Les v.a. singulières en dimension n sont en revanche importantes à considérer ; cela correspond en pratique à des v.a.r. concentrés sur une variété de dimension $< n$.

(sommation discrète ou intégrale, suivant le cas). Les deux formules sont équivalentes, mais la dernière exprime clairement $\mathbb{E}(X)$ comme une valeur moyenne de X, c'est-à-dire une sommation des valeurs de x pondérées par la distribution de probabilité $p(x)$.

3.5 Tout le reste

On dispose ainsi de tous les outils indispensables pour aborder les grands théorèmes du calcul des probabilités : la *loi (faible) des grands nombres* et la convergence vers la loi gaussienne ou *théorème central limite*.

Avec la même approche, une introduction aux *processus aléatoires* est possible, où l'on expose les notions de *stationnarité* et d'*ergodicité* en liaison avec la loi *forte* des grands nombres [13]. Les notions plus avancées – et plus difficiles, comme celles résultant des différentes définitions de convergence (étroite ou presque sûre), de la construction de Kolmogorov ou des lois du tout ou rien (comme Borel-Cantelli) – sont ainsi exposées à la fin et non au début.

Par ailleurs, l'espace des v.a.r. de carré intégrable (L^2) est naturellement muni d'un produit scalaire, pour lequel c'est un espace de Hilbert ; mais on peut se rassurer : la notion d'espace de Hilbert (essentiellement son caractère complet) n'est utile que pour résoudre des problèmes d'approximation aux moindres carrés et peut être passé sous silence dans une première approche.

Enfin, ne fantasmons pas : les mouvement browniens, les martingales, le calcul différentiel stochastique, l'intégrale de Itô ... ne seront accessibles qu'à un niveau bien supérieur. Les probabilités enseignées en CPGE resteront vraisemblablement des apprentissages fondamentaux faisant suite au lycée (voir annexe A de ce document).

4 Les probabilités : un monde à part ?

L'étude des probabilités passe par l'introduction d'un nouveau langage propre qu'il s'agit d'assimiler. Comme tous les langages d'une spécialité mathématique, celui des probabilités utilise un vocabulaire spécifique souvent constitué de mots courants auxquels on donne un sens précis : *expérience, événement, probabilité, espérance, moments, lois...*

Il n'y a là, me semble-t-il, rien de très différent de ce qui se passe dans d'autres domaines mathématiques, si ce n'est que ce vocabulaire est peut être plus proche du langage courant et donne lieu à des raisonnements plus intuitifs. Qu'on en juge en comparant aux notions classiques de *corps, anneau, idéal,*

13. Une première introduction aux chaînes de Markov est également possible (voir l'exposé de Roger MANSUY).

trace, base, noyau, relèvement, polarisation... qui sont clairement plus éloignées du sens courant ! Ce rapport des probabilités avec le langage humain constitue même une force de la théorie qui est rendue ainsi plus accessible à l'ingénieur pour d'innombrables applications importantes.

C'est aussi ce langage et ces applications qui font la richesse des probabilités et qui la rendent incontournable dans les enseignements. Il faut souligner d'ailleurs qu'actuellement les Grandes Écoles sont un peu pris à la gorge pour « faire passer » en quelques semaines trois années de programme en probabilités [14].

L'absence des probabilités en CPGE suscite en effet depuis très longtemps une incompréhension unanime et notoire des chercheurs, y compris en sciences humaines. Le réel apprentissage des mathématiques de nos étudiants commençant en Sup et s'arrêtant au mieux à la fin de la première année de Grande École, les CPGE sont responsables d'une partie très importante de la formation d'un ingénieur ; raison de plus pour ne pas y négliger l'enseignement des probabilités... Il n'y a en effet aucune raison objective de penser que les commerciaux – qui ont des cours de probabilités en classes préparatoires – soient amenés à utiliser plus souvent la loi des grands nombres que les ingénieurs, bien au contraire !

On peut malgré tout craindre une sorte d'isolement. Les probabilités ne sont-elles pas ce cheveu sur la soupe, ce pan des mathématiques appliquées (pour ne pas dire impures) qui étaient si souvent absentes des traités de mathématiques ? Ne pourrait-on craindre une scission entre probabilités et le reste des mathématiques ?

Tout dépend peut-être de la façon d'enseigner. L'expérience du mathématicien, en tout cas, démontre clairement que les probabilités sont entièrement imbriquées dans les reste des mathématiques : analyse, algèbre, arithmétique, géométrie.... Je présente ci-dessus deux exemples simples dans chacun de ces domaines. Comme l'écrivait Cédric VILLANI dans une conférence LIESSE il y a un an, « les théorèmes ne se mettent pas dans des cases ».

4.1 Deux exemples en analyse

Théorème (Weierstraß). *Toute fonction continue sur un segment est limite uniforme d'une suite de polynômes.*

Démonstration probabiliste (Bernstein, 1912). On se ramène à une fonction f continue sur $[0,1]$; soit X une v.a. binomiale de paramètre p et de longueur n et

14. On peut d'ailleurs en dire autant de l'intégration et de l'analyse de Fourier, ainsi que de l'analyse complexe qui est carrément abandonnée à Télécom ParisTech.

posons

$$B_n(p) = \mathbb{E}\Big(f\big(\tfrac{X}{n}\big)\Big).$$

Clairement $B_n(p) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} f(\frac{k}{n})$ est un polynôme en p. Pour $\varepsilon > 0$ donné, majorons la différence $|B_n(p) - f(p)|$:

$$\begin{aligned}
|B_n(p) - f(p)| &= \Big|\mathbb{E}\Big(f\big(\tfrac{X}{n}\big) - f(p)\Big)\Big| \leqslant \mathbb{E}\Big(\Big|f\big(\tfrac{X}{n}\big) - f(p)\Big|\Big) \\
&= \mathbb{E}\Big(\Big|f\big(\tfrac{X}{n}\big) - f(p)\Big| 1_{\left|\frac{X}{n}-p\right|<\varepsilon}\Big) + \mathbb{E}\Big(\Big|f\big(\tfrac{X}{n}\big) - f(p)\Big| 1_{\left|\frac{X}{n}-p\right|\geqslant\varepsilon}\Big) \\
&\leqslant \max_{\left|\frac{k}{n}-p\right|<\varepsilon} \Big|f\big(\tfrac{k}{n}\big) - f(p)\Big| + 2\|f\|_\infty \mathbb{P}\Big(\Big|\frac{X}{n} - p\Big| \geqslant \varepsilon\Big)
\end{aligned}$$

où on a utilisé le fait que f est bornée. La fonction f étant continue sur le compact $[0,1]$, elle y est uniformément continue, et le premier terme peut être rendu arbitrairement petit, indépendamment de p. Par l'inégalité de Bienaymé-Chebyshev, le deuxième terme est majoré par $2\|f\|_\infty \frac{p(1-p)}{n\varepsilon^2} \leqslant \frac{\|f\|_\infty}{2n\varepsilon^2}$ qui tend uniformément vers zéro. Ainsi $B_n \to f$ uniformément sur $[0,1]$. □

Théorème (formule de Stirling)**.**

$$\frac{n^{n+1/2}}{n!} e^{-n} \xrightarrow{n\to\infty} \frac{1}{\sqrt{2\pi}}.$$

Démonstration probabiliste. Soit $(X_n)_n$ une suite i.i.d. de v.a. poissoniennes de paramètre 1 (égal à la moyenne et la variance). La v.a. $S_n = X_1 + X_2 + \cdots + X_n$ suit une loi de Poisson de paramètre n, et par le théorème central limite, $\frac{S_n - n}{\sqrt{n}}$ converge en loi vers $Y \sim \mathcal{N}(0,1)$. Comme la fonction $x \mapsto x^-$ est continue,

$$\mathbb{E}\Big\{\Big(\frac{S_n - n}{\sqrt{n}}\Big)^-\Big\} \xrightarrow{n\to\infty} \mathbb{E}(Y^-) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} y e^{-y^2/2}\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}}$$

Mais

$$\mathbb{E}\Big\{\Big(\frac{S_n - n}{\sqrt{n}}\Big)^-\Big\} = \sum_{k=0}^{n} \Big(\frac{n-k}{\sqrt{n}}\Big) \frac{n^k}{k!} e^{-n} = \sum_{k=0}^{n} \frac{n^{k+1/2}}{k!} e^{-n} - \sum_{k=1}^{n} \frac{n^{k-1/2}}{(k-1)!} e^{-n}$$

Tous les termes s'éliminent en cascade sauf $\dfrac{n^{n+1/2}}{n!} e^{-n}$. □

4.2 Deux exemples en algèbre

Le premier en combinatoire, l'autre en algèbre linéaire.

Théorème (Sperner, 1928)**.** *On peut trouver au plus*

$$\max_k \binom{n}{k} = \binom{n}{\lfloor n/2 \rfloor}$$

parties de $\{1, 2, \ldots, n\}$ *dont aucune n'est contenue dans aucune autre.*

Démonstration probabiliste. Soit $\mathscr{A}$ une telle famille de parties et $\mathscr{C}_\sigma$ la chaîne :

$$\mathscr{C}_\sigma = \big\{\varnothing, \{\sigma_1\}, \{\sigma_1, \sigma_2\}, \ldots, \{1, 2, \ldots, n\}\big\}$$

où la permutation σ de $\{1, 2, \ldots, n\}$ est choisie au hasard suivant une distribution uniforme. Soit X le nombre de parties de $\mathscr{A}$ dans la chaîne $\mathscr{C}_\sigma$: clairement $X \in \{0, 1\}$ sinon il y aurait au moins deux parties dans $\mathscr{A}$ qui seraient incluses l'une dans l'autre. Or

$$X = |\mathscr{A} \cap \mathscr{C}_\sigma| = \sum_{A \in \mathscr{A}} 1_{A \in \mathscr{C}_\sigma}$$

d'où

$$\mathbb{E}(X) = \sum_{A \in \mathscr{A}} \mathbb{E}(1_{A \in \mathscr{C}_\sigma}) = \sum_{A \in \mathscr{A}} \mathbb{P}(A \in \mathscr{C}_\sigma)$$

et puisque $\mathscr{C}_\sigma$ ne contient qu'un seul ensemble de cardinal donné $|A|$, qui est choisi au hasard parmi toutes les $\binom{n}{|A|}$ parties de même cardinal $= |A|$,

$$\mathbb{P}(A \in \mathscr{C}_\sigma) = \frac{1}{\binom{n}{|A|}}.$$

Finalement, puisque $X \leqslant 1$,

$$\frac{|\mathscr{A}|}{\max_k \binom{n}{k}} \leqslant \sum_{A \in \mathscr{A}} \frac{1}{\binom{n}{|A|}} = \mathbb{E}(X) \leqslant 1. \quad \square$$

Théorème (Ky Fan, 1950)**.** *Pour toutes matrices* $A, B \in S_n^{++}(\mathbb{R})$ *(symétriques définies positives) et tous réels positifs* λ, μ *tels que* $\lambda + \mu = 1$,

$$\det(\lambda A + \mu B) \geqslant (\det A)^\lambda (\det B)^\mu.$$

Démonstration probabiliste (Cover, Thomas, 1988). Soit X_0 et X_1 deux vecteurs gaussiens centrées de matrices de covariance respectives A et B, Θ une v.a. binaire à valeurs dans $\{0, 1\}$, indépendante de (X_0, X_1), de loi de Bernoulli (λ, μ). La matrice de covariance du vecteur aléatoire $Y = X_\Theta$ est

$$\mathbb{E}({}^t Y Y) = \mathbb{E}\big\{\mathbb{E}({}^t X_\Theta X_\Theta | \Theta)\big\} = \lambda \mathbb{E}({}^t X_0 X_0) + \mu \mathbb{E}({}^t X_1 X_1) = \lambda A + \mu B.$$

On utilise des résultats sur l'entropie de vecteurs aléatoires définis par des densités [3]. À matrice de covariance fixée, l'entropie est maximale pour une loi normale :

$$h(Y) \leqslant h(Z)$$

où Z suit une loi normale de matrice de covariance $= \lambda A + \mu B$. Par ailleurs, en appliquant l'inégalité de l'information :

$$h(Y) \geqslant h(Y|\Theta) = \lambda h(X_0) + \mu h(X_1).$$

L'entropie d'un vecteur gaussien X de densité f_X et de matrice de covariance C se calcule explicitement :

$$h(X) = -\int f_X \ln f_X = \tfrac{1}{2}\mathbb{E}({}^t X C^{-1} X) + \ln\sqrt{(2\pi)^n \det C} = \frac{n}{2}\ln(2\pi e) + \frac{1}{2}\ln\det C.$$

On remplace alors les entropies par leurs valeurs dans l'inégalité $\lambda h(X_0) + \mu h(X_1) \leqslant h(Z)$. □

4.3 Deux exemples en arithmétique

Commençons par un résultat célèbre de Hardy et Ramanujan, qui établit que l'ordre de grandeur de $\omega(n)$, le nombre de diviseurs premiers distincts de n, est $\ln\ln n$. Ainsi, un nombre choisi au hasard entre 1000 et 500000000 n'aura usuellement que 2 ou 3 facteurs premiers distincts [15].

Théorème (Hardy & Ramanujan, 1920)**.** *Pour tout* $\varepsilon > 0$*,*

$$\frac{1}{n}\left|\left\{N \leqslant n\,;\ (1-\varepsilon)\ln\ln N < \omega(N) < (1+\varepsilon)\ln\ln N\right\}\right| \xrightarrow{n\to\infty} 1.$$

Hardy avait apparemment une sainte horreur des probabilités (jugées trop vagues) et pour les mathématiques appliquées en general. Mais Turan a grandement simplifié la preuve de Hardy-Ramanujan en interprétant leur résultat comme l'énoncé d'une loi des grands nombres (il existe d'ailleurs un théorème central limite qui affine le résultat, étudié par Erdős et Kac en 1939). Hardy et Wright, dans leur célèbre traité de théorie des nombres, ont repris la preuve de Turan ci-dessous... en prenant soin d'effacer toute trace des probabilités !

15. Hardy aimait à citer l'anecdote suivante sur Ramanujan : *« I remember once going to see [Ramanujan] when he was lying ill at Putney. I had ridden in taxi cab number 1729 and remarked that the number seemed to me rather a dull one, and that I hoped it was not an unfavorable omen. "No, Hardy, no Hardy," he replied, "it is a very interesting number; it is the smallest number expressible as the sum of two cubes in two different ways. »* Selon Hardy, un nombre tel que $1729 = 7\cdot 13\cdot 19$, $\omega(1729) = 3$, est assez typique et donc « terne ». La réponse de Ramanujan est d'un tout autre registre : $1729 = 1^3 + 12^3 = 9^3 + 10^3$...

Démonstration probabiliste (Turan, 1934). Soit N une variable aléatoire entière uniforme sur $\{1,2,\dots,n\}$ où $n \geqslant 2$. Sa factorisation en nombre premiers s'écrit : $N = \prod_p p^{X_p}$, et on a $\omega(N) = \sum_{p\leqslant n} X_p^*$ où $X_p^* = \min(X_p, 1)$. Puisque qu'il y a $\lfloor \frac{n}{p} \rfloor$ multiples de p inférieurs ou égaux à n, X_p^* suit une loi de Bernoulli de paramètre $P(X_p^* = 1) = P(X_p \geqslant 1) = \frac{1}{n}\lfloor \frac{n}{p} \rfloor$. Sachant que [16] $\sum_{p\leqslant n} \frac{1}{p} = \ln\ln n + O(1)$, il vient $E(\omega(N)) = \sum_p E(X_p^*) = \sum_p \frac{1}{n}\lfloor \frac{n}{p} \rfloor = \ln\ln n + O(1)$, car enlever les crochets revient à commettre une erreur $\leqslant \frac{1}{n}\sum_{p\leqslant n} 1 \leqslant 1$. De plus, grâce à l'encadrement $\frac{1}{k} - \frac{1}{n} \leqslant \frac{1}{n}\lfloor \frac{n}{k} \rfloor \leqslant \frac{1}{k}$:

$$\begin{aligned}
\mathrm{Var}(\omega(N)) &= \sum_{p\leqslant n} E(X_p^*) - E^2(X_p^*) + \sum_{\substack{p,p'\leqslant n\\ p\neq p'}} E(X_p^* X_{p'}^*) - E(X_p^*)E(X_{p'}^*) \\
&= \sum_{p\leqslant n} \frac{1}{n}\left\lfloor \frac{n}{p} \right\rfloor - \left(\frac{1}{n}\left\lfloor \frac{n}{p} \right\rfloor\right)^2 + \sum_{\substack{p,p'\leqslant n\\ p\neq p'}} \frac{1}{n}\left\lfloor \frac{n}{pp'} \right\rfloor - \frac{1}{n}\left\lfloor \frac{n}{p} \right\rfloor \frac{1}{n}\left\lfloor \frac{n}{p'} \right\rfloor \\
&\leqslant \sum_{p\leqslant n} \frac{1}{p} + \sum_{p,p'\leqslant n} \frac{1}{pp'} - \left(\frac{1}{p} - \frac{1}{n}\right)\left(\frac{1}{p'} - \frac{1}{n}\right) = O(\ln\ln n).
\end{aligned}$$

Par l'inégalité de Bienaymé-Chebyshev

$$\mathbb{P}\left(\left|\frac{\omega(N) - E(\omega(N))}{\ln\ln n}\right| \geqslant \varepsilon\right) \leqslant \frac{\mathrm{Var}(\omega(N))}{(\ln\ln n)^2\varepsilon^2} \xrightarrow{n\to\infty} 0$$

Il est facile de voir qu'on peut remplacer $E(\omega(N)$ par $\ln\ln n$, et même par $\ln\ln N$, car $\ln\ln n - 1 < \ln\ln N \leqslant \ln\ln n$ pour tout $N > n^{1/e}$, le reste donnant une contribution $\leqslant \mathbb{P}(N \leqslant n^{1/e}) \to 0$. □

En 1948, Shannon utilise un argument de moyenne d'ensemble pour prouver l'existence (sans preuve constructive) d'au moins un « bon » code correcteur d'erreurs pour les communications numériques arbitrairement fiables [3]. Le même argument peut être utilisé pour prouver des résultats plus élémentaires. Par exemple :

Théorème (Erdős, 1965)**.** *Tout ensemble $A = \{a_1, a_2, \dots, a_n\}$ de n entiers non nuls contient $> n/3$ entiers $\{a_{i_k}\}_k$ tels que $a_{i_k} + a_{i_l} \neq a_{i_m}$ pour tous k, l, m.*

Démonstration probabiliste. Soit p un nombre premier de la forme $p = 3k+2$, plus grand [17] que tous les $2|a_i|$ et soit X un v.a. uniforme à valeurs dans $\mathbb{Z}_p^*$

16. Voir [1] pour une preuve « probabiliste ».

17. Il y en a bien une infinité de nombres premiers de la forme $3k+2$: car s'il n'y en avait qu'un nombre fini $p_1, p_2, \dots, p_N$, le nombre impair $M = 6p_1p_2\cdots p_N - 1$ n'aurait que des facteurs premiers impairs distincts des p_i, donc de la forme $3k+1$, d'où $M = -1 \bmod 3 = 1 \bmod 3$, ce qui est impossible.

(entiers non nuls modulo p). Puisque $a_i \neq 0 \bmod p$, la v.a. $X_i = a_i \cdot X \bmod p$ à valeurs dans $\mathbb{Z}_p^*$ suit aussi une loi uniforme. Soit **B** le sous-ensemble aléatoire de A constitué des entiers a_i tels que $X_i \in \{k+1, k+2, \dots, 2k+1\}$. Sa taille moyenne est

$$\mathbb{E}(|\mathbf{B}|) = \mathbb{E}\Big(\sum_{i=1}^{n} 1_{k<X_i<=2k+1}\Big) = \sum_{i=1}^{n} \mathbb{P}(k < X_i <= 2k+1) = \sum_{i=1}^{n} \frac{k+1}{p-1} = \frac{n(k+1)}{3k+1} > \frac{n}{3}$$

donc il existe au moins une valeur $X = x$ conduisant à un ensemble B de taille $> n/3$. Les éléments $a_i \in B$ sont ceux pour lequels $a_i x \bmod p \in \{k+1, k+2, \dots, 2k+1\}$. Si on avait $a_{i_k} + a_{i_l} = a_{i_m}$ pour trois entiers de B, en multipliant par x modulo p on trouverait deux entiers $\in \{k+1, k+2, \dots, 2k+1\}$ de somme (modulo $3k+2$) égale à un autre entier $\in \{k+1, k+2, \dots, 2k+1\}$, ce qui est impossible. □

4.4 Deux exemples en géométrie

Théorème (Gram, 1874)**.** *Soit α_i ($0 \leqslant i \leqslant n$) la somme des angles solides intérieurs des i-faces*[18] *d'un polyhèdre convexe en n dimensions. Alors*

$$\sum_{i=0}^{n} (-1)^i \alpha_i = 0.$$

En particulier ($n = 2$), on retrouve que la somme des angles d'un polygone à N côtés dans le plan est $\alpha_0 = \alpha_1 - \alpha_2 = \frac{N}{2} - 1$, soit $(N-2)\pi$ radians.

Démonstration probabiliste (Welzl, 1994). Soit P le polyhèdre en question, f_i son nombre de i-faces pour $i = 0$ à n (ainsi $f_n = 1$ et $\alpha_{n-1} = f_{n-1}/2$).

Soit H un hyperplan dont la direction est choisie au hasard (le vecteur normal à H est choisi comme un point de la sphère S^n selon une distribution uniforme) et considérons le projeté orthogonal du polyhèdre P sur H ; c'est un autre polyhèdre P' à $n-1$ dimensions. Soit F_i' son nombre de i-faces ($i = 0$ à $n-1$). La loi de la direction de projection étant uniforme, la probabilité qu'une i-face de P ($i < n-1$) ne soit *pas* projetée sur une i-face de P' est égale au double de l'angle solide intérieur. En sommant les probabilités complémentaires pour chacune des f_i i-faces de P, on trouve le nombre moyen de i-faces de P' :

$$\mathbb{E}(F_i') = f_i - 2\alpha_i \qquad (i \leqslant n-2)$$

18. Les 0-faces sont les sommets, les 1-faces sont les arêtes, etc. ; la n-face est l'hyper-volume du polyhèdre. L'angle solide est supposé normalisé à 1 (mesure de l'angle total). Ainsi $\alpha_n = 1$, et α_{n-1} est égal à la moitié du nombre d'hyperplans qui délimitent le polyhèdre.

La relation d'Euler-Poincaré pour P' s'écrit $\sum_{i=0}^{n-1}(-1)^i F'_i = 1$ et celle pour P s'écrit $\sum_{i=0}^{n}(-1)^i f_i = 1$. Puisque $F'_{n-1} = 1$, $f_n = 1$ et $\alpha_{n-1} = f_{n-1}/2$:

$$\begin{aligned} 0 &= \sum_{i=0}^{n-1}(-1)^i \mathbb{E}(F'_i) - 1 = \sum_{i=0}^{n-2}(-1)^i (f_i - 2\alpha_i) + (-1)^{n-1} - 1 \\ &= \sum_{i=0}^{n-1}(-1)^i (f_i - 2\alpha_i) + (-1)^{n-1} - \sum_{i=0}^{n}(-1)^i f_i = -2\sum_{i=0}^{n-1}(-1)^i \alpha_i. \end{aligned}$$ □

Théorème (Zubkov, 1979). *Dans l'espace* $\mathbb{R}^n$ *de base canonique* $(e_1, e_2, \dots, e_n)$*, soit* $C = [0,1]^n$ *l'hypercube unité de* $\mathbb{R}^n$*,* H_t *l'hyperplan d'équation* $x_1 + x_2 + \cdots + x_n = t$*, et* S *le simplexe défini par l'enveloppe convexe des* $n+1$ *points* 0 *et* $e_1, 2e_2, \dots ne_n$*. Alors les volumes* $(n-1)$*-dimensionnels suivants sont égaux :*

$$\mathrm{vol}(C \cap H_t) = \mathrm{vol}(S \cap H_t).$$

Démonstration probabiliste. Soit $X_1, X_2, \dots, X_n$ des v.a. i.i.d. exponentielles de paramètre $\lambda > 0$, réordonnons les X_i par ordre croissant : $X_{(1)} \leqslant X_{(2)} \leqslant \cdots \leqslant X_{(n)}$ et posons $(Y_1, Y_2, \dots, Y_n) = (X_{(1)}, X_{(2)} - X_{(1)}, \dots, X_{(n)} - X_{(n-1)})$. La loi exponentielle étant sans mémoire : $\mathbb{P}(X_i \geqslant t + x | X_i \geqslant t) = \mathbb{P}(X_i \geqslant x)$, le vecteur $(X_{(2)} - X_{(1)}, X_{(3)} - X_{(1)}, \dots, X_{(n)} - X_{(1)})$ suit la même loi que le vecteur ordonné de $n-1$ v.a. exponentielles i.i.d. de paramètre λ, indépendamment de $X_{(1)}$. De proche en proche, on voit que Y_i est indépendant de $Y_1, Y_2, \dots, Y_{i-1}$ et que $\mathbb{P}(Y_i \geqslant y) = \mathbb{P}(X_{(i)} \geqslant X_{(i-1)} + y) = \mathbb{P}(\min X_j \geqslant y)$ où le minimum porte sur les $n-i+1$ v.a. exponentielles i.i.d. restantes de paramètre λ, d'où $\mathbb{P}(Y_i \geqslant y) = e^{-(n-i+1)\lambda y}$: $(Y_1, Y_2, \dots, Y_n)$ suit donc la même loi que $(\frac{X_n}{n}, \dots, \frac{X_2}{2}, X_1)$ et

$$\mathbb{P}(\forall i, X_i \leqslant 1) = \mathbb{P}(X_{(n)} \leqslant 1) = \mathbb{P}(Y_1 + Y_2 + \cdots + Y_n \leqslant 1) = \mathbb{P}\Big(\frac{X_n}{n} + \frac{X_{n-1}}{n-1} + \cdots + X_1 \leqslant 1\Big)$$

Cela se réécrit (en divisant par λ^n) :

$$\int_C e^{-\lambda(x_1 + \cdots x_n)}\,\mathrm{d}x_1 \cdots \mathrm{d}x_n = \int_S e^{-\lambda(x_1 + \cdots x_n)}\,\mathrm{d}x_1 \cdots \mathrm{d}x_n$$

ou encore

$$\int_0^{+\infty} \mathrm{vol}(C \cap H_t) e^{-\lambda t}\,\mathrm{d}t = \int_0^{+\infty} \mathrm{vol}(S \cap H_t) e^{-\lambda t}\,\mathrm{d}t$$

pour tout $\lambda > 0$, ce qui montre que la transformée de Laplace des deux volumes (qui sont clairement des fonctions de t continues à support compact $\subset \mathbb{R}_+$) sont égales. On conlut par injectivité de la transformée de Laplace [19]. □

19. La transformée de Laplace de f est $F(\lambda) = \int_0^\infty f(t) e^{-\lambda t}\,\mathrm{d}t$ ($\lambda > 0$). Il existe des preuves élémentaires que cette transformée est injective pour des fonctions f continues à support compact : Par exemple, si $F = 0$, on obtient par dérivation que les moments de f sont tous nuls, donc $\int Pf = 0$ et $\int f^2 = \int f(f - P)$ pour tout polynôme P ; le théorème d'approximation de Weierstrass permet alors de montrer que $f = 0$.

5 Les probabilités : ennuyeuses et trop abstraites ?

Les souvenirs de lycée en probabilités laissent parfois un sentiment d'ennui. On se souvient de raisonnements flous sur les cartes à jouer, les tirages de boules dans les urnes et autres types de dénombrements pas très joyeux : le titre « *Les probabilités sans les boules* » d'un recueil d'exercices de Terminale par Gérard FRUGIER est symptomatique à cet égard. Peut-être que les arbres pondérés préconisés aujourd'hui par les réformateurs des programmes pourraient avoir plus d'intérêt (on pense par exemple à leur utilisation pour des procédés de codage comme l'algorithme de Huffman).

Il y a également un sentiment d'inachevé lorsqu'on se souvient des raisonnements de pure statistique descriptive (comme les écarts-type, inter-quartiles et autres rectangles à moustache) pour introduire l'« intuition » des probabilités dans le Secondaire. Il ressort une grande impression de pauvreté au regard de la vraie théorie. Cela pose aussi le problème de l'enseignement des *statistiques* : doivent-elles être enseignés avant, après (comme une continuation naturelle) ou en même temps que la théorie des probabilités ? Quelle place accorder au problèmes pratiques de modélisation chez les étudiants ? Réciproquement, les probabilités enseignées servent-elles vraiment à faire comprendre des notions pratiques de modélisation mathématique ?

Tout semble être une question d'équilibre entre la théorie et les applications. Même à un niveau élémentaire, les probabilités peuvent être bien appliquées : des simulations, par exemple à l'aide du logiciel `scilab` accompagnent régulièrement les cours de première année d'Écoles d'ingénieurs [20]. Par la suite, certains problèmes délicats de simulation (à commencer par les méthodes de type « Monte Carlo ») nécessitent déjà un bon bagage théorique en Probabilités [21]. Par ailleurs, il est toujours possible de rendre la théorie des probabilités très abstraite sans référence apparente avec la réalité ou les simulations : mais cela se fait généralement à un niveau bien supérieur.

Il est vrai que le calcul numérique a peu ou prou disparu des épreuves écrites de concours (à cause des progrès technologiques et de la peur des tricheries) et on peut le regretter. Il reste quand même les TIPE dont le rôle formateur est important et n'est pas remis en question. Les probabilités pourront donc aider le futur ingénieur à comprendre les principes de la modélisation mathématique du monde qui nous entoure – ce serait un comble si elles ne le permettaient pas !

Il faudra bien entendu conserver à l'esprit le fait que l'enseignement des probabilités doit être utile aux sciences physiques : les probabilités ne se retrouvent évidemment pas qu'en algèbre linéaire ou en calcul différentiel et intégral, mais

20. Voir par exemple le témoignage exposé de Sylvie MÉLÉARD à l'Ecole Polytechnique.
21. Voir l'exposé d'Eric MOULINES.

aussi dans de nombreuses théories physiques [22]. Tout comme on peut accepter que, comme pour le reste du programme en mathématiques, les probabilités sont également riches de beaux raisonnements, et contribuent à leur manière à l'apprentissage de la logique, du raisonnement, de la démontration, et de la rigueur.

6 Les probabilités : probable inflation aux concours ?

Certains professeurs de classes préparatoires ont, semble-t-il, une sainte frayeur des probabilités aux concours. On craint une dérive inflationniste qui rendrait omniprésente les probabilités abstraites dans toutes les épreuves de mathématiques [23].

Cependant, si l'on regarde de près les épreuves actuelles des concours, on peut s'apercevoir que les probabilités y sont déjà présentes, bien que cachées. En ce qui concerne le concours que je connais le mieux (le concours commun Mines-Ponts), on y trouve dans les années récentes :

- une allusion à un théorème ergodique (épreuve filière PC, 2011)
- une preuve du théorème central limite (épreuve PC/PSI, 2010)
- le problème des moments et la loi log-normale (épreuve MP, 2009)
- des matrices aléatoires (épreuve PSI, 2009)
- des matrices stochastiques (épreuve PC/PSI 2007, MP 2006)
- des séries génératrices de v.a. entières (loi de Poisson) (épreuve PC/PSI 2004)
- …

sans compter le nombre important de sujets de TIPE qui se basent d'ores et déjà sur des notions de probabilités.

Une des raisons probables de cette présence « cachée » est le besoin important des probabilités dans la plupart des Écoles d'Ingénieurs. De ce point de vue, introduire explicitement les probabilités serait un moyen d'arrêter l'hypocrisie. Par exemple, si un programme raisonnable en probabilités existait en classes préparatoires, il est certain que le problème des moments et la loi log-normale de l'épreuve MP de 2009 serait rendu caduc ou traité en moins d'un quart d'heure [24].

De plus, les probabilités sont parfois, comme on l'a vu ci-dessus, des outils très utiles à d'autres domaines mathématiques (analyse, algèbre, géométrie,

22. Voir l'exposé d'Alain MARUANI.

23. Ce serait également la raison pour laquelle il serait envisagé d'éviter d'enseigner un minimum de théorie de la mesure en CPGE (voir l'exposé de Laurent DECREUSEFOND).

24. C'est à peine le temps qu'il faut pour traiter du même problème en exercice de première année à l'Ecole Polytechnique.

arithmétique...). Il est donc parfois possible de gagner du temps grâce à elles !

7 En guise de conclusion

L'introduction des probabilités dans les programmes des classes préparatoires aux grandes écoles semble acquise. J'espère que la réflexion exposée ici montre que cela n'apporte pas que des inconvénients, surtout au regard des besoins du futur ingénieur : il faudra donc s'adapter bon gré mal gré. On peut même souhaiter des influences favorables pour les autres domaines enseignées, en mathématiques ou dans les sciences physiques. Par ailleurs, le nouveau mode de raisonnement introduit enrichit nécessairement la palette du candidat aux concours.

Le principal problème pratique des enseignants sera leur formation, dès 2012-2013. Les Grandes Écoles seront naturellement mises à contribution par le biais de stages LIESSE. Les manuels actuels posent également problème : ceux qui ne sont pas simplement d'un niveau très basique (BAC ou prépas commerciales) apparaissent parfois trop abstraites au niveau master/Grandes Écoles. Des références intermédiaires sont souhaitables (c'est une des raisons qui m'ont poussé à rédiger le livre [4] à partir de notes de cours).

Références

[1] P. Billingsley, *Probability and measure*, J. Wiley & Sons, 1995.

[2] H. Cramér, *Mathematical Methods of Statistics*, Princeton Univ. Press, 1946 (réimpression 1999).

[3] O. Rioul, *Théorie de l'information et du codage*, Hermes-Science Lavoisier, 2007.

[4] O. Rioul, *Théorie des probabilités*, Hermes-Science Lavoisier, 2008.

Annexe

A Demandez le programme

A.1 Première S (septembre 2011)

- Intuition basée sur la « statistique descriptive » et des simulations sur logiciel exploitant la moyenne et écart-type, médiane et écart inter-quartile d'une série statistique.
- Approche heuristique de la loi des grands nombres pour interpréter la moyenne et la variance d'une loi de probabilité discrète dans le cas d'un grand nombre N de répétitions d'expériences « identiques et indépendantes ».
- Lois discrètes à nombre fini de possibilités, surtout la loi binaire (schéma Bernoulli) menant par répéition à des lois binomiales *via* des raisonnements sur des « arbres pondérés » ; de manière annexe, on voit la loi géométrique (*tronquée* !).
- Intervalle de « fluctuation » pour réaliser un test d'hypothèse (sans le dire).

A.2 Terminale S (septembre 2012)

Le programme est beaucoup plus ambitieux :

- Le raisonnement sur les arbres pondérés est là encore privilégié et étendu aux probabilités conditionnelles $\mathbb{P}(A|B)$ noté $P_B(A)$, événements indépendants, formule des probabilités totales, simulation de marche aléatoire.
- Et surtout, les lois à densité sont introduites (à partir de la notion intuitive d'aire pour définir les intégrales), pour des densités continues sur un intervalle borné $[a, b]$. On définit à cette occasion une v.a.r. X comme une fonction de l'univers Ω dans $\mathbb{R}$, définissant une probabilité par la formule $\mathbb{P}(X(\omega) \in J) = \int_J p(x)\,\mathrm{d}x$, d'espérance $\mathbb{E}(X) = \int_a^b x p(x)\,\mathrm{d}x$.
- Exemples : loi uniforme sur $[a, b]$, loi exponentielle avec propriété de loi sans mémoire, loi normale , ces deux exemples étant pourtant sur des intervalle non bornés.
- Méthode de Monte-Carlo, calculs admis d'intégrales donnant les moyenne et variance.
- Convergence de la loi binomiale vers la loi de Gauss (Théorème de Moivre) et intervalles de confiance, notamment à 5% et 1%, règles des σ, 2σ et 3σ. On insiste bien sur les distinctions entre intervalles de fluctuation et de confiance (l'intervalle de confiance est aléatoire).

A.3 Maths Sup. (septembre 2013)

… ?…

A.4 Maths Spé. (septembre 2014)

… ?…

A.5 Programme typique en Grande École

Ce qui est prévisible en CPGE constitue tout ou partie du programme suivant (en **gras** les choses plus difficiles qui resteront probablement aux Grandes Ecoles) :

- Probabilité sur un espace dénombrable, loi d'une v.a. (probabilité image)
- Conditionnement et indépendance ensembliste, formule de Poincaré
- Lois de Bernoulli, binomiale, géométrique, de Poisson, etc. Lois conditionnelles, v.a. indépendantes et leur somme.
- Moments et série génératrice. **Entropie**.
- **Existence et unicité d'une mesure de probabilité sur une tribu borélienne (voire jusqu'à la théorie de la mesure de Lebesgue)**
- Fonctions de répartition, lois à densité, simulation par inversion de la fonction de répartition (voire méthode du rejet).
- Espérance **(en lien avec la notion d'intégrale à rapport à une mesure)**
- Variables aléatoires intégrables et de carré intégrable, variance et covariance, corrélation. Régression linéaire.
- Lois uniforme, exponentielle, normale, gamma, de Cauchy, Bêta, de Pareto, etc.
- Inégalités de Bienaymé-Chebyshev, Cauchy-Schwarz, **Jensen, Hölder, Minkowski...**
- **Espérance conditionnelle : cas discret ou cas à densité, voire jusqu'au conditionnement par rapport à une tribu. Méthode des moindres carrés.**
- Vecteurs aléatoires (lien avec Fubini-Tonelli), matrice de covariance, vecteurs gaussiens
- **Densités conditionnelles, produits de convolution.**
- Calcul de loi et recherche de densité par la méthode de la fonction de répartition ou par le théorème de la loi image.
- **Espace probabilisé produit infini pour modéliser des expériences indépendantes en nombre infini, les deux lemmes de Borel-Cantelli.**
- Convergence en loi, en probabilité, en moyenne (en relation avec le théorème de convergence dominée), presque sûre
- Loi faible **et forte** des grands nombres, méthode de Monte-Carlo
- **Fonctions caractéristiques ou transformée de Laplace, théorème de Paul**

Lévy.

– Théorème central limite, intervalles de confiance
– **Introduction aux marches aléatoires, processus de branchement, files d'attente, chaînes (discrètes) de Markov.**

Pour les Grandes Écoles elles-même, il y aura nécessité de réorganisation complète de leur enseignement dans le domaine en **septembre 2015**. Les Grandes Écoles (conformément à leur souhait dans la plupart des cas) conserveront leurs enseignements actuels de première année mais bénéficient de bases avancées leur permettant d'aller plus vite et plus loin (lignes en gras ci-dessus).

Eric Moulines

Professeur à Télécom ParisTech, médaille d'argent CNRS

Chaînes de Markov finies et simulation

La simulation est un sujet extrêmement important en pratique. Dans de nombreuses situations, on est amené à simuler des variables dans des espaces de très grande dimension (de quelques centaines à des millions). Pour résoudre de tels problèmes, les méthodes d'échantillonnage par chaînes de Markov s'avèrent indispensables. Nous décrirons dans cet exposé une façon de construire des chaînes permettant d'échantillonner des lois dans des espaces de dimension quelconque, l'algorithme de Metropolis-Hastings. Nous verrons le rôle joué par la réversibilité des chaînes (caractère autoadjoint dans l'espace L2 de la loi stationnaire). Nous verrons comment les propriétés spectrales permettent d'approcher les vitesses de convergence.

Présenté par **Gersende Fort**

Directeur de recherche au CNRS LTCI

Méthodes de Monte Carlo par Chaînes de Markov

Eric MOULINES

Telecom Paris Tech
CNRS - LTCI

présenté par Gersende FORT (LTCI, CNRS & Telecom Paris Tech)

collaborations : S. Le Corff (PhD, Telecom ParisTech) et P. Priouret (Professeur, Paris 6).

I. Un problème de restauration d'images

Modèle d'Ising : énergie

- Un **système de spins** est une distribution de probabilité sur $\mathsf{X} = \{0,1\}^{|V|}$, où V est l'ensemble des sommets d'un graphe $G = (V, E)$. E est l'ensemble des arêtes du graphe d'adjacence. La valeur $x_v \in \{0,1\}$ est le **spin** au sommet v.
- Le **modèle d'Ising** est le modèle de spin le plus couramment étudié. Dans ce système, l'**énergie** d'une configuration $\boldsymbol{x}$ est définie par

$$H(\boldsymbol{x}) = -J \sum_{v,w \in V, v \sim w} \mathbb{1}_{\{x_v = x_w\}} - \sum_{v \in V} \left\{ h_{v,1} \mathbb{1}_{\{1\}}(x_v) + h_{v,0} \mathbb{1}_{\{0\}}(x_v) \right\}$$

où $v \sim w$ signifie que $(v,w) \in E$ et $\{(h_{v,1}, h_{v,0})\}_{v \in V}$ est le champ externe, supposé anisotrope
- Nous considérons dans la suite le cas **ferromagnétique** où $J > 0$. Dans ce modèle, l'énergie augmente comme le nombre de paires (v,w) de sites voisins dont les spins sont différents...

Modèle d'Ising : distribution de Gibbs

- La **distribution de Gibbs** associée à l'énergie H est la distribution π sur X définie par

$$\pi(\boldsymbol{x}) = \frac{1}{Z(\beta)} \mathrm{e}^{-\beta H(\boldsymbol{x})}, \qquad (\beta \geq 0) .$$

- $Z(\beta)$, la **fonction de partition** est la constante de normalisation,

$$Z(\beta) \stackrel{\text{def}}{=} \sum_{\boldsymbol{x} \in \mathsf{X}} \mathrm{e}^{-\beta H(\boldsymbol{x})} .$$

- β est l'inverse de la température.
  - A température infinie ($\beta = 0$), la fonction d'énergie H ne joue aucun rôle et π est la loi uniforme sur X.
  - Lorsque la température diminue, la distribution rend plus **probable** les configurations de faibles énergies.

Modèle d'Ising: influence de la température

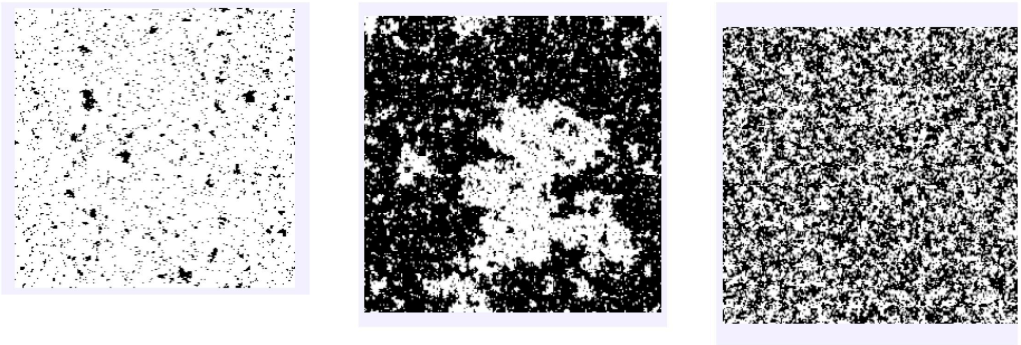


Fig.: Température croissante

Transition de phase

- Le cas le plus souvent utilisé en pratique est le cas d'un cristal ferromagnétique invariant par translation sur un réseau d-dimensionnel
- Dans ses travaux de doctorat, Ising (1924) a étudié le modèle en dimension 1. Dans ce cas, la corrélation $\mathrm{Cov}_\pi(X_v,X_w)$ décroît exponentiellement vite avec la distance des sites $|v-w|$, indépendamment de la température.
- En dimension 2, Peierls (1930) a montré l'existence d'une transition de phase entre une phase désordonnée et une phase ordonnée. Le système est désordonné pour des petites valeurs de β et ordonné pour de grandes valeurs de β, au sens où $\mathrm{Cov}_\pi(X_v,X_w) \geq C > 0$ pour tout $(v,w) \in V \times V$.

Modèle de Potts

- Le modèle de Potts est une généralisation naturelle du modèle d'Ising au cas où $\mathsf{X} = \{1,2,\ldots,Q\}^{|V|}$.
- La fonction d'énergie est alors donnée par

$$H(\boldsymbol{x}) = -J \sum_{v,w \in V, v \sim w} \mathbb{1}_{\{x_v = x_w\}} + \sum_{v \in V} \sum_{j=1}^{Q} h_{v,j} \mathbb{1}_{\{j\}}(x_v) \ .$$

  L'énergie diminue lorsque deux voisins ont la même valeur. $h_{v,j}$ donne l'influence d'un champ externe.
- La distribution de Gibbs associée au modèle de Potts est donnée par

$$\pi(\boldsymbol{x}) = \frac{1}{Z(\beta)} \mathrm{e}^{-\beta H(\boldsymbol{x})} \ ,$$

  où $Z(\beta)$ est la **fonction de partition** $\beta \geq 0$.

Modèle élémentaire d'image bruitée

- $\boldsymbol{Y} = \{Y_v\}$ des observés sur un réseau régulier V à valeurs dans $\{0,1\}^{|V|}$.
- On modélise $Y_v = X_v \oplus Z_v$ où
  - (i) $\{Z_v\}_{v \in V}$ est une famille indépendante de variables de même loi de Bernoulli.
  - (ii) Le bruit $\{Z_v\}_{v \in V}$ est indépendant de l'image $\{X_v\}_{v \in V}$.
  - (iii) $\{X_v\}_{v \in V}$ est un modèle d'Ising ferromagnétique (sans champ externe).
  - (iv) $\oplus$ est l'addition modulo 2.

$$p(\boldsymbol{y} \mid \boldsymbol{x}) = \frac{1}{Z_1(h)} \exp\left(-h \sum_{v \in V} \mathbb{1}_{\{y_v = x_v\}}\right)$$

$$p(\boldsymbol{x}) = \frac{1}{Z_2(\beta)} \exp\left(-\beta \sum_{v,w \in V, v \sim w} \mathbb{1}_{\{x_v = x_w\}}\right)$$

Restauration Bayésienne

$$\begin{aligned} p(\boldsymbol{x}|\boldsymbol{y}) &= \frac{p(\boldsymbol{x},\boldsymbol{y})}{p(\boldsymbol{y})} \\ &= \frac{p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})}{p(\boldsymbol{y})} \\ &\propto \exp\left(-\beta \sum_{v,w\in V, v\sim w} \mathbb{1}_{\{x_v=x_w\}} - h\sum_{v\in V} \mathbb{1}_{\{y_v=x_v\}}\right) \end{aligned}$$

la loi de l'image **non bruitée** conditionnellement à l'image **bruitée** est un modèle d'Ising avec un champ anisotrope, **dépendant de l'image bruitée**.

Objectif de la restauration d'image

- **Observation:** une réalisation $\{\boldsymbol{y}_v\}_{v\in V}$ d'une image entachée d'un bruit additif.
- **Objectif:** Calculer
  1. La moyenne a posteriori de l'image non bruitée conditionnellement à l'image bruitée
  $$\mathbb{E}\left[\boldsymbol{X} \,|\, \boldsymbol{Y} = \boldsymbol{y}\right] = \sum_{\boldsymbol{x}} \boldsymbol{x} p(\boldsymbol{x} \mid \boldsymbol{y})$$
  2. Le maximum a posteriori
  $$\hat{\boldsymbol{x}} = \operatorname{argmax}_{\boldsymbol{x}} p(\boldsymbol{x} \mid \boldsymbol{y}) \; .$$

II. Méthodes de Monte Carlo par Chaînes de Markov (MCMC)

Méthode de Monte Carlo par Chaînes de Markov (MCMC)

- La méthode MCMC consiste à construire une chaîne de Markov $\{X_n, n \in \mathbb{N}\}$ de noyau de transition P dont π soit la loi stationnaire.
- Les chaînes de Markov sont très simples à simuler
  1. **Simuler l'état initial X_0 suivant la loi initiale ξ.**
  2. **Simuler conditionnellement $X_{n+1} \sim P(X_n, \cdot)$ en utilisant le noyau de transition P**
- Lorsque la chaîne est **ergodique**, la loi de X_n donnée par $\xi_n = \xi P^n$ converge vers la loi stationnaire.
- Le théorème ergodique de Birkhoff montre aussi que la moyenne d'une réalisation de la chaîne converge vers la loi d'équilibre

$$\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) \to \int f \; d\pi \; .$$

Algorithmes de Hastings-Metropolis

- **Objectif:** Construire un noyau de transition admettant π comme loi stationnaire.
- **Idée:** De façon assez étonnante, il existe une façon universelle de procéder, introduite par Metropolis (circa 1950) et raffinée par Hastings (1970)
- **Algorithme:**
  1. **Proposer un candidat Y_{k+1} suivant un noyau de proposition de densité $q(X_k,\cdot)$**
  2. **Accepter ce candidat avec une probabilité égale à**

$$\alpha(X_k,Y_{k+1}) = \min\left(1, \frac{\pi(Y_{k+1})q(Y_{k+1},X_k)}{\pi(X_k)q(X_k,Y_{k+1})}\right) .$$

- **Important:** le rapport d'acceptation ne dépend que du **rapport** $\pi(y)/\pi(x)$ et ne requiert pas la connaissance de la constante de normalisation !

Matrice de transition de la chaîne

- Matrice de transition de la chaîne

$$p(x,x') = q(x,x')\alpha(x,x') + \mathbb{1}_x(x') \sum_{\bar{x} \in \mathsf{X}} (1 - \alpha(x,\bar{x}))q(x,\bar{x})$$

- Le taux d'acceptation vérifie

$$\begin{aligned}\pi(x)q(x,x')\alpha(x,x') &= \pi(x)q(x,x')\left(1 \wedge \frac{\pi(x')q(x',x)}{\pi(x)q(x,x')}\right) \\ &= \pi(x)q(x,x') \wedge \pi(x')q(x',x) \\ &= \pi(x')q(x',x)\alpha(x',x)\end{aligned}$$

- La matrice de transition satisfait la **condition de balance détaillée** par rapport à π

$$\pi(x)p(x,x') = \pi(x')p(x',x)$$

Réversibilité, balance détaillée et équilibre

$$\pi(x)p(x,x') = \pi(x')p(x',x)$$

- π est une loi stationnaire pour $P = (p(x,x'))$:

$$\begin{aligned} \sum_{x' \in \mathsf{X}} \pi(x')p(x',x) &= \pi(x) \sum_{x' \in \mathsf{X}} p(x,x') \\ &= \pi(x) \end{aligned}$$

  car la matrice p est stochastique.
- Pour $f,g \in \mathcal{L}^2(\mathsf{X},\pi)$,

$$\begin{aligned} \langle f,Pg \rangle = \sum_{x \in \mathsf{X}} f(x)Pg(x)\pi(x) &= \sum_{x \in \mathsf{X}} \sum_{x' \in \mathsf{X}} f(x)g(x')p(x,x')\pi(x) \\ &= \sum_{x \in \mathsf{X}} \sum_{x' \in \mathsf{X}} f(x)g(x')p(x',x)\pi(x') \\ &= \sum_{x' \in \mathsf{X}} Pf(x')g(x')\pi(x') = \langle Pf,g \rangle \end{aligned}$$

Algorithmes de Hastings-Metropolis à marche aléatoire symétrique

- $Y_{k+1} = X_k + Z_{k+1}$ où $Z_{k+1} \sim_{\text{i.i.d.}} \bar{q}$, et $\bar{q}$ est **symétrique**
- $q(x,x') = q(x',x) = \bar{q}(x'-x) = \bar{q}(x-x')$ et le taux d'acceptation est donné par

$$\alpha(x,x') = 1 \wedge \frac{\pi(x')}{\pi(x)}$$

- ... marche aléatoire de loi d'incrément q biaisée.
- Le noyau est **réversible** par rapport à π et la loi stationnaire de la chaîne est π.

Algorithme de Gibbs

- Supposons que $\mathsf{X} = \mathsf{X}_1 \times \cdots \times \mathsf{X}_m$ est un espace produit: $\boldsymbol{x} = (x_1, \ldots, x_m)$, où $x_k \in \mathsf{X}_k$. Pour tout $k \in \{1, \ldots, m\}$, nous notons x_k le $k$ième composant de $\boldsymbol{x}$ et par $\boldsymbol{x}_{-k} = \{x_\ell\}_{\ell \neq k}$ les autres composants.
- Soit $\pi_k(\cdot|\boldsymbol{x}_{-k})$ la probabilité conditionnelle définie par

$$\pi_k(x_k \mid \boldsymbol{x}_{-k}) = \frac{\pi(x_1, x_2, \ldots, x_m)}{\sum_{x'_k \in \mathsf{X}_k} \pi(x_1, \ldots, x'_k, \ldots, x_m)}$$

- L'algorithme de Gibbs consiste à mettre à jour une configuration $\xi^i = (\xi^i_1, \ldots, \xi^i_m)$ de la façon suivante:
pour $k = 1, 2, \ldots, m$: Echantillonner X_k^{i+1} de loi
$\pi_k(\cdot|X_1^{i+1}, \ldots, X_{k-1}^{i+1}, X_{k+1}^i, \ldots, X_m^i)$.
- **Réversibilité** Chaque noyau élémentaire m de l'algorithme de Gibbs est réversible par rapport à π et donc admet π comme loi d'équilibre.

Réversibilité de l'algorithme de Gibbs

- Appelons K_k le noyau de transition de la $k$ième étape de Gibbs

$$K_k(\boldsymbol{x}, \boldsymbol{x}') = \mathbb{1}_{\{x_{-k}\}}(\boldsymbol{x}'_{-k}) \pi_k(x'_k \mid \boldsymbol{x}_{-k}) \,.$$

- Pour toutes fonctions $f, g \in \mathcal{L}^2(\mathsf{X}, \pi)$,

$$\begin{aligned}
\langle f, K_k g \rangle &= \sum_{\boldsymbol{x}, \boldsymbol{x}'} f(\boldsymbol{x}) g(\boldsymbol{x}') \pi(\boldsymbol{x})\, K_k(\boldsymbol{x}, \boldsymbol{x}') \\
&= \sum_{\boldsymbol{x}} \left\{ f(\boldsymbol{x}) \pi(\boldsymbol{x}) \sum_{x'_k} g([x'_k, \boldsymbol{x}_{-k}]) \pi_k(x'_k \mid \boldsymbol{x}_{-k}) \right\} \\
&= \sum_{[x'_k, \boldsymbol{x}_{-k}]} \left\{ g([x'_k, \boldsymbol{x}_{-k}]) \pi(\boldsymbol{x}'_k, \boldsymbol{x}_{-k}) \sum_{x_k} f(x_k, \boldsymbol{x}_{-k}) \pi_k(x_k | \boldsymbol{x}_{-k}) \right\} \\
&= \langle K_k f, g \rangle
\end{aligned}$$

comme $\pi([x_k, \boldsymbol{x}_{-k}]) \pi_k(x'_k \mid \boldsymbol{x}_{-k}) = \pi_k(x_k | \boldsymbol{x}_{-k}) \pi([x'_k, \boldsymbol{x}_{-k}])$,

Gibbs comme cas particulier de HM

- Loi de proposition

$$q_k(\boldsymbol{x},\boldsymbol{x}') = \mathbb{1}_{\boldsymbol{x}_{-k}}(\boldsymbol{x}'_{-k})\pi_k(x'_k \mid \boldsymbol{x}_{-k})$$

- Rapport d'acceptation

$$\begin{aligned}\alpha(\boldsymbol{x},[x'_k,\boldsymbol{x}_{-k}]) &= 1 \wedge \frac{\pi([x'_k,\boldsymbol{x}_{-k}])\pi_k(x_k \mid \boldsymbol{x}_{-k})}{\pi(\boldsymbol{x})\pi_k(x'_k \mid \boldsymbol{x}_{-k})} \\ &= 1\end{aligned}$$

car $\pi([x_k,\boldsymbol{x}_{-k}])\pi_k(x'_k \mid \boldsymbol{x}_{-k}) = \pi_k(x_k|\boldsymbol{x}_{-k})\pi([x'_k,\boldsymbol{x}_{-k}])$,

- Pour l'algorithme de Gibbs, les mouvements sont acceptés avec une probabilité 1 !

Combinaison Metropolis-Gibbs

- Si l'on ne sait pas "simuler" suivant les lois conditionnelles $\pi_k(\cdot|\boldsymbol{x}_{-k})$ on peut utiliser un algorithme de Metropolis en gardant le cadre "Gibbs".
- Il suffit de proposer la k-ième composante x_k suivant un noyau de proposition sur X_k, $q_k(x_k,\cdot)$ et d'accepter ou refuser le mouvement avec la probabilité

$$\alpha(\boldsymbol{x},[x'_k,\boldsymbol{x}_{-k}]) = 1 \wedge \frac{q_k(x'_k,x_k)\pi_k(x_k \mid \boldsymbol{x}_{-k})}{q_k(x_k,x'_k)\pi_k(x'_k \mid \boldsymbol{x}_{-k})}$$

- Base de très nombreux algorithmes MCMC !...

Noyaux Hybrides

- Supposons que $K_1, \ldots, K_m$ sont des matrices markoviennes admettant π comme loi stationnaire Alors
  (a) $K_{\text{syst}} = K_1 K_2 \cdots K_m$ et
  (b) $K_{\text{rand}} = \sum_{i=1}^m \alpha_i K_i$, où $\alpha_i > 0$ pour $i = 1, \ldots, m$ $\sum_{i=1}^m \alpha_i = 1$,

  admettent π comme loi stationnaire.
- Si $K_1, \ldots, K_m$ sont π-réversible, K_{rand} est aussi π-reversible mais K_{syst} n'est pas nécessairement réversible.

Simulation dans un modèle d'Ising: Gibbs

- Loi conditionnelle d'un spin au sommet v

$$\pi_v(x_v \mid \boldsymbol{x}_{-v}) = \frac{\mathrm{e}^{-S_v(x_v, \boldsymbol{x}_{-v})}}{\mathrm{e}^{-S_v(0, \boldsymbol{x}_{-v})} + \mathrm{e}^{-S_v(1, \boldsymbol{x}_{-v})}}$$

  où

$$S_v(x_v, \boldsymbol{x}_{-v}) \stackrel{\text{def}}{=} \beta \sum_{u : u \sim v} \mathbb{1}_{\{x_v = x_u\}} + h_{v,0} \mathbb{1}_{\{0\}}(x_v) + h_{v,1} \mathbb{1}_{\{1\}}(x_v) \ .$$

  ... ne dépend que des valeurs des spins des sommet voisins de v dans le graphe
- **Algorithme:**
  1. Choisir un sommet **au hasard** ou suivant un **balayage** prédéfini.
  2. Mettre à jour la valeur du spin en simulant suivant la loi conditionnelle.

Simulation dans un modèle d'Ising: MH dans Gibbs

1. Choisir un sommet v **au hasard** ou suivant un **balayage** prédéfini.
2. Proposer $x'_v = x_v \oplus 1$
3. Calculer la différence d'énergie

$$\Delta S_v(\boldsymbol{x}) = S_v(x_v \oplus 1, \boldsymbol{x}_{-v}) - S_v(x_v, \boldsymbol{x}_{-v}) \ .$$

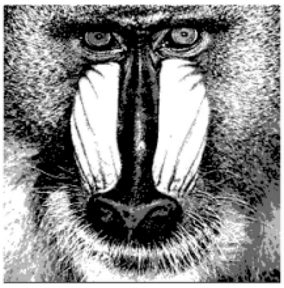
4. Accepter la nouvelle configuration avec une probabilité

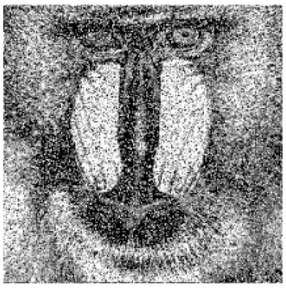
$$1 \wedge \mathrm{e}^{-\Delta S_v(\boldsymbol{x})} \ .$$

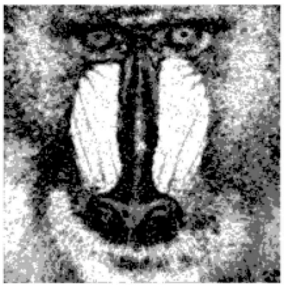
On accepte toujours une configuration d'énergie plus basse...

Modèle d'Ising: influence de la température

Modèle d'Ising: influence de la température







III. Convergence d'une chaîne de Markov

Variation totale

- Si X est fini ou dénombrable, pour ξ une mesure signée finie sur $(\mathsf{X},\mathcal{P}(\mathsf{X}))$ on définit
$$\|\xi\|_{\mathrm{TV}} = \sum_{x \in \mathsf{X}} |\xi(x)| \ .$$
- On dit qu'une matrice markovienne est **uniformément ergodique** si et seulement si
$$\lim_{k\to\infty} \sup_{x \in \mathsf{X}} \left\| P^k(x,\cdot) - \pi \right\|_{\mathrm{TV}} = 0$$
- On cherche à quantifier le plus précisément possible $\left\| P^k(x,\cdot) - \pi \right\|_{\mathrm{TV}}$ pour obtenir une **garantie** de convergence !

III-1. Convergence d'une chaîne de Markov Approche opérateur

Coefficient de Dobrushin

Définition (Coefficient de Dobrushin)

Notons

$\mathbb{M}_0(\mathsf{X},\mathcal{X})$ l'ensemble des mesures signées finies de masse nulle.

Soit P une matrice markovienne sur X. Le **coefficient de Dobrushin** $\Delta(P)$ de P est donné par

$$\Delta(P) \stackrel{\text{def}}{=} \sup \left\{ \|\xi P\|_{\mathrm{TV}} : \xi \in \mathbb{M}_0(\mathsf{X},\mathcal{X}),\ \|\xi\|_{\mathrm{TV}} \leq 1 \right\} \ .$$

Notons que $\Delta(P) \leq 1$... Le coefficient de Dobrushin est vraiment intéressant lorsqu'il existe $m \in \mathbb{N}$ tel que $\Delta(P^m) < 1$!

Rque : $\Delta(P) \leq 1$ car

$$\begin{aligned}
\|\xi P\|_{\mathrm{TV}} &= \sum_y |\xi P(y)| = \sum_y \left| \sum_x \xi(x)\ P(x,y) \right| \\
&\leq \sum_y \sum_x |\xi(x)|\ P(x,y) = \sum_x \left(\sum_y P(x,y) \right) |\xi(x)| \\
&\leq \sum_x |\xi(x)| = \|\xi\|_{\mathrm{TV}}
\end{aligned}$$

donc

$$\Delta(P) \stackrel{\text{def}}{=} \sup \left\{ \|\xi P\|_{\mathrm{TV}} : \xi \in \mathbb{M}_0(\mathsf{X},\mathcal{X}),\ \|\xi\|_{\mathrm{TV}} \leq 1 \right\} \leq 1 \ .$$

sous-multiplicativité

Proposition

Si P et R sont des matrices markoviennes, alors $\Delta(PR) \leq \Delta(P)\Delta(R)$.

Méthodes de Monte Carlo par Chaînes de Markov
└ Convergence d'une chaîne de Markov
└ Approche opérateur : Coefficient de Dobrushin

sous-multiplicativité

Proposition

Si P et R sont des matrices markoviennes, alors $\Delta(PR) \leq \Delta(P)\Delta(R)$.

Démonstration.

Soit $\xi \in \mathbb{M}_0(\mathsf{X},\mathcal{X})$. Alors $\xi P(\mathsf{X}) = \xi(\mathsf{X}) = 0$ et donc $\xi P \in \mathbb{M}_0(\mathsf{X},\mathcal{X})$. En posant $\nu = \xi P$, nous avons $\|\nu R\|_{\mathrm{TV}} \leq \Delta(R)\|\nu\|_{\mathrm{TV}}$. Par conséquent:

$$\|\xi PR\|_{\mathrm{TV}} = \|(\xi P)R\|_{\mathrm{TV}} \leq \Delta(R)\|\xi P\|_{\mathrm{TV}} \leq \Delta(P)\Delta(R)\|\xi\|_{\mathrm{TV}} \ .$$

$\square$

Ergodicité uniforme

Notons

$\mathbb{M}_1(\mathsf{X},\mathcal{X})$ l'ensemble des probabilités sur $(\mathsf{X},\mathcal{X})$.

Théorème

Soit P une matrice markovienne telle que pour un m entier, $\Delta(P^m) \leq \rho < 1$. Alors, P admet une unique probabilité invariante π. De plus, pour tout $\nu \in \mathbb{M}_1(\mathsf{X},\mathcal{X})$,

$$\|\nu P^n - \pi\|_{\mathrm{TV}} \leq \rho^{\lfloor n/m \rfloor} \ \|\nu - \pi\|_{\mathrm{TV}} \ ,$$

où $\lfloor u \rfloor$ est la partie entière de u.

Démonstration.

Pour $n \in \mathbb{N}$, on note $q = \lfloor n/m \rfloor$, et $r = n - qm \in \{0,\ldots,m-1\}$. La sous-additivité du coefficient de Dobrushin et la propriété $\Delta(R) \leq 1$ montre que pour toute probabilité ν,ν'

$$\begin{aligned}\|\nu P^n - \nu' P^n\|_{\mathrm{TV}} &\leq \Delta(P^n)\, \|\nu - \nu'\|_{\mathrm{TV}} \\ &\leq \Delta(P^r)[\Delta(P^m)]^q\, \|\nu - \nu'\|_{\mathrm{TV}} \leq \rho^{\lfloor n/m \rfloor}\, \|\nu - \nu'\|_{\mathrm{TV}} \ .\end{aligned}$$

En prenant $\nu' = \nu P$, et en notant $\|\nu - \nu P\|_{\mathrm{TV}} \leq 2$, nous avons donc

$$\left\|\nu P^n - \nu P^{n+1}\right\|_{\mathrm{TV}} \leq 2\rho^{\lfloor n/m \rfloor} \ ,$$

qui montre que la série $\sum_n \left\|\nu P^n - \nu P^{n+1}\right\|_{\mathrm{TV}}$ est convergente. Comme $(\mathbb{M}_1(\mathsf{X},\mathcal{X}), \|\cdot\|_{\mathrm{TV}})$ est un espace de Banach (facile) à, ceci montre que la suite $\{\nu P^n : n \in \mathbb{N}\}$ converge vers une mesure signée π, dont on montre qu'elle est une probabilité (facile) . □

Condition de Doeblin

Définition (Doeblin condition)

Une matrice Markovienne P satisfait la **condition de Doeblin** s'il existe un entier $m \geq 1$, $\epsilon > 0$, et une mesure de probabilité μ sur X telle que, pour tout $x \in \mathsf{X}$ et $A \in \mathcal{X}$,

$$P^m(x,A) \geq \epsilon\mu(A) \ .$$

De Doeblin à Dobrushin

1. Etape 1 : établir que

$$\Delta(P) = \frac{1}{2} \sup_{x,y} \|P(x,\cdot) - P(y,\cdot)\|_{\mathrm{TV}} \ .$$

2. Etape 2 : établir que pour toute probabilités ν,ν',

$$\|\nu - \nu'\|_{\mathrm{TV}} = \sum_{x \in \mathsf{X}} |\nu(x) - \nu'(x)| = 2\left(1 - \sum_{x \in \mathsf{X}} \nu(x) \wedge \nu'(x)\right) \ .$$

De Doeblin à Dobrushin (suite)

- En posant $\nu = P^m(x,\cdot)$ et $\nu' = P^m(x',\cdot)$, nous avons

$$\Delta(P^m) = 1 - \sum_{y \in \mathsf{X}} P^m(x,y) \wedge P^m(x',y)$$

- Si la matrice P satisfait la condition de Doeblin,

$$P^m(x,y) \wedge P^m(x',y) \geq \epsilon\mu(y)$$
$$\sum_{y \in \mathsf{X}} P^m(x,y) \wedge P^m(x',y) \geq \epsilon \ ,$$

  car μ est une probabilité.
- Par conséquent, le coefficient de Dobrushin est borné par

$$\Delta(P^m) \leq 1 - \epsilon \ .$$

  La vitesse de convergence est **explicite** et dépend d'une condition aisément vérifiable.

Cas particulier d'une matrice markovienne primitive

- Lorsque l'espace d'état X est fini, nous avons

$$\Delta P^m \leq 1 - |\mathsf{X}| \min_{(x,x') \in \mathsf{X}^2} P^m(x,x')$$

- Par conséquent, s'il existe un entier m tel que $P^m(x,x') > 0$ pour tout $(x,x') \in \mathsf{X}^2$, alors la chaine est uniformément géométriquement ergodique: il existe une unique probabilité $\pi \in \mathbb{M}_1(\mathsf{X},\mathcal{X})$ invariante et, pour tout $x \in \mathsf{X}$,

$$\|P^n(x,\cdot) - \pi\|_{\mathrm{TV}} \leq 2 \left(1 - |\mathsf{X}| \min_{(x,x') \in \mathsf{X}^2} P^m(x,x')\right)^{\lfloor n/m \rfloor} \ .$$

Démonstration étape 2

Démonstration.

Découle de l'identité élémentaire $|a-b| = a+b-2(a \wedge b)$:

$$\|\nu - \nu'\|_{\mathrm{TV}} = \sum_{x \in \mathsf{X}} (\nu(x) + \nu'(x) - 2(\nu(x) \wedge \nu'(x))) = 2\left(1 - \sum_{x \in \mathsf{X}} \nu(x) \wedge \nu'(x)\right) .$$

□

III-2. Convergence d'une chaîne de Markov Approche par couplage

Approche par couplage

- Une alternative à la démonstration de l'ergodicité uniforme sous l'hypothèse de condition de Doeblin cas $m = 1$ pour simplifier

$$P(x,A) \geq \epsilon \mu(A) \ .$$

- On introduit le noyau résiduel

$$R(x,x') \stackrel{\text{def}}{=} \frac{P(x,x') - \epsilon\mu(x')}{1-\epsilon}$$

- On construit un processus bivarié $\{(X_n, X'_n), n \geq 0\}$ de sorte que le **temps de couplage** est fini avec probabilité 1

$$T \stackrel{\text{def}}{=} \inf\{n \geq 0 : X_q = X'_q \qquad \forall\, q \geq n\} \ .$$

Construction du processus bivarié

- Initialisation : $X_0 \sim \nu \qquad X'_0 \sim \pi \qquad d_0 = 0$
- Etant donné l'état courant (X_n, X'_n, d_n)
  - Si $d_n = 1$ alors $X_{n+1} \sim P(X_n, \cdot) \quad X'_{n+1} = X_{n+1} \quad d_{n+1} = 1.$
  - Si $d_n = 0$,
    - avec probabilité ϵ, $\quad X_{n+1} = X'_{n+1} \sim \mu \qquad d_{n+1} = 1$
    - sinon, $X_{n+1} \sim R(X_n, \cdot) \quad X'_{n+1} \sim R(X'_n, \cdot) \quad d_{n+1} = 0$

Propriétés de ce processus bivarié

- $T = \inf\{n \geq 1 : d_n = 1\}$.
- $X_n \sim \nu P^n \qquad X'_n \sim \pi P^n = \pi.$

Par suite: (inégalité de couplage)

$$
\begin{aligned}
\|\nu P^n - \pi\|_{\mathrm{TV}} &= \sum_y |\nu P^n(y) - \pi(y)| = \sum_y |\mathbb{P}(X_n = y) - \mathbb{P}(X'_n = y)| \\
&= \sum_y |\mathbb{P}(X_n = y, T \leq n) - \mathbb{P}(X'_n = y, T \leq n) \\
&\quad + \mathbb{P}(X_n = y, T > n) - \mathbb{P}(X'_n = y, T > n)| \\
&= \sum_y |\mathbb{P}(X_n = y, T > n) - \mathbb{P}(X'_n = y, T > n)| \\
&\leq 2\ \mathbb{P}(T > n) \\
&\leq 2\ (1-\epsilon)^n .
\end{aligned}
$$

IV. Conclusion

Conclusion

- Utilisation des chaînes de Markov (discrètes) pour la simulation Monte Carlo.
- Convergence des méthodes de simulation Monte Carlo par Chaînes de Markov (par ex. via approche opérateur, approche couplage).
- Illustration : application (simple) au traitement statistique des images.

Yves Guiard

Directeur de recherche au CNRS LTCI

Les probabilités comme un artisanat

Quelques remarques d'un chercheur en psychologie expérimentale à propos de l'induction et de la description statistiques

Dans les sciences empiriques, tout le monde n'a pas la chance de pouvoir expérimenter sur des grands nombres. Comment tirer une loi générale de mesures obtenues chez un échantillon de quelques dizaines de sujets humains ? C'est le vieux problème épistémologique de l'induction, dont on a de bonnes raison de penser, depuis Sir Karl (Popper), qu'elle est logiquement impossible. Compliquons la situation avec une bonne dose de bruit statistique : on débouche sur le problème de la décision statistique inductive formalisé par Sir Ronald (Fisher). J'évoquerai quelques unes des objections qui ont été avancées, sans grand impact, contre l'artisanat du test statistique à la Fisher que les sciences humaines ont largement dogmatisé. J'aborderai également certaines questions rencontrées au cours de mon étude du conflit vitesse/précision dans le mouvement humain, qui ont trait aux thèmes de la mesure et de la statistique descriptive.

Références :

Guiard, Y., Olafsdottir, H.B. (2011). On the measurement of movement difficulty in the standard approach to Fitts' law.

Guiard, Y., Olafsdottir, H.B., Perrault, S.T. (2011). Fitts' law as an explicit time/error trade-off. Proceedings of CHI 2011, ACM Conference on Human Factors in Computing Systems 1619-28.

Les probabilités comme un artisanat

Quelques remarques d'un chercheur en psychologie expérimentale à propos de l'induction et de la description statistiques

Yves Guiard
LTCI CNRS
Telecom-ParisTech - Departement INFRES

Plan

1. Les probabilités dans la pratique de la recherche empirique

- Epistémologie de la science empirique, pas des maths
- Théorie de la connaissance vs. histoire des sciences : Popper vs. Kuhn
- Réalisme scientifique: du scepticisme humien au relativisme cognitif postmoderne
  - Le canular de Sokal, Sokal et Bricmont (1997)
  - Le générateur post-moderniste
- Drame de l'induction, nécessaire et impossible
  - Popper: corroboration impossible, la réfutation seule inférence valide

Plan

1. Les probabilités dans la pratique de la recherche empirique

- Epistémologie de la science empirique, pas des maths
- Théorie de la connaissance vs. histoire des sciences : Popper vs. Kuhn
- Réalisme scientifique: du scepticisme humien au relativisme cognitif postmoderne
  - Le canular de Sokal, Sokal et Bricmont (1997)
  - Le générateur post-moderniste
- Drame de l'induction, nécessaire et impossible
  - Popper: corroboration impossible, la réfutation seule inférence valide
  - Meehl (1997) : en fait même la réfutation est un problème
  - Sokal et Bricmont : réalisme modeste et enquête rationnelle
- Disparition du modus tollens dans le raisonnement probabiliste

2. Probabilité conditionnelle et statistiques inductives

- Probabilités vs. fréquences naturelles (Hoffrage et al., 2000)
- Test d'hypothèse nulle et signification statistique
  - Le principe avec un exemple (Lowry, VassarStats)
  - L'objection: *p*(DIH) ≠ *p*(HID (Cohen, 1994; Pollard & Richardson, 1987)
  - Inférence fisherienne vs. bayesienne : de l'habitude à la tradition et au dogme

3. Equivoques de la formule *y* = f(*x*/*z*) pour un expérimentaliste

- Terminologie: taxonomie des variables
  - Niveaux de la mesure (Stevens, 1946)
  - Variables manipulées vs. enregistrées
  - Variables, paramètres, constantes
- Contexte : conflit vitesse/précision et "loi" de Fitts
- Signe "=" relation symétrique? Le problème du *paradigme expérimental*
- Combien de quantités indépendantes dans x/z? description *polaire* vs. *cartésienne*
- *x*/*z* vs. *z*/*x*: danger de l'inversion des rapports

4. Deux problèmes de statistiques descriptives

- Représenter une loi empirique: courbe moyenne vs. courbe limite
- Histogramme vs. courbe selon le niveau de la mesure en x et en y

Epistémologie de la science empirique, pas des maths

Notamment à cause du problème partout crucial de l'*induction* du particulier à l'universel

- Possibilité de la *preuve par induction à* l'intérieur du domaine mathématique (par ex. la démonstration du 'dernier théorème' de Fermat par Andrew Wiles en 1993)
- Impossible dans les sciences de la nature

Théorie de la connaissance vs. histoire des sciences

Karl Popper vs. Thomas Kuhn

Réalisme scientifique

Du scepticisme humien au relativisme cognitif post-moderne

Le canular de Sokal (1996)

« Impostures intellectuelles » (Sokal & Bricmont, 1997)

Le *Générateur post-moderne*

Programme de Andrew Bulhak (1996)

http://www.elsewhere.org/pomo/

Grammaire formelle basée sur un réseau de transitions récursif

Drame de l’induction, nécessaire et impossible

Le problème de la science empirique : découvrir les lois générales du monde à partir de données ou d’observations particulières.
Théories générales
Données particulières

Induction
Type d’inférence conduisant du particulier au général (extrapolation, généralisation)

Déduction
L’ensemble des règles d’inférence de la logique. Toute logique est déductive.

Test empirique des théories
Des théories on dérive des prédictions *déductivement*
Test de ces prédictions : exercice *inductif* (des données vers les hypothèses)
Validation impossible, seule la réfutation est possible (Popper)

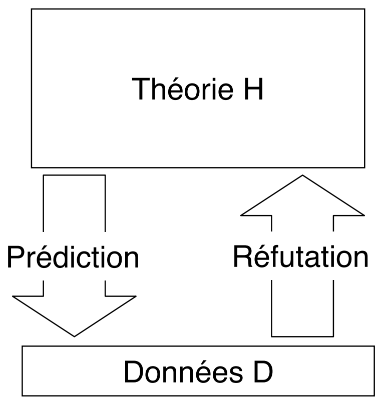
Drame de l’induction, nécessaire et impossible

Le problème de la science empirique : découvrir les lois générales du monde à partir de données ou d’observations particulières.
Théories générales
Données particulières

Induction
Type d’inférence conduisant du particulier au général (extrapolation, généralisation)

Déduction
L’ensemble des règles d’inférence de la logique. Toute logique est déductive.

Test empirique des théories
Des théories on dérive des prédictions *déductivement*
Test de ces prédictions : exercice *inductif* (des données vers les hypothèses)
Validation impossible, seule la réfutation est possible (Popper)



Cycle de la science empirique selon Popper

Karl Popper: corroboration impossible, la réfutation est la seule induction valide

Deductive Inference Possibilities for the Hypothetical Argument: If p, then q

Figure	Form	Name	Conclusion
I	If p then q p $\therefore q$	Modus ponens ("Establishing mode")	Valid Trivial
II	If p then q $\sim p$ $\therefore \sim q$	Denying the antecedent	Invalid Non trivial
III	If p then q q $\therefore p$	Affirming the consequent	Invalid Non trivial
IV	If p then q $\sim q$ $\therefore \sim p$	Modus tollens ("Destroying mode")	Valid Non trivial

Meehl (néo-poppérien) : OK pour le modus tollens, mais *en pratique*,
la réfutation est bien difficile!

On teste généralement une théorie accompagnée de nombreuse suppositions auxiliaires

T: Main substantive theory of interest;
A_x: Auxiliary theories relied on in the experiment;
C_p: *Ceteris paribus* clause ("other things being equal");
A_i: Instrumental auxiliaries (devices relied on for control and observation);
C_n: Realized particulars (conditions were as the experimenter reported);
O_1, O_2: Observations or statistical summaries of observations;

then the logical structure of a test of a theory is the conceptual formula:

$$(T \cdot A_x \cdot C_p \cdot A_i \cdot C_n) \vdash (O_1 \supset O_2)$$

where dots " $\cdot$ " are conjunctions ("and"), turnstile " $\vdash$ " is deductive derivability (entailment, "follows that . . ."), and the horseshoe " $\supset$ " is the material conditional ("If . . . then . . .").

Paul E. Meehl (1997). The problem is epistemology, not statistics. URL

Découverte de Neptune par Le Verrier et Adams (1846)

Anomalies constatées par les astronomes dans la course de certains corps célestes

Deux options:

1) Théorie de Newton est fausse (modus tollens)

2) Les observations sont fausses → calcul *ad hoc* d'un possible corps susceptible de perturber localement la trajectoire prédite

En somme:

Pour tester nos théories du monde, on ne dispose que de l'arme de la réfutation empirique, et de surcroît cette arme marche mal.

Toutefois

Epistémologie de Popper *darwiniste*: sélection naturelle ≈ réfutation stochastique

Sokal et Bricmont (2004): Plaidoyer pour un réalisme modeste

Disparition du modus tollens dans le raisonnement probabiliste

Disparition du modus tollens dans le raisonnement probabiliste

Raisonnement déterministe

Sous H_0 ce résultat est impossible

Or, j'ai obtenu ce résultat

Donc H_0 est faux — VALIDE (Modus tollens)

Idem, langage probabiliste

Sous H_0 ce résultat est extrêmement improbable

Or, j'ai obtenu ce résultat

Donc H_0 est très probablement faux — INVALIDE

Il est impossible qu'un martien soit député de la Corrèze	si p alors ¬q
François Hollande est député de la Corrèze	q
Donc François Hollande n'est pas martien	donc ¬p

VALIDE (modus tollens)

Il est impossible qu'un français soit député de la Corrèze (FAUX)
François Hollande est député de la Corrèze
Donc François Hollande n'est pas français

VALIDE (modus tollens)

Discours déterministe avec p=0 → probabiliste avec *p très proche* de zéro

Il est *hautement improbable* qu'un français soit député de la Corrèze (VRAI, $p \approx 10^{-9}$)
François Hollande est député de la Corrèze
Donc François Hollande n'est probablement pas français

INVALIDE

Sous H_0, un résultat aussi extrême que celui-ci est hautement improbable (VRAI)
J'ai obtenu ce résultat
Donc H_0 est probablement faux.

Même raisonnement INVALIDE que

Il est hautement improbable qu'un français soit député de la Corrèze (VRAI)
François Hollande est député de la Corrèze
Donc François Hollande n'est probablement pas français

Ce paralogisme sous-tend la méthode usuelle du test d'hypothèse nulle (Pollard & Richardson, 1987).

Plan

1. **Les probabilités dans la pratique de la recherche empirique**
    - Epistémologie : maths vs. science empirique
    - Théorie de la connaissance vs. histoire des sciences : Popper vs. Kuhn
    - Réalisme scientifique: du scepticisme humien au relativisme cognitif postmoderne
        - Le canular de Sokal, Sokal et Bricmont (1997)
        - Le générateur post-moderniste
    - Drame de l'induction, nécessaire et impossible
        - Popper: corroboration impossible, la réfutation seule inférence valide
        - Meehl (1997) : en fait même la réfutation est un problème
        - Sokal et Bricmont : réalisme modeste et enquête rationnelle
    - Disparition du modus tollens dans le raisonnement probabiliste
2. **Probabilité conditionnelle et statistiques inductives**
    - Probabilités vs. fréquences naturelles (Hoffrage et al., 2000)
    - Test d'hypothèse nulle et signification statistique
        - Le principe avec un exemple (Lowry, VassarStats)
        - L'objection: $p(D|H) \neq p(H|D)$ (Cohen, 1994; Pollard & Richardson, 1987)
        - Inférence fisherienne vs. bayesienne : de l'habitude à la tradition et au dogme
3. **Equivoques de la formule $y = f(x/z)$ pour un expérimentaliste**
    - Terminologie: taxonomie des variables
        - Niveaux de la mesure (Stevens, 1946)
        - Variables manipulées vs. enregistrées
        - Variables, paramètres, constantes
    - Contexte : conflit vitesse/précision et "loi" de Fitts
    - Signe "=" relation symétrique? Le problème du *paradigme expérimental*
    - Combien de quantités indépendantes dans x/z? description *polaire* vs. *cartésienne*
    - *x*/*z* vs. *z*/*x*: danger de l'inversion des rapports
4. **Deux problèmes de statistiques descriptives**
    - Représenter une loi empirique: courbe moyenne vs. courbe limite
    - Histogramme vs. courbe selon le niveau de la mesure en x et en y

Fréquences vs. Probabilités: Hoffrage et al. (2000)

Dépistage du cancer colo-rectal (test Hemoccult Abbott)

Hoffrage, U., Lindsay, S., Hertwig, R., & Gigerenzzer, G. (2000). Communicating statistical information. *Science 290*, 2261-2262. *La Recherche HS N°13*, 2003.

			HYPOTHESE	
			Malade	Non malade
			M	**¬M**
TEST	positif	**m**		
	négatif	**¬m**		

- Incidence $p(M) = .003$ (chez les plus de 50 ans)
- Sensibilité $p(m|M) = .5$
- Faux positifs $p(m|\neg M) = .03$

***Question*:** $\boldsymbol{p(M|m) = ?}$

			HYPOTHESE	
			Malade	Non malade
			M	**¬M**
TEST	positif	**m**		
	négatif	**¬m**		

- Incidence $p(M) = .003$ (chez les plus de 50 ans)
- Sensibilité $p(m|M) = .5$
- Faux positifs $p(m|\neg M) = .03$

***Réponse*:** $\boldsymbol{p(M|m) = .048}$

≈5% des tests positifs sont des bonnes alarmes p(**M**Im)

question ***médicale*** dans des conditions techniques spécifiées

≠

p(**m**IM)

question ***technique*** dans des conditions médicales spécifiées

Pourtant on tend à confondre les deux

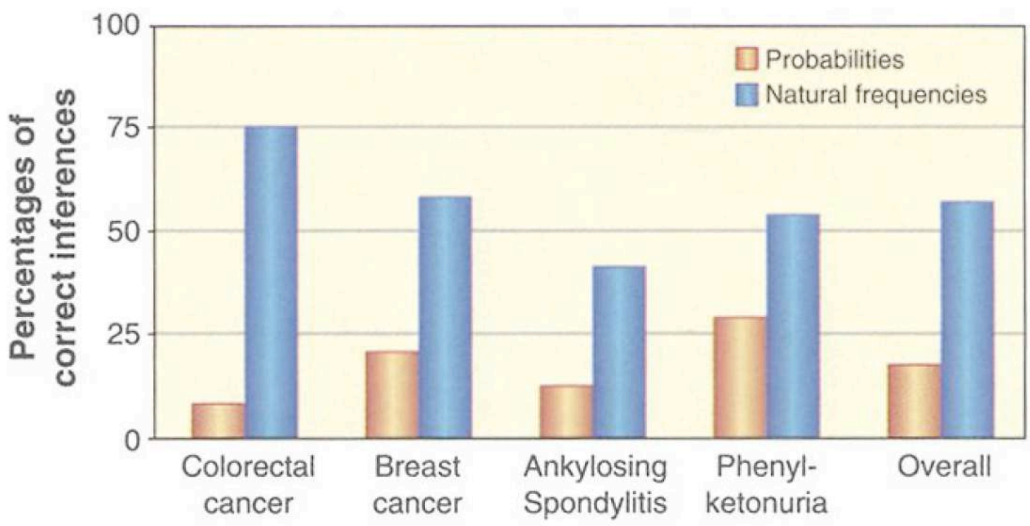


Fig. 1. Interpreting statistics. Medical students' percentage of correct inferences in four realistic diagnostic tasks.

Option #1: Probabilités

$p(M) = .003$
$p(m|M) = .5$
$p(m|\neg M) = .03$

calculer $p(M|m)$

			HYPOTHESE	
			Malade	Non malade
			M	**¬M**
TEST	positif	**m**		
	négatif	**¬m**		

10 000

$$p(M|m) = \frac{p(M) \cdot p(m|M)}{p(M) \cdot p(m|M) + p(\neg M) \cdot p(m|\neg M)}$$

$$= \frac{(.003)(.5)}{(.003)(.5) + (.997)(.03)} = .048$$

Option #2: Fréquences naturelles

$p(M) = .003$
$p(m|M) = .5$
$p(m|\neg M) = .03$

calculer $p(M|m)$

			HYPOTHESE	
			Malade	Non malade
			M	**¬M**
TEST	positif	**m**		
	négatif	**¬m**		

10 000

$\boxed{p(M) = .003}$

$p(m|M) = .5$

$p(m|\neg M) = .03$

Fréquences naturelles

calculer $p(M|m)$

HYPOTHESE

			Malade **M**	Non malade **¬M**	
TEST	positif	**m**			
	négatif	**¬m**			
			$\boxed{30}$	9 970	$\boxed{10\ 000}$

$p(M) = .003$

$\boxed{p(m|M) = .5}$

$p(m|\neg M) = .03$

Fréquences naturelles

calculer $p(M|m)$

HYPOTHESE

			Malade **M**	Non malade **¬M**	
TEST	positif	**m**	$\boxed{15}$		
	négatif	**¬m**	15		
			$\boxed{30}$	9 970	10 000

$p(M) = .003$
$p(m/M) = .5$
$\boxed{p(m/\neg M) = .03}$

Fréquences naturelles

calculer $p(M/m)$

			HYPOTHESE		
			Malade	Non malade	
			M	**¬M**	
TEST	positif	**m**	15	$\boxed{299}$	
	négatif	**¬m**	15	9 671	
			30	$\boxed{9\,970}$	10 000

$p(M) = .003$
$p(m/M) = .5$
$p(m/\neg M) = .03$

Fréquences naturelles

calculer $p(M/m)$

			HYPOTHESE		
			Malade	Non malade	
			M	**¬M**	
TEST	positif	**m**	15	299	314
	négatif	**¬m**	15	9 671	9 686
			30	9 970	10 000

$p(M) = .003$
$p(m/M) = .5$
$p(m/\neg M) = .03$

Fréquences naturelles

$\boxed{\text{calculer } p(M/m)}$

			HYPOTHESE		
			Malade	Non malade	
			M	**¬M**	
TEST	positif	**m**	$\boxed{15}$	299	$\boxed{314}$
	négatif	**¬m**	15	9 671	9 686
			30	9 970	10 000

15/314 = .048 probabilité de la bonne alerte

$p(M) = .003$
$p(m/M) = .5$
$p(m/\neg M) = .03$

Fréquences naturelles

calculer $p(M/m)$

			HYPOTHESE		
			Malade	Non malade	
			M	**¬M**	
TEST	positif	**m**	15	$\boxed{299}$	$\boxed{314}$
	négatif	**¬m**	15	9 671	9 686
			30	9 970	10 000

15/314 = .048 probabilité de la bonne alerte

299/314 = .952 proba de la *fausse* alerte

Fréquences naturelles

			HYPOTHESE		
			Malade	Non malade	
			M	**¬M**	
TEST	positif	**m**	15	299	314
	négatif	**¬m**	15	9 671	9 686
			30	9 970	10 000

Test

Utile p (m&M) = 15/10 000 = **0.15%**

Fréquences naturelles

			HYPOTHESE		
			Malade	Non malade	
			M	**¬M**	
TEST	positif	**m**	15	299	314
	négatif	**¬m**	15	9 671	9 686
			30	9 970	10 000

Test

Utile p (m&M) = 15/10 000 = **0.15%**

Sans suite* p [(¬m et ¬M) ou (¬m et M)] = 9 686/10 000 = **96.9%**

* à tort ou à raison

Fréquences naturelles

			HYPOTHESE		
			Malade	Non malade	
			M	**¬M**	
TEST	positif	**m**	15	299	314
	négatif	**¬m**	15	9 671	9 686
			30	9 970	10 000

Test

Utile *p* (m&M) = 15/10 000 = **0.15%**

Sans suite* *p* [(¬m et ¬M) ou (¬m et M)] = 9 686/10 000 = **96.9%**

Préjudiciable *p* (m &¬M) = 299/10 000 = **3%**

Intérêt de l'étude de Hoffrage et al. (2000)

- **Illustration d'une fait mathématique de base**

Si

$$p(H|D) = \frac{p(D|H) \cdot p(H)}{p(D)}$$

alors

$$p(H|D) = p(D|H) \Longleftrightarrow p(H) = p(D)$$

- **Illustration d'un fait bien documenté de la psychologie cognitive**

Difficulté de distinguer les probabilités conditionnelles

- **Une suggestion pédagogique**

Raisonner de préférence sur les fréquences naturelles

Intérêt de l'étude de Hoffrage et al. (2000)

- **Une énigme**

Pourquoi les fréquences 'naturelles' sont-elles tellement mieux maîtrisées que les probabilités?

Explication de Hoffrage et al.

Probabilités	Fréquences
$p(M\|m) = \dfrac{p(M)\cdot p(m\|M)}{p(M)\cdot p(m\|M) + p(\neg M)\cdot p(m\|\neg M)}$ $= \dfrac{(.003)(.5)}{(.003)(.5)+(.997)(.03)} = .048$	$p(M\|m) = \dfrac{a}{a+b}$ $= \dfrac{15}{15+299} = .048$

Intérêt de l'étude de Hoffrage et al. (2000)

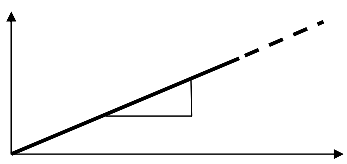
Autre élément d'explication

Même écriture, le rapport $\dfrac{y}{x} = p$ avec $0 \leqslant p \leqslant 1$

Mais

Probabilité: 1 nombre

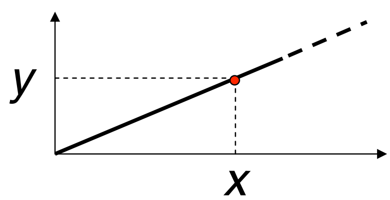
$$y = px$$



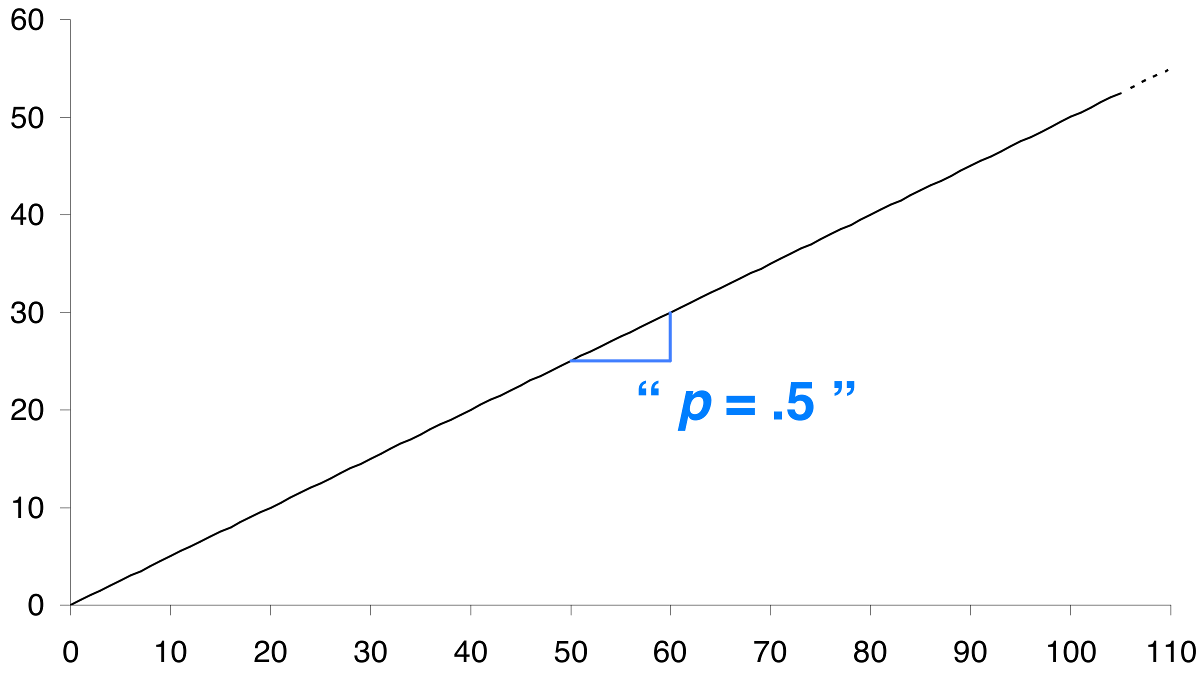
$p = .5, \quad p = ½, \quad p = 50\%$

Fréquence 'naturelle': 2 nombres

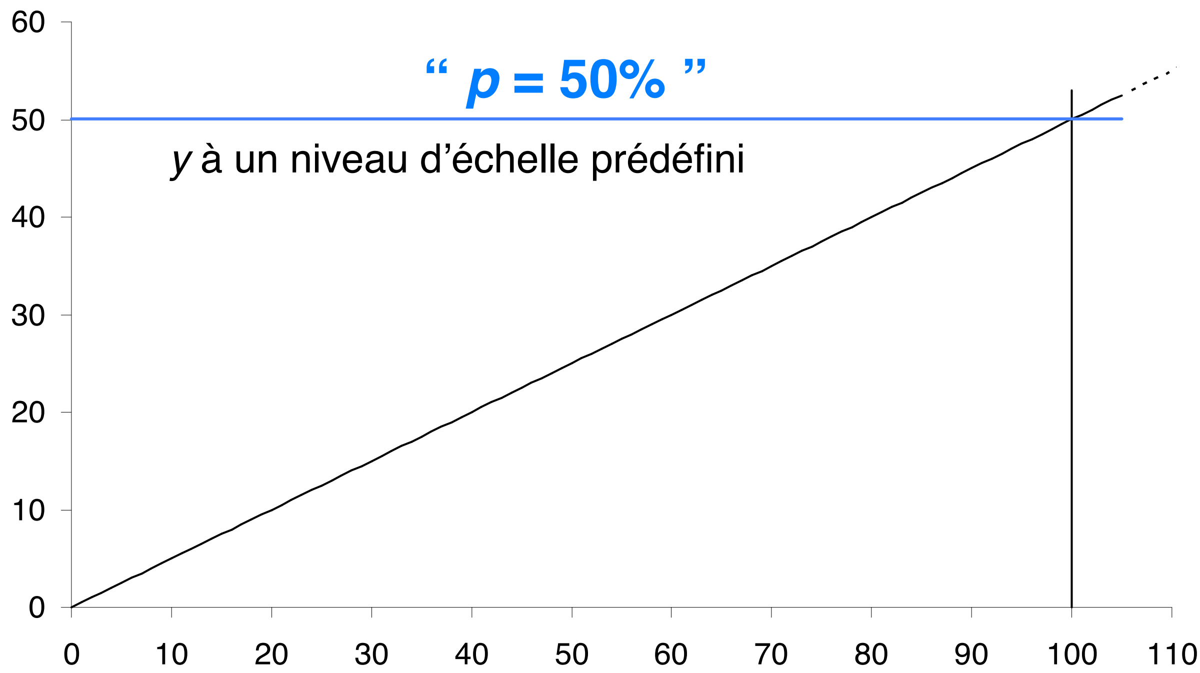
$$y = px$$



$p = 15/30$

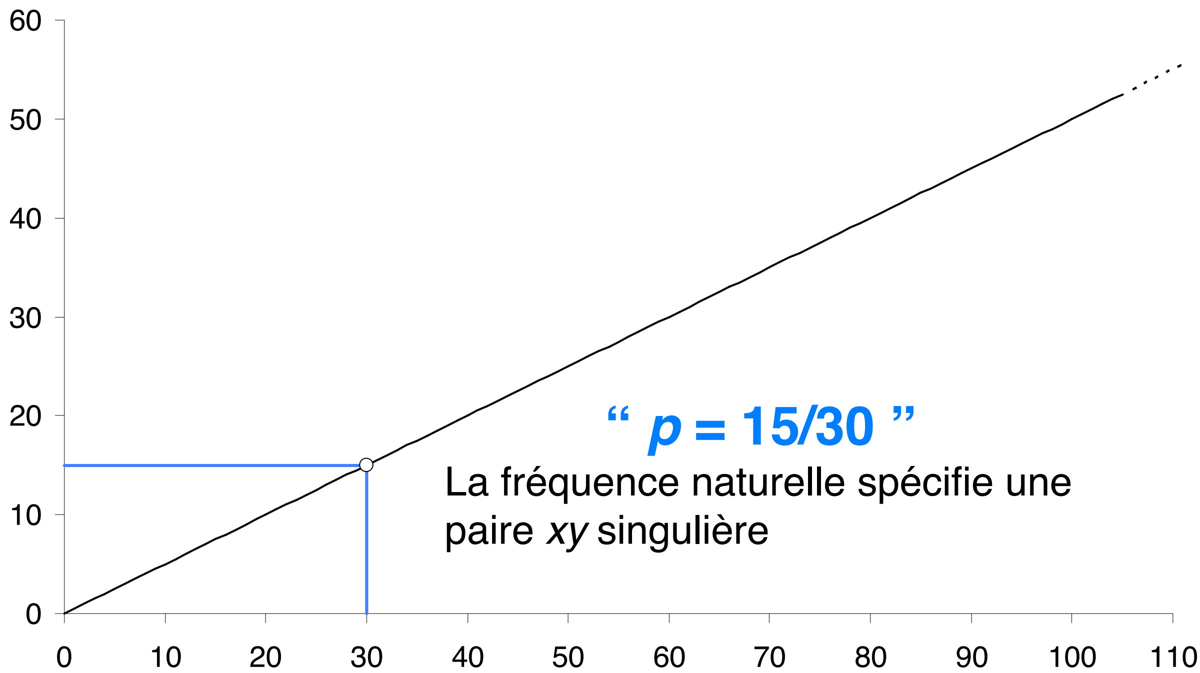


60
50
40
30
20
10
0
0
10
20
30
40
50
60
70
80
90
100
110
“ *p* = .5 ”

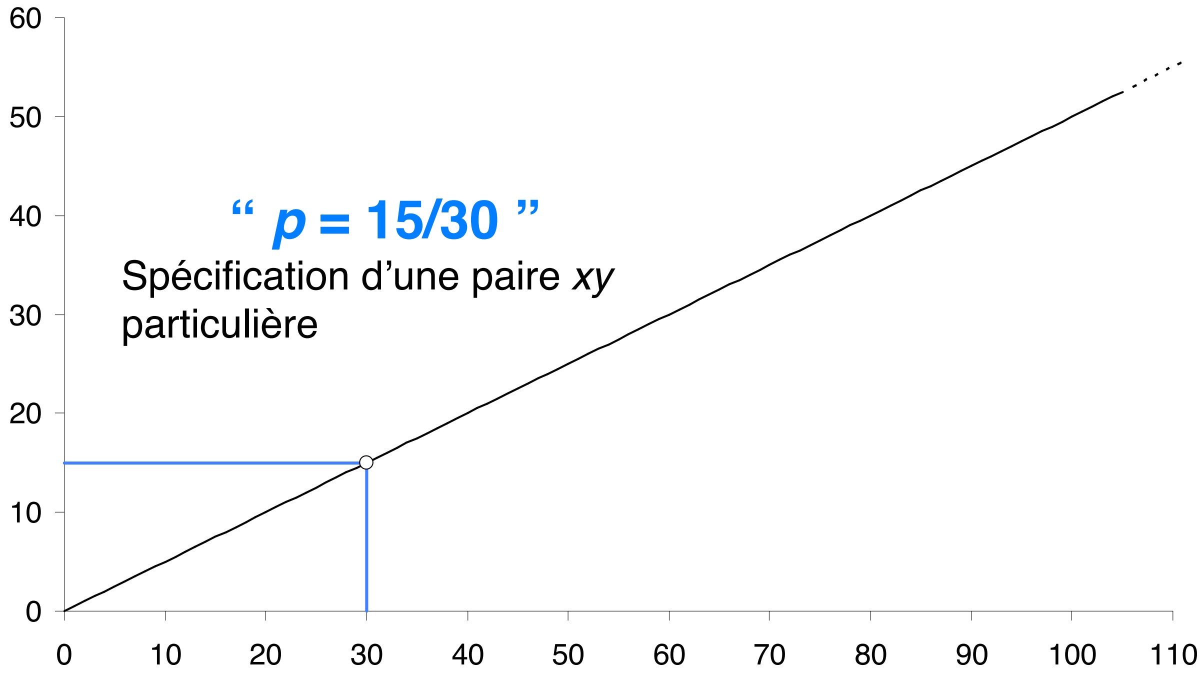


60
50
40
30
20
10
0
0
10
20
30
40
50
60
70
80
90
100
110
“ *p* = 50% ”
y à un niveau d’échelle prédéfini

9
8
7
6
5
4
3
2
1
0
0 1 2 3 4 5 6 7 8 9
“ p = 1/2 ”
la plus petite échelle où x et y restent entiers



60
50
40
30
20
10
0
0 10 20 30 40 50 60 70 80 90 100 110
“ p = 15/30 ”
La fréquence naturelle spécifie une paire xy singulière



Test d'hypothèse nulle et signification statistique

- *Concepts and Applications of Inferential Statistics*
- *VassarStats: Website for Statistical Computation*

par Richard Lowry

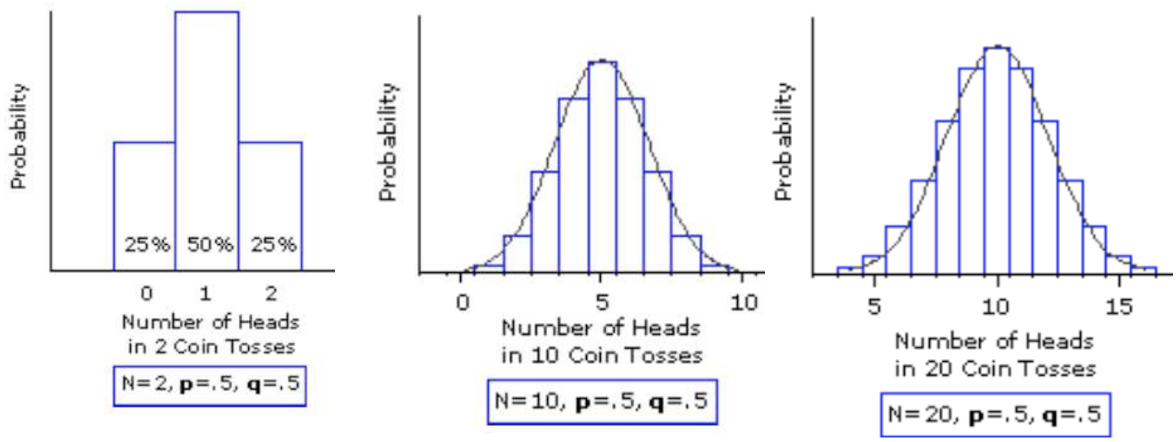
Test d'hypothèse nulle et signification statistique

Possible scénario
Test du pouvoir paranormal d'un médium prétendant être capable de biaiser mentalement un jeu de pile ou face vers l'événement PILE.

$p = q = .5$	les probabilités de PILE et FACE sous H_0
k	nombre observé de PILE
$N = 100$	nombre de lancers

Si $k < 50$, il a évidemment perdu, mais si $k > 0$?
Mettons $k = 59$

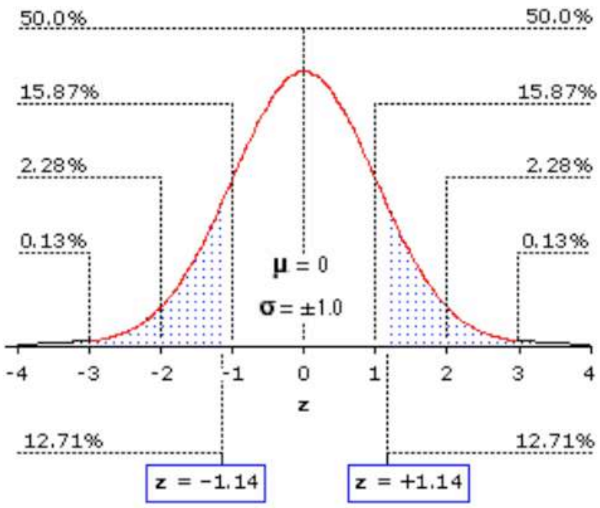
Binomial Sampling Distributions



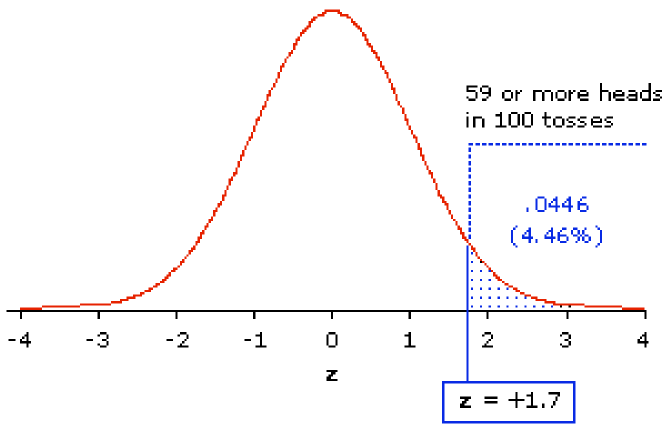
Test d’hypothèse nulle et signification statistique

Distribution normale réduite

***N* infini *et z* = (*x* – μ) / σ**



Test d’hypothèse nulle et signification statistique



Avec p = .5, q = .5, N = 100 et k = 59 on trouve z = +1.7

Table du z

$p(\text{DIH}_0)$ = .0446

Moins de 5% de chances d’obtenir par hasard une déviation au moins aussi grande
Statistiquement significatif → Convention: on rejette H_0

Plan

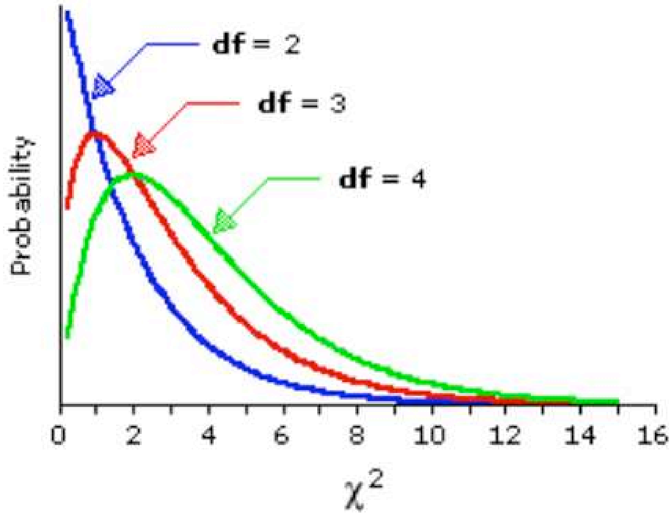
1. **Les probabilités dans la pratique de la recherche empirique**
   - Epistémologie : maths vs. science empirique
   - Théorie de la connaissance vs. histoire des sciences : Popper vs. Kuhn
   - Réalisme scientifique: du scepticisme humien au relativisme cognitif postmoderne
     - Le canular de Sokal, Sokal et Bricmont (1997)
     - Le générateur post-moderniste
   - Drame de l'induction, nécessaire et impossible
     - Popper: corroboration impossible, la réfutation seule inférence valide
     - Meehl (1997) : en fait même la réfutation est un problème
     - Sokal et Bricmont : réalisme modeste et enquête rationnelle
   - Disparition du modus tollens dans le raisonnement probabiliste
2. **Probabilité conditionnelle et statistiques inductives**
   - Probabilités vs. fréquences naturelles (Hoffrage et al., 2000)
   - Test d'hypothèse nulle et signification statistique
     - Le principe avec un exemple (Lowry, VassarStats)
     - L'objection: $p(D|H) \neq p(H|D$ (Cohen, 1994; Pollard & Richardson, 1987)
     - Inférence fisherienne vs. bayesienne : de l'habitude à la tradition et au dogme
3. **Equivoques de la formule *y* = f(*x*/*z*) pour un expérimentaliste**
   - Terminologie: taxonomie des variables
     - Niveaux de la mesure (Stevens, 1946)
     - Variables manipulées vs. enregistrées
     - Variables, paramètres, constantes
   - Contexte : conflit vitesse/précision et "loi" de Fitts
   - Signe "=" relation symétrique? Le problème du *paradigme expérimental*
   - Combien de quantités indépendantes dans x/z? description *polaire* vs. *cartésienne*
   - *x*/*z* vs. *z*/*x*: danger de l'inversion des rapports
4. **Deux problèmes de statistiques descriptives**
   - Représenter une loi empirique: courbe moyenne vs. courbe limite
   - Histogramme vs. courbe selon le niveau de la mesure en x et en y

		VARIABLE			
		MANIPULEE déterministe *Valeurs choisies par l'expérimentateur*		**ENREGISTREE aléatoire** *Valeurs lues sur un instrument de mesure*	
		Variable 'indépendante'	Paramètre expérimental	Variable 'dépendante'	Paramètre d'ajustement
NIVEAU DE LA MESURE S.S. Stevens (1946)	**Echelle nominale** *trier*				
	Echelle **ordinale** *ordonner*				
	Echelle **à intervalles égaux** *unité définie*				
	Echelle **de rapport** *zéro physique*				

Plan

1. **Les probabilités dans la pratique de la recherche empirique**
   - Epistémologie : maths vs. science empirique
   - Théorie de la connaissance vs. histoire des sciences : Popper vs. Kuhn
   - Réalisme scientifique: du scepticisme humien au relativisme cognitif postmoderne
     - Le canular de Sokal, Sokal et Bricmont (1997)
     - Le générateur post-moderniste
   - Drame de l'induction, nécessaire et impossible
     - Popper: corroboration impossible, la réfutation seule inférence valide
     - Meehl (1997) : en fait même la réfutation est un problème
     - Sokal et Bricmont : réalisme modeste et enquête rationnelle
   - Disparition du modus tollens dans le raisonnement probabiliste
2. **Probabilité conditionnelle et statistiques inductives**
   - Probabilités vs. fréquences naturelles (Hoffrage et al., 2000)
   - Test d'hypothèse nulle et signification statistique
     - Le principe avec un exemple (Lowry, VassarStats)
     - L'objection: $p(D|H) \neq p(H|D$ (Cohen, 1994; Pollard & Richardson, 1987)
     - Inférence fisherienne vs. bayesienne : de l'habitude à la tradition et au dogme
3. **Equivoques de la formule $y = f(x/z)$ pour un expérimentaliste**
   - Terminologie: taxonomie des variables
     - Niveaux de la mesure (Stevens, 1946)
     - Variables manipulées vs. enregistrées
     - Variables, paramètres, constantes
   - Contexte : conflit vitesse/précision et "loi" de Fitts
   - Signe "=" relation symétrique? Le problème du *paradigme expérimental*
   - Combien de quantités indépendantes dans x/z? description *polaire* vs. *cartésienne*
   - x/z vs. z/x: danger de l'inversion des rapports
4. **Deux problèmes de statistiques descriptives**
   - Représenter une loi empirique: courbe moyenne vs. courbe limite
   - Histogramme vs. courbe selon le niveau de la mesure en x et en y

Chi-Square Sampling Distributions for df=2, 3, and 4



2. Etude en cours du pointage (*simple rapid aiming*)

Plan pour moi

Loi de Fitts

- aperçu tâche, le modèle mathématique admis

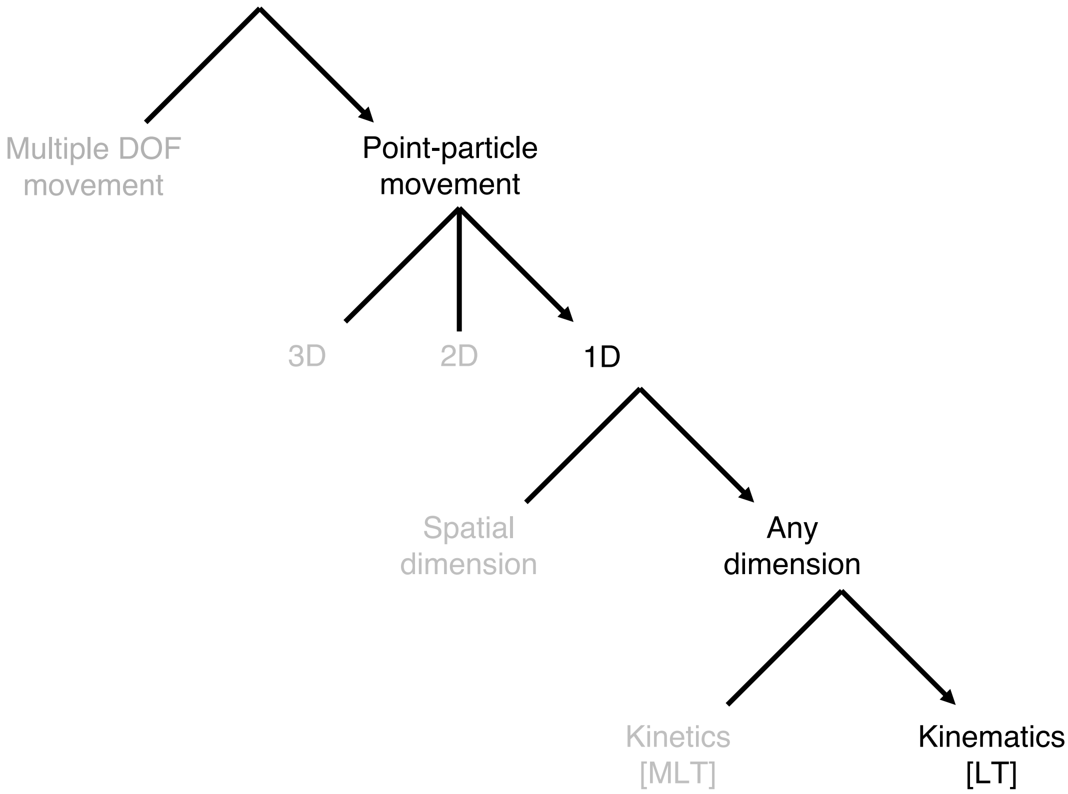
Signe égal et causalité $y = f(x)$ vs. $x = f^{-1}(y)$

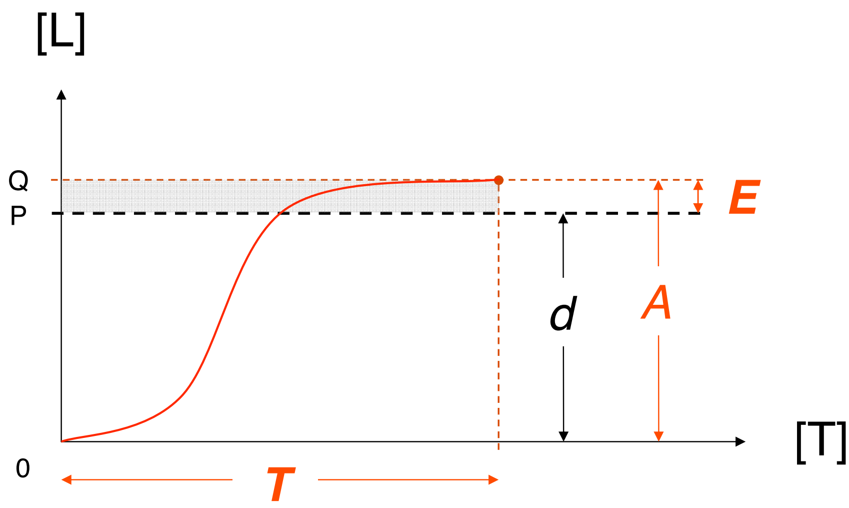
Inversion des rapports et niveau de la mesure σ_E / μ_A vs. μ_A / σ_E

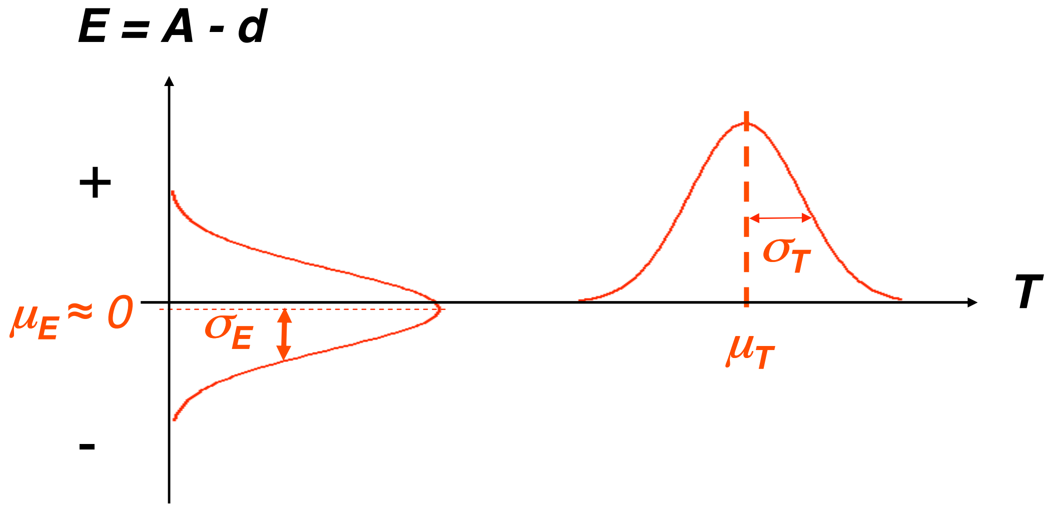
Interprétation polaire vs. cartésienne des écritures fractionaires

La loi comme une courbe limite: front convexe de performance

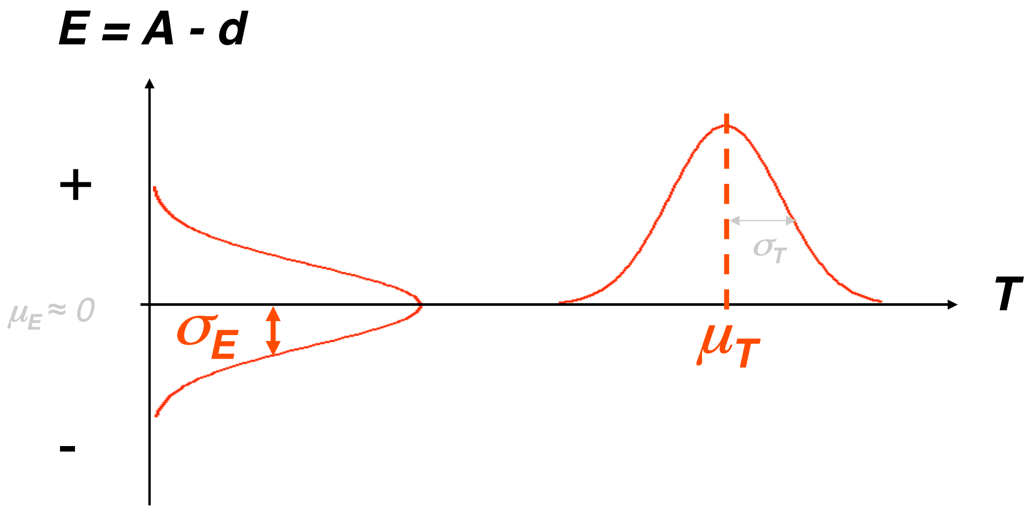
Inégalité de Jensen


Multiple DOF movement
Point-particle movement
3D
2D
1D
Spatial dimension
Any dimension
Kinetics [MLT]
Kinematics [LT]

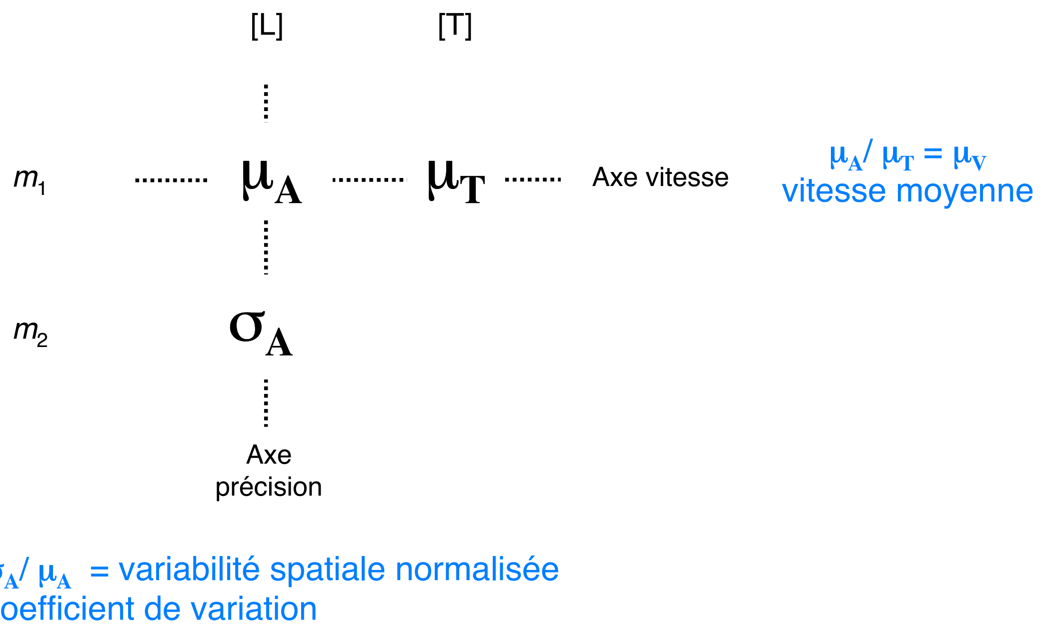

[L]
Q
P
E
d
A
0
T
[T]



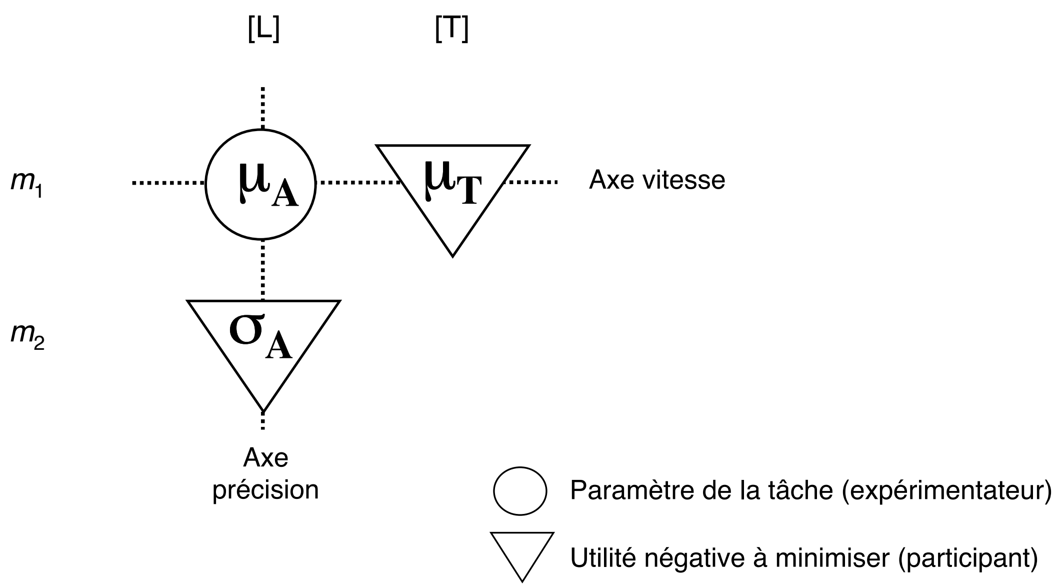
E = A - d
+
-
μE ≈ 0
σE
σT
μT
T



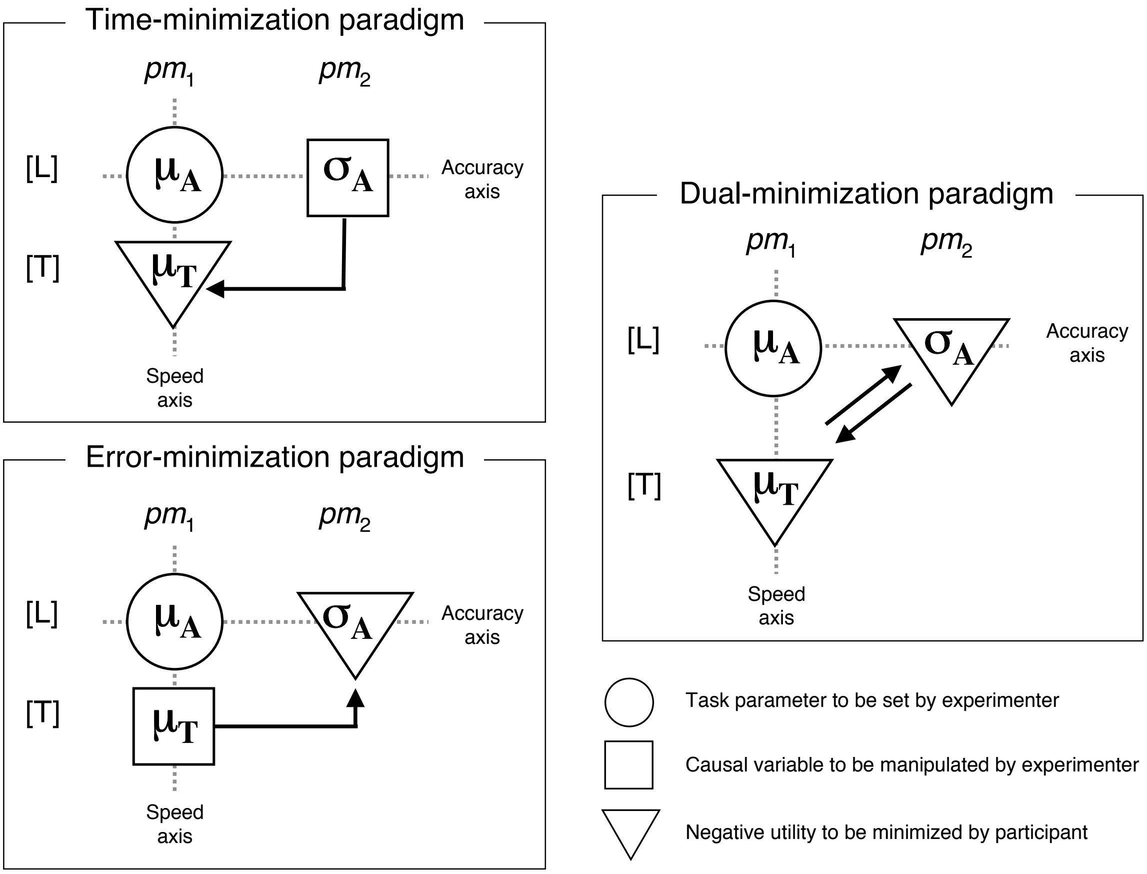
E = A - d
+
-
μE ≈ 0
σE
σT
μT
T

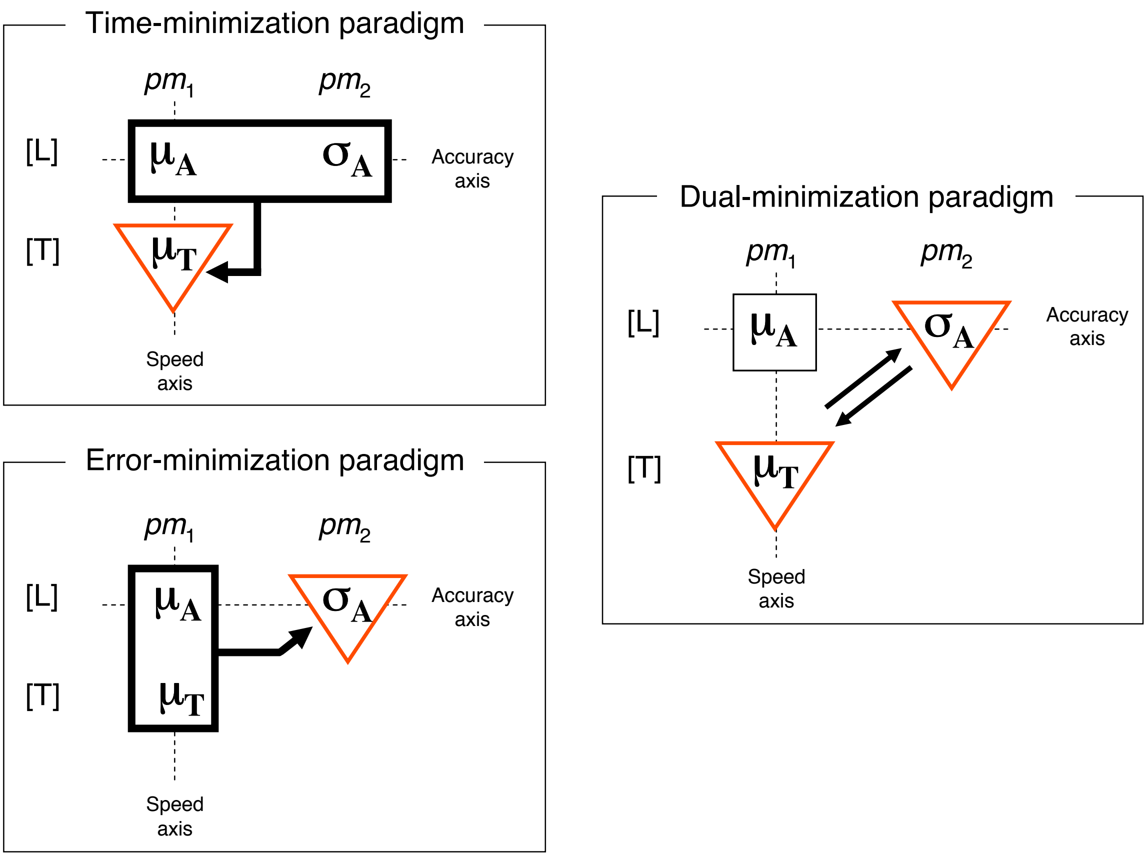


[L]
[T]
m_1
μ_A
μ_T
Axe vitesse
$\mu_A / \mu_T = \mu_V$
vitesse moyenne
m_2
σ_A
Axe
précision
σ_A / μ_A = variabilité spatiale normalisée
coefficient de variation



[L]
[T]
m_1
μ_A
μ_T
Axe vitesse
m_2
σ_A
Axe
précision
Paramètre de la tâche (expérimentateur)
Utilité négative à minimiser (participant)


Time-minimization paradigm
pm1
pm2
[L]
[T]
μA
σA
μT
Accuracy axis
Speed axis
Error-minimization paradigm
pm1
pm2
[L]
[T]
μA
σA
μT
Accuracy axis
Speed axis
Dual-minimization paradigm
pm1
pm2
[L]
[T]
μA
σA
μT
Accuracy axis
Speed axis
Task parameter to be set by experimenter
Causal variable to be manipulated by experimenter
Negative utility to be minimized by participant


Time-minimization paradigm
pm1
pm2
[L]
[T]
μA
σA
μT
Accuracy axis
Speed axis
Error-minimization paradigm
pm1
pm2
[L]
[T]
μA
σA
μT
Accuracy axis
Speed axis
Dual-minimization paradigm
pm1
pm2
[L]
[T]
μA
σA
μT
Accuracy axis
Speed axis

Références

Hoffrage, U., Lindsay, S., Hertwig, R., & Gigerenzzer, G. (2000). Communicating statistical information. *Science 290*, 2261-2262. *La Recherche HS N°13*, 2003.

Stages LIESSE
2012
à Télécom ParisTech

Mathématiques, physique, informatique et leurs interactions
avec les sciences et technologies de l'information.

Journées de formation à destination des professeurs de Classes Préparatoires aux Grandes Écoles

- ***Enseigner les probabilités en grande École... et bientôt en Classes Préparatoires ?***

  mercredi 9 et **jeudi 10 mai 2012**

  - journées Télécom-UPS -

  (mercreci 9 mai à 17h30 : conférence invitée à l'occasion du bicentenaire de la disparition d'Henri Poincaré)

- ***Représenter l'information sous forme quantique : des concepts fondamentaux aux applications pour le traitement de l'information.***

  vendredi 11 mai 2012

- ***Les enjeux scientifiques et technologiques de l'imagerie satellitaire et de l'imagerie médicale.***

  lundi 14 et **mardi 15 mai 2012**

www.telecom-paristech.fr/liesse/

Contact : liesse@telecom-paristech.fr

Télécom ParisTech
46 rue Barrault
75013 Paris
www. telecom-paristech.fr