

Probabilités pour le tronc commun.

Thierry MEYRE

Préparation à l'agrégation externe. Université Paris Diderot.
Année 2012-2013

Ce poly est téléchargeable en ligne à l'adresse
<http://www.proba.jussieu.fr/pageperso/meyre>

BIBLIOGRAPHIE.

- BIL** Probability and Measure. Billingsley. Wiley.
- BAR** Probabilité. Barbe, Ledoux. Belin.
- COT** Exercices de Probabilités. Cottrell et al. Editions Cassini.
- DAC**¹ cours Probabilités et Statistiques 1. Problèmes à temps fixe.
Dacunha-Castelle, Dufflo. 2e édition. Editions Masson.
- DAC**¹ exo Exercices de Probabilités et Statistiques 1. Problèmes à temps fixe.
Dacunha-Castelle, Dufflo. 3e tirage corrigé. Editions Masson.
- DUR** Probability Theory and Examples. Durrett. Wadsworth.
- FOA** Calcul des Probabilités. Foata, Fuchs. Editions Dunod.
- FEL**¹ An Introduction to Probability Theory and its Applications. Feller.
Wiley 3rd edition. Volume I
- FEL**² An Introduction to Probability Theory and its Applications. Feller.
Wiley 2nd edition. Volume II
- JP** L'essentiel en théorie des probabilités. Jacod et Protter.
Editions Cassini.
- MAZ** Calcul de Probabilités. Exercices et problèmes corrigés.
Mazliak. Editions Hermann.
- MET** Notions fondamentales de la théorie des probabilités. Métivier. Dunod.
- NEV** Bases mathématiques du calcul des probabilités. Neveu. Masson.
- OUV** Probabilités (2 tomes). Ouvrard. Cassini
- REV** Probabilités. Revuz. Editions Hermann. Collection Méthodes.
- REV** int Mesure et Intégration. Revuz. Editions Hermann. Collection Méthodes.
- RS** Statistique en action. Rivoirard et Stoltz. Editions Vuibert.
- ROS** Initiation aux Probabilités. Ross. Presses polytechniques et universitaires romandes. Troisième édition.
- RUD** Analyse réelle et complexe. Rudin. Editions Masson.

Chapitre 1

Un modèle probabiliste simple : le modèle additif (Cours)

L'objet de la théorie des Probabilités est de construire un modèle permettant d'étudier les phénomènes dépendant du hasard. Nous allons procéder en deux étapes en présentant dans ce chapitre un modèle simple, mais dont le champ d'application sera limité, puis en le raffinant dans le chapitre suivant pour aboutir à notre modèle définitif.

1.1 Préliminaire : structure de clan

1.1.1 Définition

Définition 1.1.1 Soit E un ensemble. On appelle clan sur E (ou algèbre de Boole¹ de parties de E) un ensemble $\mathcal{C} \subset \mathcal{P}(E)$ tel que les trois conditions suivantes soient satisfaites :

1. $E \in \mathcal{C}$
2. $A \in \mathcal{C} \implies A^c = E - A \in \mathcal{C}$
3. $(A, B) \in \mathcal{C}^2 \implies A \cup B \in \mathcal{C}$

De cette définition nous déduisons les propriétés immédiates suivantes :

1. $\emptyset \in \mathcal{C}$
2. $(A, B) \in \mathcal{C}^2 \implies A \cap B \in \mathcal{C}$
3. $(A_1, \dots, A_n) \in \mathcal{C}^n \implies \bigcup_{i=1}^n A_i \in \mathcal{C}$ et $\bigcap_{i=1}^n A_i \in \mathcal{C}$

1. George Boole (1815-1864) logicien britannique

4. Si A et B sont des ensembles de \mathcal{C} , alors leur différence $A - B = A \cap B^c$ est encore un ensemble de \mathcal{C} . Un cas particulier est celui où $B \subset A$: on parle alors de différence propre.
5. Si A et B sont des ensembles de \mathcal{C} , alors leur différence symétrique $A \Delta B = (A \cup B) - (A \cap B) = (A - B) \cup (B - A)$ est encore un ensemble de \mathcal{C} .

Exemples:

1. L'ensemble $\mathcal{P}(E)$ de toutes les parties de E est le plus grand clan sur E (au sens de l'inclusion). On l'appelle *clan total* sur E .
2. L'ensemble $\{\emptyset, E\}$ est le plus petit clan sur E . On l'appelle *clan grossier* (ou *trivial*) sur E .
3. Sur \mathbb{R} , l'ensemble \mathcal{C} des unions finies d'intervalles de la forme

$$]-\infty, a[\quad , \quad [a, b[\quad \text{ou} \quad [b, +\infty[$$

est un clan.

Le lecteur pourra montrer d'abord la stabilité de \mathcal{C} par intersection finie avant de prouver la stabilité par passage au complémentaire.

1.1.2 Clans engendrés

Proposition 1.1.2 *L'intersection d'une famille non vide $(\mathcal{C}_i)_{i \in I}$ de clans sur E est encore un clan sur E .*

Corollaire 1.1.3 *Soit \mathcal{E} un ensemble quelconque de parties de E . Il existe un plus petit clan sur E (au sens de l'inclusion) qui contienne \mathcal{E} ; on l'appelle *clan engendré par \mathcal{E}* et on le note $\mathcal{C}(\mathcal{E})$.*

Démonstration: La famille des clans sur E contenant \mathcal{E} n'est pas vide car elle contient $\mathcal{P}(E)$. L'intersection de tous les clans sur E contenant \mathcal{E} est donc encore un clan, il contient bien sûr \mathcal{E} et est manifestement le plus petit au sens de l'inclusion ayant cette propriété. □

1.2 Un modèle à trois ingrédients.

1.2.1 L'ensemble des résultats possibles.

Qu'est-ce que le hasard ? Comme nous enseignons les mathématiques et non la philosophie, nous allons adopter une définition empirique, c'est-à-dire basée sur l'expérience.

Nous dirons qu'une expérience (scientifique ou autre, de la vie courante par exemple) dépend du hasard et nous l'appellerons *expérience aléatoire* – du mot latin *alea* qui signifie jet de dés – si elle produit un résultat que l'on ne connaît pas à l'avance mais qui en revanche appartient à un ensemble connu à l'avance. Cet ensemble, appelé *ensemble des résultats possibles* est noté traditionnellement Ω . Il peut être plus ou moins grand suivant l'expérience considérée.

Exemples :

1. Jeu de pile ou face. $\Omega = \{p, f\}$.
2. Jet de dé à 6 faces. $\Omega = \{1, 2, \dots, 6\}$.
3. On joue à pile ou face de façon répétée et l'on s'intéresse au rang d'apparition du premier pile. $\Omega = \mathbb{N}^*$.
4. Durée de vie d'une batterie de voiture. $\Omega = \mathbb{R}_+$ ou $\Omega = [0, T]$ avec T suffisamment grand.
5. Courbe d'évolution de la température sur une journée dans une station météo. $\Omega = C^0([0, T], \mathbb{R})$.

Dans les exemples 1 à 3, Ω est fini ou dénombrable : c'est le cas *discret*. Dans les exemples 4 et 5, Ω est infini non dénombrable : c'est le cas *continu*.

L'ensemble Ω est le premier ingrédient du modèle mathématique que nous allons construire pour représenter le mieux possible notre expérience aléatoire. Nous allons mettre en oeuvre deux ingrédients supplémentaires pour construire notre modèle additif.

1.2.2 L'ensemble des évènements.

Une fois faite notre expérience aléatoire, nous nous trouvons devant un certain résultat $\omega \in \Omega$. À part peut-être le cas où le cardinal de Ω , noté $\#\Omega$, est "très petit", il arrive très souvent que ce ne soit pas le résultat précis ω qui nous intéresse mais que l'on cherche plutôt à répondre à la question : "est-ce que ω appartient à tel sous-ensemble donné de Ω ?"

Exemple: Durée de vie d'un ordinateur.

Si le constructeur garantit la machine pendant 3 ans, il a bien sûr d'abord étudié la question suivante : "quelles sont les chances pour que $\omega > 3$?" De la réponse à cette question dépendra le prix auquel il va vous facturer la garantie. Dans cette exemple, la question intéressante s'écrit donc : " $\omega \in A$?" avec $A =]3, +\infty[\subset \Omega = \mathbb{R}_+$.

Nous appellerons *évènement* un sous-ensemble A de Ω tel qu'un observateur de l'expérience est capable de répondre à la question " $\omega \in A$?" Par exemple, Ω est *l'évènement certain* tandis que \emptyset est *l'évènement impossible*.

Dès que nous allons manipuler des évènements, certains opérateurs ensemblistes vont apparaître :

- L'évènement contraire de A est son complémentaire $A^c = \Omega - A$.
- L'évènement "le résultat est dans A_1 ou A_2 " s'écrit $A_1 \cup A_2$
- L'évènement "le résultat est dans A_1 et A_2 " s'écrit $A_1 \cap A_2$

Un observateur de l'expérience étant donné, il est donc naturel de supposer que l'ensemble des parties de Ω dont il peut discerner la réalisation ou la non-réalisation après l'expérience a une **structure de clan**. Nous noterons dans la suite \mathcal{C} le clan des évènements discernables par notre observateur.

Remarque: Plusieurs observateurs distincts peuvent être associés à une même expérience. Par exemple, si le résultat de l'expérience aléatoire qui nous intéresse est le temps (mesuré en secondes) mis par un athlète pour courir un 100 mètres (prenons $\Omega = [0; 15]$) et si un premier arbitre réalise un chronométrage manuel tandis que le second se réfère à un dispositif électronique, nous avons $\mathcal{C}_1 \subsetneq \mathcal{C}_2$. Ainsi, l'évènement $A = [0; 9,57]$ qui se traduit en langage courant par "il a battu le record du monde masculin² de 100 mètres sur piste" vérifie $A \in \mathcal{C}_2$ mais $A \notin \mathcal{C}_1$.

1.2.3 La probabilité

Nous avons tous l'intuition que certains évènements ont plus de chances de se produire que d'autres. Par exemple, dans une loterie, il est plus probable de tirer un billet perdant que de tirer le gros lot.

Pour préciser cette intuition, nous souhaitons associer à un évènement donné un nombre réel qui mesure les chances qu'il se produise ; par exemple, si l'évènement se produit avec une chance sur deux (pile ou face avec une pièce équilibrée), nous lui associerons le nombre $1/2$.

Nous appellerons donc probabilité une application $P : \mathcal{C} \rightarrow [0, 1]$ vérifiant certaines conditions que nous allons préciser maintenant.

Tout d'abord, Ω étant l'évènement certain, nous demanderons $P(\Omega) = 1$ (en langage courant, il y a "100 chances sur 100" qu'il se produise).

Ensuite, il est naturel d'exiger la propriété d'*additivité* :

$$A \in \mathcal{C}, B \in \mathcal{C}, A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B).$$

2. le Jamaïcain Usain Bolt a réalisé un chrono de 9,58 secondes le 16 août 2009

Notons que la condition $A \cap B = \emptyset$, i.e. les deux évènements A et B sont *disjoints*, est indispensable pour ne pas compter deux fois les mêmes résultats possibles.

La loi empirique des grands nombres

On désigne ainsi une idée intuitive qui vient «justifier» l'axiome d'additivité pour une probabilité :

Si nous pouvons répéter une même expérience un grand nombre de fois dans des conditions identiques et que nous notons $\varphi_n(A)$ le nombre de fois où le résultat a appartenu à l'évènement A au cours des n premières répétitions de l'expérience, nous nous attendons à ce que le rapport $\varphi_n(A)/n$ (fréquence de réalisation de l'évènement A au cours des n premières répétitions de l'expérience) tende vers une certaine limite lorsque n augmente.

Intuitivement, c'est cette limite que nous aurions envie d'appeler «probabilité que l'évènement A se réalise». Si nous admettons qu'une telle limite existe pour tout évènement A et que nous la notons $P(A)$, il est facile de constater que nous avons ainsi défini une application $P : \mathcal{A} \rightarrow [0, 1]$ telle que $P(\Omega) = 1$ et additive car $A \cap B = \emptyset \Rightarrow \varphi_n(A \cup B) = \varphi_n(A) + \varphi_n(B)$.

1.2.4 Le modèle additif

Nous pouvons donc modéliser une expérience aléatoire par un triplet (Ω, \mathcal{C}, P) , où Ω est un ensemble, \mathcal{C} un clan sur Ω et P une application additive de \mathcal{C} dans $[0,1]$ telle que $P(\Omega) = 1$: c'est le *modèle additif*.

1.3 Espaces de probabilités finis

Le modèle additif s'avère en général suffisant lorsque l'expérience aléatoire considérée ne mène qu'à un nombre fini de résultats possibles. Nous étudions maintenant ce cas $\#\Omega < +\infty$ qui apparaît dans de nombreuses situations concrètes : jeux de hasard, enquêtes par sondage, contrôle de fabrication etc.

1.3.1 Description d'une probabilité quelconque

Lorsque Ω est fini, nous prendrons pratiquement toujours $\mathcal{C} = \mathcal{P}(\Omega)$, qui est le seul clan sur Ω contenant tous les singletons : ce clan d'évènements correspond à un observateur qui discerne le résultat précis ω de l'expérience.

Proposition 1.3.1 Soit $n \in \mathbb{N}^*$ et $\Omega = \{\omega_1, \dots, \omega_n\}$ un ensemble fini. Il existe une bijection entre l'ensemble des probabilités P sur $(\Omega, \mathcal{P}(\Omega))$ et l'ensemble des n -uplets $(p_k)_{1 \leq k \leq n} \in (\mathbb{R}_+)^n$ tels que :

$$\sum_{1 \leq k \leq n} p_k = 1.$$

Cette bijection est donnée par :

$$\forall k \in \{1, \dots, n\} \quad p_k = P(\{\omega_k\}).$$

Démonstration: Dans le sens direct, il est évident que si P est une probabilité sur $(\Omega, \mathcal{P}(\Omega))$, alors le n -uplet $(p_k)_{1 \leq k \leq n}$ ainsi défini a les propriétés voulues.

Réciproquement, un tel n -uplet étant donné, s'il existe une probabilité P qui convienne, alors elle satisfait nécessairement la relation suivante par additivité :

$$\forall A \in \mathcal{P}(\Omega) \quad P(A) = \sum_{k/\omega_k \in A} p_k \quad (1.1)$$

Il est facile de conclure en vérifiant que l'on a bien défini ainsi une probabilité sur $(\Omega, \mathcal{P}(\Omega))$. \square

Exemples:

1. Le cas le plus simple d'espace de probabilité fini non trivial est donné par $\#\Omega = 2$ et correspond à la modélisation du jeu de pile ou face (avec une pièce éventuellement biaisée). Il y a alors bijection entre les probabilités sur $(\Omega, \mathcal{P}(\Omega))$ et le segment $[0, 1]$: celle-ci est donnée par la probabilité $p \in [0, 1]$ que la pièce tombe sur pile. Formellement, on écrit en général $\Omega = \{0, 1\}$ (pile étant représenté par 1 et face par 0) et la probabilité définie par $P(\{1\}) = p$ et donc $P(\{0\}) = 1 - p$ est appelée probabilité de Bernoulli de paramètre $p \in [0, 1]$.
2. Pour tout $n \in \mathbb{N}^*$ et tout $p \in [0, 1]$, on définit sur $\Omega = \{0, 1, \dots, n\}$ (qui est donc de cardinal $n + 1$) la probabilité binomiale de paramètres n et p , notée $P = B_{n,p}$, par

$$\forall k \in \{0, 1, \dots, n\} \quad B_{n,p}(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}$$

Il s'agit bien d'une probabilité puisque la formule du binôme de Newton nous donne :

$$\sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = (p + 1 - p)^n = 1$$

Nous précisons ci-dessous dans quel type de situation cette probabilité apparaît.

1.3.2 La probabilité uniforme

Un cas particulier important qui se présente souvent dans la pratique est celui où tous les résultats possibles de l'expérience ont même chance de réalisation. Si $\Omega = \{\omega_1, \dots, \omega_n\}$ et $p_k = P(\{\omega_k\})$, $1 \leq k \leq n$, ceci s'écrit :

$$\forall k \in \{1, \dots, n\} \quad p_k = \frac{1}{n} = \frac{1}{\#\Omega}$$

Nous dirons alors que nous sommes en situation d'*équiprobabilité* ou encore que nous considérons la *probabilité uniforme* sur Ω . L'égalité (1.1) nous donne alors immédiatement :

$$\forall A \in \mathcal{P}(\Omega) \quad P(A) = \frac{\#A}{\#\Omega}.$$

C'est le fameux rapport entre le nombre de résultats favorables et le nombre de résultats possibles, très utilisé dans l'enseignement secondaire. Calculer une probabilité revient alors à résoudre un problème de dénombrement – ce qui n'est pas forcément simple.

Lorsqu'on utilise l'expression courante "tirer au hasard" sans autre précision, il est sous-entendu que c'est avec la probabilité uniforme que le tirage a lieu. Nous allons maintenant en voir deux exemples importants.

Échantillon (ordonné)

Considérons une population finie (i.e. un ensemble fini) $\mathcal{S} = \{s_1, \dots, s_N\}$. Nous appellerons *échantillon* de taille n de cette population tout n -uplet $(s_{i_1}, \dots, s_{i_n}) \in \mathcal{S}^n$.

Notons que l'ordre compte et que nous pouvons retrouver plusieurs fois le même élément : un tel n -échantillon modélise un tirage **avec remplacement** de n éléments dans la population \mathcal{S} .

Nous prenons donc ici $\Omega = \mathcal{S}^n$ et nous supposons que nous sommes en situation d'équiprobabilité. Autrement dit, nous avons :

$$\forall \omega \in \Omega, \quad P(\{\omega\}) = \frac{1}{\#\Omega} = \frac{1}{N^n}$$

Le lecteur pourra utiliser ce modèle de tirage avec remplacement pour résoudre l'exercice 2.3.2 page 25.

Lien avec la probabilité binomiale

Supposons maintenant que la population \mathcal{S} soit divisée en 2 catégories \mathcal{S}_1 et \mathcal{S}_2 de cardinaux respectifs N_1 et N_2 ($N_1 + N_2 = N$). Si nous tirons “au hasard” (équiprobabilité) un n -échantillon de \mathcal{S} , quelle est la probabilité qu’il contienne exactement n_1 éléments de \mathcal{S}_1 et n_2 éléments de \mathcal{S}_2 ($n_1 + n_2 = n$) ?

Pour répondre à cette question, nous comptons les “cas favorables”. Pour construire un n -échantillon ayant la propriété précédente, nous commençons par choisir les places des éléments de 1ère catégorie (n_1 places parmi n donc $\binom{n}{n_1}$ possibilités) puis nous remplissons ces places avec des éléments de \mathcal{S}_1 ($N_1^{n_1}$ possibilités) et enfin nous remplissons les n_2 places restantes avec des éléments de \mathcal{S}_2 ($N_2^{n_2}$ possibilités). Finalement, le nombre de cas favorables vaut $\binom{n}{n_1} N_1^{n_1} N_2^{n_2}$ et la probabilité recherchée est :

$$\frac{\binom{n}{n_1} N_1^{n_1} N_2^{n_2}}{N^n} = \frac{\binom{n}{n_1} N_1^{n_1} (N - N_1)^{n-n_1}}{N^n}$$

Si nous posons $p = \frac{N_1}{N}$ (probabilité qu’un élément tiré au hasard dans \mathcal{S} soit de 1ère catégorie), la probabilité précédente s’écrit encore :

$$\binom{n}{n_1} p^{n_1} (1-p)^{n-n_1}$$

Nous retrouvons ici la *probabilité binomiale de paramètres n et p* définie sur l’espace fini $\Omega = \{0, 1, \dots, n\}$ muni du clan total $\mathcal{C} = \mathcal{P}(\Omega)$.

Généralisation

Nous pouvons généraliser tout ce qui précède à r catégories et nous sommes alors conduits à définir la *probabilité multinomiale de paramètres n et $p_1 \geq 0, \dots, p_r \geq 0, p_1 + \dots + p_r = 1$* , notée $M_{n;p_1, \dots, p_r}$, sur l’espace fini $\{0, 1, \dots, n\}^r$ comme suit : $\forall (k_1, \dots, k_r) \in \{0, 1, \dots, n\}^r$

$$\begin{aligned} M_{n;p_1, \dots, p_r}((k_1, \dots, k_r)) &= \frac{n!}{k_1! k_2! \dots k_r!} p_1^{k_1} p_2^{k_2} \dots p_r^{k_r} \text{ si } k_1 + \dots + k_r = n \\ &= 0 \text{ sinon.} \end{aligned}$$

Sous-population

Nous appelons sous-population de taille n de \mathcal{S} tout sous-ensemble de \mathcal{S} de cardinal n . Contrairement à ce qui se passait pour un n -échantillon, l’ordre ne compte donc pas et nous ne pouvons pas retrouver plusieurs fois

le même élément : une sous-population de taille n modélise un tirage **sans remplacement** de n éléments dans la population \mathcal{S} .

Supposons que Ω soit l'ensemble de ces sous-populations de taille n et que nous soyons en situation d'équiprobabilité. Nous avons alors :

$$\forall \omega \in \Omega, \quad P(\{\omega\}) = \frac{1}{\#\Omega} = \frac{1}{\binom{N}{n}}$$

Probabilité hypergéométrique

Si nous supposons encore que la population \mathcal{S} est divisée en 2 catégories (en reprenant les notations du paragraphe précédent) et que l'on tire une sous-population de taille n au hasard, quelle est la probabilité qu'elle contienne exactement n_1 éléments de 1ère catégorie (et donc $n_2 = n - n_1$ éléments de 2ème catégorie) ?

Pour répondre à cette question, nous construisons une sous-population de taille n vérifiant la propriété précédente en réunissant simplement une sous-population de taille n_1 de \mathcal{S}_1 et une sous-population de taille n_2 de \mathcal{S}_2 . Il y a donc $\binom{N_1}{n_1} \binom{N_2}{n_2}$ cas favorables et la probabilité cherchée vaut :

$$\frac{\binom{N_1}{n_1} \binom{N_2}{n_2}}{\binom{N}{n}} = \frac{\binom{N_1}{n_1} \binom{N-N_1}{n-n_1}}{\binom{N}{n}},$$

avec la convention $\binom{n}{k} = 0$ si $k < 0$ ou $k > n$.

Cette probabilité est appelée *hypergéométrique*.

Approximation binomiale de la probabilité hypergéométrique

Lorsque la taille N de la population \mathcal{S} est très grande par rapport à n , on s'attend à ce qu'il y ait peu de différence entre les procédures avec ou sans remplacement. Pour formaliser cela, fixons les valeurs de n et $n_1 \leq n$, supposons que $N \rightarrow +\infty$ et que $N_1 = \varphi(N)$ avec :

$$\lim_{N \rightarrow +\infty} \frac{\varphi(N)}{N} = p \in]0, 1[.$$

Dans ce cas, la probabilité hypergéométrique précédente s'écrit :

$$\frac{N_1(N_1 - 1) \cdots (N_1 - n_1 + 1)}{n_1!} \frac{N_2 \cdots (N_2 - n_2 + 1)}{n_2!} \frac{n!}{N \cdots (N - n + 1)}$$

ou encore, en réordonnant les facteurs :

$$\frac{n!}{n_1!n_2!} \frac{N_1(N_1-1)\cdots(N_1-n_1+1)}{N(N-1)\cdots(N-n_1+1)} \frac{N_2(N_2-1)\cdots(N_2-n_2+1)}{(N-n_1)(N-n_1-1)\cdots(N-n+1)}$$

Il est facile de constater que lorsque $N \rightarrow +\infty$, cette quantité tend vers $\binom{n}{n_1} p^{n_1} (1-p)^{n-n_1}$, c'est-à-dire vers la probabilité binomiale $B_{n,p}(\{n_1\})$.

Dans la pratique, il est souvent commode de prendre (quand $n \ll N$) pour valeur approchée d'une probabilité hypergéométrique la probabilité binomiale correspondante, qui est beaucoup moins lourde à calculer numériquement (voir exercice 2.3.3 page 25).

1.4 Probas conditionnelles et indépendance

1.4.1 Définition des probabilités conditionnelles.

La notion de probabilité conditionnelle apparaît naturellement lorsqu'on possède une information partielle sur le résultat d'une expérience aléatoire.

Exercice : Dans une urne qui contient 100 sujets de concours, indiscernables au toucher, un candidat plonge une main fébrile. Chaque sujet aborde deux thèmes au programme, selon la répartition suivante :

- trente sujets arithmétique et calcul des probabilités, que nous noterons s_1, \dots, s_{30} ;
- vingt sujets géométrie et calcul différentiel, s_{31}, \dots, s_{50} ;
- dix sujets arithmétique et calcul différentiel, s_{51}, \dots, s_{60} ;
- quarante sujets géométrie et calcul des probabilités, s_{61}, \dots, s_{100} .

Or le candidat a fait l'“impasse” sur les probabilités.

1. Quelles sont ses chances de tirer un sujet sans calcul des probabilités ?
2. Sans le regarder, ce candidat remet le sujet tiré à l'examineur, qui lui annonce : “Il y a de la géométrie dans votre sujet”. À combien évaluez-vous les chances que ce candidat échappe aux probabilités ?

Corrigé : 1) Pour modéliser cette expérience aléatoire, nous considérons l'espace probabilisé $(\Omega, \mathcal{P}(\Omega), P)$, où $\Omega = \{s_1, \dots, s_{100}\}$ et P est la probabilité uniforme sur Ω puisque les billets sont indiscernables au toucher.

L'évènement qui nous intéresse est $A := \{s_{31}, \dots, s_{60}\}$ et nous calculons :

$$P(A) = \frac{\#A}{\#\Omega} = \frac{30}{100} = 0,3$$

2) Notons B l'évènement « Il y a de la géométrie dans le sujet tiré », *i.e.* $B := \{s_{31}, \dots, s_{50}, s_{61}, \dots, s_{100}\}$. Puisque nous sommes munis de cette information donnée par l'examineur sur le résultat du tirage, il est naturel que nous réévaluions les chances que l'évènement A se produise.

En effet, notre ensemble de résultats possibles est désormais plus réduit que l'ensemble Ω que nous avons considéré *a priori*, c'est-à-dire avant que l'information ne nous soit donnée : il s'agit maintenant de B .

Tous les sujets contenant de la géométrie étant indiscernables au toucher, nous restons en situation d'équiprobabilité. Désormais, l'évènement qui nous intéresse se produira si et seulement si le résultat de l'expérience est un élément de $A \cap B$, puisque nous savons déjà qu'il est élément de B . Notre nouvelle évaluation de la probabilité que A se produise, sachant que B est réalisé, vaut donc :

$$\frac{\#(A \cap B)}{\#B} = \frac{20}{60} = \frac{1}{3} \simeq 0,333$$

Il est d'usage de noter cette probabilité $P(A|B)$, ce qui se lit « probabilité de A sachant B » ou encore « probabilité conditionnelle de A par rapport à B ».

Notons qu'en divisant l'expression précédente au numérateur et au dénominateur par $\#\Omega$, nous pouvons l'écrire sous la forme :

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

Puisque $P(A|B) = \frac{1}{3} > P(A)$, l'information donnée par l'examineur est plutôt une bonne nouvelle pour le candidat...

De façon plus générale, nous avons la proposition suivante :

Proposition 1.4.1 *Soit (Ω, \mathcal{A}, P) un espace de probabilité et $E \in \mathcal{A}$ un évènement tel que $P(E) > 0$. L'application $P^E : \mathcal{A} \rightarrow [0, 1]$ définie par :*

$$\forall A \in \mathcal{A} \quad P^E(A) = \frac{P(A \cap E)}{P(E)}$$

est une probabilité. On l'appelle probabilité conditionnelle à l'évènement E ou probabilité sachant E . On note sa valeur sur un évènement A comme suit :

$$P(A|E) = P^E(A) = \frac{P(A \cap E)}{P(E)}$$

Démonstration: C'est bien une application de \mathcal{A} dans $[0, 1]$ puisqu'on a :

$$A \cap E \subset E \Rightarrow 0 \leq P(A \cap E) \leq P(E).$$

De plus, on vérifie immédiatement : $P^E(\Omega) = 1$. Enfin, l'additivité de P^E résulte du calcul suivant, avec $(A, B) \in \mathcal{A}^2$ tels que $A \cap B = \emptyset$:

$$\begin{aligned} \frac{P((A \uplus B) \cap E)}{P(E)} &= \frac{P((A \cap E) \uplus (B \cap E))}{P(E)} \\ &= \frac{P(A \cap E) + P(B \cap E)}{P(E)} = \frac{P(A \cap E)}{P(E)} + \frac{P(B \cap E)}{P(E)} \end{aligned}$$

□

Remarque : Dans cette définition de la probabilité conditionnelle, l'idée est que, étant donnée l'information partielle E , l'évènement intéressant n'est plus l'évènement A de départ mais plutôt $A \cap E$. L'application $A \mapsto P(A \cap E)$ est bien positive et σ -additive mais n'a pas la bonne masse totale (i.e. la valeur 1 sur Ω). Pour en faire une probabilité, il suffit de la renormaliser en divisant par $P(E)$ mais ceci n'est possible que si $P(E) > 0$.

Retour à la loi empirique des grands nombres.

Nous pouvons «justifier» la définition des probabilités conditionnelles en utilisant la loi empirique des grands nombres. En reprenant les notations utilisées précédemment sur ce sujet, voyons comment faire apparaître la probabilité conditionnelle $P(A|B)$.

Puisque nous nous plaçons sous le conditionnement 'B est réalisé', parmi les n premières répétitions de l'expérience, seules nous intéressent celles qui ont cette propriété : leur cardinal est $\varphi_n(B)$.

Parmi ces $\varphi_n(B)$ expériences, la proportion de celles qui vérifient la propriété 'A est réalisé' nous est donnée par le quotient suivant :

$$\frac{\varphi_n(A \cap B)}{\varphi_n(B)}$$

Notre intuition est que lorsque n tend vers l'infini, et donc $\varphi_n(B)$ aussi (puisque $P(B) > 0$), ce quotient va tendre vers la probabilité conditionnelle de A sachant B . Or, d'après la loi empirique des grands nombres, comme ce quotient s'écrit encore :

$$\frac{\varphi_n(A \cap B)/n}{\varphi_n(B)/n},$$

il admet pour limite lorsque n tend vers l'infini :

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

Nous retrouvons bien la définition de la probabilité conditionnelle $P(A|B)$.

Exemple : Dans une loterie, il y a 20 billets gagnants sur un total de 100 billets indiscernables au toucher. Pour obtenir le gros lot, il faut d'abord tirer un billet gagnant puis il faut faire tourner une roue divisée en 5 parties égales dont une seule donne droit au gros lot. Quelle est la probabilité d'obtenir celui-ci ?

Écrire le modèle serait un peu plus long : $\Omega = \{b_1, \dots, b_{100}\} \times \{r_1, \dots, r_5\}$ etc. Pour gagner du temps, passons directement au calcul, en écrivant les évènements informellement :

$$A = \{\text{le joueur gagne le gros lot}\}; E = \{\text{le joueur tire un billet gagnant}\}.$$

Nous pouvons alors mener le calcul de la façon suivante :

$$P(A) = P(A \cap E) = P(A|E)P(E) = \frac{1}{5} \times \frac{20}{100} = \frac{4}{100}.$$

Nous venons d'utiliser la formule $P(A \cap E) = P(A|E)P(E)$; plus généralement nous avons la proposition suivante.

Proposition 1.4.2 Soient $(A_1, \dots, A_n) \in \mathcal{A}^n$ avec $P(A_1 \cap \dots \cap A_{n-1}) > 0$. Alors, on a l'égalité :

$$P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap \dots \cap A_{n-1})$$

Démonstration: Le membre de droite s'écrit par définition d'une probabilité conditionnelle :

$$P(A_1) \frac{P(A_1 \cap A_2)}{P(A_1)} \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} \cdots \frac{P(A_1 \cap \dots \cap A_n)}{P(A_1 \cap \dots \cap A_{n-1})}$$

et l'on conclut par simplification. \square

1.4.2 Formule de Bayes.

Proposition 1.4.3 (Formule des probas totales) Soient E_1, \dots, E_n des évènements tels que $P(E_i) > 0$ pour tout $1 \leq i \leq n$ et formant une partition de l'ensemble Ω (i.e. $\bigsqcup_{1 \leq i \leq n} E_i = \Omega$), alors on a pour tout $A \in \mathcal{A}$:

$$P(A) = \sum_{i=1}^n P(A|E_i)P(E_i)$$

Démonstration: Ceci résulte des égalités successives :

$$\begin{aligned} P(A) &= P\left(A \cap \left(\bigsqcup_{1 \leq i \leq n} E_i\right)\right) = P\left(\bigsqcup_{1 \leq i \leq n} (A \cap E_i)\right) \\ &= \sum_{1 \leq i \leq n} P(A \cap E_i) = \sum_{1 \leq i \leq n} P(A|E_i)P(E_i) \end{aligned}$$

□

Notons en particulier que la formule des probabilités totales se réécrit dans le cas $n = 2$:

Soient deux évènements E et A avec $0 < P(E) < 1$. Alors, on a l'égalité :

$$P(A) = P(A|E)P(E) + P(A|E^c)P(E^c)$$

Nous déduisons de la formule des probabilités totales le théorème suivant :

Théorème 1.4.4 (Formule de Bayes) Si E_1, \dots, E_n sont des évènements de probabilités strictement positives, formant une partition de l'ensemble Ω et si A est un évènement tel que $P(A) > 0$, alors on a l'égalité :

$$P(E_1|A) = \frac{P(A|E_1)P(E_1)}{\sum_{i=1}^n P(A|E_i)P(E_i)}$$

Démonstration: Il suffit d'écrire les égalités :

$$P(E_1|A) = \frac{P(A \cap E_1)}{P(A)} = \frac{P(A|E_1)P(E_1)}{P(A)}$$

puis d'appliquer la formule des probabilités totales au dénominateur. □

Remarque : Revenons au cas particulier $n = 2$. La formule de Bayes s'écrit alors :

$$P(E|A) = \frac{P(A|E)P(E)}{P(A|E)P(E) + P(A|E^c)P(E^c)}$$

Notons qu'il n'est donc pas possible de calculer $P(E|A)$ à partir de la seule donnée de $P(A|E)$; il faut connaître d'autres quantités pour faire ce calcul.

Exercice : Une maladie rare touche 1 personne sur 10000 dans la population française. Quand cette maladie est présente, un test sanguin permet de la détecter dans 99% des cas. En revanche, ce test produit des faux positifs dans 1 cas sur 1000.

Le test d'une personnes est positif. Quelle est la probabilité qu'elle soit vraiment atteinte de la maladie? Que pensez-vous de la qualité de ce test sanguin?

Solution : Définissons les évènements :

$$E = \{\text{la personne est atteinte de la maladie}\}; A = \{\text{le test est positif}\}$$

L'énoncé nous fournit les données suivantes :

$$P(E) = 10^{-4}, P(A|E) = 0,99, P(A|E^c) = 10^{-3}$$

La formule de Bayes nous permet alors de calculer :

$$P(E|A) = \frac{0,99 \times 10^{-4}}{0,99 \times 10^{-4} + 10^{-3} \times 0,9999} \sim \frac{1}{11}$$

Conclusion : À moins de vouloir provoquer beaucoup de panique inutile, le test est bon pour la poubelle!

1.5 Indépendance d'évènements.

Définissons les évènements $A = \{\text{au moins un ascenseur du bâtiment Chevaleret sera en panne demain}\}$ et $B = \{\text{l'indice CAC40 va s'effondrer demain}\}$. Intuitivement, la réalisation (ou non-réalisation) de l'évènement A n'a aucune influence sur les chances que l'évènement B se produise, ce qui s'écrit $P(A|B) = P(A)$ ou encore $P(A \cap B) = P(A)P(B)$. De tels évènements seront dits *indépendants*.

Définition 1.5.1 Deux évènements A et B sont dits *indépendants* s'ils vérifient l'égalité :

$$P(A \cap B) = P(A)P(B)$$

La notion d'indépendance est essentielle en calcul des probabilités; elle fera souvent partie de nos hypothèses en modélisation. Par exemple, pour étudier la durée de vie d'un circuit électronique constitué de plusieurs composants élémentaires, nous supposerons souvent que ces différents composants tombent en panne indépendamment les uns des autres.

Remarque: Il est facile de vérifier que si le couple (A, B) est constitué d'évènements indépendants, alors il en est de même pour les couples suivants :

$$(A, B^c), (A^c, B), (A^c, B^c).$$

Exemple: On tire une carte au hasard dans un jeu de 52 et l'on considère les évènements $A = \{\text{c'est un roi}\}$ et $B = \{\text{c'est un coeur}\}$. Nous calculons facilement : $P(A \cap B) = 1/52$, $P(A) = 4/52 = 1/13$, $P(B) = 13/52 = 1/4$. Nous en déduisons que les évènements A et B sont indépendants, conformément à la définition précédente.

Plus généralement, dans un tel tirage, la figure et la couleur de la carte sont indépendantes.

Définition 1.5.2 Soient $(A_1, \dots, A_n) \in \mathcal{A}^n$ des évènements.

Nous dirons que ces évènements sont 2 à 2 indépendants si pour tout $(i, j) \in \{1, \dots, n\}$ avec $i \neq j$, on a $P(A_i \cap A_j) = P(A_i)P(A_j)$.

Nous dirons que ces évènements sont mutuellement indépendants (ou encore indépendants dans leur ensemble) si pour toute sous-famille d'indices $\{i_1, \dots, i_k\} \subset \{1, \dots, n\}$, on a :

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k})$$

Exemple.

Une population \mathcal{S} de cardinal N est divisée en 2 catégories de cardinaux respectifs N_1, N_2 . Nous tirons un n -échantillon "au hasard" (probabilité uniforme) et nous considérons les évènements suivants pour $1 \leq i \leq n$:

$$A_i = \{\text{Le } i\text{-ème objet tiré est de la première catégorie}\}$$

Nous allons montrer que les évènements $A_i, 1 \leq i \leq n$ sont mutuellement indépendants.

Nous prenons donc un entier $1 \leq k \leq n$ et des indices $i_1 < i_2 < \dots < i_k$ et nous calculons la probabilité $P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k})$. Le cardinal de l'ensemble des n -échantillons qui ont un objet de première catégorie aux places i_1, \dots, i_k vaut $N_1^k N^{n-k}$ donc :

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \frac{N_1^k N^{n-k}}{N^n} = p^k,$$

où l'on a posé $p = N_1/N$. En prenant le cas particulier $k = 1$, nous en déduisons que pour tout indice $1 \leq i \leq n$, nous avons $P(A_i) = p$ d'où finalement le résultat cherché :

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k}).$$

Remarque: Des évènements mutuellement indépendants sont clairement indépendants 2 à 2 mais la réciproque est fautive, comme nous allons le voir dans le contre-exemple ci-dessous. Autrement dit, la propriété de mutuelle indépendance est plus forte que la propriété d'indépendance 2 à 2.

Lorsqu'un énoncé dit que des évènements A_1, \dots, A_n sont indépendants sans autre précision, c'est toujours l'indépendance mutuelle, c'est-à-dire la plus forte, qu'il faut comprendre.

Contre-exemple: On jette une pièce équilibrée 2 fois de suite et l'on considère les évènements : $A = \{\text{on obtient pile au premier jet}\}$, $B = \{\text{on obtient pile au second jet}\}$, $C = \{\text{les résultats des deux jets sont différents}\}$.

Il est facile de calculer :

$$P(A) = P(B) = P(C) = 1/2$$

et

$$P(A \cap B) = P(A \cap C) = P(B \cap C) = 1/4.$$

Nous en déduisons que les évènements A, B et C sont indépendants 2 à 2. En revanche, ces évènements ne sont pas mutuellement indépendants puisque :

$$P(A \cap B \cap C) = P(\emptyset) = 0 \neq P(A)P(B)P(C).$$

□

Le lecteur pourra démontrer à titre d'exercice la proposition suivante.

Proposition 1.5.3 *Des évènements $(A_1, \dots, A_n) \in \mathcal{A}^n$ sont mutuellement indépendants si et seulement si on a*

$$P\left(\bigcap_{i=1}^n C_i\right) = \prod_{i=1}^n P(C_i),$$

pour tous $C_i \in \{\Omega, A_i, A_i^c, \emptyset\}$, $1 \leq i \leq n$.

Nous allons maintenant mettre en évidence une situation générale dans laquelle la probabilité binomiale apparaît naturellement. Le lecteur notera que l'exemple ci-dessus (tirage d'un n -échantillon au hasard) en était un cas particulier.

Proposition 1.5.4 *Soient $A_i, 1 \leq i \leq n$, des évènements indépendants ayant tous la même probabilité $p \in [0, 1]$. Pour tout $0 \leq k \leq n$, la probabilité que k exactement d'entre eux soient réalisés est égale à :*

$$B_{n,p}(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}$$

Démonstration: Notons Θ l'ensemble des parties de $\{1, \dots, n\}$ qui sont de cardinal k , de sorte que $\#\Theta = \binom{n}{k}$.

L'évènement dont nous voulons calculer la probabilité est le suivant :

$$\bigcup_{\theta \in \Theta} \left[\left(\bigcap_{i \in \theta} A_i \right) \cap \left(\bigcap_{i \notin \theta} A_i^c \right) \right]$$

Un instant de réflexion montre qu'il s'agit d'une union d'évènements deux à deux disjoints, si bien que la probabilité cherchée s'écrit :

$$\sum_{\theta \in \Theta} P \left[\left(\bigcap_{i \in \theta} A_i \right) \cap \left(\bigcap_{i \notin \theta} A_i^c \right) \right]$$

À cause de l'hypothèse d'indépendance des $A_i, 1 \leq i \leq n$, cette probabilité s'écrit encore :

$$\sum_{\theta \in \Theta} p^k (1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k}$$

□

Après avoir défini l'indépendance d'une famille finie d'évènements, nous terminons cette section en passant au cas d'une famille infinie dénombrable.

Définition 1.5.5 Une suite d'évènements $(A_n)_{n \in \mathbb{N}}$ est dite indépendante si pour toute sous-famille finie d'indices $\{i_1, \dots, i_k\} \subset \mathbb{N}$ (avec $k \in \mathbb{N}^*$), on a l'égalité :

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k})$$

On peut se demander si une telle suite existe. Nous verrons plus loin que la réponse est positive avec le modèle d'un jeu de pile ou face infini.

Chapitre 2

Espaces de probabilités finis.

2.1 Algèbres de parties

2.1.1 Indicatrices

Soit A une algèbre sur Ω ; on définit la fonction $\mathbf{1}_A : \Omega \rightarrow \{0, 1\}$ par :

$\mathbf{1}_A(\omega) = 0$ si $\omega \notin A$, $\mathbf{1}_A(\omega) = 1$ si $\omega \in A$.

Montrer les égalités suivantes : $\mathbf{1}_{A \cap B} = \inf(\mathbf{1}_A, \mathbf{1}_B) = \mathbf{1}_A \mathbf{1}_B$,

$\mathbf{1}_{A \cup B} = \sup(\mathbf{1}_A, \mathbf{1}_B) = \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_A \mathbf{1}_B$,

$\mathbf{1}_{A^c} = 1 - \mathbf{1}_A$,

ainsi que l'équivalence : $\forall \omega \in \Omega, \mathbf{1}_A(\omega) \leq \mathbf{1}_B(\omega) \Leftrightarrow A \subset B$.

On rappelle que la différence symétrique est définie par :

$A \Delta B = (A \cup B) \setminus (A \cap B) = (A^c \cap B) \cup (A \cap B^c)$.

Montrer que :

$\mathbf{1}_{A \Delta B} = |\mathbf{1}_A - \mathbf{1}_B| = \mathbf{1}_A + \mathbf{1}_B \pmod{2} = \mathbf{1}_A + \mathbf{1}_B - 2 \mathbf{1}_A \mathbf{1}_B$.

En déduire que la différence symétrique est commutative, associative et que l'intersection

est distributive par rapport à la différence symétrique.

2.1.2 Algèbre engendrée par une partition.

Soit (P_1, \dots, P_n) une partition de Ω . Décrire l'algèbre \mathcal{A} qu'elle engendre.

2.1.3 Réunions d'algèbres.

Montrer que la réunion d'une suite croissante d'algèbres est une algèbre. Est-ce que la réunion d'une suite quelconque d'algèbres est une algèbre ?

2.1.4 *Description de l'algèbre engendrée.

Soit \mathcal{E} une famille de parties d'un ensemble Ω . Montrer que l'algèbre engendrée par \mathcal{E} est égale à l'ensemble des réunions finies d'intersections finies d'éléments de \mathcal{E} ou de complémentaires d'éléments de \mathcal{E} :

$$\mathcal{A}(\mathcal{E}) = \left\{ \bigcup_{i=1}^n \bigcap_{j=1}^{m_i} C_{i,j} \text{ avec } \forall i, j \quad C_{i,j} \in \mathcal{E} \text{ ou } C_{i,j}^c \in \mathcal{E} \right\}$$

Remarque: Il est impossible de décrire de façon analogue la σ -algèbre engendrée par \mathcal{E} . Qu'est-ce qui ne se généralise pas dans la démonstration précédente ?

2.2 Formule de Poincaré.

[FOA 17]

2.2.1 Démonstration par récurrence.

Soit \mathcal{A} une algèbre sur Ω et P une probabilité sur \mathcal{A} . On se donne une famille finie (A_1, \dots, A_n) d'éléments de \mathcal{A} .

Etablir par récurrence la formule suivante, dite de Poincaré :

$$P(\cup_{i=1}^n A_i) = \sum_{1 \leq i \leq n} P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) + \dots + (-1)^{n-1} P(A_1 \cap \dots \cap A_n).$$

2.2.2 Application : Indicatrice d'Euler

[FOA 37-38]

On choisit au hasard un nombre parmi $\{1, \dots, n\}$, tous les choix étant équiprobables. Soit p un entier inférieur à n . On note A_p l'événement "le nombre choisi est divisible par p ".

- Calculer $P(A_p)$ lorsque p divise n .
- Soit $n = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$ la décomposition en facteurs premiers de l'entier n . Que représente l'évènement $(A_{p_1} \cup \dots \cup A_{p_r})^c$?
- On note $\phi(n)$ le nombre d'entiers strictement inférieurs à n et premiers avec n . Montrer que l'on a

$$\phi(n) = n \prod_{p \text{ premier, } p|n} \left(1 - \frac{1}{p}\right).$$

2.2.3 Application : Facteur distrait.

[FOA 38-39]

Un facteur dispose de n lettres adressées à n destinataires distincts. Il fait la distribution au hasard entre ces n personnes. Quelle est la probabilité de l'événement "une lettre au moins parvient à son destinataire" ?

En donner une approximation pour n assez grand.

2.3 Equiprobabilité.

2.3.1 Trois dés.

[COT 5-6]

On jette trois dés à 6 faces bien équilibrés.

- Calculer la probabilité d'obtenir au moins un as.
- Que vaut la probabilité d'obtenir au moins deux faces portant le même chiffre ?
- Calculer la probabilité que la somme des points marqués sur les trois faces soit paire.

2.3.2 Happy birthday to you !

[FOA 31]

Quelle est la probabilité que dans votre promotion deux agrégatifs au moins aient leur anniversaire le même jour ?

Remarque : Pour simplifier, on ne s'occupera pas du 29 Février.

2.3.3 Contrôle de fabrication.

[REV 10-11]

Un lot de 10000 objets en comporte certains qui sont défectueux dans une proportion de 1 sur 20. On prélève une sous-population de 20 objets. Que vaut la probabilité qu'elle en contienne n défectueux ? (en donner la valeur exacte puis une approximation).

En particulier, donner une valeur approchée de la probabilité pour que cette sous-population ne contienne aucun objet défectueux.

Le résultat est-il conforme à votre intuition ?

2.3.4 Boules et cases.

[DAC¹ exo page 7]

Soit Ω l'ensemble des configurations que l'on peut obtenir en répartissant r boules indiscernables dans n cases numérotées.

1. Montrer que $\#\Omega$ est égal au nombre de solutions $(r_1, \dots, r_n) \in \mathbb{N}^n$ de l'équation : $r_1 + r_2 + \dots + r_n = r$.
2. Montrer que ce nombre vaut C_{n+r-1}^r de 3 façons différentes :
 - (a) par récurrence sur n à l'aide de l'identité : $C_n^p = C_{n-1}^{p-1} + C_{n-2}^{p-1} + \dots + C_{p-1}^{p-1}$.
 - (b) en développant en série entière les deux membres de l'égalité : $\frac{1}{(1-t)^n} = \left(\sum_{k=0}^{\infty} t^k \right)^n$.
 - (c) en identifiant Ω à l'ensemble des façons d'ordonner en ligne r boules indiscernables et $n-1$ séparations (indiscernables aussi).
3. On répartit au hasard r boules indiscernables dans n cases. Calculer la probabilité qu'aucune case ne soit vide.

Chapitre 3

Probabilités conditionnelles. évènements indépendants.

3.1 Probabilités conditionnelles.

3.1.1 Maladie rare.

[COT 11]

On considère une certaine maladie qui touche 1/1000e de la population. Un laboratoire d'analyse de sang assure avec une fiabilité de 99% la détection de cette maladie lorsqu'elle est effectivement présente. Cependant, le test indique aussi un résultat faussement positif pour 0,2% des personnes réellement saines à qui on l'applique.

Quelle est la probabilité qu'une personne soit vraiment malade sachant que son test est positif? Commenter le résultat.

3.1.2 Le rouge et le noir.

[ROS 66-67]

On considère 3 cartes à jouer de même forme mais de couleurs différentes : la première est noire des deux côtés, la seconde rouge des deux côtés, tandis que la troisième a une face noire et une face rouge.

On mélange les trois cartes au fond d'un chapeau puis on en tire une carte au hasard, dont on ne montre qu'une face. Sachant que cette face est rouge, quelle est la probabilité que l'autre face soit noire ?

3.1.3 Loi de succession de Laplace.

[COT 15-16]

On dispose de $N + 1$ urnes, numérotées de 0 à N . L'urne numéro k contient k boules rouges et $N - k$ boules blanches. On choisit une urne au hasard.

Sans connaître son numéro, on en tire n fois de suite une boule, avec remise après chaque tirage. Quelle est la probabilité que le $(n + 1)^e$ tirage donne encore une boule rouge sachant que, au cours des n premiers tirages, seules des boules rouges ont été tirées? Calculer la limite de cette probabilité lorsque $N \rightarrow \infty$.

3.2 évènements indépendants.

3.2.1 Événement auto-indépendant.

Montrer qu'un évènement A est indépendant de lui-même ssi $P(A) = 0$ ou $P(A) = 1$.

3.2.2 Bruit qui court.

[COT 13]

La personne I_1 reçoit l'information 0 ou 1 et la transmet telle quelle à I_2 avec la probabilité p , I_2 de même à I_3 , etc... I_n la transmet au monde entier. On suppose que les n personnes I_1, \dots, I_n sont indépendantes. Quelle est la probabilité p_n que le monde reçoive la bonne information? Calculer $\lim_{n \rightarrow \infty} p_n$.

3.2.3 Indicatrice d'Euler (bis).

On choisit au hasard un nombre parmi $\{1, \dots, n\}$, tous les choix étant équiprobables. Soit p un entier inférieur à n . On note A_p l'évènement "le nombre choisi est divisible par p ".

1. Calculer $P(A_p)$ lorsque p divise n .
2. On suppose que (p_1, \dots, p_k) sont des diviseurs premiers distincts de n , montrer que les évènements A_{p_1}, \dots, A_{p_k} sont indépendants.
3. On note $\phi(n)$ le nombre d'entiers strictement inférieurs à n et premiers avec n . Montrer que l'on a

$$\phi(n) = n \prod_{p \text{ premier, } p|n} \left(1 - \frac{1}{p}\right).$$

Corrigé :

1. Puisque p divise n (notation : $p|n$), il existe un entier m tel que $n = mp$; on a alors : $A_p = \{p, 2p, \dots, mp\}$. On en déduit : $P(A_p) = \frac{m}{n} = \frac{1}{p}$.
2. Considérant une sous-famille A_{p_1}, \dots, A_{p_l} , nous remarquons que

$$\bigcap_{j=1}^l A_{p_{i_j}} = A_{\prod_{j=1}^l p_{i_j}},$$

puisque les p_{i_j} sont premiers. Pour la même raison, $(\prod_{j=1}^l p_{i_j})|n$ donc on peut appliquer a. pour obtenir :

$$P(\bigcap_{j=1}^l A_{p_{i_j}}) = \frac{1}{\prod_{j=1}^l p_{i_j}} = \prod_{j=1}^l P(A_{p_{i_j}}),$$

ce qui prouve l'indépendance demandée.

3. Prenons ici p_1, \dots, p_k **tous** les diviseurs premiers de n . D'après b), on a :

$$P(\bigcap_{i=1}^k A_{p_i}^c) = \prod_{p \text{ premier, } p|n} \left(1 - \frac{1}{p}\right).$$

Or un entier appartient à $\bigcap_{i=1}^k A_{p_i}^c$ **ssi** il n'est divisible par aucun des diviseurs premiers de n , autrement dit **ssi** il est premier avec n . En passant aux probabilités, on en déduit :

$$\prod_{p \text{ premier, } p|n} \left(1 - \frac{1}{p}\right) = \frac{\phi(n)}{n}.$$

3.2.4 Loi de Hardy-Weinberg.

[REV 18](non corrigé)

Les gènes se présentent le plus souvent en paires et sous deux formes alléliques A,a, ce qui donne trois génotypes AA,Aa,aa. Ces génotypes s'expriment par différents phénotypes ; par exemple, dans le cas de l'albinisme, les individus aa sont albinos, contrairement aux individus AA ou Aa.

Chaque individu reçoit au hasard un gène de chacun de ses parents et donc chaque allèle composant le gène d'un des parents a la probabilité 1/2 de passer à l'enfant. Les génotypes des parents sont indépendants. On note p, q, r les probabilités qu'un adulte, dans une population donnée, ait les génotypes AA,Aa,aa.

1. En posant $\theta = p + q/2$, calculer les probabilités P, Q, R qu'un enfant ait les génotypes AA,Aa,aa.

2. Montrer que $Q^2 = 4PR$.
3. A quelle condition sur p, q, r a-t-on $P = p, Q = q, R = r$? Montrer qu'elle est toujours réalisée dès la 2ème génération.

Corrigé :

1. Nous prenons pour espace Ω l'ensemble des triplets (génotype de l'enfant, génotype du père, génotype de la mère).

L'évènement "L'enfant a un génotype AA" s'écrit donc :

$$\{(AA, AA, AA); (AA, Aa, AA); (AA, AA, Aa); (AA, Aa, Aa)\}$$

et, d'après les lois de la génétique rappelées dans l'énoncé, admet pour probabilité :

$$P = p^2 + pq/2 + pq/2 + q^2/4 = \theta^2.$$

Un raisonnement symétrique (en échangeant les lettres A et a) nous donne :

$$R = r^2 + rq/2 + rq/2 + q^2/4 = (1 - \theta)^2.$$

Enfin, puisque $P + Q + R = 1$, on obtient : $Q = 2\theta(1 - \theta)$.

2. Cette relation résulte immédiatement de ce qui précède.
3. D'après la question précédente, on a nécessairement $q^2 = 4pr$.

Réciproquement, si cette condition est vérifiée, les égalités obtenues dans la première question nous donnent :

$$P = p^2 + pq + q^2/4 = p^2 + pq + pr = p(p + q + r) = p$$

$$\text{et } R = r^2 + rq + q^2/4 = r^2 + rq + rp = r(p + q + r) = r.$$

Il en résulte $Q = q$ puisqu'on a : $P + Q + R = p + q + r = 1$.

Finalement, l'égalité $q^2 = 4pr$ est une condition nécessaire et suffisante de "stationnarité" des probabilités des différents génotypes au cours des générations.

Ainsi, le résultat du 2) nous dit que dans tous les cas, la stationnarité s'établit dès la deuxième génération.

Chapitre 4

Complément de cours : Modèle définitif.

4.1 Espaces de probabilités.

4.1.1 Le modèle probabiliste dans le cas général.

Le modèle développé dans le chapitre 1 s'appliquait bien au cas Ω fini, qui couvre un certain nombre de situations courantes : jeux de hasard, sondages d'opinion, contrôles de qualité etc. Néanmoins, même avec un jeu aussi simple que le pile ou face, on ne peut pas toujours se restreindre au cas $\#\Omega < +\infty$: si nous étudions par exemple le rang d'apparition du premier pile lors de jets successifs, on est amené à prendre $\Omega = \mathbb{N}$ puisqu'il est impossible de borner à l'avance ce rang par un entier fixé.

Lorsque le temps intervient, nous voyons apparaître naturellement des espaces encore plus "gros" (infinis non dénombrables) :

- durée de vie d'un composant électronique : $\Omega = \mathbb{R}_+$
- évolution du cours d'un actif financier : $\Omega = C^0([0, T], \mathbb{R})$
- déplacement d'une particule : $\Omega = C^0([0, T], \mathbb{R}^3)$

Or, lorsque Ω est infini, le modèle du chapitre 1 présente des insuffisances : la structure d'algèbre pour l'ensemble \mathcal{A} des événements et l'additivité d'une probabilité sont des hypothèses trop faibles pour nous permettre de mener effectivement des calculs, comme nous allons le voir sur le simple exemple suivant.

Si nous revenons au rang du premier pile lors de jets successifs d'une pièce équilibrée, une première question qui se pose est de savoir si ce rang est fini ; autrement dit, peut-on obtenir face indéfiniment ? Intuitivement, cela paraît extrêmement improbable et l'on s'attend à ce que la probabilité en soit nulle. Plus précisément, il est naturel de faire le raisonnement suivant :

L'évènement $A_n = \{ \text{on a obtenu face lors des } n \text{ premiers jets} \}$ est de probabilité $1/2^n$ et donc l'évènement $A = \{ \text{on obtient face indéfiniment} \}$ a pour probabilité $\lim 1/2^n = 0$.

Pour rendre ce raisonnement rigoureux, il faudrait déjà que l'ensemble

$$A = \bigcap_{n \in \mathbb{N}^*} A_n$$

soit effectivement un évènement, ce qui n'est pas assuré par l'hypothèse \mathcal{A} algèbre : nous avons besoin de supposer que \mathcal{A} est une σ -algèbre (ou tribu). En outre, nous voudrions pouvoir écrire que :

$$A = \lim \downarrow A_n \implies P(A) = \lim \downarrow P(A_n)$$

En passant aux évènements complémentaires, cela revient à demander la propriété de continuité à gauche : $P(\lim \uparrow B_n) = \lim \uparrow P(B_n)$.

Or, pour une application additive comme P , cette propriété équivaut à la σ -additivité (le lecteur pourra le démontrer à titre d'exercice).

Pour pouvoir mener nos calculs, nous allons donc imposer des contraintes supplémentaires sur les éléments constitutifs de notre espace de probabilité : \mathcal{A} doit être une tribu et P une application σ -additive.

Autrement dit, nous remplaçons la définition du chapitre 1 par la définition suivante, que nous utiliserons dans tout le reste de ce cours :

Définition 4.1.1 *Un espace de probabilité est un espace mesuré (Ω, \mathcal{A}, P) de masse totale 1, c'est-à-dire tel que $P(\Omega) = 1$.*

Notons que dans le cas $\#\Omega < +\infty$, cette définition est équivalente à celle du chapitre 1. En effet, dans ce cas, $\mathcal{A} \subset \mathcal{P}(\Omega)$ est fini si bien que \mathcal{A} algèbre et \mathcal{A} σ -algèbre sont équivalents, de même que P additive et P σ -additive.

Exemple. Construisons un modèle probabiliste qui corresponde à l'expérience "tirer un nombre réel au hasard (uniformément) entre 0 et 1" – ce qu'est censée faire la touche RAND d'une calculatrice.

Nous avons donc clairement $\Omega = [0, 1]$.

Pour tous réels $0 \leq a < b \leq 1$, nous demanderons à notre observateur d'être capable de répondre à la question : "est-ce que le résultat ω de l'expérience appartient à l'intervalle $[a, b]$?". Nous choisissons donc de munir $\Omega = [0, 1]$ de la plus petite tribu qui contienne tous les intervalles $[a, b]$, $0 \leq a < b \leq 1$, c'est-à-dire $\mathcal{A} = \mathcal{B}([0, 1])$.

Enfin, l'idée d'un tirage uniforme (sous-entendue dans le langage courant lorsqu'on dit "tirer au hasard" sans autre précision) nous amène à supposer

que nous affecterons la même probabilité à deux sous-intervalles de $[0, 1]$ de même longueur (qui sont donc chacun un translaté de l'autre). La théorie de la mesure nous dit alors qu'il existe une unique mesure de probabilité sur $([0, 1], \mathcal{B}([0, 1]))$ possédant cette propriété : la mesure de Lebesgue. Nous achevons donc la construction de notre modèle en prenant $P = \lambda|_{[0,1]}$.

Exercice : Montrer que la proposition 1.4.1 reste valable dans le cadre de notre modèle définitif. Essentiellement, ceci consiste à vérifier que si P est supposée σ -additive, alors P^E conserve cette propriété.

4.1.2 Limites supérieure et inférieure. Lemme de Borel-Cantelli.

Une σ -algèbre \mathcal{A} est stable par les opérations ensemblistes suivantes, qui sont susceptibles d'interprétations intéressantes en probabilité :

Définition 4.1.2 Soit (A_n) une suite d'évènements de \mathcal{A} . Alors la limite supérieure de cette suite est définie comme suit :

$$\overline{\lim}A_n = \bigcap_n \bigcup_{k \geq n} A_k \in \mathcal{A}.$$

De même, la limite inférieure de cette suite est définie par :

$$\underline{\lim}A_n = \bigcup_n \bigcap_{k \geq n} A_k \in \mathcal{A}.$$

Interprétation Les trois propriétés suivantes sont équivalentes :

1. $\omega \in \overline{\lim}A_n$.
2. Il existe une sous-suite infinie de la suite (A_n) dont tous les éléments contiennent ω .
3. Il existe une suite strictement croissante d'entiers $(n_k(\omega))_{k \in \mathbb{N}}$ telle que, pour tout $k \in \mathbb{N}$, $\omega \in A_{n_k(\omega)}$.

De façon symétrique, nous avons équivalence entre les trois propriétés suivantes :

1. $\omega \in \underline{\lim}A_n$.
2. $\omega \in A_n$ pour tous les indices n sauf un nombre fini d'entre eux.
3. ω appartient à tous les A_n à partir d'un certain rang, ce qui s'écrit :
 $\exists N(\omega) \in \mathbb{N} \quad \forall n \geq N(\omega) \quad \omega \in A_n$

Dans les exercices, il est important de tenir compte de la dépendance en ω de la suite $(n_k(\omega))_{k \in \mathbb{N}}$ comme de l'entier $N(\omega)$!

Exemple 4.1.3 Dans un jeu de pile ou face infini, si nous considérons les évènements $A_n = \{\text{on a obtenu pile au } n^{\text{e}} \text{ coup}\}$, nous avons :

$$\begin{aligned}\overline{\lim} A_n &= \{\text{on a tiré pile une infinité de fois}\} \\ \underline{\lim} A_n &= \{\text{on a tiré face un nombre fini de fois}\}\end{aligned}$$

Propriétés Pour toute suite (A_n) , les deux propriétés suivantes sont immédiates et le lecteur pourra démontrer la troisième à titre d'exercice :

1. $\underline{\lim} A_n \subset \overline{\lim} A_n$.
2. $(\underline{\lim} A_n)^c = \overline{\lim} A_n^c$.
3. $\mathbf{1}_{\underline{\lim} A_n} = \underline{\lim} \mathbf{1}_{A_n}$ et $\mathbf{1}_{\overline{\lim} A_n} = \overline{\lim} \mathbf{1}_{A_n}$

Une probabilité étant une mesure finie, nous pourrons toujours utiliser la propriété de *continuité à droite* : pour toute suite $(A_n) \in \mathcal{A}^{\mathbb{N}}$ décroissante au sens de l'inclusion (i.e. $\forall n \in \mathbb{N} A_{n+1} \subset A_n$), nous avons

$$P(\lim \downarrow A_n) = \lim \downarrow P(A_n)$$

Pour terminer cette section, nous allons illustrer l'utilisation de la continuité à droite de P par deux propositions et un exemple.

Proposition 4.1.4 Pour toute suite $(A_n) \in \mathcal{A}^{\mathbb{N}}$, nous avons les inégalités :

$$P(\underline{\lim} A_n) \leq \underline{\lim} P(A_n) \leq \overline{\lim} P(A_n) \leq P(\overline{\lim} A_n)$$

En particulier, si (A_n) est convergente de limite A , alors $P(A_n) \rightarrow P(A)$.

Démonstration: Nous ne démontrerons que la dernière inégalité. En effet, la première se prouve de façon symétrique (ou par passage au complémentaire) et la seconde est évidente.

$$P(\overline{\lim} A_n) = P\left(\lim_n \downarrow \bigcup_{k \geq n} A_k\right) = \lim_n \downarrow P\left(\bigcup_{k \geq n} A_k\right)$$

Or nous avons clairement la minoration :

$$P\left(\bigcup_{k \geq n} A_k\right) \geq \sup_{k \geq n} P(A_k)$$

En notant que le membre de droite est décroissant en n , nous en déduisons :

$$P(\overline{\lim} A_n) \geq \lim_n \downarrow \sup_{k \geq n} P(A_k) = \overline{\lim} P(A_n)$$

□

Exemple: Revenons à l'exemple 4.1.3 en admettant provisoirement que nous pouvons construire un modèle du pile ou face infini avec une pièce bien équilibrée, ce que nous ferons plus tard. En utilisant la *sous- σ -additivité* de P , nous pouvons écrire :

$$P((\overline{\lim} A_n)^c) = P(\underline{\lim} A_n^c) = P\left(\bigcup_n \bigcap_{k \geq n} A_k^c\right) \leq \sum_n P\left(\bigcap_{k=n}^{+\infty} A_k^c\right)$$

En utilisant maintenant la continuité à droite de P , nous calculons :

$$P\left(\bigcap_{k=n}^{+\infty} A_k^c\right) = P\left(\lim_N \downarrow \bigcap_{k=n}^N A_k^c\right) = \lim_N \downarrow P\left(\bigcap_{k=n}^N A_k^c\right)$$

Puisque la pièce est bien équilibrée, on calcule facilement que le dernier terme vaut $\lim_N \frac{1}{2^{N-n+1}} = 0$. Finalement, nous avons prouvé que $P((\overline{\lim} A_n)^c) = 0$, c'est-à-dire qu'avec probabilité 1, on obtient une infinité de piles. La pièce étant équilibrée, pile et face jouent des rôles symétriques si bien que nous pouvons affirmer que face apparaîtra également une infinité de fois avec probabilité 1.

Le résultat suivant, appelé *lemme de Borel-Cantelli*, est riche en applications malgré sa simplicité.

Lemme 4.1.5 *Soit $(A_n) \in \mathcal{A}^{\mathbb{N}}$ t.q. $\sum_n P(A_n) < +\infty$. Alors $P(\overline{\lim} A_n) = 0$.*

Démonstration: Par continuité à droite et sous- σ -additivité de P , nous avons

$$P(\overline{\lim} A_n) = \lim_n \downarrow P\left(\bigcup_{k \geq n} A_k\right) \leq \lim_n \downarrow \sum_{k \geq n} P(A_k)$$

Le dernier terme qui apparaît étant le reste d'une série convergente, sa limite est nulle. \square

4.2 Variables aléatoires.

Dans tout ce qui suit, on a un espace de probabilité sous-jacent (Ω, \mathcal{A}, P) .

4.2.1 Définition et premières propriétés

Définition 4.2.1 *On appelle variable aléatoire réelle (v.a.r.) toute application mesurable $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Pour tout entier $d \geq 2$, on appelle vecteur aléatoire de dimension d (\vec{v} .a. de dim. d) toute application mesurable $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.*

Les trois propositions suivantes découlent des propriétés générales des fonctions mesurables :

Proposition 4.2.2 *Si X et Y sont des variables aléatoires à valeurs dans \mathbb{R} alors, pour tout $(\lambda, \mu) \in \mathbb{R}^2$, $\lambda X + \mu Y$ et XY sont aussi des variables aléatoires à valeurs dans \mathbb{R} .*

Si (X_n) est une suite de variables aléatoires réelles, alors $\inf X_n$, $\sup X_n$, $\underline{\lim} X_n$, $\overline{\lim} X_n$ sont aussi des variables aléatoires réelles.

L'ensemble de convergence (dans $\overline{\mathbb{R}}$) de la suite (X_n) , qui s'écrit :

$$\{\omega \in \Omega, \quad \underline{\lim} X_n(\omega) = \overline{\lim} X_n(\omega)\}$$

est un évènement et, si la suite (X_n) converge (dans $\overline{\mathbb{R}}$) en tout point $\omega \in \Omega$, $\lim X_n$ est une variable aléatoire réelle.

Proposition 4.2.3 *Soit $X : (\Omega, \mathcal{A}) \rightarrow (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ une application. Nous avons l'équivalence :*

$$X \text{ v.a.r.} \Leftrightarrow \forall a \in \mathbb{Q} \quad \{X \leq a\} \in \mathcal{A}$$

Proposition 4.2.4 *Soit $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ une application. Alors, X est un $\vec{v}.a.$ de dim. d si et seulement si, pour tout pavé $\prod_{i=1}^d]a_i, b_i[$ à extrémités rationnelles, nous avons :*

$$\left\{ X \in \prod_{i=1}^d]a_i, b_i[\right\} \in \mathcal{A}$$

Cette dernière proposition nous permet de préciser la description d'un $\vec{v}.a.$ de dim. d . Pour $1 \leq i \leq d$, notons $\pi_i : \mathbb{R}^d \rightarrow \mathbb{R}$ la i ème projection canonique définie par $\pi_i(x_1, \dots, x_i, \dots, x_d) = x_i$. Nous posons alors $X_i = \pi_i \circ X$ de sorte que $X = (X_1, \dots, X_d)$. Le lecteur pourra démontrer à titre d'exercice :

Proposition 4.2.5 *Soit $X = (X_1, \dots, X_d)$ une application de (Ω, \mathcal{A}) dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Nous avons l'équivalence suivante :*

$$X \vec{v}.a. \text{ de dim. } d \Leftrightarrow \forall i = 1, \dots, d \quad X_i \text{ v.a. à valeurs dans } \mathbb{R}$$

4.2.2 Classes d'équivalence de variables aléatoires pour l'égalité presque sûre

4.2.3 Loi d'une variable aléatoire

Rappelons ici une définition et une propriété bien connues en théorie de la mesure.

Définition 4.2.6 Une v.a.r. est dite étagée si elle ne prend qu'un nombre fini de valeurs.

Proposition 4.2.7 Toute v.a.r. positive est limite croissante d'une suite de variables étagées.

Si X est un vecteur aléatoire à valeurs dans \mathbb{R}^d , la classe $X^{-1}(\mathcal{B}(\mathbb{R}^d))$ est une tribu incluse dans \mathcal{A} . C'est en fait la plus petite tribu sur Ω qui rend l'application $X : \Omega \rightarrow \mathbb{R}^d$ mesurable : on l'appelle *tribu engendrée par la variable X* et on la note $\sigma(X)$. Cette tribu est associée à un observateur qui est capable après l'expérience de nous renseigner sur la valeur prise par la variable X ... et c'est tout ! La proposition suivante décrit exactement le type de v.a. sur laquelle cet observateur pourra nous renseigner.

Proposition 4.2.8 Une v.a. Y à valeurs dans $\mathbb{R}^{d'}$ est $\sigma(X)$ -mesurable si et seulement si $\exists f : \mathbb{R}^d \rightarrow \mathbb{R}^{d'}$ borélienne telle que $Y = f(X)$.

Démonstration: Le sens (\Leftarrow) est immédiat par composition d'applications mesurables.

Passons au sens (\Rightarrow). Puisqu'un vecteur aléatoire Y à valeurs dans $\mathbb{R}^{d'}$ n'est autre qu'un d' -uplet $(Y_1, \dots, Y_{d'})$ constitué par des variables aléatoires à valeurs dans \mathbb{R} , il suffit de faire la preuve dans le cas $d' = 1$.

Supposons donc que Y est une variable aléatoire à valeurs dans \mathbb{R} . En écrivant $Y = Y_+ - Y_-$, nous constatons que nous pouvons même nous ramener au cas $Y \geq 0$.

Nous savons d'après la proposition 4.2.7 qu'il existe alors une suite (Y_n) de variables étagées $\sigma(X)$ -mesurables telles que $Y = \lim \uparrow Y_n$. Or une fonction étagée $\sigma(X)$ -mesurable s'écrit sous la forme :

$$\sum_{i=1}^n \alpha_i \mathbf{1}_{X^{-1}(A_i)} = \left(\sum_{i=1}^n \alpha_i \mathbf{1}_{A_i} \right) \circ X,$$

où A_1, \dots, A_n sont des boréliens de \mathbb{R}^d . Nous pouvons donc écrire $Y_n = f_n \circ X$ avec $f_n : \mathbb{R}^d \rightarrow \mathbb{R}$ borélienne. Si nous posons $f = \overline{\lim} f_n$, alors f est borélienne et nous avons $Y = f(X)$. \square

4.2.4 Lois de probabilités discrètes

Une variable aléatoire X suit la loi binomiale $B(n, p)$ (où $n \in \mathbb{N}^*$ et $p \in [0, 1]$) si et seulement si :

$$\forall k = 0, \dots, n, \quad P(X = k) = C_n^k p^k (1-p)^{n-k}$$

La variable aléatoire X est alors presque sûrement à valeurs dans $\{0, \dots, n\}$. Étant bornée, X est intégrable et nous calculons son espérance comme suit :

$$E[X] = \sum_{k=0}^n k C_n^k p^k (1-p)^{n-k} = \sum_{k=1}^n \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k}$$

ce qui s'écrit encore :

$$E[X] = \sum_{k=1}^n n p C_{n-1}^{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)}$$

Un simple changement d'indice nous donne alors :

$$E[X] = n p \sum_{j=0}^{n-1} C_{n-1}^j p^j (1-p)^{n-1-j} = n p.$$

Nous avons déjà rencontré la loi binomiale $B(n, p)$ dans le chapitre 1 où p représentait la proportion d'objets de première catégorie lors d'un tirage avec remise : le résultat que nous venons d'obtenir s'interprète facilement dans ce contexte.

4.2.5 Exercice proposé : Taux de panne

[COT 16-18]

Soit T une variable aléatoire prenant ses valeurs dans \mathbb{N} telle que :

$$\forall n \in \mathbb{N} \quad P(T \geq n) > 0$$

On appelle taux de panne la suite $(\theta_n)_{n \in \mathbb{N}}$ définie par :

$$\forall n \in \mathbb{N} \quad \theta_n = P(T = n \mid T \geq n)$$

1. Calculer $P(T = n)$ en fonction des θ_k , $k \leq n$.
2. Etablir qu'une suite de réels θ_k est un taux de panne associé à une certaine variable aléatoire à valeurs dans \mathbb{N} **ssi** :

$$\forall k \in \mathbb{N} \quad 0 \leq \theta_k < 1 \quad \text{et} \quad \sum_{k=0}^{\infty} \theta_k = \infty$$

3. Montrer que T suit une loi géométrique **ssi** $(\theta_n)_{n \in \mathbb{N}}$ est constante.
Rappel : T suit une loi géométrique de paramètre $a \in]0, 1[$ si

$$\forall k \in \mathbb{N}, \quad P(T = k) = a(1-a)^k$$

Chapitre 5

Fonctions génératrices. Loi binomiale. Loi de Poisson.

5.1 Problème sur les fonctions génératrices :

On considère une variable aléatoire X à valeurs dans \mathbb{N} .

1. Montrer que pour tout $s \in [-1, 1]$, la variable aléatoire s^X est intégrable (on pose $s^0 = 1$ pour tout s) et que l'on a : $E[s^X] = \sum_{n=0}^{\infty} P(X = n) s^n$.

On notera $G_X(s) = E[s^X]$ et l'on appellera G_X la fonction génératrice de X .

2. Montrer que X et Y ont même loi si et seulement si $G_X = G_Y$.
3. Calculer les fonctions génératrices des lois suivantes :
 - (a) Bernoulli de paramètre $p \in [0, 1]$.
 - (b) Géométrique de paramètre $a \in]0, 1[$.
 - (c) Poisson de paramètre λ .
4. Si X et Y sont deux variables indépendantes, calculer G_{X+Y} en fonction de G_X et G_Y . En déduire :
 - (a) la fonction génératrice d'une loi binomiale de paramètres $(n, p) \in \mathbb{N}^* \times [0, 1]$.
 - (b) la loi de la somme de deux variables indépendantes suivant des lois de Poisson de paramètres respectifs λ et μ .
5. Montrer que si X est intégrable, alors G_X est dérivable sur $[-1, 1]$, de dérivée : $G'_X(s) = E[Xs^{X-1}]$.
En déduire les espérances des lois introduites dans la question 3).

6. On considère maintenant une suite de v.a. $(X_n)_{n \in \mathbb{N}}$ à valeurs dans \mathbb{N} , i.i.d., et une v.a. N à valeurs dans \mathbb{N} , indépendante de la suite $(X_n)_{n \in \mathbb{N}}$.
- (a) On définit la v.a. X_N par : $\forall \omega \in \Omega$, $X_N(\omega) = X_{N(\omega)}(\omega)$. Calculer G_{X_N} .
- (b) On définit la suite de variables aléatoires $(S_n)_{n \in \mathbb{N}}$ par : $S_0 = 0$, $S_n = X_1 + X_2 + \dots + X_n$. Calculer la fonction génératrice de S_N en fonction de G_{X_1} et G_N .

Corrigé :

1. Pour tout $s \in [-1, 1]$, $|s^X| \leq 1$ donc s^X est intégrable. En appliquant le théorème de convergence dominée de Lebesgue à la suite de fonctions $f_m : \omega \mapsto \sum_{n=0}^m \mathbf{1}_{\{X=n\}}(\omega) s^n$ qui converge simplement sur Ω vers $f : \omega \mapsto s^{X(\omega)}$ avec $\forall m \forall \omega$, $|f_m(\omega)| \leq 1$, on obtient :

$$E[s^X] = \lim_{m \rightarrow \infty} E[f_m] = \lim_{m \rightarrow \infty} \sum_{n=0}^m P(X = n) s^n = \sum_{n=0}^{\infty} P(X = n) s^n.$$

2. L'égalité du 1) nous donne le développement en série entière de G_X sur $[-1, 1]$; en particulier, G_X admet des dérivées à tous les ordres au point $s = 0$ et l'on a, pour tout $n \in \mathbb{N}$:

$$P(X = n) = \frac{G_X^{(n)}(0)}{n!}$$

Il en résulte que $G_X = G_Y \implies P_X = P_Y$; la réciproque est évidente. La fonction génératrice de X caractérise donc la loi de X .

3. (a) $G_X(s) = 1 - p + ps$.
 (b) $G_X(s) = \sum_{n=0}^{+\infty} (1-a)a^n s^n = \frac{1-a}{1-as}$.
 (c) $G_X(s) = \sum_{n=0}^{+\infty} e^{-\lambda} \frac{\lambda^n}{n!} s^n = e^{-\lambda(1-s)}$.
4. On a par définition : $G_{X+Y}(s) = E[s^{X+Y}] = E[s^X s^Y]$.
 Si X et Y sont indépendantes, on en déduit : $G_{X+Y}(s) = E[s^X] E[s^Y] = G_X(s) G_Y(s)$.

- (a) Une récurrence immédiate prouve que si X_1, \dots, X_n sont des v.a. à valeurs dans \mathbb{N} indépendantes, alors $G_{X_1 + \dots + X_n}(s) = G_{X_1}(s) \dots G_{X_n}(s)$. En particulier, si X_1, \dots, X_n suivent la loi de Bernouilli de paramètre p , alors $X_1 + \dots + X_n$ suit la loi binomiale de paramètres (n, p) , dont la fonction génératrice vaut donc (en utilisant 3)a) : $G(s) = (1 - p + ps)^n$.

- (b) Si $X \sim \text{Poisson}(\lambda)$ et $Y \sim \text{Poisson}(\mu)$ sont indépendantes, alors 3)b) nous permet d'écrire :

$$G_{X+Y}(s) = G_X(s)G_Y(s) = e^{-\lambda(1-s)}e^{-\mu(1-s)} = e^{-(\lambda+\mu)(1-s)}.$$

On reconnaît la fonction génératrice de la loi $\text{Poisson}(\lambda+\mu)$, ce qui d'après 2) prouve que $X+Y$ suit une loi de Poisson de paramètre $(\lambda+\mu)$.

5. On veut appliquer le théorème de dérivation sous l'intégrale (ici, l'espérance)

à $G_X : s \mapsto E[s^X]$.

La dérivée de $s \mapsto s^X$ vaut Xs^{X-1} qui admet un majorant intégrable indépendant

de s :

$$\forall s \in [-1, 1] \quad |Xs^{X-1}| \leq |X|.$$

On en conclut que G_X est dérivable sur $[-1, 1]$ (c'est-à-dire dérivable sur $] -1, 1[$, dérivable à droite en -1 et dérivable à gauche en 1) et que sa dérivée vaut $G'_X(s) = E[Xs^{X-1}]$.

En particulier, $E[X] = G'_X(1)$.

Si $X \sim \text{Bernouilli}(p)$, $G'_X(s) = p$ donc $E[X] = p$.

Si $X \sim \text{Geom}(a)$, $G'_X(s) = \frac{a(1-a)}{(1-as)^2}$ donc $E[X] = \frac{a}{1-a}$.

Enfin, pour $X \sim \text{Poisson}(\lambda)$, $G'_X(s) = \lambda e^{-\lambda(1-s)}$ d'où $E[X] = \lambda$.

6. (a) En utilisant la partition de Ω formée par les événements $\{N = n\}$, $n \in \mathbb{N}$, on peut écrire : $E[s^{X_N}] = E[\sum_{n=0}^{+\infty} \mathbf{1}_{\{N=n\}} s^{X_n}]$. Comme dans le 1), on en déduit :

$$E[s^{X_N}] = \sum_{n=0}^{+\infty} P(N = n)E[s^{X_n}] = \sum_{n=0}^{+\infty} P(N = n)G_{X_n}(s).$$

Comme les v.a. X_n ont toutes même loi, elles ont aussi même fonction génératrice, notée $G_X(s)$; on a donc : $G_{X_N}(s) = E[s^{X_N}] = G_X(s) \sum_{n=0}^{+\infty} P(N = n) = G_X(s)$.

Ceci prouve que X_N suit la loi commune aux X_n .

- (b) En procédant comme au a), on trouve : $E[s^{S_N}] = \sum_{n=0}^{+\infty} P(N = n)G_{S_n}(s)$.

Comme les $(X_n)_{n \in \mathbb{N}}$ sont indépendantes, le 4) nous donne : $G_{S_n}(s) = (G_X(s))^n$. Il s'en suit :

$$G_{S_N}(s) = \sum_{n=0}^{+\infty} P(N = n)(G_X(s))^n = G_N \circ G_X(s)$$

(bien défini car $|G_X(s)| \leq E[|s^X|] \leq 1$).

5.2 Loi binomiale. Loi de Poisson.

5.2.1 Stone-Weierstrass par les polynomes de Bernstein.

[FEL² 219-220]

1. On suppose donnée une famille de v.a.r. $(B_n^\theta, n \in \mathbb{N}^*, \theta \in [0, 1])$ telle que B_n^θ suit une loi binomiale de paramètres (n, θ) pour tous $n \in \mathbb{N}^*, \theta \in [0, 1]$.

Soit u une application continue de $[0, 1]$ dans \mathbb{R} .

Pour tout $n \in \mathbb{N}^*$, on définit l'application $u_n : [0, 1] \rightarrow \mathbb{R}$ par la formule :

$$u_n(\theta) = E[u(B_n^\theta/n)]$$

En utilisant Bienaymé-Chebyshev montrer que la suite d'applications $(u_n)_{n \in \mathbb{N}^*}$ converge uniformément vers u sur $[0, 1]$.

2. En déduire que toute application continue de $[0, 1]$ dans \mathbb{R} est limite uniforme d'une suite de polynomes que l'on explicitera.

5.2.2 *Théorème de Raikov.

[LUK 243ss]

Soient M et N deux v.a. à valeurs dans \mathbb{N} indépendantes et telles que $M + N$ suit une loi de Poisson de paramètre λ .

Le but de cet exercice est de montrer que M et N suivent chacune une loi de Poisson (constater la réciprocity avec ce qui a été démontré dans le Problème 4)b)).

1. Montrer que pour tout $k \in \mathbb{N}$, on a : $P[M = k] \leq \frac{\lambda^k}{k!}$ et en déduire que l'on définit une fonction entière (= holomorphe sur \mathbb{C}) par la formule : $\forall z \in \mathbb{C}, g_M(z) = E[z^M]$. On définit g_N de façon similaire.
2. Montrer que g_M et g_N ne s'annulent jamais. Puisque \mathbb{C} est simplement connexe, on peut donc définir des fonctions entières γ_M et γ_N telles que $g_M = \exp(\gamma_M)$ et $g_N = \exp(\gamma_N)$.
3. Montrer que pour tout $|z| \geq 1$, on a $|g_M(z)| \leq e^{\lambda(|z|-1)}$.
En déduire que pour tout $z \in \mathbb{C}$, on a : $Re\gamma_M(z) \leq \lambda(|z| - 1)_+$.
4. Conclure en utilisant le théorème de Liouville cité dans la section suivante.

Corrigé :

1. Pour tout $k \in \mathbb{N}$, on a :

$$P[M = k] \leq P[M + N \geq k] = e^{-\lambda} \sum_{i \geq k} \frac{\lambda^i}{i!} = \frac{\lambda^k}{k!} e^{-\lambda} \sum_{j \geq 0} \frac{\lambda^j k!}{(k+j)!} \leq \frac{\lambda^k}{k!}.$$

On en déduit que pour tout $r \geq 0$, $\sum_{k \geq 0} P[M = k] r^k \leq \sum_{k \geq 0} \frac{(\lambda r)^k}{k!} < +\infty$ et donc la série entière $\sum_{k \geq 0} P[M = k] z^k$ admet un rayon de convergence infini ; on a ainsi prouvé que g_M est une fonction entière.

2. Par indépendance de M et N , on a, pour tout $z \in \mathbb{C}$:

$$g_M(z)g_N(z) = E[z^{M+N}] = e^{\lambda(z-1)} \neq 0,$$

ce qui prouve que g_M et g_N ne s'annulent jamais.

Puisque \mathbb{C} est clairement simplement connexe, la théorie des fonctions holomorphes nous donne l'existence de γ_M et γ_N (voir [RUD 252-253]).

3. Pour tout $z \in \mathbb{C}$ tel que $|z| \geq 1$, on a :

$$|E[z^M]| \leq E[|z|^M] \leq E[|z|^{M+N}] = e^{\lambda(|z|-1)}.$$

D'autre part, si $|z| \leq 1$, il est immédiat que $|E[z^M]| \leq 1$.

En distinguant ces deux cas, on obtient l'inégalité demandée pour $Re\gamma_M(z) = \log |g_M(z)|$.

4. D'après ce qu'on vient de montrer, on a :

$$\sup_{z \in \mathbb{C}} \frac{Re\gamma_M(z)}{1 + |z|} \leq \lambda < \infty.$$

Le théorème de Liouville nous dit alors que γ_M n'est autre qu'une fonction affine.

Il existe donc deux constantes $\alpha, \beta \in \mathbb{C}$ telles que, pour tout $z \in \mathbb{C}$:

$$g_M(z) = \exp(\alpha(z-1) + \beta).$$

Or $g_M(1) = 1$ d'où $\beta = 0$ et $g'_M(1) = E[M] = \alpha$.

En conclusion, on a prouvé que :

$$g_M(z) = e^{\alpha(z-1)};$$

on reconnaît la fonction génératrice caractérisant la loi de Poisson de paramètre $\alpha = E[M]$.

5.2.3 Approximation d'une binomiale par Poisson.

1. Soit $\lambda > 0$ et une suite $(p_n)_{n \in \mathbb{N}^*}$ telle que $p_n \sim \lambda/n$. Montrer que pour tout $k \in \mathbb{N}$, on a :

$$\lim_{n \rightarrow \infty} C_n^k p_n^k (1 - p_n)^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!}.$$

2. Une usine fabrique des vis avec une proportion de 0,5% de pièces défectueuses.

J'ai acheté un lot de 1000 vis ; quelle est la loi approximative du nombre de vis ayant un défaut de fabrication ?

5.3 Théorème de Liouville.

Soit $n_0 \in \mathbb{N}$ et F une fonction entière telle que :

$$\sup_{z \in \mathbb{C}} \frac{\operatorname{Re} F(z)}{1 + |z|^{n_0}} < +\infty.$$

Alors F est un polynôme de degré n_0 au plus.

Remarque : Ceci est une version un peu modifiée du théorème classique de Liouville, qui s'écrit avec $|F(z)|$ à la place de $\operatorname{Re} F(z)$.

Démonstration : On a : $F(z) = \sum_n a_n z^n$, série entière de rayon de convergence infini.

Notons $a_n = u_n + iv_n$ ($u_n \in \mathbb{R}, v_n \in \mathbb{R}$) et $z = re^{it}$ de sorte que :

$$\operatorname{Re} F(z) = \sum_n (u_n r^n \cos nt - v_n r^n \sin nt).$$

Par convergence uniforme de la série $\sum_n a_n z^n$ sur le cercle de centre 0 et de rayon r , il vient d'une part : $2\pi u_0 = \int_0^{2\pi} \operatorname{Re} F(re^{it}) dt$,

d'autre part et successivement, pour tout $n \geq 1$, $\pi u_n r^n = \int_0^{2\pi} \operatorname{Re} F(re^{it}) \cos ntdt$,
 $-\pi v_n r^n = \int_0^{2\pi} \operatorname{Re} F(re^{it}) \sin ntdt$, $\pi a_n r^n = \int_0^{2\pi} \operatorname{Re} F(re^{it}) e^{-int} dt$, $\pi |a_n| r^n \leq \int_0^{2\pi} |\operatorname{Re} F(re^{it})| dt$. On en déduit les inégalités :

$$|a_n| r^n + 2\operatorname{Re} F(0) \leq \frac{1}{\pi} \int_0^{2\pi} (|\operatorname{Re} F(re^{it})| + \operatorname{Re} F(re^{it})) dt \leq 4 \max(0, \sup_{|z|=r} \operatorname{Re} F(z)).$$

Ainsi, posant $M = \sup_{z \in \mathbb{C}} \frac{\operatorname{Re} F(z)}{1+|z|_0^n}$, on obtient :

$$\frac{|a_n|r^n + 2\operatorname{Re} F(0)}{1+r^{n_0}} \leq 4M_+.$$

On en conclut immédiatement : $|a_n| = 0$ pour tout $n > n_0$.

Chapitre 6

Théorème des classes monotones.

Définitions préliminaires : Soit Ω un ensemble.

On appelle classe monotone une classe $\mathcal{M} \subset \mathcal{P}(\Omega)$ telle que :

- $\Omega \in \mathcal{M}$
- $A \in \mathcal{M}, B \in \mathcal{M}, A \subset B \Rightarrow B - A \in \mathcal{M}$
- $(A_n) \in \mathcal{M}^{\mathbb{N}}, A_n \nearrow A \Rightarrow A \in \mathcal{M}$.

Remarque : La notation $A_n \nearrow A$ signifie que la suite (A_n) est croissante au sens de l'inclusion et que $A = \cup_{n \in \mathbb{N}} A_n$.

On appelle π -système une classe $\mathcal{C} \subset \mathcal{P}(\Omega)$ stable par intersection finie :
 $A \in \mathcal{C}, B \in \mathcal{C} \Rightarrow A \cap B \in \mathcal{C}$.

6.1 Démonstration du théorème par étapes.

6.1.1 Ce qui manque à une classe monotone pour être une tribu.

Montrer qu'on a l'équivalence suivante pour toute classe $\mathcal{C} \subset \mathcal{P}(\Omega)$:
 \mathcal{C} tribu $\Leftrightarrow \mathcal{C}$ classe monotone et π -système.

6.1.2 Classe monotone engendrée.

Soit $\mathcal{C} \subset \mathcal{P}(\Omega)$ une classe quelconque de parties de Ω .

Montrer qu'il existe une plus petite classe monotone contenant \mathcal{C} .

On la notera $\mathcal{M}(\mathcal{C})$.

6.1.3 Théorème des classes monotones.

[REV int 48-49]

Le but de cet exercice est de démontrer le théorème suivant :

Si \mathcal{C} est un π -système, alors $\mathcal{M}(\mathcal{C}) = \sigma(\mathcal{C})$.

1. Montrer l'implication : $\mathcal{M}(\mathcal{C}) \pi\text{-système} \Rightarrow \mathcal{M}(\mathcal{C}) = \sigma(\mathcal{C})$.
Il nous reste donc à montrer que $\mathcal{M}(\mathcal{C})$ est un π -système pour pouvoir conclure, ce qui est l'objet des deux questions suivantes.
2. Soit $A \in \mathcal{C}$; on considère l'ensemble $\mathcal{D}_A = \{B \in \mathcal{M}(\mathcal{C}), B \cap A \in \mathcal{M}(\mathcal{C})\}$.
Montrer que $\mathcal{D}_A = \mathcal{M}(\mathcal{C})$.
3. Soit $B \in \mathcal{M}(\mathcal{C})$; définir \mathcal{D}_B suivant 2) et conclure.

6.2 Applications.

6.2.1 Egalité de lois.

1. Soient P_1 et P_2 deux mesures de probabilité sur l'espace mesurable (Ω, \mathcal{A}) .
Montrer que l'ensemble $\mathcal{M} = \{A \in \mathcal{A}, P_1(A) = P_2(A)\}$ est une classe monotone.
2. En déduire que si P_1 et P_2 coïncident sur un π -système qui engendre \mathcal{A} , alors $P_1 = P_2$.
3. Soient X et Y deux variables aléatoires réelles telles que

$$\forall x \in \mathbb{Q}, \quad P(X \leq x) = P(Y \leq x).$$

Montrer que X et Y ont même loi.

6.2.2 Critère d'indépendance.

Soit (Ω, \mathcal{A}, P) un espace de probabilité et $\mathcal{C}_1, \dots, \mathcal{C}_n$ des π -systèmes inclus dans \mathcal{A} et contenant tous l'élément Ω .

Montrer que les sous-tribus $\sigma(\mathcal{C}_1), \dots, \sigma(\mathcal{C}_n)$ sont indépendantes **ssi** :

$$\forall i \in \{1, \dots, n\}, \quad \forall C_i \in \mathcal{C}_i, \quad P\left(\bigcap_{i=1}^n C_i\right) = \prod_{i=1}^n P(C_i).$$

Indication : On pourra introduire l'ensemble :

$$\mathcal{D}_1 = \left\{ B \in \mathcal{A}, \quad P\left[B \cap \left(\bigcap_{i=2}^n C_i\right)\right] = P[B] \prod_{i=2}^n P(C_i) \right\}$$

puis des ensembles $\mathcal{D}_2, \mathcal{D}_3 \dots$ judicieusement choisis.

Question subsidiaire :

Dans le cas $n = 2$, on peut ôter la condition "et contenant tous l'élément Ω ". Pourquoi?

Chapitre 7

Outils pour le calcul d'une loi (Cours)

Soit X un vecteur aléatoire d -dimensionnel de loi P_X supposée connue et $f : \mathbb{R}^d \rightarrow \mathbb{R}^{d'}$ une application borélienne. Nous définissons le vecteur aléatoire d' -dimensionnel $Y := f(X)$. L'objet de ce chapitre est de mettre en évidence plusieurs méthodes qui nous permettront de déterminer la loi de Y .

Lemme 7.0.1 *La loi du vecteur aléatoire $Y = f(X)$ est la mesure image de la loi P_X par l'application borélienne f .*

Démonstration: Pour tout $B \in \mathcal{B}(\mathbb{R}^{d'})$, nous avons par définition même :

$$P_Y(B) = P(Y^{-1}(B)) = P((f \circ X)^{-1}(B))$$

Or il est immédiat de vérifier que $(f \circ X)^{-1}(B) = X^{-1}(f^{-1}(B))$, si bien que :

$$P_Y(B) = P_X(f^{-1}(B)) = (P_X)_f(B)$$

□

7.1 Fonctions de répartition.

Définition et proposition 7.1.1 *Soit μ une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On appelle fonction de répartition associée à μ l'application $F_\mu : \mathbb{R} \rightarrow [0, 1]$ définie par :*

$$\forall x \in \mathbb{R} \quad F_\mu(x) = \mu(] - \infty, x])$$

L'application F_μ est croissante, continue à droite, telle que $F(-\infty) = 0$ et $F(+\infty) = 1$.

Réciproquement, la théorie de la mesure nous fournit le résultat suivant.

Théorème 7.1.2 *Soit $F : \mathbb{R} \rightarrow \mathbb{R}$ une application croissante, continue à droite, telle que $F(-\infty) = 0$ et $F(+\infty) = 1$. Il existe une unique mesure de probabilité μ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que :*

$$\forall x \in \mathbb{R} \quad \mu(] - \infty, x]) = F(x)$$

Rappelons que le résultat d'existence se démontre grâce au théorème de prolongement de Carathéodory tandis que celui d'unicité est une conséquence du théorème des classes monotones. □

Une fonction de répartition F_μ étant croissante, elle admet en tout point $x \in \mathbb{R}$ une limite à gauche :

$$F_\mu(x-) := \lim_{\substack{y \rightarrow x \\ y < x}} F_\mu(y) = \sup_{y < x} F_\mu(y)$$

Pour x fixé, choisissons une suite strictement croissante (y_n) qui tend vers x . Nous avons alors, en utilisant la continuité à gauche d'une mesure,

$$F_\mu(x-) := \lim F_\mu(y_n) = \lim \uparrow \mu(] - \infty, y_n]) = \mu(\lim \uparrow] - \infty, y_n])$$

si bien que finalement :

$$F_\mu(x-) = \mu(] - \infty, x[) \tag{7.1}$$

$$F_\mu(x) - F_\mu(x-) = \mu(\{x\}) \tag{7.2}$$

Un réel $x \in \mathbb{R}$ tel que $\mu(\{x\}) > 0$ est appelé *masse ponctuelle* de la mesure μ . Si μ est une mesure finie, l'ensemble D de ses masses ponctuelles est dénombrable au plus. En effet, en notant pour tout $n \in \mathbb{N}^*$:

$$D_n := \left\{ x \in \mathbb{R}, \quad \mu(\{x\}) \geq \frac{1}{n} \right\},$$

nous avons

$$D = \bigcup_{n \in \mathbb{N}^*} D_n \text{ et } \#D_n \leq n\mu(E) < +\infty$$

Définition et proposition 7.1.3 *Une mesure de probabilité μ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est dite diffuse si elle n'admet aucune masse ponctuelle. Ceci équivaut à la continuité de sa fonction de répartition F_μ .*

Démonstration: Puisqu'une fonction de répartition est toujours continue à droite, l'équivalence résulte immédiatement de l'égalité (7.2). \square

Exemple: L'escalier de Lebesgue [BP 288-291] est la fonction de répartition d'une mesure de probabilité μ diffuse. En outre, $\mu(C) = 0$, où l'on a noté C l'ensemble de Cantor, qui est λ -négligeable. Une telle mesure, à la fois diffuse et étrangère à la mesure de Lebesgue, est dite *singulière*. \square

Un type de mesure diffuse que nous rencontrerons beaucoup plus couramment dans le cadre de notre programme est constitué par les mesures de la forme $\mu = p \cdot \lambda$, où p est une densité de probabilité, autrement dit les mesures de probabilité *absolument continues* par rapport à la mesure de Lebesgue.

Proposition 7.1.4 *Soit $p : \mathbb{R} \rightarrow \mathbb{R}_+$ une application borélienne positive et μ une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Il y a équivalence entre les deux propositions suivantes :*

1. *L'application p est une densité de probabilité et $\mu = p \cdot \lambda$.*
2. *La fonction de répartition de la mesure μ s'écrit :*

$$\forall x \in \mathbb{R} \quad F_\mu(x) = \int_{-\infty}^x p(t) dt \quad (7.3)$$

Démonstration: L'implication (1. \Rightarrow 2.) est immédiate.

Pour l'implication réciproque, on déduit d'abord de $F_\mu(+\infty) = 1$ que

$$\int_{-\infty}^{+\infty} p(t) dt = 1,$$

ce qui nous permet d'affirmer que p est bien une densité de probabilité. D'après l'implication (1. \Rightarrow 2.), le membre de droite de (7.3) est la fonction de répartition de la mesure $p \cdot \lambda$, ce qui nous permet de réécrire notre hypothèse sous la forme $F_\mu = F_{p \cdot \lambda}$. D'après le théorème 7.1.2 (partie unicité) ceci implique $\mu = p \cdot \lambda$. \square

Remarque: Un théorème (difficile) de dérivation dû à Lebesgue affirme :

$$(7.3) \Rightarrow F'_\mu(x) = p(x) \lambda(dx)\text{-presque partout.}$$

Pour prouver qu'une mesure μ admet une certaine densité, il est néanmoins insuffisant de dériver sa fonction de répartition. Par exemple, l'escalier de Lebesgue nous mènerait par cette méthode à une «densité» nulle!

Chapitre 8

Outils pour le calcul d'une loi.

8.1 Fonctions de répartition.

8.1.1 Espérance et f.r.

[REV 39] (donne des indications)

1. Soit X une v.a.r. positive p.s. Montrer les égalités suivantes :

$$E[X] = \int_0^{+\infty} (1 - F_X(u)) \, du = \int_0^{+\infty} P(X > u) \, du = \int_0^{+\infty} P(X \geq u) \, du.$$

2. Si X n'est plus supposée positive p.s. , montrer que son espérance existe **ssi**

$$\int_0^{+\infty} (1 - F_X(u) + F_X(-u)) \, du < \infty$$

et qu'alors

$$E[X] = \int_0^{+\infty} (1 - F_X(u) - F_X(-u)) \, du.$$

8.1.2 Extrema de variables exponentielles.

1. Soient X_1, \dots, X_n des v.a.r. suivant des lois exponentielles de paramètres respectifs $\lambda_1, \dots, \lambda_n$. On suppose que ces v.a.r. sont indépendantes , ce qui équivaut à :

$\forall (x_i)_{1 \leq i \leq n} \in \mathbb{R}^n$, les évènements $\{X_i \leq x_i\}, 1 \leq i \leq n$, sont mutuellement indépendants.

Calculer la loi de $X = \min_{1 \leq i \leq n} X_i$.

2. On considère maintenant une suite $(Y_n)_{n \in \mathbb{N}^*}$ de variables indépendantes et de même loi exponentielle de paramètre λ . Pour tout $n \in \mathbb{N}^*$, on définit $M_n = \max_{1 \leq i \leq n} Y_i$.

Pour $\epsilon \in]0, 1[$ arbitraire, calculer les quantités suivantes :

$$P(Y_n > (1 + \epsilon)\lambda^{-1} \log n) \quad , \quad P(M_n \leq (1 - \epsilon)\lambda^{-1} \log n) .$$

En appliquant le lemme de Borel-Cantelli, en déduire que $P(d\omega)$ -p.s. on a la convergence suivante lorsque n tend vers $+\infty$:

$$\frac{M_n(\omega)}{\log n} \rightarrow \frac{1}{\lambda}$$

8.1.3 Transformation par quantile.

[DAC 37-38]

Soit X une v.a.r. de fonction de répartition F .

1. On définit la "pseudo-inverse" de F par :

$$\forall t \in [0, 1], \quad F^{-1}(t) = \inf\{x \in \mathbb{R}, F(x) \geq t\} \in \bar{\mathbb{R}}.$$

Montrer que $F(F^{-1}(t)) \geq t$, avec égalité lorsque F est continue.

Montrer que $F^{-1}(F(x)) \leq x$, avec égalité lorsque F est strictement croissante.

2. Montrer que $F^{-1}(F(X)) = X$ p.s.
3. On suppose F continue. Montrer que la v.a. $F(X)$ suit la loi uniforme sur $[0, 1]$.
4. On ne suppose plus F continue. On se donne une v.a.r. $U \sim Unif([0, 1])$. Montrer que $F^{-1}(U)$ suit la loi de f.r. F .

8.1.4 Variables annésiques.

[COT 91-92]

Soit T une v.a. à valeurs réelles telle que, pour tous $s, t \geq 0$, on ait :

$$P(T > t + s) = P(T > t)P(T > s).$$

Le but de cet exercice est de montrer que, soit $P(T > 0) = 0$, soit T suit une loi exponentielle.

1. Montrer que si $P(T > 0) > 0$, alors pour tout $t > 0$, $P(T > t) > 0$.

2. On définit alors l'application $f(t) = \log P(T > t)$, $t > 0$.
Montrer que $f(x) = xf(1)$ pour tout x rationnel positif, puis pour tout x réel positif.
3. Conclure.

8.2 Lois marginales.

8.2.1 Calcul avec une densité particulière.

Soit (X, Y) un couple aléatoire à valeurs dans \mathbb{R}^2 , admettant pour densité :

$$p(x, y) = \frac{1}{\pi} \frac{ye^{-y}}{x^2 + y^2} \mathbf{1}_{\mathbb{R}_+^*}(y).$$

Vérifier que p est bien une densité et calculer la loi marginale de Y .

8.2.2 Densité gaussienne en dimension 2.

Soit (X, Y) un vecteur aléatoire à valeurs dans \mathbb{R}^2 de densité :

$$p(x, y) = \frac{1}{2\pi\sigma\tau\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma^2} - 2\rho\frac{xy}{\sigma\tau} + \frac{y^2}{\tau^2}\right)\right\},$$

où σ et τ sont deux réels positifs strictement et $|\rho| < 1$.

Montrer que $X \sim \mathcal{N}(0, \sigma^2)$ et $Y \sim \mathcal{N}(0, \tau^2)$.

8.3 Changement de variables.

8.3.1 Loi du χ^2 à un degré de liberté.

[COT 83 et 86]

Soit $X \sim \mathcal{N}(0, 1)$. Quelle est la loi de $Y = X^2$?

8.3.2 Loi gaussienne dans \mathbb{R} .

Soit $X \sim \mathcal{N}(0, 1)$. Quelle est la loi de $Y = m + \sigma X$, où $m \in \mathbb{R}$ et $\sigma > 0$?

8.3.3 Avec une loi de Cauchy.

[COT 52-53]

Soit X une v.a.r. suivant une loi de Cauchy de paramètre 1.

1. Quelle est la loi de aX , où $a \in \mathbb{R}^*$?
2. Montrer que $Y = \log |X|$ admet la densité : $p(y) = \frac{1}{\pi \cosh y}$.

8.3.4 Simulation d'une loi gaussienne dans \mathbb{R}^2 par la méthode de Box-Müller.

[OUV² 67-68]

Soit $U = (U_1, U_2)$ un vecteur aléatoire de loi uniforme sur $[0, 1]^2$. Calculer la loi du vecteur aléatoire défini comme suit :

$$(X, Y) = (\sqrt{-2 \log U_1} \cos(2\pi U_2), \sqrt{-2 \log U_1} \sin(2\pi U_2)).$$

Chapitre 9

Variables aléatoires indépendantes.

9.1 Cas où il existe une densité (changement de variable).

9.1.1 Lois Gamma et Bêta

[COT 83-87]

1. (a) Montrer que pour tout $a > 0$, on a : $0 < \int_0^{+\infty} e^{-x} x^{a-1} dx < +\infty$.
On notera cette intégrale $\Gamma(a)$.
- (b) Montrer que pour tout $a > 0$, on a $\Gamma(a+1) = a\Gamma(a)$ et en déduire la valeur de $\Gamma(n)$ pour $n \in \mathbb{N}^*$.
- (c) Montrer que pour tous $(\lambda, a) \in (\mathbb{R}_+^*)^2$, l'application définie comme suit :

$$g_{\lambda,a}(x) = \frac{\lambda^a}{\Gamma(a)} e^{-\lambda x} x^{a-1} \mathbf{1}_{\mathbb{R}_+^*}(x)$$

est une densité de probabilité.

On appelle loi Gamma de paramètres λ et a et l'on note $\gamma(\lambda, a)$ la loi de densité $g_{\lambda,a}$.

Reconnaître la loi $\gamma(\lambda, 1)$.

- (d) Constater que la loi du χ^2 à un degré de liberté est une loi γ particulière et en déduire que $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.
2. (a) Montrer que pour tous $(p, q) \in (\mathbb{R}_+^*)^2$, on a :

$$0 < \int_0^1 x^{p-1} (1-x)^{q-1} dx < +\infty.$$

On notera cette intégrale $\beta(p, q)$.

- (b) Pour $q > 1$, établir une relation entre $\beta(p, q)$ et $\beta(p + 1, q - 1)$; en déduire la valeur de $\beta(p, q)$ pour $p \in \mathbb{N}^*, q \in \mathbb{N}^*$.

On appelle loi Bêta(p, q) de première espèce la loi de densité : $\frac{1}{\beta(p, q)} x^{p-1} (1-x)^{q-1} \mathbf{1}_{]0,1[}(x)$.

3. On considère deux v.a.r. indépendantes X et Y de lois respectives $\gamma(\lambda, a)$ et $\gamma(\lambda, b)$. Montrer que $S = X + Y$ et $T = \frac{X}{X+Y}$ sont indépendantes et préciser leurs lois.

Montrer qu'on a la relation :

$$\beta(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

4. Soit $X = (X_1, X_2, \dots, X_n)$ un vecteur aléatoire dont les composantes sont i.i.d. de loi $\mathcal{N}(0, 1)$. Quelle est la loi de $\|X\|^2$?

On l'appelle loi du χ^2 à n degrés de liberté.

9.1.2 Gauss et Cauchy.

[FOA 201]

1. On considère deux v.a.r. indépendantes S et T de lois respectives $\mathcal{N}(0, \sigma^2)$ et $\mathcal{N}(0, \tau^2)$.

Montrer que $\frac{S}{T}$ suit une loi de Cauchy de paramètre $\frac{\sigma}{\tau}$.

2. En déduire que si Z est une v.a.r. de Cauchy, alors $\frac{1}{Z}$ aussi.

9.1.3 Avec des lois exponentielles.

Soient $X \sim \text{Exp}(\lambda)$ et $Y \sim \text{Exp}(\mu)$ deux v.a.r. indépendantes ($\lambda > 0, \mu > 0$).

On pose $U = \inf(X, Y)$ et $V = |X - Y|$. Montrer que U et V sont indépendantes.

Remarque : Vous pourrez trouver dans [OUV² 75-79] une caractérisation de la loi exponentielle par une propriété de ce type, s'énonçant comme suit :

Si X et Y sont deux v.a.r. indépendantes de même loi ν , de densité f telle que $f(x) > 0$ ssi $x \in \mathbb{R}_+^*$ et si l'on définit $U = \inf(X, Y)$ et $V = |X - Y|$, on a l'équivalence suivante :

U et V sont indépendants ssi ν est une loi exponentielle.

9.1.4 Densité d'un couple non indépendant.

[COT 49-51]

Soient X et Y deux v.a.r. indépendantes de même densité $d(x) = \frac{1}{x^2} \mathbf{1}_{[1,+\infty[}(x)$, dite densité de Pareto de paramètre $(1,1)$.

On pose $(U, V) = (XY, \frac{X}{Y})$.

1. Calculer la loi du couple aléatoire (U, V) et montrer, en considérant le support de cette loi, que les v.a.r. U et V ne sont pas indépendantes.
2. Calculer les lois marginales.

9.1.5 Produit de variables de Cauchy.

[COT 52]

Soient X et Y deux v.a.r. i.i.d. de loi $C(1)$. Montrer que XY admet pour densité :

$$\frac{2}{\pi^2} \frac{\log|x|}{x^2 - 1}.$$

9.2 Cas général.

9.2.1 Diagonale négligeable.

[COT 53]

Soient X et Y deux v.a. à valeurs dans \mathbb{R}^2 et indépendantes.

Montrer que $P(X = Y) = 0$ dès que P_X ou P_Y est diffuse.

9.2.2 V.a.r. auto-indépendante.

Montrer qu'une v.a.r. X est indépendante d'elle-même si et seulement si elle est constante presque sûrement.

Indication : On pourra étudier l'allure de la fonction de répartition F_X en appliquant le résultat de l'exercice 3.2.1.

9.2.3 Statistique d'ordre.

[COT 54-55]

Soient X_i , $1 \leq i \leq n$, des v.a.r. i.i.d. de loi μ supposée diffuse.

Nous noterons Σ_n le groupe des permutations de $\{1, 2, \dots, n\}$.

1. Montrer que $P(d\omega)$ —presque sûrement, il existe une unique permutation aléatoire $\sigma(\omega) \in \Sigma_n$ telle que :

$$X_{\sigma(\omega)(1)}(\omega) < X_{\sigma(\omega)(2)}(\omega) < \dots < X_{\sigma(\omega)(n)}(\omega).$$

On notera en abrégé : *p.s.*, $X_{\sigma(1)} < X_{\sigma(2)} < \cdots < X_{\sigma(n)}$.

2. Montrer que σ est une variable aléatoire à valeurs dans le groupe Σ_n , de loi uniforme sur Σ_n . (On supposera que tout ensemble P -négligeable est élément de la tribu \mathcal{A} .)
3. On définit le vecteur aléatoire $Y = (X_{\sigma(1)}, \cdots, X_{\sigma(n)})$.

Montrer que les v.a. σ et Y sont indépendantes et que la loi de Y vaut $n! \mathbf{1}_{D_n} \mu^{\otimes n}$, où $D_n = \{(x_1, \cdots, x_n) \in \mathbb{R}^n, x_1 < \cdots < x_n\}$.

9.2.4 Processus de Poisson.

[COT 97] puis [REV 267-268]

1. Soient $(X_n)_{n \in \mathbb{N}^*}$ une suite de v.a.r. indépendantes de même loi exponentielle de paramètre λ .

On définit une nouvelle suite de v.a.r. par la formule : $\forall n \in \mathbb{N}^*$, $S_n = \sum_{i=1}^n X_i$. Montrer que, pour tout $n \in \mathbb{N}^*$, le vecteur aléatoire (S_1, \cdots, S_n) admet la densité suivante :

$$\lambda^n e^{-\lambda s_n} \mathbf{1}_{0 < s_1 < \cdots < s_n}$$

2. On définit une famille $(N_t)_{t \in \mathbb{R}_+}$ de v.a.r. par la formule suivante :

$$\forall t \geq 0, N_t = \sum_{n=1}^{\infty} \mathbf{1}_{S_n \leq t}.$$

Soient $0 < t_1 < \cdots < t_k$.

Montrer que les v.a.r. $N_{t_1}, N_{t_2} - N_{t_1}, \cdots, N_{t_k} - N_{t_{k-1}}$ sont indépendantes et suivent des lois de Poisson de paramètres respectifs $\lambda t_1, \lambda(t_2 - t_1), \cdots, \lambda(t_k - t_{k-1})$.

La famille $(N_t)_{t \in \mathbb{R}_+}$ est appelée processus de Poisson.

Chapitre 10

Fonctions caractéristiques (cours).

Rappelons que la loi d'une v.a. X à valeurs dans \mathbb{R}^d est déterminée de façon unique par les quantités $E[f(X)]$, $f : \mathbb{R}^d \rightarrow \mathbb{R}$ borélienne bornée (prendre $f = \mathbf{1}_A$, $A \in \mathcal{B}(\mathbb{R}^d)$ pour le constater).

Un résultat essentiel de ce chapitre est qu'il suffit de connaître ces quantités pour $f_t(x) = \exp(it \cdot x)$, où t est un paramètre dans \mathbb{R}^d , pour que la loi soit déterminée.

Cette transformation injective d'une loi en une application $\phi_X : \mathbb{R}^d \rightarrow \mathbb{C}$, appelée *fonction caractéristique* de X , n'est autre que la transformation de Fourier (appliquée à la loi de X).

Nous utiliserons donc ici les résultats établis dans le chapitre 10 du poly d'intégration, auquel le lecteur est invité à se reporter.

Une application importante des fonctions caractéristiques est la démonstration du *théorème-limite central*, que nous verrons plus tard.

10.1 Fonction caractéristique d'une variable aléatoire.

Nous traduisons en langage probabiliste les notions relatives à la transformation de Fourier.

Définition 10.1.1 Soit X une v.a. à valeurs dans \mathbb{R}^d . On appelle *fonction caractéristique (f.c.)* de X l'application $\phi_X : \mathbb{R}^d \rightarrow \mathbb{C}$ définie par :

$$\forall u \in \mathbb{R}^d, \quad \phi_X(u) = E[e^{iu \cdot X}]$$

D'après le théorème de transfert, ceci s'écrit encore :

$$\forall u \in \mathbb{R}^d, \quad \phi_X(u) = \int_{\mathbb{R}^d} e^{iu \cdot x} dP_X(x)$$

Autrement dit, la f.c. d'une variable aléatoire n'est autre que la transformée de Fourier de sa loi. En particulier, elle est toujours bornée par 1 (qui est ici la masse totale), uniformément continue et telle que $\phi_X(0) = 1$.

Le lemme de Riemann-Lebesgue se traduit comme suit : si X admet une densité, alors $\phi_X \in C_0(\mathbb{R}^d)$.

Le lecteur trouvera dans l'exercice 11.2.1 une liste de lois "classiques" dont il est élémentaire de calculer la transformée de Fourier.

Le calcul peut être nettement plus difficile et demander par exemple d'appliquer la méthode des résidus. Nous allons voir maintenant un exemple important qui fait appel à la théorie des fonctions holomorphes mais qui ne nécessite pas un calcul de résidus.

Proposition 10.1.2 *Soit $X \sim \mathcal{N}(m, \sigma^2)$; alors la f.c. de X est donnée par la formule :*

$$\forall u \in \mathbb{R}, \quad \phi_X(u) = \exp\left(ium - \frac{u^2\sigma^2}{2}\right)$$

Démonstration: Il suffit de démontrer que dans le cas $m = 0, \sigma^2 = 1$, c'est-à-dire $X \sim \mathcal{N}(0, 1)$, on a $\phi_X(u) = \exp(-u^2/2)$. En effet, si ce résultat est acquis, on a $Y = m + \sigma X \sim \mathcal{N}(m, \sigma^2)$ et

$$\phi_Y(u) = E[\exp(iu(m + \sigma X))] = \exp(ium)\phi_X(\sigma u),$$

ce qui nous donne le résultat général.

Revenons donc au cas centré réduit et écrivons la densité de X :

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Nous avons démontré dans le cours d'intégration :

$$\forall z \in \mathbb{C}, \quad \int_{\mathbb{R}} e^{zx} g(x) dx = e^{\frac{z^2}{2}}$$

Il suffit de prendre $z = iu$ pour obtenir la valeur de $\phi_X(u)$.

Remarque: Nous obtenons comme sous-produit de cette démonstration la valeur des moments de la loi $\mathcal{N}(0, 1)$ en développant $\exp(z^2/2)$ en série entière et en identifiant les coefficients :

$$\sum_{n=0}^{+\infty} \frac{z^n}{n!} \int_{\mathbb{R}} x^n g(x) dx = \sum_{p=0}^{+\infty} \frac{z^{2p}}{2^p p!}$$

d'où :

$$\forall p \in \mathbb{N}, \quad E[X^{2p+1}] = 0 \text{ et } E[X^{2p}] = \frac{(2p)!}{2^p p!}$$

□

10.2 Produit de convolution et transformée de Fourier.

10.2.1 Définition.

Dans la pratique, on s'intéresse souvent à la loi d'une somme de variables aléatoires indépendantes. Par exemple, si la durée de vie d'un certain matériel suit une loi μ , l'instant où l'on devra effectuer le n^e remplacement s'écrit $X_1 + \dots + X_n$, les variables X_i étant indépendantes et de même loi μ . Notons au passage que nous avons déjà résolu ce genre de problème lorsque $\mu = B(p)$: nous obtenons la loi $B(n, p)$.

Les variables aléatoires indépendantes que l'on additionne ne sont pas forcément de même loi : on peut penser par exemple à la durée de montage d'un matériel dont la fabrication requiert plusieurs étapes indépendantes.

Connaissant les lois de plusieurs variables aléatoires indépendantes à valeurs dans \mathbb{R}^d , nous pouvons calculer - au moins théoriquement ! - la loi de leur somme grâce à la notion suivante.

Définition 10.2.1 Soient μ et ν des lois de probabilités sur \mathbb{R}^d ; on appelle produit de convolution de μ et ν et l'on note $\mu \star \nu$ la loi image de $\mu \otimes \nu$ par l'application borélienne $s : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ définie par $s(x, y) = x + y$.

$$\mu \star \nu = (\mu \otimes \nu)_s$$

Ainsi, $\mu \star \nu$ est une loi de probabilité sur \mathbb{R}^d .

Proposition 10.2.2 Si X et Y sont deux variables aléatoires d -dimensionnelles indépendantes, alors la loi de $X + Y$ est égale au produit de convolution des lois de X et Y :

$$P_{X+Y} = P_X \star P_Y$$

Démonstration: Cette proposition résulte immédiatement de la définition précédente puisque $X + Y = s(X, Y)$ et $(X, Y) \sim P_X \otimes P_Y$ d'où

$$P_{X+Y} = (P_X \otimes P_Y)_s = P_X \star P_Y$$

□

10.2.2 F.c. d'une somme de v.a. indépendantes

Proposition 10.2.3 Si X et Y sont deux variables aléatoires d -dimensionnelles indépendantes, alors la fonction caractéristique de leur somme vaut :

$$\forall u \in \mathbb{R}^d, \quad \phi_{X+Y}(u) = \phi_X(u)\phi_Y(u)$$

Démonstration: Comme $P_{X+Y} = P_X \star P_Y$, nous avons $\widehat{P_{X+Y}} = \widehat{P_X} \widehat{P_Y}$, ce qui nous donne l'égalité voulue.

Nous pouvons également faire une preuve directe en écrivant :

$$\forall u \in \mathbb{R}^d \quad \phi_{X+Y}(u) = E [e^{iu \cdot (X+Y)}] = E [e^{iu \cdot X} e^{iu \cdot Y}]$$

Par indépendance des variables X et Y , nous en déduisons (par Fubini-Lebesgue dans \mathbb{C}) :

$$\phi_{X+Y}(u) = E [e^{iu \cdot X}] E [e^{iu \cdot Y}] = \phi_X(u) \phi_Y(u)$$

□

En quoi ce résultat nous aide-t-il à calculer des produits de convolution? Nous allons le voir maintenant sur un exemple important, en admettant provisoirement l'injectivité de la transformation de Fourier des lois, que nous prouverons dans le paragraphe suivant.

Proposition 10.2.4 *La somme de deux variables gaussiennes indépendantes est encore gaussienne. Plus précisément, on a :*

$$\mathcal{N}(m_1, \sigma_1^2) \star \mathcal{N}(m_2, \sigma_2^2) = \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2).$$

Démonstration: Soient X et Y deux v.a.r. de lois respectives $\mathcal{N}(m_1, \sigma_1^2)$ et $\mathcal{N}(m_2, \sigma_2^2)$. Grâce à la proposition précédente, nous calculons :

$$\begin{aligned} \forall u \in \mathbb{R} \quad \phi_{X+Y}(u) &= \exp\left(ium_1 - \frac{u^2 \sigma_1^2}{2}\right) \exp\left(ium_2 - \frac{u^2 \sigma_2^2}{2}\right) \\ &= \exp\left(iu(m_1 + m_2) - \frac{u^2(\sigma_1^2 + \sigma_2^2)}{2}\right) \end{aligned}$$

Nous reconnaissons la transformée de Fourier de la loi $\mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$. Par injectivité de la transformation de Fourier, ceci prouve que $X + Y$ suit la loi $\mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

□

Nous terminons ce paragraphe en considérant brièvement la notion de loi infiniment divisible, qui est hors programme.

Définition 10.2.5 *Une loi de probabilité μ est dite infiniment divisible si pour tout entier $n \in \mathbb{N}^*$, il existe une loi ν_n telle que $\mu = (\nu_n)^{\star n}$.*

On peut caractériser assez facilement les lois indéfiniment divisibles sur \mathbb{N} – ceci constitue un développement possible dans la leçon *Loi binomiale, loi de Poisson. Applications*. Nous aurons besoin pour cela d'une définition préliminaire.

Définition 10.2.6 On appelle loi de Poisson composée sur \mathbb{N} toute loi de probabilité d'une variable aléatoire de la forme

$$S = \sum_{i=0}^N X_i,$$

où $(X_i)_{i \in \mathbb{N}}$ est une suite de v.a. à valeurs dans \mathbb{N} indépendantes identiquement distribuées et N est une variable de Poisson indépendante de la suite $(X_i)_{i \in \mathbb{N}}$.

Remarque: La notation précédente signifie que, pour tout $\omega \in \Omega$,

$$S(\omega) = \sum_{i=0}^{N(\omega)} X_i(\omega).$$

On vérifie que S est bien une variable aléatoire en l'écrivant sous la forme :

$$S = \sum_{n=0}^{+\infty} \mathbf{1}_{\{N=n\}} \sum_{i=0}^n X_i.$$

Théorème 10.2.7 Les lois infiniment divisibles sur \mathbb{N} sont exactement les lois de Poisson composées.

Démonstration: Le lecteur trouvera la preuve, qui utilise la notion de fonctions génératrices, dans [FEL¹ 288-290] \square

10.3 Caractérisation de l'indépendance

Nous avons démontré l'injectivité de la transformation de Fourier sur les mesures bornées (donc sur les lois de probabilité). En voici une application.

Proposition 10.3.1 Les variables aléatoires $X_j, 1 \leq j \leq n$, à valeurs respectivement dans $\mathbb{R}^{d_j}, 1 \leq j \leq n$ sont indépendantes si et seulement si :

$$\forall u_j \in \mathbb{R}^{d_j}, \quad 1 \leq j \leq n, \quad \phi_{(X_1, \dots, X_n)}(u_1, \dots, u_n) = \phi_{X_1}(u_1) \cdots \phi_{X_n}(u_n) \quad (10.1)$$

Démonstration: Si les variables X_j sont indépendantes, nous pouvons écrire (en utilisant le théorème de Fubini pour des applications à valeurs complexes) :

$$E \left[\exp \left(i \sum_{j=1}^n u_j \cdot X_j \right) \right] = E \left[\prod_{j=1}^n \exp(iu_j \cdot X_j) \right] = \prod_{j=1}^n E [\exp(iu_j \cdot X_j)],$$

d'où l'égalité cherchée.

Réciproquement, supposons que la condition (10.1) soit satisfaite et notons $\mu_j = P_{X_j}$, la loi de X_j , ainsi que $d = \sum_{j=1}^n d_j$. D'après le théorème de Fubini, la loi $\mu = \mu_1 \otimes \cdots \otimes \mu_n$ admet pour transformée de Fourier au point $u = (u_1, \cdots, u_n) \in \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_n} \simeq \mathbb{R}^d$:

$$\hat{\mu}(u_1, \cdots, u_n) = \int_{\mathbb{R}^d} e^{i \sum_{j=1}^n u_j \cdot x_j} d\mu(x_1, \cdots, x_n) = \prod_{j=1}^n \int_{\mathbb{R}^{d_j}} e^{iu_j \cdot x_j} d\mu_j(x_j),$$

ce qui s'écrit encore :

$$\hat{\mu}(u_1, \cdots, u_n) = \prod_{j=1}^n \hat{\mu}_j(u_j) = \prod_{j=1}^n \phi_{X_j}(u_j) = \phi_{(X_1, \cdots, X_n)}(u_1, \cdots, u_n)$$

Par injectivité de la transformation de Fourier, la loi du vecteur aléatoire (X_1, \cdots, X_n) est donc égale à $\mu = \mu_1 \otimes \cdots \otimes \mu_n$, ce qui s'écrit encore :

$$P_{(X_1, \cdots, X_n)} = \mu_1 \otimes \cdots \otimes \mu_n = P_{X_1} \otimes \cdots \otimes P_{X_n},$$

d'où l'indépendance des variables X_j , $1 \leq j \leq n$. □

Remarque: Le lecteur trouvera un raffinement de ce résultat dans l'exercice 11.2.5

Terminons ce chapitre par une application de la formule d'inversion de Fourier.

Exercice : Calculer la fonction caractéristique de la loi de Laplace, de densité $\frac{a}{2} \exp(-a|x|)$, avec $a > 0$, puis par inversion de Fourier en déduire la fonction caractéristique de la loi de Cauchy. On pourra prouver ainsi que $C(a) \star C(b) = C(a+b)$. En particulier, la loi de Cauchy est infiniment divisible.

Chapitre 11

Fonctions caractéristiques.

11.1 Propriétés générales.

11.1.1 Dérivabilité des f.c.

Soit X une v.a. à valeurs dans \mathbb{R}^d appartenant à $L^p(\Omega, \mathcal{A}, P)$, i.e. telle que

$$\forall i = 1, 2, \dots, d, E[|X_i|^p] < +\infty.$$

Montrer que sa f.c. ϕ_X est de classe C^p et déterminer ses dérivées successives.

11.1.2 F.c. périodiques.

1. Si une v.a. X est à valeurs dans $a\mathbb{Z}$ pour un réel $a \neq 0$, montrer que sa f.c. ϕ_X est périodique de période $\frac{2\pi}{a}$.

2. Soit Y une v.a.r.; montrer qu'on a l'équivalence :

$$\exists t_0 \neq 0, |\phi_Y(t_0)| = 1 \Leftrightarrow Y \text{ est p.s. à valeurs dans une progression arithmétique de la forme } a\mathbb{Z} + b.$$

11.1.3 Loi symétrique et f.c.

Soit X une v.a. à valeurs dans \mathbb{R}^d ; on dit que sa loi est symétrique si $P_{-X} = P_X$. Montrer que ϕ_X est à valeurs réelles **ssi** la loi de X est symétrique.

11.2 F.c. de lois particulières.

11.2.1 Quelques f.c. de lois “classiques”.

Vérifier dans chaque cas que la v.a.r. X admet pour fonction caractéristique la fonction ϕ_X proposée :

1. $X \sim \text{Bernoulli}(p)$; $\phi_X(u) = pe^{iu} + 1 - p$.
2. $X \sim \text{Binomiale}(n, p)$; $\phi_X(u) = (pe^{iu} + 1 - p)^n$.
3. $X \sim \mathcal{P}(\lambda)$; $\phi_X(u) = \exp(\lambda(e^{iu} - 1))$.
4. X suit une loi géométrique de paramètre $p \in]0, 1[$ i.e.
 $P(X = n) = p(1 - p)^n, n \in \mathbb{N}$;

$$\phi_X(u) = \frac{p}{1 - (1 - p)e^{iu}}.$$

5. $X \sim \text{Exp}(\lambda)$; $\phi_X(u) = \frac{1}{1 - iu/\lambda}$.
6. $X \sim \text{Unif}([a, b])$; $\phi_X(u) = \frac{e^{ibu} - e^{iau}}{iu(b-a)}$.

11.2.2 Développement en base 2.

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de v.a.r. i.i.d. de loi de Bernoulli symétrique :
 $P(X_n = 1) = P(X_n = -1) = \frac{1}{2}$.

1. Montrer que $\sum_{n \geq 1} \frac{X_n}{2^n}$ est convergente p.s.
 2. Calculer la f.c. de $\sum_{k=1}^n \frac{X_k}{2^k}$ et en déduire la loi de $\sum_{n \geq 1} \frac{X_n}{2^n}$.
- Indication* : On montrera que $\prod_{k=1}^n \cos\left(\frac{t}{2^k}\right) = \frac{\sin t}{2^n \sin\left(\frac{t}{2^n}\right)}$.

11.2.3 Laplace et Cauchy.

1. On appelle loi de Laplace de paramètre $a > 0$ la loi de densité $\frac{a}{2} \exp(-a|x|)$.
Calculer sa f.c.
2. En déduire la f.c. de la loi de Cauchy de paramètre a .
3. Que vaut le produit de convolution $C(a) * C(b)$?
4. Soient $Y \sim \text{Exp}(\alpha)$ et $Z \sim \mathcal{N}(0, 1)$ deux v.a.r. indépendantes ; on pose $X = \sqrt{Y}Z$. Calculer la f.c. de X et en déduire sa loi.

11.2.4 Convolées de lois.

1. Calculer la convolée $\mathcal{P}(\lambda) * \mathcal{P}(\mu)$ par deux méthodes : directement et par calcul de f.c.
2. Retrouver l'égalité $\text{Exp}(\lambda)^{*n} = \gamma(\lambda, n)$ en utilisant les f.c.

11.2.5 Une caractérisation de l'indépendance.

Montrer que les variables aléatoires $X_j, 1 \leq j \leq n$, à valeurs respectivement dans $\mathbb{R}^{d_j}, 1 \leq j \leq n$ sont indépendantes si et seulement si $\phi_{(X_1, \dots, X_n)}$ est à variables séparables, i.e. il existe des applications $f_j : \mathbb{R}^{d_j} \rightarrow \mathbb{C}, 1 \leq j \leq n$, telles que :

$$\forall u_j \in \mathbb{R}^{d_j}, \quad 1 \leq j \leq n, \quad \phi_{(X_1, \dots, X_n)}(u_1, \dots, u_n) = f_1(u_1) \cdots f_n(u_n)$$

11.2.6 Une caractérisation de la loi de Gauss.

1. Soient X et Y deux v.a.r. i.i.d. de loi $\mathcal{N}(0, 1)$. Calculer la f.c. du couple $(X + Y, X - Y)$. En déduire que $X + Y$ et $X - Y$ sont indépendantes et préciser leurs lois marginales.
2. Nous supposons maintenant que X et Y sont i.i.d. de loi μ centrée réduite (i.e. $E[X] = 0$ et $E[X^2] = 1$) et telle que $X + Y$ et $X - Y$ sont indépendantes.

On note ϕ la f.c. de la loi μ .

(a) Montrer que $\phi(2t) = (\phi(t))^3 \phi(-t)$.

(b) En déduire que ϕ ne s'annule jamais.

Indication : Raisonner par l'absurde et construire une suite $(t_n) \rightarrow 0$ telle que $\phi(t_n) = 0$ pour tout $n \in \mathbb{N}$.

(c) Montrer que μ est une loi symétrique.

Indication : Poser $p(t) = \phi(t)/\phi(-t)$ et prouver que $p \equiv 1$ en partant de l'égalité $p(2t) = p(t)^2$ et en utilisant un développement limité de ϕ en 0 à l'ordre 2.

(d) Montrer en conclusion que $\mu = \mathcal{N}(0, 1)$.

Chapitre 12

Convergence dans L^p .

Convergence en probabilité.

12.1 Moments d'une v.a.r. Convergence dans L^p .

12.1.1 Espérances et variances de lois "classiques".

Vérifier dans chaque cas que la v.a.r. X admet bien l'espérance et la variance indiquées :

1. $X \sim \text{Bernoulli}(p)$; $E[X] = p$, $\text{Var}X = p(1-p)$.
2. $X \sim \text{Binomiale}(n, p)$; $E[X] = np$, $\text{Var}X = np(1-p)$.
3. $X \sim \mathcal{P}(\lambda)$; $E[X] = \lambda$, $\text{Var}X = \lambda$.
4. $X \sim \mathcal{N}(m, \sigma^2)$; $E[X] = m$, $\text{Var}X = \sigma^2$.
5. $X \sim \text{Exp}(\lambda)$; $E[X] = \frac{1}{\lambda}$, $\text{Var}X = \frac{1}{\lambda^2}$.
6. $X \sim \text{Unif}([a, b])$; $E[X] = \frac{a+b}{2}$, $\text{Var}X = \frac{(b-a)^2}{12}$.

12.1.2 Produit infini.

Si $(X_n)_{n \in \mathbb{N}^*}$ est une suite i.i.d. de v.a.r. réduites, montrer que la suite $(Z_n)_{n \in \mathbb{N}^*}$ définie par :

$$Z_n = \prod_{i=1}^n (a + bX_i)$$

est convergente en moyenne quadratique dès que $a^2 + b^2 < 1$.

Que se passe-t-il si $a^2 + b^2 > 1$?

12.1.3 Une CNS de convergence dans L^2 .

Montrer que la suite $(X_n)_{n \in \mathbb{N}^*}$ est convergente en moyenne quadratique **ssi**

$$\lim_{m,n \rightarrow +\infty} E[X_m X_n] \text{ existe.}$$

12.1.4 Séries convergentes dans L^2 .

[REV 139](non corrigé)

1. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a.r. indépendantes, centrées et de carrés intégrables. Montrer que la série $\sum_{n \in \mathbb{N}} X_n$ converge dans L^2 **ssi** la série réelle positive $\sum_{n \in \mathbb{N}} E[X_n^2]$ est convergente.
2. On suppose maintenant que pour tout $n \in \mathbb{N}$, X_n ne prend p.s. que deux valeurs a_n et $-b_n$, avec $0 < a_n \leq b_n$. Donner une CNS pour que la série $\sum_{n \in \mathbb{N}} X_n$ converge dans L^2 . Sous cette condition, la série converge-t-elle en un autre sens ?
3. Soit $(Y_n)_{n \in \mathbb{N}}$ une suite de v.a.r. indépendantes et telles que, pour tout $n \in \mathbb{N}$, Y_n suit une loi exponentielle de paramètre $\frac{1}{\lambda_n}$, $\lambda_n > 0$.
Donner une CNS pour que la série $\sum_{n \in \mathbb{N}} Y_n$ converge dans L^2 .

12.2 Convergence en probabilité.

12.2.1 Métrique de la convergence en probabilité.

[BAR 121]

Montrer que $d(X, Y) = E[|X - Y| \wedge 1]$ est une distance sur l'espace L^0 des (classes d'équivalences de) v.a.r. p.s. finies.

Montrer que la convergence en probabilité est définie par cette distance, ainsi que les suites de Cauchy en probabilité.

Que pouvez-vous dire de l'espace métrique (L^0, d) ?

12.2.2 où l'on revisite Lebesgue.

[FOA 212]

1. Montrer que $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers X **ssi** de toute sous-suite, on peut extraire une (sous-)sous-suite qui converge p.s. vers X .
2. En déduire que dans le théorème de convergence dominée de Lebesgue, il suffit de supposer la convergence en probabilité au lieu de la convergence p.s.

3. Démontrer le théorème suivant :

Si $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers X et si $f : \mathbb{R} \rightarrow \mathbb{R}$ est continue, alors

la suite $(f(X_n))_{n \in \mathbb{N}}$ converge en probabilité vers $f(X)$.

Chapitre 13

Convergence en loi. Théorème-limite central.

13.1 Convergence en loi.

13.1.1 Le cas discret.

[COT 110]

1. Montrer qu'une suite de probabilités portées par \mathbb{Z} , notée $(P_n)_{n \in \mathbb{N}}$, converge étroitement vers une probabilité P ssi la suite $(P_n(k))_{n \in \mathbb{N}}$ converge vers $P(k)$ pour tout $k \in \mathbb{Z}$.
2. Montrer que pour tout paramètre $\lambda > 0$, la suite de lois $Bin(n, \frac{\lambda}{n})$ converge étroitement vers la loi $\mathcal{P}(\lambda)$.
3. Montrer que toute limite en loi de variables de Poisson est une variable de Poisson.

13.1.2 Probabilité d'un intervalle.

[COT 111-113]

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a.r. qui converge en loi vers une v.a.r. X .

Soient deux suites réelles $(a_n)_{n \in \mathbb{N}}$ et $(b_n)_{n \in \mathbb{N}}$ de limites respectives a et b avec $a \leq b$ et $P(X = a) = P(X = b) = 0$.

Montrer qu'on alors :

$$\lim_{n \rightarrow \infty} P(X_n \in [a_n, b_n]) = P(X \in [a, b]).$$

Indication : On pourra supposer dans un premier temps que la f.r. X est continue.

Remarque : On peut montrer que ceci reste exact lorsqu'on remplace les intervalles $[a_n, b_n]$ et $[a, b]$ par des intervalles ouverts ou semi-ouverts, bornés ou non.

13.1.3 Théorème de Scheffé.

[COT 110-111] ou [FOA 216]

Soient $(X_n)_{n \in \mathbb{N}}$ et X des v.a.r. admettant des densités respectives $(f_n)_{n \in \mathbb{N}}$ et f .

1. On suppose que dx -presque partout, on a $\lim_{n \rightarrow \infty} f_n(x) = f(x)$.
 - (a) Montrer que la suite $(f_n)_{n \in \mathbb{N}}$ converge vers f dans $L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, où λ désigne la mesure de Lebesgue sur \mathbb{R} .
Indication : $|f_n - f| = f_n + f - 2 \inf(f, f_n)$.
 - (b) En déduire que la suite $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X .
2. Montrer que la réciproque est fautive en prenant $f_n(x) = (1 - \cos(2\pi nx)) \mathbf{1}_{[0,1]}(x)$ et $f(x) = \mathbf{1}_{[0,1]}(x)$.

13.1.4 Moyenne empirique de Cauchy.

1. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a.r. i.i.d. suivant une loi $C(\alpha)$ et $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$.
 Montrer que pour tout $n \in \mathbb{N}$ et tout $p \in \mathbb{N}^*$, les v.a.r. $\bar{X}_{n+p} - \bar{X}_n$ et $\frac{2}{1+n/p} X_1$ ont même loi.
2. La suite $(\bar{X}_n)_{n \in \mathbb{N}}$ converge-t-elle en loi, en probabilité, dans L^p ?

13.1.5 Avec des variables uniformes.

[FOA 223]

Soit $(X_n)_{n \in \mathbb{N}}$ une suite i.i.d. de variables uniformes sur $[0,1]$. On pose :

$$U_n = n \inf_{1 \leq i \leq n} X_i.$$

Montrer que la suite $(U_n)_{n \in \mathbb{N}}$ converge en loi vers T , où T est une v.a. exponentielle dont on précisera le paramètre.

13.1.6 Vrai ou faux ?

[FOA 221]

Soient $(X_n)_{n \in \mathbb{N}}, (Y_n)_{n \in \mathbb{N}}, X, Y$ des vecteurs aléatoires à valeurs dans R^d .

1. On suppose que $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X$ et $Y_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} Y$. Peut-on en déduire les relations suivantes ?

(a) $X_n - X \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} 0$

(b) $X_n + Y_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X + Y$

(c) $X_n \cdot Y_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X \cdot Y$ (produit scalaire)

2. On suppose que $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} a$ et $Y_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} Y$.

Montrer que $(X_n, Y_n) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} (a, Y)$, $X_n + Y_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} a + Y$ et $X_n \cdot Y_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X \cdot Y$

13.1.7 Une moyenne empirique paramétrée.

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de v.a.r. indépendantes telles que :

$$\forall n \in \mathbb{N}^* \quad P(X_n = n^\alpha) = P(X_n = -n^\alpha) = \frac{1}{2}.$$

On pose $\bar{X}_n = \sum_{i=1}^n X_i$.

1. Montrer que pour $\alpha < \frac{1}{2}$, la suite $(\bar{X}_n)_{n \in \mathbb{N}^*}$ converge vers 0 en moyenne quadratique.
2. Etudier la convergence en loi de $(\bar{X}_n)_{n \in \mathbb{N}^*}$ pour $\alpha = \frac{1}{2}$.
3. (a) Montrer que si $(U_n)_{n \in \mathbb{N}^*}$ converge en loi vers U et si $(\lambda_n)_{n \in \mathbb{N}^*}$ est une suite réelle telle que $(|\lambda_n|)_{n \in \mathbb{N}^*}$ tende en décroissant vers 0, alors la suite $(\lambda_n U_n)_{n \in \mathbb{N}^*}$ converge en loi vers 0.
- (b) Etudier la convergence en loi de la suite :

$$Z_{n,\beta} = \frac{1}{n^\beta} \sum_{i=1}^n X_i, \quad n \in \mathbb{N}^*.$$

En déduire que pour $\alpha > \frac{1}{2}$, la suite $(\bar{X}_n)_{n \in \mathbb{N}^*}$ ne converge pas en loi.

13.1.8 Suite gaussienne.

[COT 224-225]

Montrer que toute limite en loi d'une suite de v.a.r. gaussiennes est nécessairement gaussienne. Que valent son espérance et sa variance ?

13.1.9 Suite totale dans \mathcal{C}_0 .

Pour tout $a > 0$ et tout $b \in \mathbb{R}^d$, on définit l'application $g_{a,b} : \mathbb{R}^d \rightarrow \mathbb{R}$ par :

$$\forall x \in \mathbb{R}^d, \quad g_{a,b}(x) = \exp(-a|x|^2 + b \cdot x).$$

Montrer que l'ensemble dénombrable des applications $g_{a,b}$, $a \in \mathbb{Q}_+^*$, $b \in \mathbb{Q}^d$ est total dans \mathcal{C}_0 .

Indication : Appliquer le théorème de Stone-Weierstrass. On pourra prolonger $g_{a,b}$ en une application continue sur le compactifié d'Alexandrov $\mathbb{R}^d \cup \{\infty\}$ en posant $g_{a,b}(\infty) = 0$.

13.2 Théorème-limite central.

13.2.1 Coups de dés.

[FEL¹ 244]

On jette 1000 fois de suite un dé à 6 faces équilibré et l'on additionne l'ensemble des points obtenus. Montrer que le résultat a environ une chance sur deux d'appartenir à l'intervalle [3464,3536].

Indication : Si $Z \sim \mathcal{N}(0, 1)$, on a $P(|Z| \leq \alpha) = 0,5$ pour $\alpha \sim 0,6744 \dots$

13.2.2 Approximation de la loi binomiale par la loi de Gauss

[REV 165]

On considère un serveur auquel peuvent se relier n clients, qui sont supposés se connecter indépendamment les uns des autres. Pendant une période de pointe, la probabilité qu'un client donné veuille se brancher est p .

Quel doit être le nombre c_n de clients que le serveur est capable de traiter simultanément pour que la probabilité d'avoir à faire attendre au moins un client soit d'environ 2,5% ?

13.2.3 Une caractérisation de la loi de Gauss.

[COT 128-129]

Soient X et Y deux v.a.r. i.i.d. de carrés intégrables. On suppose que $\frac{X+Y}{\sqrt{2}}$ a même loi que X et Y .

1. Montrer que X et Y sont centrées.
2. Montrer que si $(X_i)_{1 \leq i \leq 2^n}$ est une famille de v.a.r. i.i.d. de même loi que X , alors $\frac{1}{\sqrt{2^n}} \sum_{i=1}^{2^n} X_i$ a même loi que X .
3. En déduire que X est gaussienne.

13.2.4 Une formule de Bernstein.

[COT 127]

Utiliser le théorème limite central pour montrer que :

$$\lim_{n \rightarrow \infty} e^{-n} \sum_{k=0}^n \frac{n^k}{k!} = \frac{1}{2}.$$

13.2.5 *TCL pour sommes aléatoires de variables i.i.d.

[COT 129-131]

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de v.a.r. i.i.d. centrées admettant une variance σ^2 ; on définit, pour tout $n \in \mathbb{N}^*$, $S_n = \sum_{m=1}^n X_m$.

Soit $(N_k)_{k \in \mathbb{N}^*}$ une suite de v.a. à valeurs dans \mathbb{N}^* , indépendantes de la suite $(X_n)_{n \in \mathbb{N}^*}$ et telles que, presque sûrement, $N_k(\omega) \rightarrow +\infty$ quand $k \rightarrow \infty$.

On définit alors, pour tout $k \in \mathbb{N}^*$, la v.a.r. $Z_k = \frac{1}{\sqrt{N_k}} S_{N_k}$ (c'est-à-dire que $Z_k(\omega) = \frac{1}{\sqrt{N_k(\omega)}} S_{N_k(\omega)}$).

Montrer que la suite $(Z_k)_{k \in \mathbb{N}^*}$ converge en loi vers la loi $\mathcal{N}(0, \sigma^2)$.

13.2.6 *Théorème de Bochner.

[COT 143-146]

Une application $f : \mathbb{R} \rightarrow \mathbb{C}$ est dite de type positif si pour tous $(t_1, \dots, t_n) \in \mathbb{R}^n$ et $(a_1, \dots, a_n) \in \mathbb{C}^n$, on a :

$$\sum_{1 \leq j \leq n, 1 \leq k \leq n} f(t_j - t_k) a_j \bar{a}_k \geq 0. (*)$$

Le but de cet exercice est de montrer qu'une fonction $\phi : \mathbb{R} \rightarrow \mathbb{C}$ est une fonction caractéristique **ssi** elle est continue, de type positif et telle que $\phi(0) = 1$.

1. Montrer que toute fonction caractéristique est de type positif.
2. Montrer que si $\phi : \mathbb{R} \rightarrow \mathbb{C}$ est de type positif, alors on a :
 $\phi(0) \geq 0$ et, pour tout $t \in \mathbb{R}$, $\phi(-t) = \bar{\phi}(t)$, $|\phi(t)| \leq \phi(0)$.
Indication : Utiliser (*) avec $n = 2$, $t_1 = 0$, $t_2 = t$ et un choix convenable de a_1 et a_2 .
3. On suppose maintenant que $\phi : \mathbb{R} \rightarrow \mathbb{C}$ est continue, de type positif et telle que $\phi(0) = 1$. Montrer que l'application ϕ_n définie par :

$$\forall t \in \mathbb{R}, \phi_n(t) = \phi(t) \exp\left(-\frac{t^2}{2n^2}\right)$$

est une fonction caractéristique.

Conclure en utilisant l'inversion de Fourier et le théorème de Paul Lévy.

Chapitre 14

Complément de cours : Applications du théorème-limite central

Référence : Pour les applications statistiques du TLC et pour l'option probas-stat, le livre suivant a été spécialement écrit en vue de la préparation au concours pour tout ce qui est en rapport avec la statistique.

Statistique en action par V. RIVOIRARD et G.STOLTZ chez Vuibert.

Préliminaire : Nous proposons ici une démonstration élémentaire (sans utilisation du logarithme complexe) du lemme suivant, qui intervient dans la preuve du théorème-limite central.

Lemme 14.0.1 Soit $(u_n)_{n \in \mathbb{N}^*}$ une suite de nombres complexes qui converge vers $z \in \mathbb{C}$. Alors, on a :

$$\lim \left(1 + \frac{u_n}{n} \right)^n = e^z$$

L'idée simple que nous allons utiliser plusieurs fois dans la démonstration de ce lemme est que si Q est un polynôme dont tous les coefficients sont des réels positifs ou nuls, alors on a l'inégalité :

$$\forall z \in \mathbb{C}, \quad |Q(z)| \leq Q(|z|)$$

En effet, en écrivant $Q(z) = \sum_{k=0}^n a_k z^k$, avec $a_k \geq 0$ pour tout $k = 0, \dots, n$ nous constatons que cette inégalité est une conséquence immédiate de l'inégalité triangulaire.

Passons maintenant à la démonstration du lemme, que nous décomposons en quatre étapes :

Étape 1 : (u_n) est une suite à valeurs réelles.

Dans ce cas, à partir d'un certain rang, on a $|u_n/n| \leq 1/2$ et nous pouvons passer au logarithme (sur \mathbb{R}_+^*) pour conclure.

Étape 2 : $\forall n \in \mathbb{N}^*, u_n = z \in \mathbb{C}$

Introduisons les polynômes :

$$P_n(z) = \sum_{k=0}^n \frac{z^k}{k!} \quad ; \quad R_n(z) = P_n(z) - \left(1 + \frac{z}{n}\right)^n$$

Nous calculons facilement :

$$R_n(z) = \sum_{k=0}^n \frac{z^k}{k!} - \sum_{k=0}^n C_n^k \frac{z^k}{n^k} = \sum_{k=2}^n \frac{z^k}{k!} \left(1 - \frac{n-1}{n} \dots \frac{n-k+1}{n}\right)$$

Nous constatons donc que le polynôme R_n est à coefficients réels positifs ou nuls et l'inégalité $|R_n(z)| \leq R_n(|z|)$ nous permet de conclure grâce à l'étape 1.

Étape 3 : La suite complexe (u_n) converge vers 0.

Il suffit de remarquer que le polynôme $(1+z)^n - 1$ est à coefficients réels positifs ou nuls pour en déduire l'inégalité :

$$\left| \left(1 + \frac{u_n}{n}\right)^n - 1 \right| \leq \left(1 + \frac{|u_n|}{n}\right)^n - 1,$$

ce qui nous permet de conclure grâce à l'étape 1.

Étape 4 : La suite complexe (u_n) converge vers $z \in \mathbb{C}^*$.

Cette hypothèse s'écrit encore :

$$\forall n \in \mathbb{N}^*, \quad u_n = z(1 + v_n) \text{ avec } \lim v_n = 0$$

Nous calculons alors :

$$\left(1 + \frac{u_n}{n}\right)^n - \left(1 + \frac{z}{n}\right)^n = \sum_{k=0}^n C_n^k \frac{z^k}{n^k} ((1 + v_n)^k - 1)$$

Toujours pour la même raison, nous en déduisons l'inégalité :

$$\left| \left(1 + \frac{u_n}{n}\right)^n - \left(1 + \frac{z}{n}\right)^n \right| \leq \sum_{k=0}^n C_n^k \frac{|z|^k}{n^k} ((1 + |v_n|)^k - 1)$$

Cette inégalité s'écrit encore :

$$\left| \left(1 + \frac{u_n}{n}\right)^n - \left(1 + \frac{z}{n}\right)^n \right| \leq \left(1 + \frac{|z|(1 + |v_n|)}{n}\right)^n - \left(1 + \frac{|z|}{n}\right)^n$$

Le membre de droite dans cette inégalité tend vers 0 en vertu de l'étape 1. Nous pouvons alors conclure grâce au résultat de l'étape 2.

14.1 Intervalles de confiance

14.1.1 Quelques notions de statistique

Pour montrer la différence entre l'objet du calcul des probabilités et celui de la statistique, prenons l'exemple d'un jeu de pile ou face dans lequel on jette la pièce (pas nécessairement équilibrée) un "grand nombre" de fois. Nous introduisons donc le *modèle canonique* :

$$\Omega = \{0, 1\}^N \quad ; \quad \mathcal{A} = \mathcal{P}(\Omega) \quad ; \quad P = B(p)^{\otimes N}, \quad p \in]0, 1[$$

Le probabiliste connaît p et cherche à en déduire des propriétés sur le résultat de l'expérience. Par exemple, si nous appelons X_i la i^e projection canonique (définie par : $\forall \omega = (\omega_1, \dots, \omega_N) \in \Omega, \quad X_i(\omega) = \omega_i$), la loi forte des grands nombres nous donne :

$$P(d\omega)\text{-p.s.}, \quad \frac{X_1(\omega) + \dots + X_N(\omega)}{N} \approx p$$

Notons que ceci est une propriété du résultat ω de l'expérience puisqu'elle s'écrit encore :

$$\frac{\omega_1 + \dots + \omega_N}{N} \approx p$$

La théorie des probabilités nous permet de dire que si l'expérience nous donne un résultat ω , il a "de fortes chances" de vérifier cette propriété.

Pour le probabiliste, la probabilité P est donc connue et il s'agit d'en déduire des propriétés du résultat ω de l'expérience. Le statisticien fait la démarche en sens inverse : il connaît le résultat ω de l'expérience et il en déduit des propriétés de la probabilité P qui est inconnue.

Ainsi, dans l'exemple précédent, supposons que le paramètre p qui intervient dans la définition de la probabilité P est inconnu mais que nous disposons du résultat de N jets de pièce successifs, c'est-à-dire que nous ayons observé un ω pour l'expérience considérée. Se basant sur la loi forte des grands nombres, le statisticien pourra proposer l'estimation suivante du paramètre p :

$$\hat{p}(\omega) = \frac{X_1(\omega) + \dots + X_N(\omega)}{N}$$

De façon plus générale, le statisticien part d'une famille de probabilités $(P_\theta)_{\theta \in \Theta}$ parmi laquelle il cherche à déterminer celle qui modélisera le mieux le phénomène observé. En *statistique paramétrique*, on suppose que l'ensemble Θ est inclus dans \mathbb{R}^d pour un certain $d \in \mathbb{N}^*$. Au vu du résultat ω de l'expérience, le statisticien propose une valeur $\hat{\theta}(\omega)$ pour estimer le paramètre θ

(ou encore pour estimer une fonction $f(\theta)$ de ce paramètre). Dans l'exemple précédent, nous avons $\Theta =]0, 1[$ et $P_\theta = B(\theta)^{\otimes N}$.

Nous demanderons à cette fonction $\hat{\theta} : \Omega \rightarrow \Theta$ d'être mesurable et nous l'appellerons un *estimateur* de θ . Dans notre exemple, nous avons donc $\hat{\theta} = \bar{X}_n$.

Notons que faire de la statistique nous demande de travailler avec toute une famille de probabilités à la fois et que nous serons donc amenés à écrire des expressions telle que :

$$\forall \theta \in \Theta \quad P_\theta(|\hat{\theta} - \theta| \geq \epsilon) < \alpha$$

Par exemple, l'estimateur $\hat{\theta}$ est dit *sans biais* s'il vérifie $E_\theta[\hat{\theta}] = \theta$ pour tout $\theta \in \Theta$, où E_θ désigne l'espérance calculée sous la probabilité P_θ .

14.1.2 L'estimateur moyenne empirique

Dans ce paragraphe, nous allons nous intéresser plus particulièrement à l'estimateur \bar{X}_N , appelé *moyenne empirique*. Nous allons constater en effet qu'il possède des propriétés intéressantes lorsque nous souhaitons estimer la moyenne d'une loi.

Supposons donc maintenant que $P_\theta = \mu_\theta^{\otimes N}$, avec μ_θ loi sur \mathbb{R} admettant une espérance $m(\theta)$. Nous souhaitons estimer la valeur de $m(\theta)$.

Dans l'exemple précédent, nous avons $\mu_\theta = B(\theta)$ donc $m(\theta) = \theta$.

Citons un autre exemple : $\Omega = \mathbb{R}^N$, $\mathcal{A} = \mathcal{B}(\mathbb{R}^n)$, $P_\theta = \text{Exp}(\theta)^{\otimes N}$, $\Theta = \mathbb{R}_+^*$; dans ce cas, $m(\theta) = 1/\theta$.

Une propriété intéressante de l'estimateur \bar{X}_N de $m(\theta)$ est qu'il est *sans biais*, c'est-à-dire qu'il donne la "bonne" valeur en moyenne en ce sens que :

$$\forall \theta \in \Theta \quad E_\theta[\bar{X}_N] = \frac{1}{N} \sum_{i=1}^N E_\theta[X_i] = m(\theta)$$

Supposons maintenant que nous pouvons répéter indéfiniment la même procédure dans des conditions identiques (par exemple jouer à pile ou face indéfiniment). Nous prenons alors plutôt pour *modèle statistique* :

$$\Omega = \mathcal{E}^{\mathbb{N}} \quad , \quad \mathcal{A} = \sigma(X_n, n \in \mathbb{N}) \quad , \quad P_\theta = \mu_\theta^{\otimes \mathbb{N}}$$

où $\mathcal{E} = \{0, 1\}$ ou \mathbb{R} ou $\mathbb{R}^d \dots$, X_n est la n^e injection canonique définie par $X_n(\omega) = \omega_n$ et la probabilité $\mu_\theta^{\otimes \mathbb{N}}$ est l'unique probabilité sous laquelle la suite $(X_n)_{n \in \mathbb{N}}$ est indépendante identiquement distribuée de loi commune μ_θ (nous admettons l'existence et l'unicité de cette probabilité, qui résultent du

théorème des classes monotones ainsi que d'un théorème d'extension énoncé par Kolmogorov).

Dans un tel modèle statistique, nous disposons d'une suite d'estimateurs de $m(\theta)$: la suite des moyennes empiriques $(\bar{X}_n)_{n \in \mathbb{N}^*}$.

Une propriété intéressante de cette suite d'estimateurs est qu'elle est *consistante* (ou *convergente*, ce qui est synonyme) c'est-à-dire qu'elle donne la "bonne" valeur asymptotiquement en ce sens que, d'après la loi forte des grands nombres :

$$\forall \theta \in \Theta \quad P_\theta(d\omega)\text{-p.s.}, \quad \bar{X}_n(\omega) \longrightarrow m(\theta)$$

C'est à ce stade qu'intervient le théorème-limite central car dans la pratique, il est totalement insuffisant d'avoir une suite d'estimateur convergente si nous n'avons aucune idée de la vitesse à laquelle elle converge. C'est toute la question du degré de précision de notre estimateur qui est en jeu et le théorème-limite central va nous permettre d'y apporter une réponse grâce à la notion d'*intervalle de confiance*.

14.1.3 Intervalles de confiance

Afin de pouvoir appliquer le théorème-limite central, nous supposons désormais que pour tout $\theta \in \Theta$, la loi μ_θ admet un moment d'ordre 2 et nous notons alors $\sigma^2(\theta)$ sa variance. Nous avons alors :

$$\text{Sous } P_\theta, \quad \frac{\sqrt{n}}{\sigma(\theta)} (\bar{X}_n - m(\theta)) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

La loi $\mathcal{N}(0, 1)$ n'ayant pas de masse ponctuelle, nous en déduisons que pour tous réels $a \leq b$, nous avons la convergence suivante lorsque n tend vers l'infini :

$$P_\theta \left[a \leq \frac{\sqrt{n}}{\sigma(\theta)} (\bar{X}_n - m(\theta)) \leq b \right] \longrightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$$

Il est facile de vérifier que pour tout $0 < \alpha \leq 1$, il existe un unique $\phi_\alpha \in \mathbb{R}_+$ tel que :

$$\frac{2}{\sqrt{2\pi}} \int_{\phi_\alpha}^{+\infty} e^{-\frac{x^2}{2}} dx = \alpha$$

Nous en déduisons que, pour tout $\theta \in \Theta$, la convergence suivante a lieu lorsque n tend vers l'infini :

$$P_\theta \left[-\phi_\alpha \leq \frac{\sqrt{n}}{\sigma(\theta)} (\bar{X}_n - m(\theta)) \leq \phi_\alpha \right] \longrightarrow 1 - \alpha$$

Cette convergence s'écrit encore :

$$\forall \theta \in \Theta, \quad P_\theta \left[m(\theta) \in \left[\bar{X}_n - \frac{\sigma(\theta)\phi_\alpha}{\sqrt{n}}, \bar{X}_n + \frac{\sigma(\theta)\phi_\alpha}{\sqrt{n}} \right] \right] \rightarrow 1 - \alpha$$

L'intervalle (aléatoire) $\left[\bar{X}_n - \frac{\sigma(\theta)\phi_\alpha}{\sqrt{n}}, \bar{X}_n + \frac{\sigma(\theta)\phi_\alpha}{\sqrt{n}} \right]$ est appelé intervalle de confiance pour $m(\theta)$ de niveau de confiance asymptotique $1 - \alpha$ (ou encore de niveau d'erreur asymptotique α). Cela signifie que, si nous affirons que $m(\theta)$ est dans cet intervalle, la probabilité que nous fassions erreur est proche de α lorsque n est "grand".

Par exemple, si $\alpha = 0,05$, les tables de la loi normale (classiques en statistique, voir par exemple les dernières pages de [DAC¹ exo]) nous fournissent la valeur $\phi_\alpha \sim 1,96$; pour $\alpha = 0,002$, nous obtenons $\phi_\alpha \sim 3,09$.

Il est intéressant de faire la comparaison entre l'information qui nous est donnée par le théorème-limite central et celle que nous fournit l'inégalité de Bienaymé-Tchebichev, à savoir :

$$P_\theta \left[-\phi_\alpha \leq \frac{\sqrt{n}}{\sigma(\theta)} (\bar{X}_n - m(\theta)) \leq \phi_\alpha \right] \geq 1 - \frac{1}{\phi_\alpha^2}$$

Nous constatons que le théorème-limite central nous donne un renseignement beaucoup plus précis. Notons cependant que l'inégalité de Bienaymé-Tchebichev est vraie même si n est petit.

Un problème subsiste concernant $\left[\bar{X}_n - \frac{\sigma(\theta)\phi_\alpha}{\sqrt{n}}, \bar{X}_n + \frac{\sigma(\theta)\phi_\alpha}{\sqrt{n}} \right]$, intervalle de confiance que nous venons de déterminer : le paramètre θ y intervient au travers du facteur $\sigma(\theta)$, or précisément θ est un paramètre inconnu que nous souhaitons évaluer (en estimant $m(\theta)$) donc nous risquons de tourner en rond !

Dans certaines situations, il est possible de majorer $\sigma(\theta)$ indépendamment de θ , ce qui nous permet de construire un intervalle de confiance dans lequel le paramètre θ n'intervient plus.

Par exemple, si $\mu_\theta = B(\theta)$, nous avons $\sigma^2(\theta) = \theta(1 - \theta) \leq 1/4$ pour tout $\theta \in \Theta = [0, 1]$. Nous en déduisons que $\left[\bar{X}_n - \frac{1}{\sqrt{n}}, \bar{X}_n + \frac{1}{\sqrt{n}} \right]$ est un intervalle de confiance pour $m(\theta) = \theta$ de niveau de confiance asymptotique 95%.

Dans le cas général, nous introduisons la *variance empirique* définie par :

$$\Sigma_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Une certaine propriété de stabilité de la convergence en loi, dite lemme de Slutsky, permet de démontrer que l'intervalle aléatoire suivant (dans lequel nous avons remplacé l'écart-type théorique $\sigma(\theta)$ par l'écart-type empirique

Σ_n , ce qui supprime toute dépendance en le paramètre θ :

$$\left[\bar{X}_n - \frac{\Sigma_n \phi_\alpha}{\sqrt{n}}, \bar{X}_n + \frac{\Sigma_n \phi_\alpha}{\sqrt{n}} \right]$$

est encore un intervalle de confiance pour $m(\theta)$ de niveau de confiance asymptotique $1 - \alpha$.

14.2 Méthodes de Monte-Carlo

On appelle méthode de Monte-Carlo¹ toute méthode visant à calculer une valeur numérique en utilisant des techniques probabilistes.

Le problème de départ peut être *a priori* de nature complètement déterministe, comme par exemple le calcul approché de l'intégrale :

$$I = \int_{[0,1]^d} f(x) dx, \text{ où } f \text{ est une application borélienne de } [0, 1]^d \text{ dans } \mathbb{R},$$

mais nous allons y introduire artificiellement un aspect aléatoire en remarquant que cette intégrale s'écrit encore, d'après le théorème de transfert,

$$I = E[f(X)] , \text{ où } X \text{ est un vecteur aléatoire uniforme sur } [0, 1]^d.$$

La loi forte des grands nombres implique alors que, si nous disposons d'une suite $(X_n)_{n \in \mathbb{N}^*}$ indépendante et de même loi uniforme sur $[0, 1]^d$:

$$\frac{f(X_1) + \dots + f(X_n)}{n} \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} I$$

Notre méthode de Monte-Carlo consistera donc à simuler un nombre "suffisamment grand" N de vecteurs aléatoires uniformes sur $[0, 1]^d$ indépendants (ce qui revient à simuler Nd variables aléatoires uniformes sur $[0, 1]$ indépendantes) et à prendre pour approximation :

$$I \approx \frac{f(X_1) + \dots + f(X_N)}{N}$$

Dans la pratique, toute la question est de savoir ce que l'on entend par N "suffisamment grand" !

Si les variables aléatoires $f(X_i)$ sont dans $L^2(\Omega, \mathcal{A}, P)$, ou de façon équivalente si $f \in L^2([0, 1]^d, \mathcal{B}([0, 1]^d), \lambda_d)$, le théorème-limite central nous donne la convergence suivante lorsque N tend vers l'infini :

$$\frac{\sqrt{N}}{\sigma} \left(\frac{f(X_1) + \dots + f(X_N)}{N} - I \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) ,$$

1. quartier de la principauté de Monaco connu pour ses jeux de hasard

où nous avons noté :

$$\sigma^2 = \text{Var}(f(X_i)) = \int_{[0,1]^d} f^2(x) dx - \left(\int_{[0,1]^d} f(x) dx \right)^2 = \int_{[0,1]^d} f^2 d\lambda_d - I^2$$

Nous constatons donc que l'erreur d'approximation commise est d'ordre de grandeur σ/\sqrt{N} .

En dimension $d = 1$, faisons la comparaison avec une méthode numérique classique d'approximation des intégrales, par exemple la *méthode des trapèzes*, qui nous donne l'approximation :

$$\int_{[0,1]} f(x) dx \approx T_N(f) := \frac{f(0) + f(1)}{N} + \frac{\sum_{i=1}^{N-1} f\left(\frac{i}{N}\right)}{N}$$

Sous l'hypothèse $f \in C^2([0, 1], \mathbb{R})$, nous avons la majoration suivante de l'erreur d'approximation :

$$\left| \int_{[0,1]} f(x) dx - T_N(f) \right| \leq \frac{1}{12N^2} \max_{[0,1]} |f''|.$$

Nous constatons donc que cette méthode numérique simple est beaucoup plus efficace que celle de Monte-Carlo lorsque f est régulière, plus précisément de classe C^2 .

En revanche, si f est peu régulière, la méthode des trapèzes risque de fournir une approximation de mauvaise qualité (faire un petit dessin pour voir sur quelle idée intuitive repose cette méthode). Il en va différemment pour la méthode de Monte-Carlo puisque le théorème-limite central n'impose qu'une hypothèse de régularité très faible sur f , qui est supposée borélienne.

Ainsi, dans le cas d'une fonction f très irrégulière, la méthode de Monte-Carlo peut donner une meilleure approximation de l'intégrale qu'une méthode numérique classique.

Plaçons-nous maintenant en dimension $d \geq 2$. Quelle est l'influence de la dimension sur l'efficacité des méthodes numériques et de la méthode de Monte-Carlo ?

On peut dire *grosso modo* qu'une méthode numérique classique d'ordre N calculera les valeurs de la fonction f en N^d points "bien choisis" dans $[0, 1]^d$ puis en déduira une approximation de l'intégrale. Une méthode de Monte-Carlo d'ordre N reposera sur la simulation de Nd variables aléatoires uniformes sur $[0, 1]$ indépendantes, pour obtenir N vecteurs aléatoires uniformes sur $[0, 1]^d$ indépendants. En les valeurs prises par ces vecteurs aléatoires, c'est-à-dire en N points de $[0, 1]^d$, l'algorithme calculera alors les valeurs prises par f puis en fera la moyenne arithmétique.

Nous constatons donc la sensibilité d'une méthode de Monte-Carlo à la dimension d est beaucoup moindre que celle d'une méthode numérique classique. Dans la pratique, dès la dimension 3 ou 4, une méthode de Monte-Carlo peut s'avérer plus efficace qu'une méthode numérique classique.

Terminons cette introduction aux méthodes de Monte-Carlo en notant que l'erreur d'approximation, d'ordre de grandeur σ/\sqrt{N} , sera d'autant plus satisfaisante que la variance σ^2 sera petite.

Il existe donc des techniques, dites de *réduction de variance*, qui consistent à modifier un peu la méthode de Monte-Carlo présentée dans ce paragraphe pour faire chuter la variance des variables aléatoires intervenant dans l'approximation de l'intégrale et donc accélérer la convergence.

14.3 Test du χ^2 d'ajustement.

Nous indiquons ici cette application du TLC pour mémoire car elle sera traitée en détail dans le chapitre suivant.

Chapitre 15

Test du χ^2 d'ajustement.

15.1 Complément de cours.

Nous considérons n répétitions d'une même expérience aléatoire qui produit un résultat dans un ensemble fini, par exemple $\{1, \dots, k\}$.

Nous avons des raisons de supposer que la loi sur $\{1, \dots, k\}$ qui gouverne cette expérience est donnée par $p = (p^1, \dots, p^k)$, avec $p_i \geq 0$ pour tout $1 \leq i \leq k$ et $p_1 + \dots + p_k = 1$, mais nous voudrions confirmer ou infirmer cette hypothèse au regard des valeurs $(x_1, \dots, x_n) = (X_1(\omega), \dots, X_n(\omega))$ observées au cours des n répétitions de l'expérience, c'est-à-dire tester l'*ajustement* de notre modèle à la réalité expérimentale.

Pour ce faire, une première étape consiste à introduire une sorte de distance entre les lois de probabilité sur l'ensemble $\{1, \dots, k\}$, l'idée étant de regarder ensuite si la loi empirique (définie ci-dessous et calculable à partir du résultat de l'expérience) est proche ou éloignée de la loi théorique p dont nous avons fait l'hypothèse.

Définition 15.1.1 Soient p et q deux lois de probabilité sur $\{1, \dots, k\}$. On appelle "distance" du χ^2 entre p et q la quantité :

$$d_{\chi^2}(p, q) = \sum_{i=1}^k \frac{(p^i - q^i)^2}{p^i}$$

Le mot distance est entre guillemets car ce n'est pas du tout une distance au sens des espaces métriques : elle n'est visiblement pas symétrique et il est facile de constater qu'elle ne vérifie pas non plus l'inégalité triangulaire. En réalité, son seul rapport avec une distance est la propriété suivante :

$$d_{\chi^2}(p, q) = 0 \Leftrightarrow p = q$$

Remarquons que cette pseudo-distance a tendance à surévaluer les différences entre p et q sur les entiers i où p_i est petit : nous chercherons à limiter ce phénomène dans la suite en imposant des conditions telles que $np_i \geq 5$ pour tout i .

la deuxième étape de construction du test qui va nous permettre de confirmer ou infirmer notre hypothèse consiste à comparer la loi théorique p avec la loi empirique \bar{p}_n que nous définissons comme suit :

Définition 15.1.2 Si (X_1, \dots, X_n) est la variable aléatoire modélisant les n répétitions de notre expérience, nous posons :

$$\forall i \in \{1, \dots, k\}, \quad N_n^i = \sum_{j=1}^n \mathbf{1}_{\{X_j=i\}}$$

Nous appelons alors loi empirique et nous notons \bar{p}_n la loi sur $\{1, \dots, k\}$ définie par :

$$\forall i \in \{1, \dots, k\}, \quad \bar{p}_n^i = \frac{N_n^i}{n}$$

Notons que la valeur de \bar{p}_n dépend du résultat ω de l'expérience, d'où le qualificatif empirique. En toute rigueur, c'est d'ailleurs $\bar{p}_n(\omega)$ (et non pas \bar{p}_n) qui est une loi de probabilité sur $\{1, \dots, k\}$.

Définition 15.1.3 On appelle χ^2 d'ajustement la variable aléatoire suivante :

$$nd_{\chi^2}(p, \bar{p}_n) = n \sum_{i=1}^k \frac{(p^i - \bar{p}_n^i)^2}{p^i} = \sum_{i=1}^k \frac{(np^i - N_n^i)^2}{np^i}$$

Nous rappelons maintenant la définition d'une loi de probabilité sur \mathbb{R} qui a été introduite dans l'Exercice 9.1.1 et identifiée à une loi Gamma, ce qui nous fournit sa densité.

Définition 15.1.4 Considérons $Z = (Z_1, \dots, Z_d)$ un vecteur aléatoire dont les composantes sont i.i.d. de loi commune $\mathcal{N}(0, 1)$.

Alors la loi de la v.a.r. $\|Z\|^2$ est appelée loi du χ^2 à d degrés de liberté et notée $\chi^2(\mathbf{d})$.

Cette loi est en fait égale à la loi $\gamma(\frac{1}{2}, \frac{d}{2})$ et admet donc pour densité :

$$g_{\frac{1}{2}, \frac{d}{2}}(x) = \frac{1}{2^{\frac{d}{2}} \Gamma(\frac{d}{2})} e^{-\frac{x}{2}} x^{\frac{d}{2}-1} \mathbf{1}_{\mathbb{R}_+^*}(x)$$

Le résultat essentiel qui va nous permettre de construire le test dit du χ^2 est le suivant :

Proposition 15.1.5 (Pearson) *Si pour tout $n \in \mathbb{N}^*$, le vecteur aléatoire (X_1, \dots, X_n) suit la loi $p^{\otimes n}$, alors la convergence suivante a lieu :*

$$nd_{\chi^2}(p, \bar{p}_n) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \chi^2(\mathbf{k} - \mathbf{1})$$

La démonstration de ce résultat nécessite la version vectorielle du théorème-limite central, qui fait partie du programme de l'option probabilités. Remarquons simplement que si (X_1, \dots, X_n) suit la loi $p^{\otimes n}$, c'est-à-dire si les variables X_i sont i.i.d. et de loi commune p , alors le théorème-limite central (en dimension 1) implique la convergence suivante :

$$\frac{N_n^i - np^i}{\sqrt{np^i}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1 - p_i)$$

Il est alors assez intuitif que la loi limite du χ^2 d'ajustement, c'est-à-dire de la variable suivante :

$$nd_{\chi^2}(p, \bar{p}_n) = \sum_{i=1}^k \left(\frac{N_n^i - np^i}{\sqrt{np^i}} \right)^2$$

soit une loi du χ^2 mais on pourrait penser que celle-ci a k degrés de liberté, alors que la vraie loi limite est $\chi^2(\mathbf{k} - \mathbf{1})$. En fait, la "perte" d'un degré de liberté peut se comprendre en constatant que nos variables ne sont pas totalement "libres" puisqu'il existe entre elles la relation linéaire suivante :

$$\sum_{i=1}^k (N_n^i - np^i) = 0$$

La proposition de Pearson nous dit donc que, si l'expérience est bien gouvernée par la loi p supposée, alors le χ^2 d'ajustement suit une loi proche de $\chi^2(\mathbf{k} - \mathbf{1})$ lorsque le nombre n de répétitions de l'expérience devient "grand".

En revanche, si l'expérience est en réalité régie par une loi $q \neq p$, alors il existe $1 \leq i \leq k$ tel que $q^i \neq p^i$ et la loi des grands nombres implique la convergence suivante lorsque n tend vers l'infini :

$$d_{\chi^2}(p, \bar{p}_n) = \sum_{i=1}^k \frac{(p^i - \bar{p}_n^i)^2}{p^i} \xrightarrow{\text{p.s.}} d_{\chi^2}(p, q) > 0$$

d'où nous déduisons le comportement asymptotique du χ^2 d'ajustement :

$$nd_{\chi^2}(p, \bar{p}_n) \xrightarrow{\text{p.s.}} +\infty \tag{15.1}$$

C'est la différence entre ces deux comportements asymptotiques qui va nous permettre de tester l'hypothèse H_0 : "L'expérience est gouvernée par la loi p " contre l'hypothèse alternative H_1 : "L'expérience est gouvernée par une loi $q \neq p$ "

Passons à la construction effective du test.

Si nous notons F_{k-1} la fonction de répartition de la loi $\chi^2(\mathbf{k} - \mathbf{1})$, nous prouvons facilement que F_{k-1} est une bijection de \mathbb{R}_+ sur $[0, 1[$ en constatant que la densité de la loi $\chi^2(\mathbf{k} - \mathbf{1})$ est strictement positive sur \mathbb{R}_+^* et nulle sur \mathbb{R}_- . Par conséquent, pour tout $\alpha \in]0, 1]$, il existe un unique $c_\alpha \in \mathbb{R}_+$ tel que $\chi^2(\mathbf{k} - \mathbf{1})(]c_\alpha, +\infty[) = \alpha$ et l'on a $c_\alpha = F_{k-1}^{-1}(1 - \alpha)$.

Si l'hypothèse H_0 est réalisée, donc si l'expérience est régie par la loi p , la proposition de Pearson implique (la loi $\chi^2(\mathbf{k} - \mathbf{1})$ n'admettant pas de masse ponctuelle) que la convergence suivante a lieu lorsque n tend vers l'infini :

$$P_p(nd_{\chi^2}(p, \bar{p}_n) > c_\alpha) \longrightarrow \alpha \quad (15.2)$$

En revanche, si c'est l'hypothèse H_1 qui est réalisée, donc si l'expérience est gouvernée par une loi $q \neq p$, alors le comportement asymptotique (15.1) entraîne la convergence suivante lorsque n tend vers l'infini :

$$P_q(nd_{\chi^2}(p, \bar{p}_n) > c_\alpha) \longrightarrow 1 \quad (15.3)$$

Nous pratiquons donc notre test comme suit :

Nous choisissons une valeur $\alpha \in]0, 1]$ (typiquement α est petit car, comme nous allons le voir, il représente le niveau d'erreur du test) et nous en déduisons la valeur c_α . Pour le résultat ω de l'expérience que nous observons, nous calculons la valeur du χ^2 d'ajustement :

$$nd_{\chi^2}(p, \bar{p}_n(\omega)) = \sum_{i=1}^k \frac{(np^i - N_n^i(\omega))^2}{np^i}$$

Nous comparons alors cette valeur à c_α pour conclure :

- Si $nd_{\chi^2}(p, \bar{p}_n(\omega)) > c_\alpha$, alors nous rejetons l'hypothèse H_0 .
- Si $nd_{\chi^2}(p, \bar{p}_n(\omega)) \leq c_\alpha$, alors nous acceptons l'hypothèse H_0 .

De façon générale, lorsque nous pratiquons un test statistique, notre conclusion peut être erronée de deux façons différentes :

Erreur de 1ère espèce : Je rejette l'hypothèse H_0 alors qu'elle est satisfaite en réalité.

Erreur de 2nde espèce : J'accepte l'hypothèse H_0 alors qu'elle n'est pas satisfaite en réalité.

Dans de nombreuses situations pratiques, ces deux types d'erreurs ne sont pas symétriques, l'erreur de première espèce étant considérée comme plus grave que l'autre. Par exemple, si je teste le câble d'un ascenseur supposé pouvoir accueillir 10 personnes (750kg) et si je note M la masse critique à partir de laquelle le câble casse, je choisirai $H_0 = "M \leq 750"$ et $H_1 = "M > 750"$. L'erreur de 1ère espèce conduirait des usagers de l'ascenseur au grand plongeon : c'est ce risque que je veux absolument maîtriser. L'erreur de 2nde espèce conduirait à des réparations inutiles sur l'ascenseur : je veux l'éviter mais elle est moins grave que la première.

Dans le test du χ^2 que nous venons de construire, la convergence (15.2) se traduit comme suit : la probabilité de commettre une erreur de 1ère espèce est asymptotiquement égale à α . On dit qu'on a construit un test de niveau d'erreur asymptotique α (ou de niveau de confiance asymptotique $1 - \alpha$). Quant à la convergence (15.3), sa traduction est plus vague : lorsque n devient grand, la probabilité de commettre une erreur de 2nde espèce devient "petite" mais nous ne maîtrisons pas la vitesse à laquelle cette convergence se produit. On dit que la puissance du test, c'est-à-dire la probabilité de rejeter l'hypothèse H_0 quand elle n'est effectivement pas satisfaite dans la réalité, tend vers 1 lorsque n tend vers l'infini.

Nous terminons cette introduction au test du χ^2 en le généralisant au cas où l'ensemble E des résultats possibles pour l'expérience est infini. Nous pouvons alors adapter notre méthode comme suit :

Nous choisissons une partition finie (E_1, \dots, E_k) de l'ensemble E . Si ν est la loi sur E supposée gouverner l'expérience, nous posons $p^i = \nu(E_i)$ et nous comptons maintenant le nombre de fois où l'on tombe dans la classe E_i au cours des n répétitions de l'expérience :

$$N_n^i(\omega) = \sum_{j=1}^n \mathbf{1}_{\{X_j(\omega) \in E_i\}}$$

Le reste du test se déroule comme précédemment.

Notons qu'avec cette méthode, nous ne testons pas réellement l'adéquation de la loi ν mais uniquement l'adéquation de ses valeurs sur les différentes classes E_i .

Le choix des classes E_i n'est pas du tout innocent et fait partie de l'art du statisticien qui pratique le test du χ^2 . Une règle d'or, liée à la remarque faite à la suite de la définition de la pseudo-distance du χ^2 , est que les effectifs théoriques np_i des différentes classes E_i ne doivent jamais être trop "petits" (noter que np_i est appelé effectif théorique car c'est l'espérance de N^i dans le cas où l'expérience est vraiment gouvernée par la loi ν). Une règle pratique

que l'on peut adopter est d'exiger que la condition suivante soit satisfaite :

$$\forall i = 1, \dots, k, \quad np^i \geq 5$$

D'autres règles pratiques existent : par exemple, on a constaté qu'il était préférable de choisir les classes E_i de sorte que leurs effectifs théoriques respectifs np_i soient à peu près tous équivalents. Nous n'insisterons pas plus sur ce sujet des règles pratiques qui concernent plus les orfèvres du test du χ^2 que les agrégatifs !

Pour des compléments sur le test du χ^2 , par exemple dans le cas de l'ajustement à une famille de lois, le lecteur peut consulter le tome 1 du livre de cours "Probabilités et statistiques" par Dacunha-Castelle et Duflo, éditions Masson, chapitre 5 intitulé 'Echantillons gaussiens, régression et analyse de la variance', paragraphe intitulé 'Le test du χ^2 ' (pages 135 à 137 dans la 2ème édition).

15.2 Agrégation de classes.

On parle d'agrégation de deux classes i et j lorsqu'on les réunit pour n'en faire plus qu'une seule. Montrer que l'agrégation diminue le χ^2 d'ajustement à une loi p donnée.

15.3 Pile ou face biaisé ou non ?

On effectue 200 jets d'une pièce de monnaie et l'on obtient 110 piles. Tester l'hypothèse H_0 : "La pièce est équilibrée." par un χ^2 d'ajustement de niveau asymptotique 5% .

Reprendre l'exercice avec 2000 jets donnant lieu à 1100 piles.

15.4 Avec une loi non discrète.

Après 1000 répétitions d'une expérience, on obtient la répartition suivante dans différentes classes :

<i>Classe</i> :	[0; 0,5]	[0,5;1]	[1; 1,5]	[1,5;2]	[2; 2,5]	[2,5;3]	[3; 3,5]	[3,5;4]
<i>Effectif</i> :	197	220	112	115	71	94	61	45

<i>Classe</i> :	[4; 4,5]	[4,5;5]	[5; 5,5]	[5,5;6]
<i>Effectif</i> :	36	24	9	16

Appliquer un test d'ajustement de niveau voisin de 1% pour la loi ν définie ainsi : soit X une v.a. uniforme sur $\{0, \dots, 6\}$ et Y une v.a. indépendante de X et uniforme sur $[0,1]$; alors ν est la loi de XY .

15.5 Test du χ^2 d'indépendance.

Un examen est ouvert à des candidats qui ont suivi des filières différentes : économie, informatique, mathématiques. On désire savoir si le choix d'une filière par un étudiant influe sur sa réussite à cet examen. Pour cela, on dispose des résultats obtenus l'année précédente par 286 étudiants d'origines diverses; on les a regroupés dans le tableau suivant :

	Eco	Info	Maths
Succès	41	59	54
Echec	21	36	75

Avec un niveau de confiance voisin de 5%, quelle est votre conclusion ?

Chapitre 16

Complément de cours sur les vecteurs gaussiens

La notion de vecteurs gaussiens ne fait pas partie du programme officiel du tronc commun ; elle n'apparaît que dans le programme de l'option Probabilités et Statistiques. Néanmoins, un nouveau titre est apparu en 2005 parmi les oraux d'Analyse et Probabilités : "Variables gaussiennes. Applications". Il paraît difficile de concevoir cette leçon en se limitant au cas de la dimension 1, qui n'est pas assez substantiel ; le but de ce complément de cours est donc de vous donner les notions essentielles relatives aux vecteurs gaussiens en dimension d .

16.1 Le cas de la dimension 1 : variable aléatoire réelle gaussienne (rappels)

- On dit que la variable aléatoire réelle (v.a.r.) X suit la loi gaussienne centrée réduite et l'on note $X \sim \mathcal{N}(0, 1)$ si X admet la densité $\frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$; on vérifie facilement qu'on a bien $E[X] = 0$, $\text{Var}X = 1$.
- On définit alors la loi gaussienne de paramètres $(m, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$ comme celle de la variable $Y = m + \sigma X$.
Si $\sigma \neq 0$, il est équivalent de dire qu'il s'agit de la loi de densité

$$g(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-m)^2}{2\sigma^2}\right)$$

mais notez bien que l'on autorise le cas $\sigma = 0$: une variable constante est considérée comme gaussienne (dite dégénérée).
On a bien sûr $E[Y] = m$, $\text{Var}Y = \sigma^2$.

FIGURE 16.1 – Graphe de la densité gaussienne centrée réduite.

- La fonction caractéristique de la loi $\mathcal{N}(m, \sigma^2)$ vaut

$$\phi_Y(t) = \exp(itm - \frac{t^2\sigma^2}{2}).$$

En particulier, nous pouvons en déduire la propriété suivante : si Y_1, Y_2 sont des variables gaussiennes indépendantes de lois respectives $\mathcal{N}(m_1, \sigma_1^2)$, $\mathcal{N}(m_2, \sigma_2^2)$ alors $Y_1 + Y_2$ est encore une variable gaussienne, de loi $\mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

- Les lois gaussiennes occupent une place centrale dans la théorie des probabilités à cause du célèbre théorème suivant :

Théorème 16.1.1 (Théorème-limite central) Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite de v.a.r. indépendantes identiquement distribuées (i.i.d.) dans $L^2(\Omega, \mathcal{A}, P)$ (i.e. $E[X_i^2] < +\infty$).

Notons $m = E[X_i]$ leur espérance commune et $\sigma^2 = \text{Var}X_i$ leur variance commune. Posons, pour tout $n \in \mathbb{N}^*$, $S_n = X_1 + \dots + X_n$.

Alors nous avons la convergence en loi suivante :

$$\frac{S_n - nm}{\sqrt{n}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

Ce théorème explique pourquoi de nombreux phénomènes naturels prennent leurs valeurs aléatoirement autour d'une valeur moyenne en suivant une dispersion gaussienne.

16.2 Le cas de la dimension d : vecteur aléatoire gaussien

16.2.1 Définition. Loi gaussienne dans \mathbb{R}^d

- Un vecteur aléatoire (X_1, \dots, X_d) est dit gaussien si toute combinaison linéaire $\sum a_i X_i$ est une v.a.r. gaussienne.

En particulier, les composantes X_i sont gaussiennes mais la réciproque est fautive : voir l'exercice 16.3.1.

Néanmoins, il y a une réciproque partielle : si les X_i sont mutuellement indépendantes et gaussiennes alors le vecteur aléatoire (X_1, \dots, X_d) est gaussien.

- Si X est un vecteur gaussien de dimension d et A une matrice de taille $d' \times d$ alors $Y = AX$ est un vecteur gaussien de dimension d' .
- Pour $X = (X_1, \dots, X_d)$ vecteur aléatoire *quelconque*, on définit son vecteur moyenne :

$$E[X] = (E[X_1], \dots, E[X_d])^*$$

et sa matrice de covariances :

$$K_X = \left[Cov(X_i, X_j) \right]_{1 \leq i, j \leq d}.$$

- Avec les définitions précédentes, la fonction caractéristique d'un vecteur aléatoire gaussien de dimension d vaut :

$$\forall u \in \mathbb{R}^d, \quad \varphi_X(u) = \exp(iu^*E[X] - \frac{1}{2}u^*K_Xu).$$

En particulier, une loi gaussienne vectorielle est déterminée par sa moyenne m et sa matrice de covariances K .

Réciproquement, pour tout $m \in \mathbb{R}^d$ et pour toute matrice K de taille $d \times d$ symétrique semi-définie positive, il existe un vecteur gaussien de moyenne m et de matrice de covariances K .

On peut donc parler de la loi $\mathcal{N}_d(m, K)$.

- Si K est inversible, alors la loi $\mathcal{N}_d(m, K)$ a pour densité :

$$f(x) = (2\pi)^{-d/2} |K|^{-1/2} \exp(-(x - m)^* K^{-1} (x - m) / 2).$$

Sinon, la loi est portée par un sous-espace affine propre de \mathbb{R}^d et ne saurait donc avoir de densité : on parle alors de loi gaussienne dégénérée.

16.2.2 Critère d'indépendance

Théorème 16.2.1 (Indépendance dans le cas gaussien) *Nous supposons que $(X_1, \dots, X_m, Y_1, \dots, Y_n)$ est un vecteur gaussien. Alors il y a équivalence entre les trois propriétés suivantes :*

1. les vecteurs $X = (X_1, \dots, X_m)$ et $Y = (Y_1, \dots, Y_n)$ sont indépendants

2. Pour tous $1 \leq i \leq m, 1 \leq j \leq n$, on a : $\text{Cov}(X_i, Y_j) = 0$
3. La matrice de covariances de (X, Y) est diagonale par blocs :

$$K_{(X,Y)} = \left(\begin{array}{c|c} K_X & 0 \\ \hline 0 & K_Y \end{array} \right).$$

††† C'est le "grand vecteur" (X, Y) qui doit être gaussien et non pas seulement chacun des sous-vecteurs X et Y : voir l'exercice 16.3.1 pour un contre-exemple.

- Tout ce qui vient d'être dit se généralise de 2 blocs à k blocs.
- En particulier, si $X = (X_1, \dots, X_d)$ est un vecteur gaussien, alors on a l'équivalence suivante :
Les v.a.r. X_1, \dots, X_d sont indépendantes **ssi** la matrice de covariances K_X est diagonale.
- Insistons sur un cas encore plus particulier :
Les v.a.r. X_1, \dots, X_d sont i.i.d. de loi commune $\mathcal{N}(0, 1)$ **ssi** $X = (X_1, \dots, X_d)$ est un vecteur gaussien centré et de matrice de covariances I_d .

16.3 Exercices corrigés.

16.3.1 Un contre-exemple.

Soit X une v.a.r. de loi $\mathcal{N}(0, 1)$ et soit $a > 0$. On pose :

$$Y^a = X \mathbf{1}_{\{|X| < a\}} - X \mathbf{1}_{\{|X| \geq a\}}.$$

1. La v.a.r. Y^a est-elle gaussienne ? Le couple (X, Y^a) est-il gaussien ?
2. Montrer qu'il existe $b > 0$ tel que $\frac{1}{\sqrt{2\pi}} \int_0^b t^2 e^{-\frac{t^2}{2}} dt = \frac{1}{4}$.
Calculer $\text{Cov}(X, Y^b)$. Les variables X et Y^b sont-elles indépendantes ?

Corrigé :

1. Pour toute fonction f borélienne bornée, on a :

$$\begin{aligned} E[f(Y^a)] &= E[f(X) \mathbf{1}_{\{|X| < a\}} + f(-X) \mathbf{1}_{\{|X| \geq a\}}] \\ &= \int_{\{|x| < a\}} f(x) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx + \int_{\{|x| \geq a\}} f(-x) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \end{aligned}$$

Le changement de variable $x' = -x$ dans la seconde intégrale nous permet alors d'obtenir :

$$E[f(Y^a)] = \int_{\mathbb{R}} f(x) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx,$$

ce qui prouve que $Y^a \sim \mathcal{N}(0, 1)$.

Puisque $X + Y^a = 2X\mathbf{1}_{\{|X| < a\}}$ et $X \neq 0$ *p.s.*, on a :

$$P(X + Y^a = 0) = P(|X| \geq a) \in]0, 1[.$$

On en conclut que $X + Y^a$ ne peut être une v.a.r. gaussienne et donc que (X, Y^a) n'est pas un couple gaussien.

2. L'application définie sur $[0, +\infty[$ par $G(u) = \frac{1}{\sqrt{2\pi}} \int_0^u t^2 e^{-\frac{t^2}{2}} dt$ est continue et strictement croissante, donc bijective sur son image $[0, \frac{1}{2}[$ (en effet, la limite à l'infini vaut $\frac{1}{2}E[X^2] = \frac{1}{2}$); on prend $b = G^{-1}(\frac{1}{4})$. Notons qu'on a alors :

$$\frac{1}{\sqrt{2\pi}} \int_0^b t^2 e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} \int_b^{+\infty} t^2 e^{-\frac{t^2}{2}} dt.$$

On calcule d'autre part :

$$\begin{aligned} \text{Cov}(X, Y^b) &= E[XY^b] = E[X^2\mathbf{1}_{\{|X| < b\}} - X^2\mathbf{1}_{\{|X| \geq b\}}] \\ &= \frac{2}{\sqrt{2\pi}} \int_0^b t^2 e^{-\frac{t^2}{2}} dt - \frac{2}{\sqrt{2\pi}} \int_b^{+\infty} t^2 e^{-\frac{t^2}{2}} dt. \end{aligned}$$

D'après la remarque précédente, cette covariance est donc nulle; néanmoins X et Y^b ne sont pas indépendantes puisque $|X| = |Y^b|$ *p.s.*

16.3.2 Indépendance gaussienne : un exemple très simple.

Soit (X, Y) un vecteur gaussien de matrice de covariance :

$$K = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad \rho \in [0, 1].$$

Montrer que $X + Y$ et $X - Y$ sont deux v.a. gaussiennes indépendantes.

Corrigé succinct : Le couple $(X + Y, X - Y)$ est gaussien en tant que transformé linéaire du couple gaussien (X, Y) .

Or, on a : $\text{Cov}(X + Y, X - Y) = \text{Var}X - \text{Var}Y = 0$, ce qui entraîne l'indépendance.

16.3.3 Théorème de Student.

Soit (X_1, \dots, X_n) un vecteur aléatoire de loi $\mathcal{N}(0, Id_n)$. On pose :

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad R_n = \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Montrer que $\bar{X}_n \sim \mathcal{N}(0, \frac{1}{n})$, $R_n \sim \chi^2(n-1)$ et que ces deux variables sont indépendantes.

Indication : La transformée de Laplace de la loi $\chi^2(n)$ est définie pour $t > -\frac{1}{2}$ et vaut $(\frac{1}{1+2t})^{n/2}$.

Corrigé succinct : Le vecteur aléatoire $(\bar{X}_n, X_1 - \bar{X}_n, \dots, X_n - \bar{X}_n)$ est gaussien en tant que transformé linéaire d'un vecteur gaussien. Or, pour tout $1 \leq k \leq n$, on a :

$$\text{Cov}(\bar{X}_n, X_k - \bar{X}_n) = \text{Cov}(\bar{X}_n, X_k) - \text{Var} \bar{X}_n = \frac{1}{n} - \frac{1}{n} = 0,$$

d'où l'indépendance entre \bar{X}_n et $(X_1 - \bar{X}_n, \dots, X_n - \bar{X}_n)$. On en déduit immédiatement l'indépendance demandée.

Notant maintenant que $R_n + n\bar{X}_n^2 = \sum_{i=1}^n X_i^2$, nous obtenons la relation suivante entre transformées de Laplace :

$$\mathcal{L}_{R_n}(t) \times \mathcal{L}_{n\bar{X}_n^2}(t) = \mathcal{L}_{\sum_{i=1}^n X_i^2}(t), \quad t > -\frac{1}{2}$$

d'où

$$\mathcal{L}_{R_n}(t) = \frac{(\frac{1}{1+2t})^{\frac{n}{2}}}{(\frac{1}{1+2t})^{\frac{1}{2}}} = \left(\frac{1}{1+2t}\right)^{\frac{n-1}{2}}$$

et donc $R_n \sim \chi^2(n-1)$.

Remarque : On définit la loi de Student $T(k)$, où k est un paramètre entier non nul, comme celle de la v.a.r. $\frac{X}{\sqrt{\frac{Y}{k}}}$, avec $X \sim \mathcal{N}(0, 1)$ et $Y \sim \chi^2(k)$ indépendantes.

Définissons la variance empirique de l'échantillon (X_1, \dots, X_n) comme la variable $S_n^2 = \frac{1}{n-1} R_n$; alors, d'après ce qui précède, la loi de $\frac{\sqrt{n}\bar{X}_n}{S_n}$ vaut $T(n-1)$.

Si l'on remplace notre n -échantillon de loi $\mathcal{N}(0, 1)$ par un n -échantillon de loi $\mathcal{N}(m, \sigma^2)$, on constate facilement que la variable $\frac{\sqrt{n}(\bar{X}_n - m)}{S_n}$ suit encore la loi $T(n-1)$. En statistique, cela permet en particulier de construire un intervalle de confiance pour m lorsque σ est inconnue [DAC¹ cours 122].

16.3.4 Limite en loi d'une suite gaussienne.

1. Montrer que toute limite en loi d'une suite de variables aléatoires gaussiennes est une variable aléatoire gaussienne de moyenne (resp. variance) la limite des moyennes (resp. variances).
2. En déduire que si (X_n) est un processus gaussien indexé par \mathbb{N}^* et tel que $X_n \xrightarrow{(P)} X$, alors on a : $X_n \xrightarrow{(L^2)} X$.

Corrigé :

1. Considérons une suite $(X_n)_{n \in \mathbb{N}^*}$ qui converge en loi vers une v.a.r. X et supposons que pour tout $n \in \mathbb{N}^*$, $X_n \sim \mathcal{N}(m_n, \sigma_n^2)$.
On a alors, pour tout $t \in \mathbb{R}$:

$$\varphi_{X_n}(t) = \exp(itm_n - \frac{t^2\sigma_n^2}{2}) \longrightarrow \varphi_X(t). \quad (16.1)$$

En particulier, on a $|\varphi_X(t)| = \lim \exp(-\frac{t^2\sigma_n^2}{2})$. Or φ_X est continue et telle que $\varphi_X(0) = 1$, ce qui nous permet de choisir $t_0 \neq 0$ tel que $\varphi_X(t_0) \neq 0$; passant au logarithme dans l'égalité précédente prise en $t = t_0$, nous en déduisons que la suite (σ_n^2) est convergente. Notons σ^2 sa limite.

D'après ce que l'on vient de montrer, la suite $\exp(\frac{t^2\sigma_n^2}{2})\varphi_{X_n}(t) = \exp(itm_n)$ est convergente pour tout $t \in \mathbb{R}$; nous allons en déduire que la suite (m_n) est convergente.

Pour cela, posons $\underline{m} = \underline{\lim} m_n$ et $\overline{m} = \overline{\lim} m_n$; dans un premier temps, nous allons montrer par l'absurde que $\overline{m} < \infty$. Si ce n'était pas le cas, on pourrait extraire une sous-suite (m_{n_k}) qui diverge vers $+\infty$.

En tout point $a \in \mathbb{R}$ tel que $P(X = a) = 0$, on aurait alors :

$$P(X_{n_k} \leq a) \longrightarrow P(X \leq a).$$

Le membre de gauche étant égal à $P(m_{n_k} + \sigma_{n_k}Y \leq a)$, avec $Y \sim \mathcal{N}(0, 1)$, on en déduit que $P(X \leq a) = 0$.

Comme a peut être choisi arbitrairement grand, cela contredit le fait que la fonction de répartition de X a pour limite 1 en $+\infty$.

On a donc $\overline{m} < \infty$ et, par un raisonnement symétrique, $\underline{m} > -\infty$.

La convergence de la suite $(\exp(itm_n))$ implique alors l'égalité :

$$\exp(it\underline{m}) = \exp(it\overline{m}).$$

On en déduit que, pour tout $t \in \mathbb{R}$,

$$t(\overline{m} - \underline{m}) \equiv 0 \pmod{2\pi},$$

ce qui n'est bien sûr possible que si $\overline{m} = \underline{m}$.

Ainsi, la suite (m_n) est convergente et nous noterons sa limite m .

La convergence (16.1) nous donne alors immédiatement, pour tout $t \in \mathbb{R}$, $\varphi_X(t) = \exp(itm - \frac{t^2\sigma^2}{2})$, ce qui prouve que $X \sim \mathcal{N}(m, \sigma^2)$.

2. On a les implications :

$$X_n \xrightarrow[n \rightarrow +\infty]{(P)} X \implies X_n - X \xrightarrow[n \rightarrow +\infty]{(P)} 0 \implies X_n - X \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} 0.$$

Pour $n \in \mathbb{N}^*$ fixé, la v.a.r. $X_n - X$ est limite en probabilité (et donc en loi) de la suite $(X_n - X_k)_{k \in \mathbb{N}^*}$; or cette suite est gaussienne puisque (X_n) est un processus gaussien indexé par \mathbb{N} . D'après la question précédente, on en déduit que la v.a.r. $X_n - X$ est gaussienne.

Comme $n \in \mathbb{N}^*$ était arbitraire, on peut de nouveau utiliser la question précédente pour dire que $(X_n - X)_{n \in \mathbb{N}^*}$ est une suite de gaussiennes tendant vers 0, donc que :

$$E[X_n - X] \longrightarrow 0 \text{ et } \text{Var}(X_n - X) \longrightarrow 0.$$

Il suffit alors décrire l'égalité :

$$E[(X_n - X)^2] = (E[X_n - X])^2 + \text{Var}(X_n - X)$$

pour en déduire que $X_n \xrightarrow{(L^2)} X$.

Table des matières

1	Le modèle additif	5
1.1	Préliminaire : structure de clan	5
1.1.1	Définition	5
1.1.2	Clans engendrés	6
1.2	Un modèle à trois ingrédients.	6
1.2.1	L'ensemble des résultats possibles.	6
1.2.2	L'ensemble des évènements.	7
1.2.3	La probabilité	8
1.2.4	Le modèle additif	9
1.3	Espaces de probabilités finis	9
1.3.1	Description d'une probabilité quelconque	9
1.3.2	La probabilité uniforme	11
1.4	Probabilités conditionnelles et indépendance	14
1.4.1	Définition des probabilités conditionnelles.	14
1.4.2	Formule de Bayes.	18
1.5	Indépendance d'évènements.	19
2	Espaces de probabilités finis.	23
2.1	Algèbres de parties	23
2.1.1	Indicatrices	23
2.1.2	Algèbre engendrée par une partition.	23
2.1.3	Réunions d'algèbres.	23
2.1.4	*Description de l'algèbre engendrée.	24
2.2	Formule de Poincaré.	24
2.2.1	Démonstration par récurrence.	24
2.2.2	Application : Indicatrice d'Euler	24
2.2.3	Application : Facteur distract.	25
2.3	Equiprobabilité.	25
2.3.1	Trois dés.	25
2.3.2	Happy birthday to you!	25
2.3.3	Contrôle de fabrication.	25

2.3.4	Boules et cases.	26
3	Probab cond. évènements indépendants.	27
3.1	Probabilités conditionnelles.	27
3.1.1	Maladie rare.	27
3.1.2	Le rouge et le noir.	27
3.1.3	Loi de succession de Laplace.	27
3.2	évènements indépendants.	28
3.2.1	Évènement auto-indépendant.	28
3.2.2	Bruit qui court.	28
3.2.3	Indicatrice d'Euler (bis).	28
3.2.4	Loi de Hardy-Weinberg.	29
4	Complément de cours : Modèle définitif.	31
4.1	Espaces de probabilités.	31
4.1.1	Le modèle probabiliste dans le cas général.	31
4.1.2	Limites supérieure et inférieure. Lemme de Borel-Cantelli.	33
4.2	Variables aléatoires.	35
4.2.1	Définition et premières propriétés	35
4.2.2	Classes d'équivalence de variables aléatoires pour l'égalité presque sûre	36
4.2.3	Loi d'une variable aléatoire	36
4.2.4	Lois de probabilités discrètes	37
4.2.5	Exercice proposé : Taux de panne	38
5	F.g. Loi binomiale. Loi de Poisson.	39
5.1	Problème sur les fonctions génératrices :	39
5.2	Loi binomiale. Loi de Poisson.	42
5.2.1	Stone-Weierstrass par les polynomes de Bernstein.	42
5.2.2	*Théorème de Raikov.	42
5.2.3	Approximation d'une binomiale par Poisson.	44
5.3	Théorème de Liouville.	44
6	Théorème des classes monotones.	47
6.1	Démonstration du théorème par étapes.	47
6.1.1	Ce qui manque à une classe monotone pour être une tribu.	47
6.1.2	Classe monotone engendrée.	47
6.1.3	Théorème des classes monotones.	47
6.2	Applications.	48
6.2.1	Egalité de lois.	48

6.2.2	Critère d'indépendance.	48
7	Outils pour le calcul d'une loi (Cours)	51
7.1	Fonctions de répartition.	51
8	Outils pour le calcul d'une loi.	55
8.1	Fonctions de répartition.	55
8.1.1	Espérance et f.r.	55
8.1.2	Extrema de variables exponentielles.	55
8.1.3	Transformation par quantile.	56
8.1.4	Variables amnésiques.	56
8.2	Lois marginales.	57
8.2.1	Calcul avec une densité particulière.	57
8.2.2	Densité gaussienne en dimension 2.	57
8.3	Changement de variables.	57
8.3.1	Loi du χ^2 à un degré de liberté.	57
8.3.2	Loi gaussienne dans \mathbb{R}	57
8.3.3	Avec une loi de Cauchy.	57
8.3.4	Simulation d'une loi gaussienne dans \mathbb{R}^2 par la méthode de Box-Müller.	58
9	Variables aléatoires indépendantes.	59
9.1	Cas où il existe une densité (changement de variable).	59
9.1.1	Lois Gamma et Bêta	59
9.1.2	Gauss et Cauchy.	60
9.1.3	Avec des lois exponentielles.	60
9.1.4	Densité d'un couple non indépendant.	61
9.1.5	Produit de variables de Cauchy.	61
9.2	Cas général.	61
9.2.1	Diagonale négligeable.	61
9.2.2	V.a.r. auto-indépendante.	61
9.2.3	Statistique d'ordre.	61
9.2.4	Processus de Poisson.	62
10	Fonctions caractéristiques (cours).	63
10.1	Fonction caractéristique d'une variable aléatoire.	63
10.2	Produit de convolution et transformée de Fourier.	65
10.2.1	Définition.	65
10.2.2	F.c. d'une somme de v.a. indépendantes	65
10.3	Caractérisation de l'indépendance	67

11 Fonctions caractéristiques.	69
11.1 Propriétés générales.	69
11.1.1 Dérivabilité des f.c.	69
11.1.2 F.c. périodiques.	69
11.1.3 Loi symétrique et f.c.	69
11.2 F.c. de lois particulières.	70
11.2.1 Quelques f.c. de lois "classiques".	70
11.2.2 Développement en base 2.	70
11.2.3 Laplace et Cauchy.	70
11.2.4 Convolées de lois.	70
11.2.5 Une caractérisation de l'indépendance.	71
11.2.6 Une caractérisation de la loi de Gauss.	71
12 Cvg dans L^p. Cvg en probabilité.	73
12.1 Moments d'une v.a.r. Convergence dans L^p	73
12.1.1 Espérances et variances de lois "classiques".	73
12.1.2 Produit infini.	73
12.1.3 Une CNS de convergence dans L^2	74
12.1.4 Séries convergentes dans L^2	74
12.2 Convergence en probabilité.	74
12.2.1 Métrique de la convergence en probabilité.	74
12.2.2 où l'on revisite Lebesgue.	74
13 Cvg en loi. Théorème-limite central.	77
13.1 Convergence en loi.	77
13.1.1 Le cas discret.	77
13.1.2 Probabilité d'un intervalle.	77
13.1.3 Théorème de Scheffé.	78
13.1.4 Moyenne empirique de Cauchy.	78
13.1.5 Avec des variables uniformes.	78
13.1.6 Vrai ou faux?	79
13.1.7 Une moyenne empirique paramétrée.	79
13.1.8 Suite gaussienne.	80
13.1.9 Suite totale dans \mathcal{C}_0	80
13.2 Théorème-limite central.	80
13.2.1 Coups de dés.	80
13.2.2 Approximation de la loi binomiale par la loi de Gauss	80
13.2.3 Une caractérisation de la loi de Gauss.	81
13.2.4 Une formule de Bernstein.	81
13.2.5 *TCL pour sommes aléatoires de variables i.i.d.	81
13.2.6 *Théorème de Bochner.	81

14 Complément de cours : Applications du TLC	83
14.1 Intervalles de confiance	85
14.1.1 Quelques notions de statistique	85
14.1.2 L'estimateur moyenne empirique	86
14.1.3 Intervalles de confiance	87
14.2 Méthodes de Monte-Carlo	89
14.3 Test du χ^2 d'ajustement.	91
15 Test du χ^2 d'ajustement.	93
15.1 Complément de cours.	93
15.2 Agrégation de classes.	98
15.3 Pile ou face biaisé ou non ?	98
15.4 Avec une loi non discrète.	98
15.5 Test du χ^2 d'indépendance.	99
16 Complément de cours sur les vecteurs gaussiens	101
16.1 Le cas de la dimension 1 : variable aléatoire réelle gaussienne (rappels)	101
16.2 Le cas de la dimension d : vecteur aléatoire gaussien	102
16.2.1 Définition. Loi gaussienne dans \mathbb{R}^d	102
16.2.2 Critère d'indépendance	103
16.3 Exercices corrigés.	104
16.3.1 Un contre-exemple.	104
16.3.2 Indépendance gaussienne : un exemple très simple. . .	105
16.3.3 Théorème de Student.	106
16.3.4 Limite en loi d'une suite gaussienne.	107