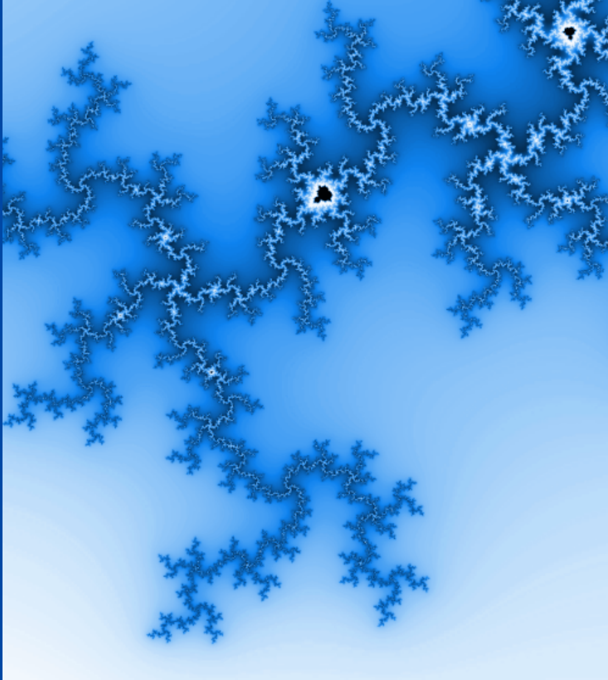
Excursion aléatoire dans l'histoire des probabilités

Rencontre annuelle de l'AMUlg

Laurent Loosveldt

16 avril 2025

LIÈGE université
Mathématique



Clause de non-responsabilité

Clause de non-responsabilité

Je suis **mathématicien**.

Clause de non-responsabilité

J'**aime beaucoup** l'histoire.

Clause de non-responsabilité

Je suis **mathématicien**.

Axiomatique de Kolmogorov (1933)

Version originale

Soit E un ensemble d'éléments $\xi, \eta, \zeta, \ldots$, que nous appellerons événements élémentaires, et $\mathscr{F}$ une famille de sous-ensembles de E, les éléments de $\mathscr{F}$ seront appelés événements aléatoires.

1. $\mathscr{F}$ est une famille d'ensembles
2. $\mathscr{F}$ contient Ω
3. À chaque ensemble A de $\mathscr{F}$ est assigné un nombre non-négatif $\mathbb{P}(A)$. Ce nombre $\mathbb{P}(A)$ est appelé probabilité de l'événement A
4. $\mathbb{P}(E) = 1$
5. Si A et B n'ont pas d'éléments en commun, alors $\mathbb{P}(A+B) = \mathbb{P}(A) + \mathbb{P}(B)$.

Un système d'ensembles $\mathscr{F}$, associé à un assignement défini de nombres $\mathbb{P}(A)$, satisfaisant les axiomes 1 à 5 est appelé un espace de probabilités.

Axiomatique de Kolmogorov (1933)

Aujourd'hui, on écrit plutôt

Soit $\mathscr{F}$ une σ-algèbre sur un espace fondamental Ω. Une mesure de probabilité sur $(\Omega, \mathscr{F})$ est une application

$$\mathbb{P} : \mathscr{F} \to [0,1]$$

telle que

(a) l'espace fondamental a une probabilité 1: $\mathbb{P}(\Omega) = 1$,

(b) σ-additivité: si $(A_j)_j$ est une suite (finie ou infinie) d'événements deux à deux disjoints de $\mathscr{F}$, alors

$$\mathbb{P}\left(\bigcup_j A_j\right) = \sum_j \mathbb{P}(A_j). \tag{1}$$

Axiomatique de Kolmogorov (1933)

σ-algèbre

Soit Ω un ensemble. Une *σ-algèbre* $\mathscr{F}$ sur Ω est une partie de $\mathscr{P}(\Omega)$ qui satisfait les conditions suivantes :

(a) $\Omega \in \mathscr{F}$,

(b) Stabilité par passage au complémentaire: si $A \in \mathscr{F}$, alors $A^c \in \mathscr{F}$,

(c) Stabilité par union dénombrable: si $(A_j)_j$ est une suite de sous-ensembles de $\mathscr{F}$, alors $\bigcup_j A_j \in \mathscr{F}$.

Axiomatique de Kolmogorov (1933)

Aujourd'hui, on écrit plutôt

Soit $\mathscr{F}$ une σ-algèbre sur un espace fondamental Ω. Une mesure de probabilité sur $(\Omega, \mathscr{F})$ est une application

$$\mathbb{P} : \mathscr{F} \to [0,1]$$

telle que

(a) l'espace fondamental a une probabilité 1: $\mathbb{P}(\Omega) = 1$,

(b) σ-additivité: si $(A_j)_j$ est une suite (finie ou infinie) d'événements deux à deux disjoints de $\mathscr{F}$, alors

$$\mathbb{P}\left(\bigcup_j A_j\right) = \sum_j \mathbb{P}(A_j). \tag{1}$$

Motivations

Motivations

- L'exemple de la géométrie et l'émergence de géométrie non-euclidienne.

Motivations

- L'exemple de la géométrie et l'émergence de géométrie non-euclidienne.
- Le 6ème problème de Hilbert :

> Traitement mathématique des axiomes en physique. Les investigations sur les fondements de la géométrie suggèrent le problème : Traiter de la même manière, au moyen d'axiomes, les sciences physiques dans lesquelles les mathématiques jouent déjà aujourd'hui un rôle important ; au premier rang figurent la théorie des probabilités et la mécanique.

Motivations

- L'exemple de la géométrie et l'émergence de géométrie non-euclidienne.
- Le 6ème problème de Hilbert :

  Traitement mathématique des axiomes en physique. Les investigations sur les fondements de la géométrie suggèrent le problème : Traiter de la même manière, au moyen d'axiomes, les sciences physiques dans lesquelles les mathématiques jouent déjà aujourd'hui un rôle important ; au premier rang figurent la théorie des probabilités et la mécanique.

  Norbert Wiener considérait que les probabilités étaient le pont entre les mathématiques et la physique

Motivations

- L'exemple de la géométrie et l'émergence de géométrie non-euclidienne.
- Le 6ème problème de Hilbert :

> Traitement mathématique des axiomes en physique. Les investigations sur les fondements de la géométrie suggèrent le problème : Traiter de la même manière, au moyen d'axiomes, les sciences physiques dans lesquelles les mathématiques jouent déjà aujourd'hui un rôle important ; au premier rang figurent la théorie des probabilités et la mécanique.

Norbert Wiener considérait que les probabilités étaient le pont entre les mathématiques et la physique
La théorie de la relativité et la mécanique quantique ont rendu le problème obsolète.

Motivations

- L'exemple de la géométrie et l'émergence de géométrie non-euclidienne.
- Le 6ème problème de Hilbert :
- Emile Borel et ensuite et surtout Henri Lebesgue avaient déjà fait une bonne partie du travail !

Deux remarques de Kolmogorov

Le système d'axiomes est consistant : $\mathscr{F} = \{\emptyset, E\}$, $\mathbb{P}(\emptyset) = 0$ et $\mathbb{P}(E) = 1$.

Deux remarques de Kolmogorov

Le système d'axiomes est consistant : $\mathscr{F} = \{\emptyset, E\}$, $\mathbb{P}(\emptyset) = 0$ et $\mathbb{P}(E) = 1$.

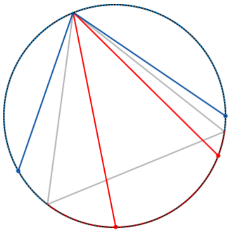
Le système d'axiomes n'est pas complet : dans divers problèmes de la théorie des probabilités, différents espaces de probabilités sont à examiner.

Paradoxe de Bertrand

Dans un cercle de rayon 1, le côté d'un triangle équilatéral inscrit a pour longueur $\sqrt{3}$. Quelle est la probabilité qu'une corde d'un cercle, choisie au hasard, soit d'une longueur supérieur à $\sqrt{3}$?

Paradoxe de Bertrand

Quelle est la probabilité qu'une corde d'un cercle, choisie au hasard, soit d'une longueur supérieur à $\sqrt{3}$? Première proposition : choix aléatoire des extrémités

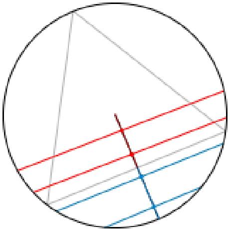


1. Choisir un point sur le cercle et tracer le triangle équilatéral inscrit dont ce point est un des sommets.
2. Choisir aléatoire un autre point sur le cercle et donc la corde qui relie ces deux points.
3. La longueur de cette corde est supérieure à $\sqrt{3}$ si sa deuxième extrémité est sur l'arc entre les deux sommets opposés à la première extrémité.

$\mathbb{P}$(avoir une corde de longeur supérieure à $\sqrt{3}) = 1/3$.

Paradoxe de Bertrand

Quelle est la probabilité qu'une corde d'un cercle, choisie au hasard, soit d'une longueur supérieur à $\sqrt{3}$?
Deuxième proposition : choix aléatoire du rayon perpendiculaire

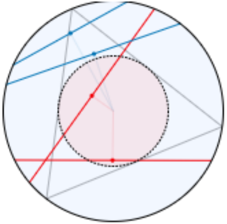


1. Chosir un rayon du cercle et tracer le triangle équilatéral inscrit dont un côtés est perpendiculaire à ce rayon.
2. On considère toutes les cordes dont le milieu est situé sur ce rayon.
3. La longueur de cette corde est supérieure à $\sqrt{3}$ si la corde est au-dessus du côté perpendiculaire du traingle équilatéral.

$\mathbb{P}$(avoir une corde de longeur supérieure à $\sqrt{3}$) $= 1/2$.

Paradoxe de Bertrand

Quelle est la probabilité qu'une corde d'un cercle, choisie au hasard, soit d'une longueur supérieur à $\sqrt{3}$?
Troisième proposition : choix aléatoire du milieu de la corde



1. Choisir un point dans le disque et considérer la corde dont il est le milieu (ou un diamètre si ce point est le centre).
2. On montre que la longueur de cette corde est supérieur à $\sqrt{3}$ si son milieu est dans le disque concentrique de rayon $1/2$.

$$\mathbb{P}(\text{avoir une corde de longeur supérieure à } \sqrt{3}) = 1/4.$$

Paradoxe de Bertrand-Explications

- la locution “Choisir une corde au hasard” est ambiguë vu qu’elle ne précise la méthode de sélection;

Paradoxe de Bertrand-Explications

- la locution “Choisir une corde au hasard” est ambiguë vu qu'elle ne précise la méthode de sélection;
- nous avons présenté trois méthodes de sélection différente;

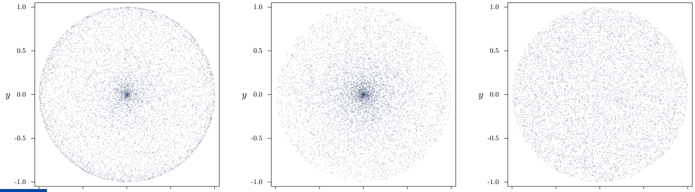
Paradoxe de Bertrand-Explications

- la locution "Choisir une corde au hasard" est ambiguë vu qu'elle ne précise la méthode de sélection;
- nous avons présenté trois méthodes de sélection différente;
- chaque méthode est associée à un espace probabilisé qui lui est propre et donc une solution qui lui est propre.

Paradoxe de Bertrand-Explications

- la locution “Choisir une corde au hasard” est ambiguë vu qu’elle ne précise la méthode de sélection;
- nous avons présenté trois méthodes de sélection différente;
- chaque méthode est associée à un espace probabilisé qui lui est propre et donc une solution qui lui est propre.

Chaque corde qui n’est pas un diamètre est entièrement déterminée par son milieu. Chaque méthode de sélection correspond à une distribution de ces milieux, qui lui est propre.



Jeux de dés

Jeux de dés

Figure: En Mésopotamie, dès le 3ème millénaire avant J.-C.

Jeux de dés

Figure: En Égypte, au 2ème millénaire avant J.-C., avec des dés équilibrés

Jeux de dés

Figure: En Grèce, au 4ème siècle avant J.-C., Aristote distingue la logique du fortuit

Jeux de dés

Figure: On a retrouvé des dés pipés datant de la Rome antique.

Jeux de dés

Figure: Au Moyen-Age, les jeux de dés étaient très populaires

Jeux de dés

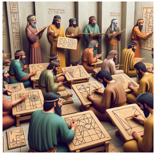
Figure: Contre l'avis de l'Eglise catholique

Jeux de dés

Jeux de dés









Jeux de dés

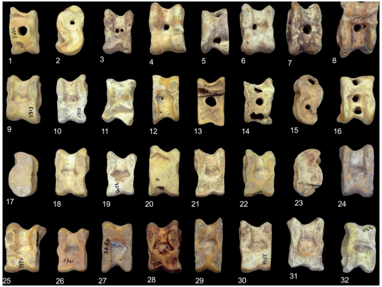


Figure: Les premiers dés n'étaient pas vraiment équilibrés

Jeux de dés

Jeux de dés

Figure: Le hasard est l'oeuvre de Dieu ou, pire, du Diable !

Et les statistiques ?

Et les statistiques ?

Figure: Le recensement des hommes lors de l'exode hors d'Égypte

Et les statistiques ?

Figure: Le dénombrement de Bethléem (selon Bruegel)

Et les statistiques ?

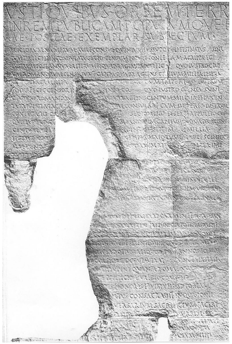


Figure: Les actes du divin Auguste (14)

Et les statistiques ?

Figure: Le livre du Jugement dernier de Guillaume le conquérant (11ème siècle)

Premières estimations

Premières estimations

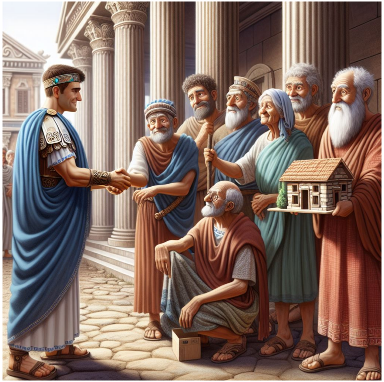


Figure: 3ème siècle : table d'estimation des rentes viagères

Premières estimations

Figure: 13ème-14ème siècle : assurance maritime

Les premiers écrits - Cardan

Figure: Jérôme Cardan (1501-1576)

Les premiers écrits - Cardan

Cardan écrit entre 1526 et 1560 *Liber de ludo aleae*, essentiellement un traité d'analyse combinatoire, accompagné de réflexions morales sur les jeux et la triche.

Les premiers écrits - Cardan

Cardan écrit entre 1526 et 1560 *Liber de ludo aleae*, essentiellement un traité d'analyse combinatoire, accompagné de réflexions morales sur les jeux et la triche. Quelques faits intéressants :

- il étudie les différentes combinaisons possibles et illustre ces affirmations via des dés (distinction de la théorie et de l'observation)

Les premiers écrits - Cardan

Cardan écrit entre 1526 et 1560 *Liber de ludo aleae*, essentiellement un traité d'analyse combinatoire, accompagné de réflexions morales sur les jeux et la triche. Quelques faits intéressants :

- il étudie les différentes combinaisons possibles et illustre ces affirmations via des dés (distinction de la théorie et de l'observation)
- il parle de dés truqués et de dés honnêtes (approche fréquentiste des probabilités)

Les premiers écrits - Cardan

Cardan écrit entre 1526 et 1560 *Liber de ludo aleae*, essentiellement un traité d'analyse combinatoire, accompagné de réflexions morales sur les jeux et la triche. Quelques faits intéressants :

- il étudie les différentes combinaisons possibles et illustre ces affirmations via des dés (distinction de la théorie et de l'observation)
- il parle de dés truqués et de dés honnêtes (approche fréquentiste des probabilités)
- "équipossibilité", "égale possibilité" (formes anciennes d'équiprobable)

Les premiers écrits - Cardan

Cardan écrit entre 1526 et 1560 *Liber de ludo aleae*, essentiellement un traité d'analyse combinatoire, accompagné de réflexions morales sur les jeux et la triche. Quelques faits intéressants :

- il étudie les différentes combinaisons possibles et illustre ces affirmations via des dés (distinction de la théorie et de l'observation)
- il parle de dés truqués et de dés honnêtes (approche fréquentiste des probabilités)
- "équipossibilité", "égale possibilité" (formes anciennes d'équiprobable)
- "Aussi il y a une règle générale, que nous devons considérer le circuit entier, et le nombre de ces lancers qui représente en combien de façons les résultats favorables peuvent se produire, et comparer ce nombre au reste du circuit, et les paris mutuels devront être posés selon cette proportion, de sorte qu'on puisse disputer en termes égaux." (nombre de cas favorables sur le nombre de cas possibles)

Les premiers écrits - Galilée

A la cour de Florence, on s'amusait régulièrement avec des jeux de dés. Dans l'un de ceux-ci, les joueurs devaient parier sur la somme des 3 dès à 6 faces. Après avoir observé de nombreuses parties, le Duc de Toscane s'adressa à Galilée (1564-1642) pour comprendre pourquoi le résultat 10 apparaissait plus régulièrement que 9.

Paradoxe du Duc de Toscane

Le Duc fit remarquer qu'il y avait autant de possibilités pour décomposer 10 et 9 comme somme de 3 nombres choisis parmi $\{1,2,3,4,5,6\}$:

$$10 = \begin{cases} 6+3+1 \\ 6+2+2 \\ 5+4+1 \\ 5+3+2 \\ 4+4+2 \\ 4+3+3 \end{cases} \qquad 9 = \begin{cases} 6+2+1 \\ 5+3+1 \\ 5+2+2 \\ 4+4+1 \\ 4+3+2 \\ 3+3+3. \end{cases}$$

Paradoxe du Duc de Toscane

Le Duc fit remarquer qu'il y avait autant de possibilités pour décomposer 10 et 9 comme somme de 3 nombres choisis parmi $\{1,2,3,4,5,6\}$:

$$10 = \begin{cases} 6+3+1 \\ 6+2+2 \\ 5+4+1 \\ 5+3+2 \\ 4+4+2 \\ 4+3+3 \end{cases} \qquad 9 = \begin{cases} 6+2+1 \\ 5+3+1 \\ 5+2+2 \\ 4+4+1 \\ 4+3+2 \\ 3+3+3. \end{cases}$$

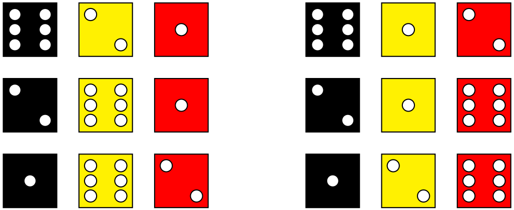
Il interrogea Galilée pour comprendre pourquoi alors est-ce que le 10 apparaissait plus souvent?

Analyse des cas favorables

$$9 = \begin{cases} 6+2+1 \\ 5+3+1 \\ 5+2+2 \\ 4+4+1 \\ 4+3+2 \\ 3+3+3. \end{cases}$$

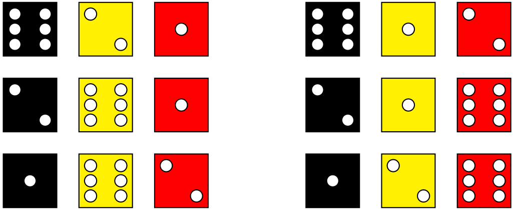
Analyse des cas favorables

$$9 = \begin{cases} \color{blue}{6+2+1} \\ 5+3+1 \\ 5+2+2 \\ 4+4+1 \\ 4+3+2 \\ 3+3+3. \end{cases}$$



Analyse des cas favorables

$$9 = \begin{cases} 6+2+1 \rightarrow \text{6 lancers} \\ 5+3+1 \\ 5+2+2 \\ 4+4+1 \\ 4+3+2 \\ 3+3+3. \end{cases}$$

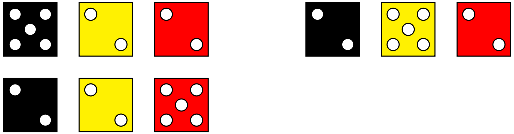


Analyse des cas favorables

$$9 = \begin{cases} 6+2+1 \rightarrow \text{6 lancers} \\ 5+3+1 \rightarrow \text{6 lancers} \\ 5+2+2 \\ 4+4+1 \\ 4+3+2 \rightarrow \text{6 lancers} \\ 3+3+3. \end{cases}$$

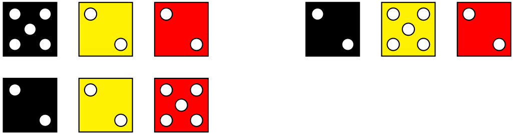
Analyse des cas favorables

$$9 = \begin{cases} 6+2+1 \textcolor{orange}{\longrightarrow \text{6 lancers}} \\ 5+3+1 \textcolor{orange}{\longrightarrow \text{6 lancers}} \\ \textcolor{blue}{5+2+2} \\ 4+4+1 \\ 4+3+2 \textcolor{orange}{\longrightarrow \text{6 lancers}} \\ 3+3+3. \end{cases}$$



Analyse des cas favorables

$$9 = \begin{cases} 6+2+1 \longrightarrow \text{6 lancers} \\ 5+3+1 \longrightarrow \text{6 lancers} \\ 5+2+2 \longrightarrow \text{3 lancers} \\ 4+4+1 \\ 4+3+2 \longrightarrow \text{6 lancers} \\ 3+3+3. \end{cases}$$

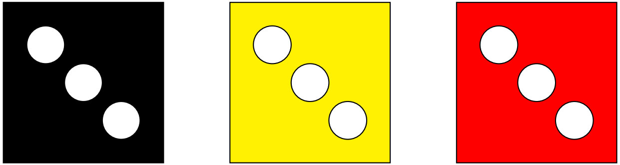


Analyse des cas favorables

$$9 = \begin{cases} 6+2+1 \rightarrow \text{6 lancers} \\ 5+3+1 \rightarrow \text{6 lancers} \\ 5+2+2 \rightarrow \text{3 lancers} \\ 4+4+1 \rightarrow \text{3 lancers} \\ 4+3+2 \rightarrow \text{6 lancers} \\ 3+3+3. \end{cases}$$

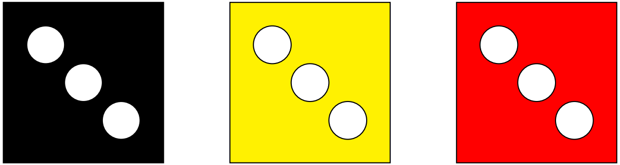
Analyse des cas favorables

$$9 = \begin{cases} 6+2+1 \rightarrow \text{6 lancers} \\ 5+3+1 \rightarrow \text{6 lancers} \\ 5+2+2 \rightarrow \text{3 lancers} \\ 4+4+1 \rightarrow \text{3 lancers} \\ 4+3+2 \rightarrow \text{6 lancers} \\ 3+3+3. \end{cases}$$



Analyse des cas favorables

$$9 = \begin{cases} 6+2+1 \rightarrow \text{6 lancers} \\ 5+3+1 \rightarrow \text{6 lancers} \\ 5+2+2 \rightarrow \text{3 lancers} \\ 4+4+1 \rightarrow \text{3 lancers} \\ 4+3+2 \rightarrow \text{6 lancers} \\ 3+3+3 \rightarrow \text{1 lancer.} \end{cases}$$



Analyse des cas favorables

$$9 = \begin{cases} 6+2+1 \rightarrow 6 \text{ lancers} \\ 5+3+1 \rightarrow 6 \text{ lancers} \\ 5+2+2 \rightarrow 3 \text{ lancers} \\ 4+4+1 \rightarrow 3 \text{ lancers} \\ 4+3+2 \rightarrow 6 \text{ lancers} \\ 3+3+3 \rightarrow 1 \text{ lancer}. \end{cases}$$

Total : 25 lancers

Analyse des cas favorables

$$9 = \begin{cases} 6+2+1 \rightarrow \text{6 lancers} \\ 5+3+1 \rightarrow \text{6 lancers} \\ 5+2+2 \rightarrow \text{3 lancers} \\ 4+4+1 \rightarrow \text{3 lancers} \\ 4+3+2 \rightarrow \text{6 lancers} \\ 3+3+3 \rightarrow \text{1 lancer.} \end{cases}$$

Total : 25 lancers

$$10 = \begin{cases} 6+3+1 \rightarrow \text{6 lancers} \\ 6+2+2 \rightarrow \text{3 lancers} \\ 5+4+1 \rightarrow \text{6 lancers} \\ 5+3+2 \rightarrow \text{6 lancers} \\ 4+4+2 \rightarrow \text{3 lancers} \\ 4+3+3. \rightarrow \text{3 lancers} \end{cases}$$

Total : 27 lancers

La correspondance de Fermat et Pascal (1654)


- Fermat (1601-1665) : discret mais prolifique tout au long de sa vie, on lui doit de nombreuses contributions en géométrie, arithmétique, optique,...
- Pascal (1623-1662) : gravement malade, sa vie oscille entre les mathématiques, la méditation religieuse et la philosophie.

La correspondance de Fermat et Pascal (1654)

- Fermat (1601-1665) : discret mais prolifique tout au long de sa vie, on lui doit de nombreuses contributions en géométrie, arithmétique, optique,...
- Pascal (1623-1662) : gravement malade, sa vie oscille entre les mathématiques, la méditation religieuse et la philosophie.

En 1654, le chevalier de Méré s'adresse à Pascal au sujet du "Problème des partis". Pascal entame alors une correspondance avec Fermat autour de ce problème.

Le problème des partis

Le problème des partis

Le problème des partis

Le problème des partis

Règles du jeu :

- chaque joueur mise une somme m d'argent sur sur un numéro entre 1 et 6;
- on lance successivement un dé;
- le premier joueur dont le numéro est apparu lors de 3 lancers remporte la mise totale $2m$.

Le problème des partis

- Pascal écrit directement une solution dans le cas le plus simple

Le joueur A a remporté 2 parties contre 1 partie pour le joueur B

Le problème des partis

- Pascal écrit directement une solution dans le cas le plus simple

Le joueur A a remporté 2 parties contre 1 partie pour le joueur B

A remporte le jeu

B égalise

Le problème des partis

- Pascal écrit directement une solution dans le cas le plus simple

Le joueur A a remporté 2 parties contre 1 partie pour le joueur B

A remporte le jeu

B égalise

$2m$ pour A et 0 pour B

Le problème des partis

- Pascal écrit directement une solution dans le cas le plus simple

Le joueur A a remporté 2 parties contre 1 partie pour le joueur B

A remporte le jeu $\rightarrow$ $2m$ pour A et 0 pour B

B égalise $\rightarrow$ m pour chaque joueur

Le problème des partis

- Pascal écrit directement une solution dans le cas le plus simple

Le joueur A a remporté 2 parties contre 1 partie pour le joueur B

A remporte le jeu → $2m$ pour A et 0 pour B

B égalise → m pour chaque joueur

Le joueur A peut *espérer* gagner $\frac{3m}{2}$ alors que le joueur peut espérer gagner $\frac{m}{2}$.

Le problème des partis

- Pascal écrit directement une solution dans le cas le plus simple.
- Ce qui importe c'est le nombre maximum de parties qu'il reste à jouer.

Le problème des partis

- Pascal écrit directement une solution dans le cas le plus simple.
- Ce qui importe c'est le nombre maximum de parties qu'il reste à jouer.
- Fermat utilise les combinaisons pour dénombrer les cas possibles et favorables, en fonction du nombre maximum de parties à jouer.

Le problème des partis

- Pascal écrit directement une solution dans le cas le plus simple.
- Ce qui importe c'est le nombre maximum de parties qu'il reste à jouer.
- Fermat utilise les combinaisons pour dénombrer les cas possibles et favorables, en fonction du nombre maximum de parties à jouer.
- Pascal regroupe et formalise ces résultats grâce au triangle de Pascal.

Le problème des partis

Mathématiquement, dans leurs raisonnements, on trouve

- la notion d'équiprobabilité
- le calcul d'espérance
- les propriétés d'additivité et de multiplication pour les événements indépendants

Le premier traité publié

Christiaan Huygens (1629-1695) lors d'un séjour à Paris en 1655 a vent des travaux de Fermat et Pascal qui éveillent son intérêt. Il publie le premier traité sur les probabilités en 1657.

Le premier traité publié

- Formalisation de l'espérance

  " Quoique dans les jeux de hasard pur les résultats soient incertains, la chance qu'un joueur a de gagner ou de perdre a cependant une valeur déterminée".
- Résolution récursive, sans combinaisons, du problème des partis.
- Il résout des problèmes dans lesquels le nombre de parties n'est pas limité, via des équations de récurrence.

Alors qui est le premier ?

si on se focalise sur

- la date d'écriture : Cardan;
- l'influence : Fermat et Pascal;
- la date de publication : Huygens.

Alors qui est le premier ?

si on se focalise sur

- la date d'écriture : Cardan;
- l'influence : Fermat et Pascal;
- la date de publication : Huygens.

Remarque : le terme *probabilité* n'apparait dans aucun de ces travaux, on parle plutôt de "nombre de chances" ou "nombre de hasards".

Alors qui est le premier ?

si on se focalise sur

- la date d'écriture : Cardan;
- l'influence : Fermat et Pascal;
- la date de publication : Huygens.

Remarque : le terme *probabilité* n'apparait dans aucun de ces travaux, on parle plutôt de "nombre de chances" ou "nombre de hasards".
À cette époque "probabilité" avait un sens juridique, il s'agissait d'apprécier les preuves, indices et témoignages lors d'un procès (en latin *probare* signifie "prouver").

Alors qui est le premier ?

si on se focalise sur

- la date d'écriture : Cardan;
- l'influence : Fermat et Pascal;
- la date de publication : Huygens.

Remarque : le terme *probabilité* n'apparait dans aucun de ces travaux, on parle plutôt de "nombre de chances" ou "nombre de hasards".
À cette époque "probabilité" avait un sens juridique, il s'agissait d'apprécier les preuves, indices et témoignages lors d'un procès (en latin *probare* signifie "prouver").
En 1662, on trouve le premier écrit dans lequel "probable" est utilisé au sens actuel.

Le traité d'Huygens reste unique en son genre jusqu'à 18ème siècle.

Le traité d'Huygens reste unique en son genre jusqu'à 18ème siècle.
Tentatives d'explication :

- les mathématiciens de la fin du 17ème siècle sont déjà bien occupés à développer le calcul différentiel et intégral;
- pas d'applications en sciences.

Le traité d'Huygens reste unique en son genre jusqu'à 18ème siècle.
Tentatives d'explication :

- les mathématiciens de la fin du 17ème siècle sont déjà bien occupés à développer le calcul différentiel et intégral;
- pas d'applications en sciences.

Figure: À la fin du 17ème siècle, Graunt, Huygens, Leibniz dressent de nombreuses tables de naissance et de mortalité.

Le traité d'Huygens reste unique en son genre jusqu'à 18ème siècle.
Tentatives d'explication :

- les mathématiciens de la fin du 17ème siècle sont déjà bien occupés à développer le calcul différentiel et intégral;
- pas d'applications en sciences.

Figure: On y constante qu'il naît légèrement plus de garçons que de filles, Arbuthnot soutient qu'il s'agit d'un signe de la divine Providence.

Pierre Rémond de Montmort (1678-1719)

- Auteur du deuxième traité (publié) concernant les probabilités, principalement consacré à des jeux de dés et de cartes dans sa première édition (1708) et complété d'un traité des combinaisons dans la seconde édition (1713)
- Large correspondance avec la famille Bernoulli.

Paradoxe de St-Pétersbourg

C'est l'extension d'un problème proposé par Nicolas Bernoulli à Montmort dans leur correspondance.

Figure: Nicolas Bernoulli

Paradoxe de St-Pétersbourg

Un joueur joue contre une banque.

- Pour pouvoir jouer, le joueur offre une mise initiale conservée par la banque;
- on lance une pièce; le jeu se poursuit tant que la pièce tombe sur *pile* et se termine quand elle tombe sur *face*, à ce moment la banque paie son gain au joueur
- le gain initial est 1 et est doublé à chaque tour : le joueur gagne 2^{n-1} si *face* apparait après n lancers

Paradoxe de St-Pétersbourg

L'espérance de gain pour le joueur est donc calculé (avec les notations modernes) via

$$\begin{aligned}
&\sum_{n=1}^{+\infty} 2^{n-1}\mathbb{P}(\text{ \textit{face} apparait pour la première fois au lancer } n)\\
&= \sum_{n=1}^{+\infty} 2^{n-1}\frac{1}{2^n}\\
&= \sum_{n=1}^{+\infty} \frac{1}{2}\\
&= \infty.
\end{aligned}$$

Paradoxe de St-Pétersbourg

Jacques Bernoulli, quelques années auparavant, était le premier à avoir utilisé des séries pour calculer des probabilités.

Figure: Jacques Bernoulli

Paradoxe de St-Pétersbourg

Alors qu'il semble extrêmement risqué de miser toute sa fortune dans ce jeu, l'espérance de gain y est pourtant infinie !

En particulier, le jeu n'est pas équitable.

Paradoxe de St-Pétersbourg

Parallèlement à cela, Jacques Bernoulli publie en 1713 un résultat qui est aujourd'hui compris dans la loi des grands nombres.

Loi (faible) des grands nombres

Soient $(X_j)_j$ des variables aléatoires intégrables i.i.d d'espérance μ. Alors pour tout $\varepsilon > 0$,

$$\lim_{n\to+\infty} \mathbb{P}\left(\left|\frac{X_1+\cdots+X_n}{n} - \mu\right| \geq \varepsilon\right) = 0.$$

Paradoxe de St-Pétersbourg

Parallèlement à cela, Jacques Bernoulli publie en 1713 un résultat qui est aujourd'hui compris dans la loi des grands nombres.

Loi (faible) des grands nombres

Soient $(X_j)_j$ des variables aléatoires intégrables i.i.d d'espérance μ. Alors pour tout $\varepsilon > 0$,

$$\lim_{n\to+\infty} \mathbb{P}\left(\left|\frac{X_1+\cdots+X_n}{n} - \mu\right| \geq \varepsilon\right) = 0.$$

Ce théorème **n'est pas applicable** lorsque la loi commune est donnée par

$$\mathbb{P}(X = 2^{k-1}) = \frac{1}{2^k},$$

qui correspond au jeu en question, puisque $\mathbb{E}[X] = +\infty$.

Paradoxe de St-Pétersbourg

Daniel Bernoulli (1700-1782), dans ses travaux, distingue l'espérance mathématique de l'espérance morale pour apporter une solution au paradoxe.

Figure: Daniel Bernoulli

Paradoxe de St-Pétersbourg

Daniel Bernoulli (1700-1782), dans ses travaux, distingue l'espérance mathématique de l'espérance morale pour apporter une solution au paradoxe.

Il s'agit de trouver une fonction "d'utilité" qui lie la différence de richesse du joueur entre avant et après le jeu et sa volonté de prendre part au jeu.

Daniel Bernoulli suggère de considérer

$$\ln(r + 2^{n-1} - c) - \ln(r).$$

où

- r est la richesse initiale du joueur;
- c est le coût d'entrée dans le jeu;
- 2^{n-1} est le gain si *pile* apparait après n lancés

Paradoxe de St-Pétersbourg

Daniel Bernoulli (1700-1782), dans ses travaux, distingue l'espérance mathématique de l'espérance morale pour apporter une solution au paradoxe.

Il s'agit de trouver une fonction "d'utilité" qui lie la différence de richesse du joueur entre avant et après le jeu et sa volonté de prendre part au jeu.

Daniel Bernoulli de considérer

$$\ln(r + 2^{n-1} - c) - \ln(r).$$

On remarque alors que l'"espérance morale"

$$\sum_{n=1}^{+\infty} \frac{1}{2^n} \big(\ln(r + 2^{n-1} - c) - \ln(r)\big) \text{ est finie.}$$

Paradoxe de St-Pétersbourg

Daniel Bernoulli (1700-1782), dans ses travaux, distingue l'espérance mathématique de l'espérance morale pour apporter une solution au paradoxe.

Il s'agit de trouver une fonction "d'utilité" qui lie la différence de richesse du joueur entre avant et après le jeu et sa volonté de prendre part au jeu.

Daniel Bernoulli de considérer

$$\ln(r+2^{n-1}-c)-\ln(r)$$

- Cramer avait suggéré quelques années plus tôt d'utiliser la fonction $\sqrt{\cdot}$ au lieu de $\log(\cdot)$;
- En économie, on étudie toujours la théorie de l'utilité.

Paradoxe de St-Pétersbourg

Cela mènera dans les travaux de Feller (1906-1970) a des généralisations de la loi des grands nombres.

Loi des grands nombres pondérées

Des variables aléatoires $(X_j)_j$ vérifie la loi des grands nombres pondérées par $(x_n)_n \subset \mathbb{R}_+$ si

$$\lim_{n\to+\infty} \mathbb{P}\Big(\Big|\frac{X_1+\cdots+X_n}{x_n} - 1\Big| \geq \varepsilon\Big) = 0.$$

- si l'espérance μ est finie, on obtient la loi des grand nombres avec le poids $(n\mu)_n$;
- dans le cas du paradoxe de St-Pétersbourg, on peut utiliser le poids $(\frac{n}{2}\log_2(n))_n$.

De Moivre et le premier TCL

Influencé par les travaux d'Huygens, Abraham de Moivre (1667-154) publie en 1718 son premier traité des probabilités.

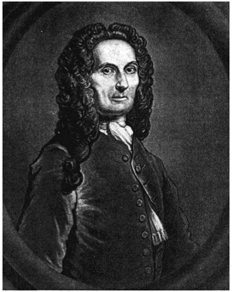


Figure: Abraham de Moivre

De Moivre et le premier TCL

Parmi d'autres choses, on trouve une version du célèbre Théorème Centrale Limite

Théorème Centrale Limite

Soient $X_1, X_2, \ldots$ des variables aléatoires i.i.d d'espérance μ et de variance σ^2. Pour tout $x \in \mathbb{R}$,

$$\lim_{n \to +\infty} \mathbb{P}\left(\frac{\frac{1}{n}\sum_{j=1}^{n} X_j - \mu}{\sqrt{\frac{\sigma^2}{n}}} \leq x \right) = \Phi(x),$$

où Φ est la fonction de répartition de la loi $\mathcal{N}(0,1)$.

On dit que $\frac{\frac{1}{n}\sum_{j=1}^{n} X_j - \mu}{\sqrt{\frac{\sigma^2}{n}}}$ converge en loi vers une loi normale $\mathcal{N}(0,1)$.

De Moivre et le premier TCL

De Moivre étudie la loi binomiale $\mathrm{Bin}(n,p)$ qui compte le nombre de succès lors de la répétition de n répétitions indépendantes d'un même schéma de Bernoulli de paramètre p : pour tout $k \in \{0,\ldots,n\}$,

$$\mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

De Moivre et le premier TCL

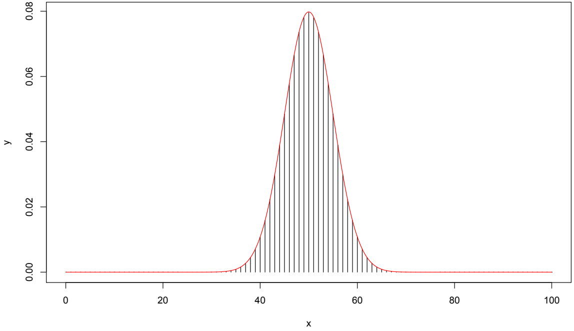


Figure: Diagramme en bâtons de la loi Bin $(100, 1/2)$

De Moivre et le premier TCL

De Moivre exploite la formule de Stirling

$$\lim_{n\to+\infty} \frac{n!}{\sqrt{2\pi n}(n/e)^n} = 1$$

pour décrire la courbe approximante :

Le logarithme du rapport d'un terme, situé à la distance d du maximum, à ce maximun est $-2\frac{d^2}{n}$

L'aiguille de Buffon

En 1733, Georges-Louis Leclerc de Buffon (1707-1488) propose une expérience aléatoire qui permet d'approximer π.

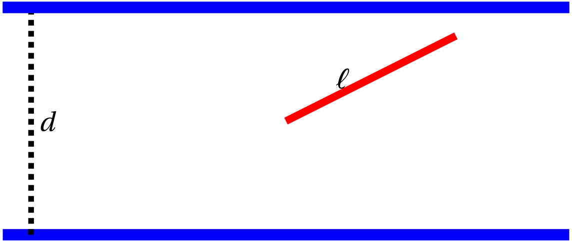
Figure: Georges-Louis Leclerc de Buffon

L'aiguille de Buffon

L'expérience consiste à lancer des aiguilles sur un parquet à lattes parallèles et compter le nombre d'aiguilles à cheval entre deux lattes.

L'aiguille de Buffon

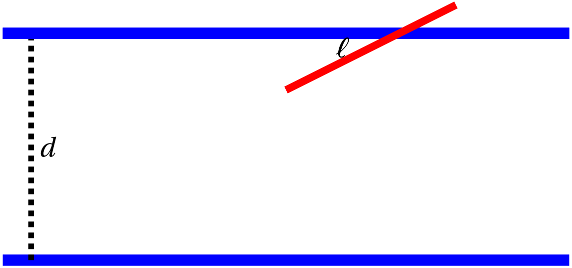
On note d la distance entre deux lattes du parquet et ℓ la longueur de l'aiguille (avec $\ell < d$).



Cette aiguille n'est pas à cheval entre deux lattes.

L'aiguille de Buffon

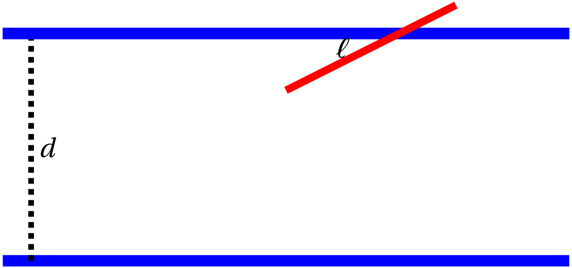
On note d la distance entre deux lattes du parquet et ℓ la longueur de l'aiguille (avec $\ell < d$).



Cette aiguille est à cheval entre deux lattes.

L'aiguille de Buffon

On note d la distance entre deux lattes du parquet et ℓ la longueur de l'aiguille (avec $\ell < d$).

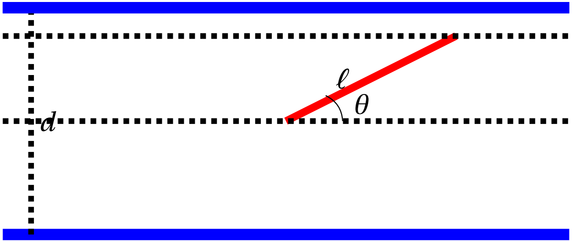


En lançant suffisamment d'aiguilles, on a

$$\pi \approx \frac{2}{\text{proportion d'aiguilles à cheval}} \frac{\ell}{d}$$

L'aiguille de Buffon

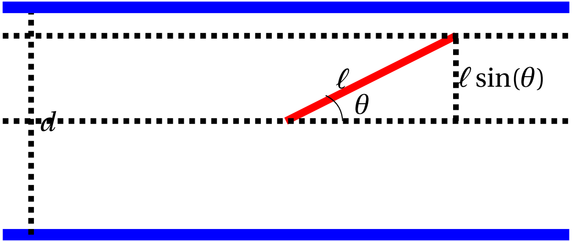
On note d la distance entre deux lattes du parquet et ℓ la longueur de l'aiguille (avec $\ell < d$).



Un aiguille tombant en formant un angle $\theta \in [0, \frac{\pi}{2}]$ avec l'horizontale

L'aiguille de Buffon

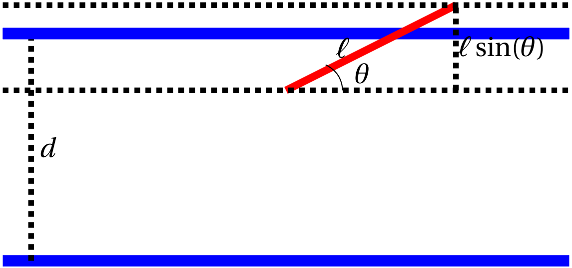
On note d la distance entre deux lattes du parquet et ℓ la longueur de l'aiguille (avec $\ell < d$).



Un aiguille tombant en formant un angle $\theta \in [0, \frac{\pi}{2}]$ avec l'horizontale a une hauteur $\ell\sin(\theta)$

L'aiguille de Buffon

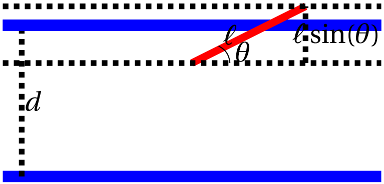
On note d la distance entre deux lattes du parquet et ℓ la longueur de l'aiguille (avec $\ell < d$).



Sa probabilité de toucher la latte du dessus est donc $\frac{\ell \sin(\theta)}{d}$

L'aiguille de Buffon

On note d la distance entre deux lattes du parquet et ℓ la longueur de l'aiguille (avec $\ell < d$).



En moyennant sur tous les angles θ possibles, on trouve que la probabilité d'intersection est donnée par

$$\frac{2}{\pi}\int_0^{\frac{\pi}{2}} \frac{\ell\sin(\theta)}{d}\,d\theta = \frac{2}{\pi}\frac{l}{d}.$$

L'aiguille de Buffon

On note d la distance entre deux lattes du parquet et ℓ la longueur de l'aiguille (avec $\ell < d$).
En moyennant sur tous les angles θ possibles, on trouve que la probabilité d'intersection est donnée par

$$\frac{2}{\pi}\int_0^{\frac{\pi}{2}} \frac{\ell \sin(\theta)}{d}\, d\theta = \frac{2}{\pi}\frac{l}{d}$$

La loi des grands nombres permet d'assurer que la proportion observée approxime (presque sûrement) π.

L'aiguille de Buffon

Quelques faits remarquables :

- étude d'un problème bidimensionnel;
- première incursion du calcul intégral en théorie des probabilités;
- il s'agit en fait de la première méthode de Monte-Carlo.

Le principe de “probabilité inverse”

Thomas Bayes (1702-1761) dans son mémoire publié à titre posthume en 1763 cherche à déterminer la probabilité avec laquelle les résultats d’expériences passées peuvent être exploités pour déduire des résultats futurs.

Figure: Thomas Bayes

Le principe de “probabilité inverse”

C'est dans ce cadre qu'il définit ce qu'on appelle aujourd'hui les probabilités conditionnelles. De son point de vue, si A et B sont deux événements,

- $\mathbb{P}(A)$ est la probabilité *a priori* de A;
- $\mathbb{P}(A|B)$ est la probabilité *a posteriori* de A sachant B.

Il note que dès lors qu'on connait la probabilité *a priori* de B (également appelée probabilité marginale), on a

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

Il énonce aussi la fameuse formule de Bayes

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

Le principe de “probabilité inverse”

L'approche de Bayes s'est développée vers ce qu'on appelle de nos jours l'inférence bayésienne.

- en statistique inférentielle *fréquentiste*, on exploite les données pour déterminer les distributions de probabilités qui caractérisent un phénomène
- en statistique inférentielle *bayésienne*, on part d'un distribution *a priori* et ce sont les observations qui vont mettre à jour cette distribution.

Le principe de “probabilité inverse”

Question : est-ce que tous les corbeaux sont noirs ?

Le principe de "probabilité inverse"

Approche fréquentiste : on capture un échantillon de corbeaux et on vérifie s'ils sont noirs ou non.

Le principe de “probabilité inverse”

Les observations doivent nous permettre d'augmenter notre niveau de confiance en l'expression logique

$$\text{“corbeau} \Rightarrow \text{noir”}$$

Le principe de “probabilité inverse”

Les observations doivent nous permettre d’augmenter notre niveau de confiance en l’expression logique

“corbeau ⇒ noir”

Le principe de “probabilité inverse”

Les observations doivent nous permettre d’augmenter notre niveau de confiance en l’expression logique

$$\text{“corbeau} \Rightarrow \text{noir”}$$

Le principe de “probabilité inverse”

Les observations doivent nous permettre d’augmenter notre niveau de confiance en l’expression logique

$$\text{“corbeau} \Rightarrow \text{noir”}$$

Le principe de “probabilité inverse”

Les observations doivent nous permettre d’augmenter notre niveau de confiance en l’expression logique “corbeau $\Rightarrow$ noir” ou de façon équivalente

$$\text{“non noir} \Rightarrow \text{non corbeau”}$$

Le principe de “probabilité inverse”

Ainsi, on peut gagner de la confiance en l’affirmation “tous les corbeaux sont noirs” sans jamais avoir vu de corbeaux?

Le principe de “probabilité inverse”

Oui, mais de manière insignifiante.

La théorie se développe

Selon les principes

La probabilité est le rapport du nombres des cas favorables à celui de tous les cas possibles.

Mais cela suppose les divers cas également possibles. S'ils ne le sont pas, on déterminera d'abord leurs possibilités respectives dont la juste appréciation est un des points les plus délicats de la théorie des hasards. Alors la probabilité sera la somme des possibilités de chaque cas favorable.

La théorie se développe

Ces principes sont par exemple énoncés par Pierre-Simon de Laplace dans sa *Théorie analytique des probabilités* (1812).

Figure: Pierre-Simon de Laplace

La théorie se développe

Ces principes suffisent pour le développement

- des études démographiques;
- l'apparition des premiers tests d'hypothèse statistique;
- l'utilisation de densité pour calculer des probabilités (première "loi continue" dans les travaux de Joseph-Louis Lagrange (1736-1813));
- le développement de la théorie des erreurs de mesure (via des courbes en cloche).

Méthode des moindres carrés

Contexte historique :
Le jour de l'an 1801, l'astronome Giuseppe Piazzi (1746-1826) découvre l'astéroïde Cérès et observe sa trajectoire jusqu'au 14 février 1801.

Figure: Giuseppe Piazzi

Sur base des observations de Piazzi, plusieurs scientifiques souhaitent prédire la trajectoire de Cérès et le localiser à nouveau.

Méthode des moindres carrés

La seule méthode suffisamment précise fut proposée par Carl-Friedrich Gauss (1777-1855).

Figure: Carl-Friedrich Gauss

Il est à noter que Legendre a proposé la même solution, parallèlement et indépendamment.

Méthode des moindres carrés

L'idée est de considérer que le phénomène d'intérêt est régi par un modèle théorique

$$f(\bullet,\theta)$$

qui dépend de paramètres θ. L'observation y_j $(1 \leq j \leq N)$ d'une occurrence $f(x_j,\theta)$ est entachée d'une erreur de mesure. La méthode des moindres carrés consiste à considérer que les paramètres θ qui décrivent le mieux le modèle sont ceux qui minimisent la quantité

$$\sum_{j=1}^{N}(y_j - f(x_j,\theta))^2.$$

Vers la loi normale

Contexte :

- on dispose de N observations $(x_1,\dots,x_N)$ d'une grandeur d'intérêt θ
- on considère que, pour tout $j \in \{1,\dots,N\}$, $x_j = \theta + y_j$ où y_j est issue d'une loi Y_j à densité f
- on suppose $Y_1,\dots,Y_N$ indépendants : leur densité jointe est donc

$$(y_1,\dots,y_N) \mapsto f(y_1)\dots f(y_N).$$

- les observations sont donc issues de la loi de densité jointe

$$(x_1,\dots,x_N) \mapsto f(x_1-\theta)\dots f(x_N-\theta).$$

Vers la loi normale

Gauss souhaite alors que la meilleure estimation de θ via $(x_1,\ldots,x_N)$ soit donnée par leur moyenne arithmétique

$$\overline{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$$

et que cette estimation satisfasse

$$\forall N \in \mathbb{N}^*, \forall x_1,\ldots,x_N, \forall \widetilde{\theta}, \quad f(x_1-\overline{x})\ldots f(x_N-\overline{x}) \geq f(x_1-\widetilde{\theta})\ldots f(x_N-\widetilde{\theta}).$$

Avec le vocabulaire actuel, on dit que $\overline{x}$ est l'estimateur obtenu par maximum de vraisemblance.

Vers la loi normale

$$\forall N \in \mathbb{N}^*, \forall x_1, \ldots, x_N, \forall \widetilde{\theta}, \quad f(x_1 - \overline{x}) \ldots f(x_N - \overline{x}) \geq f(x_1 - \widetilde{\theta}) \ldots f(x_N - \widetilde{\theta}).$$

En supposant f dérivable et positif et en posant $\Phi = \ln(f)$ et $\varphi = \Phi'$, on tire

- avec $N = 2$

  $$\forall x_1, x_2; \quad \varphi(x_1 - \overline{x}) + \varphi(x_2 - \overline{x}) = 0$$

  ou encore

  $$\forall x_1, x_2; \quad \varphi\left(\frac{x_1 - x_2}{2}\right) + \varphi\left(\frac{x_2 - x_1}{2}\right) = 0$$

  et donc φ est une fonction paire.

Vers la loi normale

$$\forall N \in \mathbb{N}^*, \forall x_1, \ldots, x_N, \forall \widetilde{\theta}, \quad f(x_1 - \overline{x}) \ldots f(x_N - \overline{x}) \geq f(x_1 - \widetilde{\theta}) \ldots f(x_N - \widetilde{\theta}).$$

En supposant f dérivable et positif et en posant $\Phi = \ln(f)$ et $\varphi = \Phi'$, on tire

- du cas $N = 2$, φ est une fonction paire.
- avec $N = 3$

$$\forall x_1, x_2, x_3; \quad \varphi\left(\frac{2x_1 - x_2 - x_3}{3}\right) + \varphi\left(\frac{2x_2 - x_1 - x_3}{3}\right) + \varphi\left(\frac{2x_3 - x_1 - x_2}{3}\right) = 0$$

ou encore, en exploitant la parité de φ,

$$\forall u, v; \quad \varphi(u) + \varphi(v) = \varphi(u + v).$$

Vers la loi normale

$$\forall N \in \mathbb{N}^*, \forall x_1, \ldots, x_N, \forall \widetilde{\theta}, \quad f(x_1 - \overline{x}) \ldots f(x_N - \overline{x}) \geq f(x_1 - \widetilde{\theta}) \ldots f(x_N - \widetilde{\theta}).$$

En supposant f dérivable et positif et en posant $\Phi = \ln(f)$ et $\varphi = \Phi'$, on tire

- du cas $N = 2$, φ est une fonction paire.
- du cas $N = 3$

$$\forall u, v; \quad \varphi(u) + \varphi(v) = \varphi(u + v).$$

Ainsi φ est linéaire, $\Phi : x \mapsto ax^2 + b$ et $f : x \mapsto \lambda e^{ax^2}$ (avec $\lambda = e^b$).

Vers la loi normale

$$\forall N \in \mathbb{N}^*, \forall x_1, \ldots, x_N, \forall \widetilde{\theta}, \quad f(x_1 - \overline{x}) \ldots f(x_N - \overline{x}) \geq f(x_1 - \widetilde{\theta}) \ldots f(x_N - \widetilde{\theta}).$$

En supposant f dérivable et positif et en posant $\Phi = \ln(f)$ et $\varphi = \Phi'$, on tire que φ est linéaire, $\Phi : x \mapsto ax^2 + b$ et $f : x \mapsto \lambda e^{ax^2}$ (avec $\lambda = e^b$).

On montre ensuite que pour tout choix de $\lambda > 0$ et $A < 0$ conduit à une solution de l'équation fonctionnelle.

Vers la loi normale

Laplace confirme dans sa *Théorie analytique des probabilités* l'importance de la loi normale dans l'estimation des erreurs de mesure en "démontrant" une version du TCL dans le cas

- de somme d'erreurs indépendantes et uniformément distribuées;
- du carré de somme d'erreurs indépendantes et uniformément distribuées.

Vers la loi normale

Au sujet des théorèmes limites, Laplace écrit

> “À mesure que les événements se multiplient, leurs probabilités respectives se développent de plus en plus; leurs résultats moyens et les bénéfices ou les pertes qui en dépendent convergent vers des limites dont ils approchent avec des probabilités toujours croissantes. La détermination de ces accroissements et de ces limites est une des parties les plus intéressantes et les plus délicates de l’analyse des hasards.”

Vers la loi normale

Pour obtenir ses versions du TCL, Laplace utilise un “prototype” de ce que nous appelons aujourd'hui la fonction caractéristique d'une variable aléatoire X, donné par

$$\varphi_X : t \mapsto \mathbb{E}[e^{itX}].$$

Laplace tente également dans ses travaux une première axiomatisation des probabilités. Outre les deux principes déjà énoncés, on y retrouve notamment la règle de multiplication pour les événements indépendants et la définition des probabilités conditionnelles.

Vers la loi normale

Il faudra attendre la fin du 19ème siècle et les travaux de l'"école russe" pour avoir une démonstration rigoureuse des théorèmes limites : Tchebychev démontre la loi des grands nombres en 1863 et Markov le TCL en 1898.

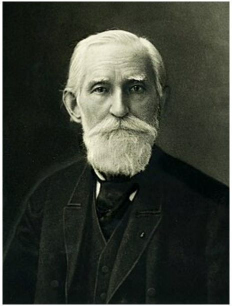


Figure: Paftouni Tchebychev (1821-1894) et Andreï A. Markov (1856-1922)

Vers la loi normale

Il faudra attendre la fin du 19ème siècle et les travaux de l'"école russe" pour avoir une démonstration rigoureuse des théorèmes limites : Tchebychev démontre la loi des grands nombres en 1863 et Markov le TCL en 1898. La version forte de la loi des grands nombres a été obtenue par Borel en 1909.

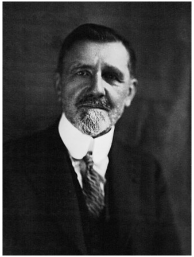


Figure: Émile Borel (1871-1956)

Vers la loi normale

On peut également citer les travaux de Paul Lévy (1886-1971) et William Feller (1906-1970) qui ont déterminé dans les années 1930 des conditions nécessaires et suffisantes pour que le TCL soit applicable.

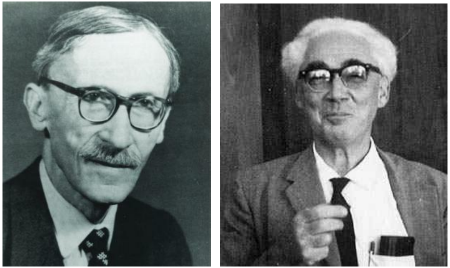


Figure: Paul Lévy (1886-1971) et William Feller (1906-1970)

La notion de variable aléatoire

La première mention de variable aléatoire (dans un cas discret) revient à Siméon Denis Poisson en 1837.

Figure: Siméon Denis Poisson (1781-1840)

La théorie de la mesure

En 1894 et 1897, Émile Borel introduit les premiers éléments de théorie de la mesure : il s'intéresse aux ensembles négligeables et mesurables

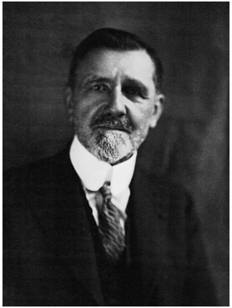


Figure: Émile Borel (1871-1956)

La théorie de la mesure

Intéressé par cette approche, Henri-Léon Lebesgue entame en 1901 une longue correspondance avec Émile Borel. L'idée de Lebesgue est de définir l'intégrale à partir de cette nouvelle théorie de la mesure

Figure: Henri-Léon Lebesgue (1875-1941)

La théorie de la mesure

C'est Gottfried Wilhelm Leibniz qui, dès 1686, s'intéressa au calcul intégral, en tant que réciproque du calcul différentiel.

Figure: Gottfried Wilhelm Leibniz (1646-1716)

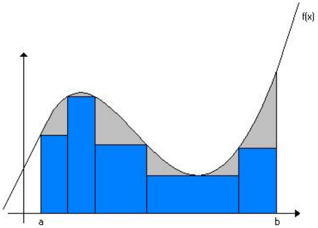
La théorie de la mesure

Jusqu'aux travaux de Lebesgue, c'était l'approche de Riemann de l'intégrale qui prédominait.

Figure: Bernhard Riemann (1826-1866)

La théorie de la mesure

Principe : on sait intégrer facilement les fonctions en escalier. Et l'intégrale est définie pour toute fonction qui va pouvoir être approchée par de telles fonctions.



La théorie de la mesure

L'intégrale de Riemann permet notamment d'intégrer les fonctions

- monotones (par morceaux)
- continues (par morceaux)
- les limites uniformes de telles fonctions

MAIS pas les limites simples de telles fonctions.

La théorie de la mesure

L'idée de Lebesgue est de ne plus se baser sur les fonctions de la forme

$$\sum_{j=1}^{J} c_j \mathbb{1}_{[a_j,b_j[} \text{ avec } a_j, b_j, c_j \in \mathbb{R}$$

mais bien sur des fonctions

$$\sum_{j=1}^{J} c_j \mathbb{1}_{A_j} \text{ avec } A_j \text{ mesurable et } c_j \in \mathbb{R}.$$

La théorie de la mesure

Cette approche

- d'intégrer beaucoup plus de fonctions (même des fonctions non localement bornées);
- coïncide avec celle de Riemann dans le cas des fonctions Riemann-intégrables;
- permet de donner une structure d'espace normé complet à l'ensemble des fonctions intégrables.

La théorie de la mesure

Au début des années 1930, Andreï Kolmogorov, qui travaille sur des généralisations de l'intégrale, voyage en Allemagne et en France.

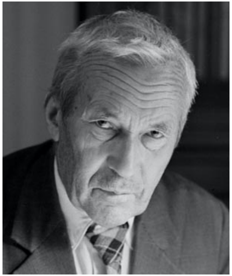


Figure: Andreï Kolmogorov (1903-1987)

La théorie de la mesure

Au début des années 1930, Andreï Kolmogorov, qui travaille sur des généralisations de l'intégrale, voyage en Allemagne et en France. Il y rencontre du beau monde.

Figure: Constantin Carathéodory, René Maurice Fréchet, Émile Borel Henri-Léon Lebesgue et Paul Lévy

La théorie de la mesure

En 1933, Andreï Kolmogorov propose son axiomatisation des probabilités. Cela lui permet de formaliser l'étude des processus stochastiques.

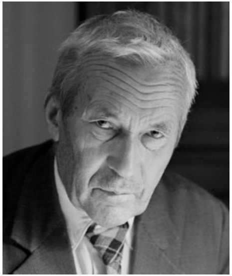


Figure: Andreï Kolmogorov (1903-1987)

La théorie de la mesure

En 1919, Richard von Mises avait proposé une axiomatisation “fréquentiste” des probabilités

Figure: Richard von Mises (1883-1953)

- la fréquence relative de la réalisation de l’événement considéré converge, c’est à dire admet une valeur-limite;
- cette valeur-limite demeure inchangée lorsqu’on remplace la suite d’observations par une sous-suite quelconque.

La théorie de la mesure

En 1938, à l'occasion du "Colloque consacré à la théorie des probabilités" (présidé par Fréchet), l'approche de von Mises fut enterrée.

- basé sur "le bon sens empirique", cette axiomatisation "relativisait" la notion de probabilité au lieu de la formaliser;
- l'axiomatique postule l'existence de la limite des fréquences;
- pensée pour la "pratique", cette approche n'était en fait... pas pratique.

Statistique inférentielle

En statistique inférentielle, on tente de répondre à des questions relatives à une population en “extrapolant” les résultats observés sur un échantillon.

On est conscient que cela comporte un risque (l'image renvoyé par l'échantillon pourrait ne pas être fidèle) et on mesure ce risque.

Statistique inférentielle

Par exemple, dans un test d'hypothèse

$$H_0 \longleftrightarrow H_1$$

où

- H_0 est l'hypothèse nulle;
- H_1 l'hypothèse alternative.

Statistique inférentielle

On est conscient que les scénarios suivants peuvent survenir

		Réalité	
		H_0 est vrai	H_1 est vrai
Décision prise	Ne pas rejeter H_0	Bonne décision	Erreur
	Rejeter H_0	Erreur	Bonne décision

Statistique inférentielle

On est conscient que les scénarios suivants peuvent survenir

		Réalité	
		H_0 est vrai	H_1 est vrai
Décision prise	NR H_0	Bonne décision	Erreur de seconde espèce
	R H_0	Erreur de première espèce	Bonne décision

Statistique inférentielle

Lors d'un test de grossesse :

$$H_0 \text{ : vous n'êtes pas enceinte contre } H_1 \text{ : vous êtes enceinte}$$

Statistique inférentielle

Lors d'un test de grossesse :

$$H_0 \ : \ \text{vous n'êtes pas enceinte contre } H_1 \ : \ \text{vous êtes enceinte}$$

Voici les deux cas où la décision du test est correcte

Statistique inférentielle

Lors d'un test de grossesse :

$$H_0 \,:\, \text{vous n'êtes pas enceinte contre } H_1 \,:\, \text{vous êtes enceinte}$$

Ceci est l'erreur de première espèce

Statistique inférentielle

Lors d'un test de grossesse :

$$H_0 \ : \ \text{vous n'êtes pas enceinte contre } H_1 \ : \ \text{vous êtes enceinte}$$

Ceci est l'erreur de seconde espèce

Statistique inférentielle

La théorie des tests statistiques se basent sur la suggestion de Neyman et Pearson.

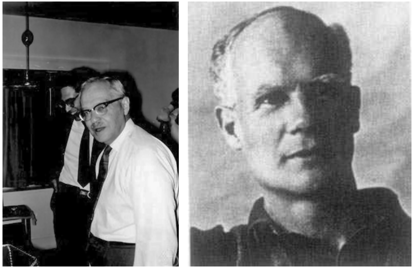


Figure: Jerzy Spława-Neyman (1894-1981) et Egon Sharpe Pearson (1895-1980)

Statistique inférentielle

La théorie des tests statistiques se base sur la suggestion de Neyman et Pearson qui consiste à

- fixer un niveau $\alpha \in]0,1[$ et veiller à ce que le risque de commettre une erreur de première espèce soit maintenu sous ce niveau α. On parle alors de test de signification α
- parmi tous les tests de niveau α, choisir, si cela existe, celui qui va minimiser le risque de commettre une erreur de seconde espèce (cfr. test (unifomément) plus puissant).

Statistique inférentielle

La théorie des tests statistiques se base sur la suggestion de Neyman et Pearson qui consiste à

- fixer un niveau $\alpha \in]0,1[$ et veiller à ce que le risque de commettre une erreur de première espèce soit maintenu sous ce niveau α. On parle alors de test de signification α
- parmi tous les tests de niveau α, choisir, si cela existe, celui qui va minimiser le risque de commettre une erreur de seconde espèce (cfr. test (unifomément) plus puissant).

Ce sont des résultats de probabilité qui permettent de réaliser de telles "constructions" !

Statistique inférentielle

Quelques pères fondateurs de la statistique inférentielle

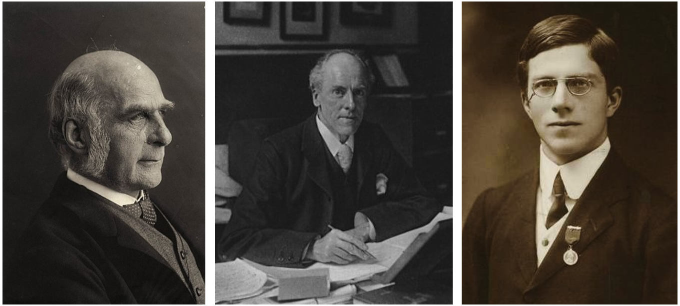


Figure: Francis Galton (1822-1911), Karl Pearson (1857-1936), Ronald Aylmer Fisher (1890-1962)

Statistique inférentielle

Ils étaient tous animés par une idéologie eugéniste.

Eugénisme : définition

Théorie cherchant à opérer une sélection sur les collectivités humaines à partir des lois de la génétique.

En militant pour l'“efficacité raciale”, Karl Pearson ira jusqu'à suggérer le recours à des politiques de stérilisation contrainte : “si toute personne qui est née a le droit de vivre, ce droit de vivre ne se convertit pas en droit de reproduire son espèce”.

Statistique inférentielle

William Sealy Gosset (1876-1937) était chef du département de statistiques à la brasserie Guinness à Dublin. N'étant pas autorisé à signer en ce nom par son entreprise, il publie ses travaux sous le nom de Student.

Figure: William Sealy Gosset alias Student