Advanced probability (preliminary version)

N.C.

November 15, 2021

Contents

I "Static" probability **4**

1 Rappels sur les fondements **5**
1.1 Axiomatique $(\Omega, \mathcal{F}, \mathbb{P})$ . . . 5
1.1.1 Tribus . . . 5
1.1.2 Mesures . . . 6
1.1.3 Mesure de Lebesgue & Axiome du choix . . . 6
1.1.4 Variables aléatoires . . . 7
1.2 Rappel d'intégration . . . 8
1.2.1 Intégration des fonctions réelles . . . 8
1.2.2 Mesures à densité . . . 8
1.2.3 Calculs de lois . . . 9
1.2.4 Moments, Espaces L^p . . . 9
1.3 Indépendance . . . 9
1.3.1 Espaces produit, mesure produit . . . 9
1.3.2 Indépendance . . . 9

2 Convergences **11**
2.1 Convergences de v.a. . . . 11
2.1.1 Convergence en probabilité et presque-sûre . . . 11
2.1.2 Convergence L^p . . . 12
2.1.3 Lois des grands nombres . . . 12
2.2 Convergences des lois . . . 12
2.2.1 Couplage . . . 12
2.2.2 Convergence en loi . . . 12
2.2.3 Cas de $\mathbb{R}$ . . . 13
2.2.4 Théorème central limite . . . 13
2.3 Compléments . . . 13

3 Conditionnement **14**
3.1 Entrée en matière . . . 14

3.2 Construction de l'espérance conditionnelle . . . 14
3.3 Propriétés . . . 15
3.4 Calculs . . . 15

II Processus **16**

4 Généralités sur les processus **17**
4.1 Processus aléatoires . . . 17
4.1.1 Loi d'un processus . . . 17
4.1.2 Filtration, temps d'arrêt . . . 17
4.1.3 Saupoudrage de processus continus . . . 17
4.2 Exemples . . . 17

5 Martingales **18**
5.1 Définitions, exemples . . . 18
5.1.1 Définition . . . 18
5.1.2 Exemples . . . 18
5.2 Théorème d'arrêt . . . 18
5.2.1 Stratégie de mise . . . 18
5.2.2 Applications directes . . . 18
5.3 Convergence des martingales . . . 19
5.3.1 Inégalités maximales . . . 19
5.3.2 Convergence p.s. . . . 19
5.4 Séries aléatoires . . . 19

6 Chaînes de Markov : I espace d'état fini **20**
6.1 Définitions générales . . . 20
6.2 Espace fini et Perron-Frobenius . . . 20

Ce document est un guide de lecture et de révision pour le cours de M1 de probabilités. Il ne peut (et ne doit pas) être considéré comme un support de cours et nous renvoyons aux excellents cours [6, 3, 2, 8, 9, 1, 4, 7] pour avoir les énoncés et les preuves des résultats. On s'appuiera en particulier sur le cours de J.F. Le Gall [6]. Les résultats précédés d'un symbole ☠ ne sont pas abordé en cours, mais sont des prolongements naturels.

Bibliography

[1] P. BARBE AND M. LEDOUX, *Probabilité*, EDP Sciences, 2021.

[2] T. BODINEAU, *Modélisation de phénomènes aléatoires*, Polycopié de cours X, (2020).

[3] T. DUQUESNE, *Théorie de l'intégration*, Polycopié de cours ENS Ulm, (2013-2014).

[4] R. DURRETT, *Probability: theory and examples*, vol. 49, Cambridge university press, 2019.

[5] O. KALLENBERG, *Foundations of Modern Probability*, Springer, New York, second ed., 2002.

[6] J.-F. LE GALL, *Intégration, porbabilités et processus aléatoires*, Polycopié de cours, (2006).

[7] S. M. ROSS AND E. A. PEKÖZ, *A second course in probability*, www. ProbabilityBookstore. com, 2007.

[8] T. TAO, *An introduction to measure theory*, vol. 126, American Mathematical Society Providence, RI, 2011.

[9] D. WILLIAMS, *Probability with martingales*, Cambridge university press, 1991.

Part I :
"Static" probability

Chapter I : Rappels sur les fondements

... c'était le fondement qui lui échappait ! (naissance de Gargantua)

Dans l'axiomatique de Kolmogorov, une variable aléatoire est une fonction mesurable ! Bien que cette construction permet de fonder proprement les bases mathématiques des probabilités, il convient pour l'intuition de considérer que le $\omega \in \Omega$ est vraiment aléatoire. Le Ω est pensé comme une boîte noire, un générateur d'aléa dont "sort" notre variable aléatoire $X \equiv X(\omega)$. Plusieurs boîtes noires, d'entrailles différentes, peuvent générer la même "fonction" aléatoire : on peut par exemple pour générer un pile ou face standard, utiliser une pièce de monnaie, un lancer de dé, ou un nombre aléatoire et uniforme sur $[0,1]$. Mais la *loi* de la variable générée est la même. La mesure de probabilité $\mathbb{P}$ sur les parties de Ω est là pour mesurer la probabilité d'apparition des événements et le but de la théorie des probabilités est de calculer les probabilités d'événements de plus en plus compliqués faisant intervenir possiblement plusieurs variables aléatoires. La plupart du temps, on suppose l'espace probabilisé et ses variables construites, mais nous allons voir que l'étape de la construction, appelée *modélisation* n'est pas toujours aisée mathématiquement : la notion de tribu est là pour restreindre les événements à qui on peut associer une probabilité, sans quoi on aboutirait à des contradictions. Dans ce chapitre nous allons revoir rapidement les bases de théorie de la mesure et d'intégration avec en mire l'application aux probabilités.

1.1 Axiomatique $(\Omega, \mathcal{F}, \mathbb{P})$

On commence par des rappels de théorie de la mesure et d'intégration (les énoncés et définitions principales sont rappelés, les preuves sont uniquement esquissées ou omises). En théorie de la mesure/analyse on utilise plutôt les notations $(E, \mathcal{A}, \mu)$ pour l'espace mesuré et $f, g, ...$ pour les fonctions. En proba/stats les notations courantes sont $(\Omega, \mathcal{F}, \mathbb{P})$ pour l'espace sous jacent et $X, Y, ...$ (en majuscule) pour les variables aléatoires.

1.1.1 Tribus

Tribus (ou σ-algèbre, σ-field en anglais) : définition, insister sur l'union *dénombrable*. L'intersection de tribus est une tribu.
Exemples : $\mathcal{P}(\Omega)$, ensemble des parties dénombrables ou co-dénombrables, tribu engendrée par une partie, tribu borélienne (cas métrique) notée $\mathcal{B}(\mathbb{R}^d)$ dans le cas de $\mathbb{R}^d$.

Ensembles mesurables=événements.
Limsup$(A_n : n \geqslant 1) = \bigcap_{n\geqslant 1} \bigcup_{m\geqslant n} A_n$ et Liminf$(A_n : n \geqslant 1) = \bigcup_{n\geqslant 1} \bigcap_{m\geqslant n} A_n$, interprétation.
Définition et **lemme de classe monotone** [6, Théorème 1.4.1].
Méthode générale pour montrer qu'une tribu satisfît une certaine propriété.

Remark 1.1 (Culture). Notez la similarité de la définition de tribu avec celle d'une topologie. Les topologies sont bien plus faciles à comprendre que les tribus : il est difficile de décrire les boréliens de $\mathbb{R}$ alors qu'il est facile de décrire ses ouverts. En revanche les tribus ne "servent pas vraiment" pour elles-mêmes, mais juste pour définir les probabilités d'ensembles sans avoir de contradiction. On peut se demander aussi ce qu'il reste si l'on laisse tomber l'union dénombrable mais que l'on conserve l'union finie dans la définition de mesure... en gros on repasse de l'intégrale de Lebesgue à celle de Riemann ou l'on perd la stabilité par limite simple qui était bien utile. Voir [8, Chapter 1.1] ou [3, Chapitre 5.2.d et Chapitre 6].

1.1.2 Mesures

Mesures (positives) : definition, σ-additivité.
Exemples : mesure de comptage, Diracs, atomique (fini et dénombrable), diffuse. Mesures finies, σ-finies, mesures de probabilité notées $\mathbb{P}$ et les mesures générales μ.
☠ il faut faire attention dès qu'une mesure n'est pas σ-finie.
Sous une mesure de proba, limite des probabilités d'événements croissants ou décroissants.
Lois usuelles sur $\mathbb{N}$: Bernoulli, Binomiale, Multinomiale $\mathbb{P}(N_1 = n_1, ..., N_k = n_k) = \frac{n!}{n_1!...n_k!} p_1^{n_1} ... p_k^{n_k}$, Géométrique, Poisson, loi de Borel–Tanner (non trivial de vérifier masse 1). **Ensembles négligeables** si inclus dans ensemble mesurable de mesure nulle = Événements presque-sûrs. Presque partout, presque sûrement. Tribu complétée $\overline{\mathcal{F}} = \sigma(\mathcal{F}, \mathcal{N})$ en ajoutant les négligeables, description.
Lemme de Borel–Cantelli sans indépendance.
Application : (anticipe un peu avec la mesure de Lebesgue). Pour tout $\varepsilon > 0$, pour Lebesgue-presque tout $x \in [0,1]$, il n'existe qu'un nombre fini de rationnels p/q avec $p \wedge q = 1$ tels que $|x - \frac{p}{q}| < q^{-2-\varepsilon}$. Mise en perspective avec le résultat de Dirichlet qui dit que si α est irrationnel alors $|\alpha - \frac{p}{q}| < q^{-2}$ pour une infinite de fractions différentes.
Caractérisation des mesures (σ-finies) via des valeurs sur un π-système générateur [6, Corollaire 1.4.2].
Approximation d'événements : si $\mathcal{F} = \sigma(\bigcup_{n\geqslant 1} \mathcal{F}_n)$ avec $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots$ alors pour tout $A \in \mathcal{F}$ et tout $\varepsilon > 0$ on trouve $A_n \in \mathcal{F}_n$ tel que $\mathbb{P}(A \Delta A_n) \leqslant \varepsilon$ (attention, le A_n dépend de la mesure de proba), voir [5, Lemma 3.16]

1.1.3 Mesure de Lebesgue & Axiome du choix

Existence de la mesure de Lebesgue (via Carathéodory) [6, Théorème 3.2.1 et Proposition 3.2.4], régularité [6, Proposition 3.2.7]. Tribu borélienne complétée $\to$ lebesguienne $\mathcal{L}$.
Existence d'un ensemble non mesurable pour Lebesgue : l'ensemble de Vitali [6, Chapitre

3.4] (plus précisément, n'importe quel ensemble de Lebesgue-mesure positive contient un ensemble non Lebesgue-mesurable).

Remark 1.2 (Culture). ☠ Le cardinal de $\mathcal{B}(\mathbb{R})$ est celui de $\mathbb{R}$ (dur, induction transfinie) alors que celui de son complétée est $2^{\mathbb{R}}$. En fait (Solovay, Shelah...), si on voulait vraiment, on pourrait laisser tomber l'axiome du choix et admettre à la place que tous les sous ensembles de $\mathbb{R}$ sont Lebesgue-mesurables... mais l'abandon de l'axiome du choix aurait des conséquences bien plus néfastes comparées à la mesurabilité de tous les sous ensembles de $\mathbb{R}$... Version géométrique : Banach–Tarski [1] Version à raconter en famille : Les nains et leurs chapeaux[2].

Exo : Régularité de la mesure de Lebesgue (mesure de Radon)

$$\sup\{\text{Leb}(K) : K \subset A \text{ compact}\} = \text{Leb}(A) = \sup\{\text{Leb}(O) : A \subset O \text{ ouvert}\}.$$

1.1.4 Variables aléatoires

Fonctions mesurables : définition, composée, **stabilité par limite simple**. Cas où les applications sont sympathiques (continues) et les tribus cibles sont boréliennes. Tribu engendrée.

Remark 1.3. On met toujours la tribu la plus fine au départ et la plus grossière (mais pas trop quand même) à l'arrivée. Car sinon ☠ on peut trouver des fonctions $f : [0,1] \to [0,1]$ de classe $\mathcal{C}^\infty$ qui ne sont pas Lebesgue-Lebesgue mesurables [3].

Approximation : Toute fonction mesurable positive est limite simple de fonctions étagées (partout, pas presque partout). Les fonctions mesurables ne sont pas si loin des fonctions continues : **théorèmes d'Egoroff et de Lusin** (énoncé).
Une variable aléatoire (v.a.) est une fonction mesurable. **Convention :** les v.a. sont notées X, Y, Z... avec des lettres majuscules.
Mesure image, loi. $\mathbb{P}_X$, espace canonique. **Comme on le verra, tous les calculs ne dépendent que de la loi (jointe) des variables aléatoires, pas de la modélisation, on aura donc tendance à "oublier" l'espace** $(\Omega, \mathcal{F}, \mathbb{P})$.
Exemple : Quelques modélisations différentes d'un pile-face.

Dans 99,9% des probabilités modernes, on peut se contenter de prendre $\Omega = [0,1]$, $\mathcal{F} = \mathcal{L}([0,1])$ et $\mathbb{P}$ la mesure de Lebesgue. De même dans 99% des cas les espaces cibles sont des espaces sympathiques comme $\mathbb{R}, \mathbb{R}^d$ ou un espace polonais (métrique séparable et complet) munis de leur tribus boréliennes.

[1] Connaissez-vous un anagramme de Banach–Tarski ? Réponse : Banach–Tarski Banach–Tarski.
[2] https : //math.stackexchange.com/questions/34569/enigma-of-wizards-dwarves-and-hats
[3] Une fonction qui envoie un Cantor de mesure positive – donc qui contient un ensemble non Lebesgue mesurable – sur un Cantor de mesure 0

1.2 Rappel d'intégration

1.2.1 Intégration des fonctions réelles

Intégration des fonctions étagées, positives (comme le sup sur les étagées plus petites), $\mathcal{L}^1$.
Convention : $\int_A \mathrm{d}\mu\infty = 0$ dès que $\mu(A) = 0$.
Théorèmes de convergence monotone, dominée et de Fatou [6, Chapitre 2]. Contre-exemples dès qu'une hypothèse est relâchée.
Exo : Expliciter les théorèmes de convergences dans le cas de $\mathbb{N}$ avec la mesure de comptage et donner des preuves élémentaires des théorèmes de convergence.

Remark 1.4. L'intégrale de Lebesgue (avec la mesure de Lebesgue) coincide avec celle de Riemann pour les fonctions Riemann-intégrables, [6, Chapitre 3.3]. Mais Riemann par stable par limite simple.

Remark 1.5. "Je dois payer une certaine somme ; je fouille dans mes poches et j'en sors des pièces et des billets de différentes valeurs. Je les verse à mon créancier dans l'ordre où elles se présentent jusqu'à atteindre le total de ma dette. C'est l'intégrale de *Riemann*. Mais je peux opérer autrement. Ayant sorti tout mon argent, je réunis les billets de même valeur, les pièces semblables, et j'effectue le paiement en donnant ensemble les signes monétaires de même valeur. C'est mon intégrale." *Lebesgue 1901*.

Espérance et moments de variables aléatoires. Cas des variables discrètes. **Cas des variables à valeurs dans** $\{0, 1, 2, ...\}$ **alors on a** $\mathbb{E}[X] = \sum_{k\geqslant 0} \mathbb{P}(X > k)$.
Application : Si $(X_i : i \geqslant 1)$ entières et de même loi avec $\mathbb{E}[X] < \infty$ alors $\limsup_{n\to\infty} n^{-1} X_n = 0$.

1.2.2 Mesures à densité

Mesures à densité (en général), absolue continuité, mesure étrangère.
Exo : si $\mu << \nu$ et ν est finie, alors $(\forall \varepsilon > 0), (\exists \delta > 0), (\forall A \in \mathcal{F})$ on a $\nu(A) \leqslant \delta \Rightarrow \mu(A) \leqslant \varepsilon$.
Indice : utiliser Borel–Cantelli.
Exemples : Exponentielles, Gaussiennes, Cauchy...
Théorème de Radon–Nikodym (cas des mesures positives σ-finies) voir [6, Théorème 4.4.1].
Cex : $\lambda <<$ Comptage sur $\mathbb{R}$ mais pas de densité.
Cas unidimensionel : Fonction de répartition $\iff$ **mesure de proba sur** $\mathbb{R}$**.** Atome, diffuse, densité. Escalier du diable. Inverse continue à droite et simulation numérique.
☠ **Théorème de différentiation de Lebesgue.**
☠ Comment vraiment générer au hasard ? Ordinateur quantique.

1.2.3 Calculs de lois

"Théorème de transfert" (mesure image) [6, Proposition 8.1.1] pour tout ϕ mesurable positive

$$\mathbb{E}[\phi(X)] = \begin{cases} \int \mathrm{d}\mathbb{P}(\omega)\phi(X(\omega)) \\ \int \mathbb{P}(\mathrm{d}\omega)\phi(X(\omega)) \\ \int \mathrm{d}\mathbb{P}\phi(X) \end{cases} = \begin{cases} \int \mathrm{d}\mathbb{P}_X(x)\phi(x) \\ \int \mathbb{P}_X(\mathrm{d}x)\phi(x) \\ \int \mathrm{d}\mathbb{P}_X\,\phi. \end{cases}$$

On utilise cette formule pour caractériser la loi image et faire des calculs (on peut restreindre la classe de fonctions, par exemple continues à support compact). **Formule de changement de variable pour les variables à densités**. Calculs de lois.
Exemples : $a\mathcal{E}(1)$, $\sigma\mathcal{N}(0,1)+m$, projection de multinomiale, $\mathrm{Poi}(\alpha)+\mathrm{Poi}(\beta)$ indépendants -> plein d'exemples en exo.
☠ Représentation des formes linéaires continues positives sur $\mathcal{C}_c(\mathbb{R}^d)$ [6, Théorème 3.6.1].

1.2.4 Moments, Espaces L^p

Inégalité de Markov (corollaire : Tchebytchev, Markov exponentiel).
Méthode du premier moment on a $\mathbb{P}(X>0) \leqslant \mathbb{E}[X]$.
Application : balls in bins, avec $(1+\varepsilon)n\log n$ boules placées uniformément et indépendamment dans n urnes, aucune vide avec grande proba.
Jensen, Hölder, Minkowski. Espaces $\mathcal{L}^p$, L^p (☠ dualité).
Cas de mesure de proba : $L^p \subset L^q$ si $1 \leqslant p \leqslant q$.
Densité (cas de $\Omega = \mathbb{R}^d$) des fonctions régulières à support compact, [6, Théorème 4.3.1]
Cas $p=2$: espace de Hilbert, Cauchy–Schwarz. Méthode du second moment.
Applications : retour sur balls in bins, avec $(1-\varepsilon)n\log n$ boules, il existe une urne vide avec grand proba.
Produit scalaire et représentation des formes linéaires continues (Riesz).
Retour sur la preuve de Radon-Nikodym.

1.3 Indépendance

1.3.1 Espaces produit, mesure produit

Espace produit, tribu produit, les sections sont mesurables.
Exo : $\mathcal{B}(\mathbb{R})\otimes\mathcal{B}(\mathcal{R}) = \mathcal{B}(\mathbb{R}^2)$ mais $\mathcal{L}(\mathbb{R})\otimes\mathcal{L}(\mathbb{R}) \neq \mathcal{L}(\mathbb{R}^2)$.
Mesure produit. **Théorème de Fubini** (esquisse de démonstration), [6, Théorème 5.3.1].
Attention à l'hypothèse de σ-additivité : $\lambda \otimes$ Comptage ne vérifie pas Fubini.

1.3.2 Indépendance

Indépendance de tribus (d'événements et de v.a). Caractérisation avec la loi jointe et l'espérance de produit de fonctionnelles. Indépendance d'une infinite de tribus.

Application : Markov exponentiel, $\mathbb{P}(X_1 + \cdots + X_n \geqslant (m+\varepsilon)n)^{1/n} \leqslant \inf_\theta \mathbb{E}[e^{\theta X}]e^{-\theta(m+\varepsilon)}$.
Lemme d'indépendance des tribus via classe monotone [6, Proposition 9.2.4].
Regroupement par paquet (possiblement infinis).
Lemme de Borel–Cantelli indépendant, contre-exemples [5, Exo 3.14]. *Applications :*

1. Le singe qui tape Shakespeare, [6, Section 9.3].
2. Il n'existe pas de loi de proba μ sur $\mathbb{Z}_{\geqslant 1}$ tel que $\mu(a \cdot \mathbb{Z}_{\geqslant 1}) = \frac{1}{a}$, voir [6, Section 9.3]
3. (come-back) $(X_i : i \geqslant 1)$ v.a.i.i.d. à valeurs dans $\{0, 1, 2, ...\}$ telles que $\mathbb{E}[X] = \infty$ alors $\limsup n^{-1} X_n = \infty$.

Loi du $0/1$ **Kolmogorov**. *Applications :*

1. Retour sur le singe qui tape Skakespeare, nouvelle preuve.
2. Limsup et liminf de la marche aléatoire simple.
3. ☠ la preuve de Schwartz du théorème de Steinhaus [Zuily-Queffelec].

☠ **Loi du** $0/1$ **de Hewitt-Savage**.
Construction de mesures. Développement dyadique : $U = \sum_{i \geqslant 1} 2^{-i} \epsilon_i$ avec ϵ_i vaiid $\mathbb{P}(\epsilon = 1) = 1 - \mathbb{P}(\epsilon = 0) = \frac{1}{2}$. Suite infinie Pile/Face id. **La multiplication des pains** et construction d'une infinité dénombrable de v.a. de lois données (court-circuit en faisant tout à partir de Lebesgue).

Remark 1.6 (Culture). En général, si $\lambda \in (0, 1)$ le nombre réel $V = \sum_{i \geqslant 1} \lambda^i \epsilon_i$ avec ϵ_i des variables de Bernoulli équilibrées indépendantes est très mal connu : pour quel λ sa loi a une densité? si elle est singulière, quelle est la dimension du support? Ce champ de recherche est appelé "Bernoulli convolution".

Chapter II : Convergences

Nous distinguerons deux types de convergence : les converges "fortes" portant sur les variables aléatoire (en particulier les variables aléatoires X_n on besoin d'être définies sur le même espace de probabilité) et les convergence des lois $\mathbb{P}_{X_n}$ où les variables n'ont pas besoin d'être définies sur le même espace. En particulier, le premier type de convergence dépend de la loi jointe des (X_n), alors que la deuxième ne dépend que de la suite des lois des X_n.

2.1 Convergences de v.a.

2.1.1 Convergence en probabilité et presque-sûre

On ne traitera que le cas des variables à valeurs dans $\mathbb{R}^d$ (muni de la norme usuelle) mais on mentionnera que l'extension au cas métrique (Polonais) ne pose aucune difficulté.

Définition convergence en proba $(\mathbb{P})$ et presque sûre (p.s.).
$p.s. \Rightarrow (\mathbb{P})$ avec contre-exemple pour la réciproque (construction ad-hoc ou suite de Bernoulli de paramètre $1/n$ indépendantes).
Lien de la convergence p.s. avec limsup et lemme de Borel–Cantelli.
Exo : Soit $(X_i : i \geqslant 1)$ des variables réelles quelconques. Montrer qu'il existe $(c_i : i \geqslant 1)$ positives telles que $\sum c_i X_i$ cv p.s.
Métrique $\mathrm{d}_{\mathrm{prob}}(X,Y) = \inf\{\varepsilon > 0 : \mathbb{P}(|X-Y| \geqslant \varepsilon) \leqslant \varepsilon\}$ ou $\tilde{\mathrm{d}}_{\mathrm{prob}}(X,Y) = \mathbb{E}[|X-Y| \wedge 1]$, avec $\mathrm{d}_{\mathrm{prob}}(X,Y) \leqslant \sqrt{\tilde{\mathrm{d}}_{\mathrm{proba}}(X,Y)}$. **Convergence $(\mathbb{P})$ vers X ssi de toute sous suite on peut extraire une sous-suite qui converge p.s. vers X** [5, Lemma 4.2]. **Complétude de** $(L^0, \mathrm{d}_{\mathrm{proba}})$ [5, Lemma 4.6]. Retour sur la complétude de L^p. *Applications :*

1. On peut remplacer la convergence p.s. par une convergence en proba dans le théorème de convergence dominée.

2. Exo : Si $X_n \to X$ et $Y_n \to Y$ en proba, alors $f(X_n) + Y_n \to f(X) + Y$ en proba pour $f : \mathbb{R}^d \to \mathbb{R}^d$ continue.

3. ☠ La convergence p.s. n'est pas métrisable (même pas topologisable): ce n'est qu'une notion de convergence (et c'est bien pratique !).

Retour sur le $0/1$: convergence en proba/p.s. de variables indépendantes.

2.1.2 Convergence L^p

Convergence L^p, cv $(L^p) \Rightarrow (\mathbb{P})$, $\mathrm{d}_{\text{proba}}(X,Y) \leqslant \|X-Y\|_p^{\frac{p}{p+1}}$. Complétude. Contre-exemples.
Uniforme intégrabilité (2 définitions équivalentes): famille $(X_i : i \in I)$ bornée dans L^1 et

$$\lim_{A\to\infty} \sup_{i\in I} \mathbb{E}[|X_i|\mathbf{1}_{|X_i|>A}] = 0 \quad \text{ou de manière équivalente} \quad \lim_{\delta\to 0} \sup_{E:\mathbb{P}(E)\leqslant\delta} \sup_{i\in I} \mathbb{E}[|X_i|\mathbf{1}_E] = 0,$$

voir [6, Proposition 12.5.1]. Exemples : famille finie, union, famille dominée, barycentre, de La Vallée Poussin (sens facile).
Théorème de Convergence dominée ultime : $(\mathbb{P}) + UI \iff L^1$, voir [6, Théorème 12.5.3].
Application : Cv (L^1) et bornée dans L^p implique convergence dans L^q pour tout $1 \leqslant q < p$.
☠ **Compacité $L^1 - L^\infty$ et Dunford–Pettis.**

2.1.3 Lois des grands nombres

Enoncé général avec convergence p.s. dans le cas L^1. Preuve de la loi faible ($\mathbb{P}$) dans le cas L^2, de la loi forte (p.s.) dans le cas L^4 via Borel–Cantelli.
Convergence dans L^1 (la famille $\{\frac{1}{n}(X_1 + ... + X_n) : n \geqslant 1\}$ est uniformément intégrable).
Réciproque à la loi forte : Soit $(X_i : i \geqslant 1)$ v.a.i.id réelles telles que $\frac{1}{n}(X_1 + \cdots + X_n)$ converge p.s. alors $\mathbb{E}[X] < \infty$ et la convergence a lieu vers $\mathbb{E}[X]$.
☠ Réciproque fausse dans le cas de la loi faible avec $\mathbb{P}(X = k) = \mathbb{P}(X = -k) = \frac{c}{k^2 \log k}$.
Applications : Sondages, Bernstein–Weierstrass, equirepartition entropie (**Théorème de Shannon–McMillan–Breiman**).
☠ Généralisation, **théorème ergodique de Birkhoff** (énoncé sans preuve).

2.2 Convergences des lois

2.2.1 Couplage

Définition d'un couplage. Idée : comparer la distribution de X et de Y en construisant des copies $\tilde{X} \sim X$ et $\tilde{Y} \sim Y$ sur un même espace de proba (afin de comparer directement les variables aléatoires comme dans la section précédente). Exemple : couplage indépendant, couplage de Bernoulli avec une uniforme.
En général si on a une métrique δ sur les v.a. on peut définir une métrique Δ sur les lois en posant

$$\Delta(\mu,\nu) = \inf\{\delta(X,Y) : (X,Y) \text{ couplage de } (\mu,\nu)\}.$$

☠ **gluing lemma** avec lois conditionnelles (admis) pour vérifier que c'est une distance.

2.2.2 Convergence en loi

Convergence étroite des mesures et convergence en loi $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ pour $f \in \mathcal{C}^b$.
La convergence en proba implique la convergence en loi.

Exemples : v.a. avec nombre fini de valeurs, Uniformes discrètes $\to$ uniformes, v.a. à densités f_n dont les densités convergent vers une densité limite (lemme de Scheffé).
Pour que $\mu_n \to \mu$ en loi, il faut et il suffit que l'une des propriétés suivantes soit vérifiée :

- $\Delta(\mu_n, \mu) \to 0$ avec Δ la métrique (Lévy-Prokhorov) associée à $\mathrm{d}_{\mathrm{prob}}$,
- $\int \mathrm{d}\mu_n f = \mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)] = \int \mathrm{d}\mu f$ pour tout f continue bornée,
- Portmanteau.

[6, Definition 10.3.1 et Proposition 10.3.2]. Equivalence avec la première propriété admise (voir plus bas dans le cas de $\mathbb{R}$).

2.2.3 Cas de $\mathbb{R}$

Caractérisation de la convergence en loi via les fonction de répartition: $\mu_n \to \mu$ ssi $F_{\mu_n} \to F_\mu$ simplement sur un sous ensemble dense de $\mathbb{R}$.
Fonction caractéristique, caractérisation de la convergence en loi **théorème de Lévy** (cas de $\mathbb{R}^d$, preuve dans le cas de $\mathbb{R}$), [6, Théorème 10.3.4]
Exo : Preuve de la loi *faible* des grands nombres avec les fonctions caractéristiques.
☠ Prokhorov, tension via les fonctions de répartition (théorème de Helly).
☠ Théorème de représentation de Skorokhod en couplant avec la même uniforme.

2.2.4 Théorème central limite

Théorème Central limite cas $d = 1$ et preuve en $d = 1$.
De Moivre–Laplace pour les variables de Bernoulli.
Il n'y a pas convergence en probabilité dans le théorème central limite et on a même

$$\liminf_{n\to\infty} \frac{X_1 + ... + X_n}{\sqrt{n}} = -\infty \quad \text{et} \quad \limsup_{n\to\infty} \frac{X_1 + ... + X_n}{\sqrt{n}} = \infty.$$

Application : $\lim_{n\to\infty} \mathrm{e}^{-n} \sum_{k=0}^{n} \frac{n^k}{k!}$, intervalles de confiance asymptotiques.

2.3 Compléments

Distance en variation totale. Si on munit l'espace cible de la distance triviale $d(x, y) = \infty \mathbf{1}_{x\neq y}$, on peut définir une notion de convergence en probabilité et une notion de convergence p.s. (égale à partir d'un certain rang). Tout marche pareil (attention, on n'est généralement plus séparable). On peut définir la convergence en loi associée, et maintenant les fonctions continues sont simplement les fonctions boréliennes. On obtient une distance super forte : la distance en variation totale

$$\mathrm{d}_{\mathrm{TV}}(\mu, \nu) = \sup_{A\in\mathcal{F}} |\mu(A) - \nu(A)| = \inf_{\substack{(X,Y)\text{couplage} \\ X\sim\mu \ \& \ Y\sim\nu}} \mathbb{P}(X \neq Y) \underset{\text{cas densite}}{=} \frac{1}{2} \int \mathrm{d}\mu |f - g|.$$

Exemple : Binomiale-Poisson.

Chapter III : Conditionnement

... Modern probability theory can be said to begin with the notion of conditioning [5, Chap. 6]

3.1 Entrée en matière

Conditionnement cas discret : Bayes, loi conditionnelle $\mathbb{P}(Y = y \mid X = x)$ qui permet d'écrire $\mathbb{P}_{(X,Y)}(\mathrm{d}x\mathrm{d}y) = \mathbb{P}_X(\mathrm{d}x)\mathbb{P}(\mathrm{d}y \mid x)$. Simulation "en deux temps". **Cas à densité**. Existence loi conditionnelle cas général (cas Polonais, admis), [6, Chapitre 11.5]: la loi conditionnelle de Y sachant X est une famille de probabilités $\nu(x, \mathrm{d}y)$ définies à égalité pour $\mathbb{P}_X$-pp x tels que

$$\mathbb{E}[\phi(X,Y)] = \int \mathrm{d}\mathbb{P}_X(x) \int \nu(x,\mathrm{d}y)\phi(x,y).$$

Définition de l'espérance conditionnelle :

$$\mathbb{E}[Y \mid X] := \int \nu(X,\mathrm{d}y)\, y,$$

défini presque partout. Si Y est réelle (ou à valeurs dans un polonais) et $\sigma(X)$-mesurable, alors c'est une fonction mesurable de X, [6, Proposition 8.1.3] ou [5, Lemma 1.13]. **Propriété caractéristique** pour tout ϕ mesurable positive

$$\mathbb{E}[\phi(X)Y] = \mathbb{E}[\phi(X)\mathbb{E}[Y \mid X]],$$

conditionnement par rapport à une tribu.

3.2 Construction de l'espérance conditionnelle

Définition via propriété caractéristique, unicité presque sûre.
Construction pour $Y \in L^1$ via Radon–Nikodym [6, Théorème 11.2.1].
Interprétation comme projection dans le cas L^2.
Cas des variables positives (pas forcément dans L^1).
Exemple dans le cas d'une tribu engendrée par un nombre fini d'événements.

3.3 Propriétés

Voir [6, Chapitre 11].
Linéarité et croissance.
Jensen, Cv monotone, dominée et Fatou conditionnels.
Propriété de la tour (tower property).
Cas indépendant.

3.4 Calculs

Retour sur le cas discret, à densité, cas indépendant.
Conditionnement gaussien: Définition d'un **Vecteur gaussien (dimension finie)**.
Matrice de covariance, fonction caractéristique, construction.
Retour sur le théorème central limite cas vectoriel.
Conditionnement gaussien.

Part II : Processus

Chapter IV : Généralités sur les processus

4.1 Processus aléatoires

4.1.1 Loi d'un processus

Cas des processus à temps discret. Tribu cylindrique, loi fini-dimensionnelles.
Théorème d'extension de Kolmogorov (admis).

4.1.2 Filtration, temps d'arrêt

Filtration, processus adapté.
Exemple : filtration engendrée, filtration dyadique.
Temps d'arrêt, opérations sur les temps d'arrêt. Tribu du passé avant T notée $\mathcal{F}_T$.
Si $\mathbb{P}(T \leqslant n + n_0 \mid \mathcal{F}_n) > \varepsilon$ alors $T < \infty$ p.s.

4.1.3 Saupoudrage de processus continus

Notion de processus à temps continu: Tribu cylindrique (version coordonnées), tribu borélienne pour la convergence uniforme sur tout compact (mais pas pour la convergence uniforme partout, ce ne serait pas séparable).

Remark 4.1. ☠ Contre-exemple : L'événement $\{f : \mathbb{R} \to \mathbb{R} \text{ est continue}\}$ n'est pas mesurable pour la tribu cylindrique sur $\mathbb{R}^{\mathbb{R}}$. Les problèmes de mesurabilité sont plus des problèmes de topologie que des problème de théorie de la mesure.

4.2 Exemples

Pas de résultats, juste des exemples !

Marche aléatoire simple sur $\mathbb{Z}$. Cas symétrique, cas non symétrique.
Urnes de Polya, d'Ehrenfest.
Processus de Bienaymé–Galton–Watson.
Problème des secrétaires, inégalité prophète.
Processus de Poisson standard.
Notion de mouvement Brownien.

Chapter V : Martingales

5.1 Définitions, exemples

5.1.1 Définition

Définition martingale, surmartingale, sousmartingale.
Interprétation en terme de jeu et d'information.
Quantité conservée : l'espérance.
Jensen et martingale.

5.1.2 Exemples

Marches aléatoires, martingales fermés, produit de variables positives.

5.2 Théorème d'arrêt

5.2.1 Stratégie de mise

Transformation de (sur-sous)martingale M avec une stratégie prévisible H (et bornée) qui amène à $H \bullet M$. Cas des temps d'arrêt. Interprétation en termes de jeu -> on ne bat pas le casino. **Théorème d'arrêt** (uniquement avec $T \wedge n$)

$$\mathbb{E}[M_{n\wedge T}] = \mathbb{E}[M_0].$$

Contre exemple quand le temps d'arrêt n'est pas borné.

5.2.2 Applications directes

- **Ruine du joueur**, proba de sortie de $[a, b]$ pour une marche symétrique, puis non symétrique. Ruine, suite. Temps de sortie : Wald via la martingale $S_n^2 - n$ et calcul de la loi exacte de T_{-1} avec les martingales exponentielles.
- Trouver deux as consécutifs dans un tas de carte.
- Le singe qui tape ABRACADABRA.
- Le problème des secrétaires.
- Contrôle des moutons de Mabinogion.

5.3 Convergence des martingales

À partir d'ici je ne m'occupe que des martingales (pas de sur ou sous-martingales).

5.3.1 Inégalités maximales

Inégalité maximale de Doob pour une martingale $(M_n : n \geqslant 0)$ pour $a > 0$

$$a \cdot \mathbb{P}(\sup_{0 \leqslant k \leqslant n} M_n \geqslant a) \leqslant \mathbb{E}[|M_n|].$$

Cas des martingales dans L^p avec $\mathbf{p > 1}$

$$\mathbb{E}[\sup_{0 \leqslant k \leqslant n} |M_n|^p] \leqslant C_p \cdot \mathbb{E}[|M_n|^p].$$

Cas de L^2: Orthogonalité dans L^2 des incréments d'une martingale.
Convergence dans L^2 si bornée dans L^2.

5.3.2 Convergence p.s.

Doob's forward convergence theorem (avec les montées) pour les martingales bornées dans L^1 (en particulier positives ou bornées).
Corollaire sur la convergence L^p pour $p > 1$.
Exemple : urne de Polya et processus de branchement.
CNS pour convergence $L^1 \iff$ U.I. $\iff$ martingale fermée. Théorème $0/1$ de Lévy: Si $\mathcal{F}_\infty = \sigma(\mathcal{F}_i : i \geqslant 0)$ alors

$$\mathbb{E}[X \mid \mathcal{F}_n] \xrightarrow[n \to \infty]{p.s.\ L^1} \mathbb{E}[X \mid \mathcal{F}_\infty].$$

5.4 Séries aléatoires

Théorème des trois séries de Kolmogorov.
Application à la loi forte des grands nombres.
Théorème de Kakutani sur les martingales produits.

Chapter VI : Chaînes de Markov

6.1 Définitions générales

Définition d'une chaîne de Markov (espace d'état au plus dénombrable).
Exemples : GW, marche sur graphe, Doudou le hamster, Erhenfest...
Espace d'état discret, temps discret. Interprétation graphique (graphe orienté).
Classes communicantes, irréductibilité, apériodicité.

6.2 Espace fini et Perron-Frobenius

Cas d'un espace d'état fini. Interprétation matricielle, matrice stochastique.
Enoncé de Perron–Frobenius, applications à la limite de P^n une matrice stochastique, positive puis primitive puis irréductible.
Convergence vers l'équilibre (vitesse exponentielle) d'une chaîne de Markov à espace d'état fini. Loi stationnaire.