

Notions de convergence probabiliste; applications aux structures combinatoires

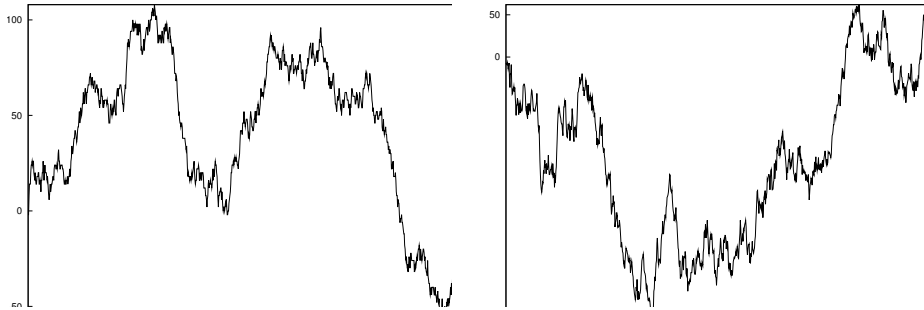
JF Marckert
CNRS, LaBRI, Université de Bordeaux

Contents

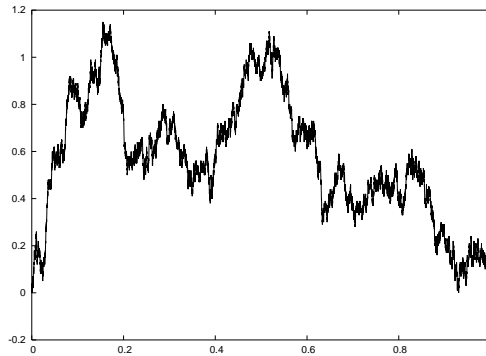
1	Convergence en loi, en probabilité, p.s., L^1, L^p, etc	6
1.1	Espaces polonais	8
2	Convergences	9
2.1	Convergence en probabilité, p.s., etc	10
2.2	Loi faible des grands nombres	11
2.3	Caractérisation des lois de probabilités sur \mathbb{R}	13
2.4	Caractérisation de la convergence en loi	14
2.5	Un aperçu de quelques méthodes pour prouver la convergence	20
2.6	Théorème local limite	21
2.7	Méthodes de martingales	22
2.8	Théorèmes de contractions	23
2.9	Convergence des lois dans \mathbb{R}^d	23
2.10	Comparaison des différents types de convergence	23
3	Convergence de structures combinatoires aléatoires	25
3.1	Convergence des chemins	27
3.1.1	Exemple de critères de convergence dans $C[0, 1]$:	29
3.2	Comment procède-t-on en pratique ?	31
3.2.1	Exemple de critères de convergence dans $D[0, 1]$:	31
3.3	Convergence des arbres et des cartes	32
3.4	Convergence des arbres normalisés	33
4	Quelques pointeurs bibliographiques	36
5	Convergence de structures aléatoires p.s. ou en proba	37
6	Appendix: Probability measure on a Polish set	38
6.1	Measure on Polish space. Characterization	38
6.2	Random variables with a values in a Polish space	41
6.3	Convergence in distribution in a Polish space	41

Quelques images avant de commencer:

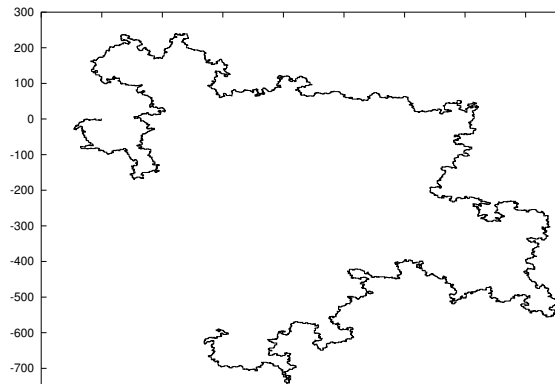
Des marches aléatoires: Simulation de 2 marches aléatoires à 10000 pas. La première a des incréments de loi $\frac{1}{2}(\delta_1 + \delta_{-1})$, la seconde de loi $\frac{1}{4}\delta_{-4} + \frac{1}{5}\delta_2 + \frac{1}{5}\delta_3 + \frac{7}{20}\delta_0$. La similarité entre les trajectoires est notable...



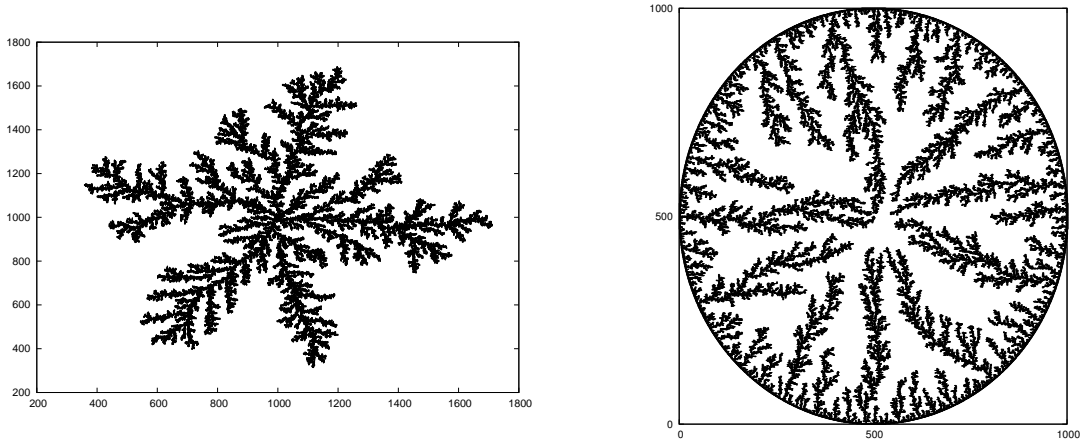
Un grand chemin de Dyck: uniforme parmi ceux à 10000 pas.



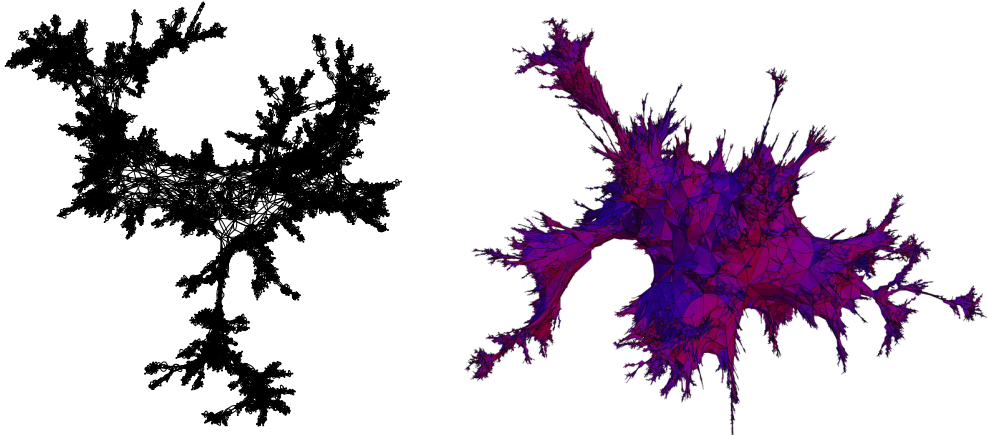
Marche aléatoire à boucles effacées, photographiée la première fois où 10 000 pas ont survécu: on fait une marche aléatoire au plus proche voisin dans \mathbb{Z}^2 ; chaque fois qu'une boucle est faite, on la retire... et on continue



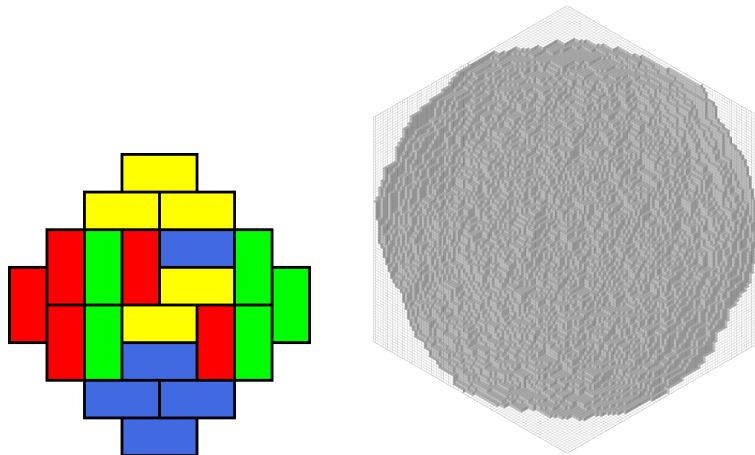
Diffusion limited aggregation (DLA): On met une cellule en $(0, 0)$; on fait ensuite partir successivement des marches aléatoires sur \mathbb{Z}^2 provenant de $+\infty$. On fige les cellules alors qu'elles arrivent à distance 1 de l'amas courant. Devinez qu'est le modèle le second modèle.



Quadrangulation uniforme choisie parmi celles à 25000 sommets / 50 000 sommets (la deuxième par Jérémie Bettinelli). Notez la différence de style.



Pavage par domino du diamant aztèque (taille 800 sur la 2ème image): (Crédit : wikipédia). Apparition du phénomène du cercle arctique.



Introduction

Les théorèmes de convergence jouent un rôle absolument central en théorie des probabilités: on a du mal à imaginer à quoi serait réduite cette théorie si elle devait en être dépourvue.

La notion même de probabilité d'un événement A est souvent perçue comme étant définie par la proportion asymptotique de réalisations de A lors d'une suite infinie d'expériences aléatoires indépendantes dont A est une issue possible, même si un choix différent a été fait pour les fondations formelles de la théorie des probabilités.

Les théorèmes de convergence... Lorsqu'on s'intéresse à la répétition d'une expérience aléatoire, ou à des caractéristiques dans une grande population, ou aux propriétés d'une grande structure combinatoire aléatoire, on observe des phénomènes de "régularisation asymptotique": c'est ainsi, qu'à l'échelle d'un pays, on observe chaque année à peu près le même nombre de naissances, de décès, d'occurrences de telle maladie (en dehors d'événements particulier) et que donc on peut fabriquer des produits financiers comme les assurances; que les désintégrations nucléaires, phénomène aléatoire au niveau atomique, ne l'est plus lorsqu'on entasse des tonnes d'uranium... on peut donc fabriquer des centrales nucléaires ou des bombes atomiques. Au niveau combinatoire, aussi: lorsqu'on simule une grande structure aléatoire, on observe dans de nombreux cas, soit un phénomène de type "régularisation, et convergence vers un objet déterministe", soit "régularisation, et convergence vers un objet aléatoire", ou encore, une convergence de paramètres (différentes notions de convergence se cachant ici sous le vocable unique de "convergence").

Dans ce petit cours, il ne va pas être possible de donner un aperçu complet des théorèmes de convergence en probabilité, loin de là, il y en a tellement... On va se contenter de revenir sur les théorèmes de convergence classiques dans \mathbb{R} (ou dans \mathbb{R}^d). Dans un deuxième temps, on parlera de la convergence des structures aléatoires. On évoquera quelques cas révélateurs des difficultés rencontrées, et on essaiera d'expliquer comment on doit s'y prendre pour démontrer que tel ou tel objet combinatoire converge en loi, et quels sont les sous-produits de cette convergence. On donnera peu de preuve, mais des théorèmes, des pointeurs vers la littérature, et on l'espère, un peu d'intuition.

Il se trouve, et c'est peut-être l'un des grands écueils de la théorie de la convergence, que les phénomènes plus ou moins intuitifs dont on parle sont un peu plus compliqués qu'il n'y paraît de prime abord, et ce pour deux raisons:

- tout d'abord, il y a plusieurs notions de convergence, et ce même dans le cas d'une suite de variables aléatoires réelles (en proba, p.s., L_p en loi)... et il ne s'agit pas seulement d'une différence de degré, de qualité de convergence: la convergence en loi est vraiment une convergence d'une autre nature que la convergence presque sûre,
- d'autre part, la théorie des probabilités étant fortement connectée avec la théorie de la mesure par la nature même des objets concernés, les critères et définitions sont souvent plus compliqués que ce qu'on souhaiterait...

Mais c'est comme cela. Le fait que les choses sont plus compliquées que ce qu'on souhaiterait, est un problème complètement général dans la vie.

1 Convergence en loi, en probabilité, p.s., L^1 , L^p , etc

Une petite remarque préliminaire: on utilise toujours les mots “aléatoire” ou “hasard” lorsqu’on fait des probabilités... Mais ce n’est qu’une façon de parler. Les probabilités sont utilisées pour modéliser des phénomènes aléatoires, mais les probabilités ne sont “que des mathématiques”. On a renoncé à définir dans la théorie le hasard: on se contente d’attribuer des poids aux différentes possibilités, et étudier comment ces poids se répartissent lorsqu’on fait des opérations plus ou moins élémentaires dans notre espace abstrait de probabilité. De ce fait, tout repose sur la théorie de la mesure qui a justement pour vocation à décrire les mesures et leur comportement sous des opérations simples, incluant au passage, le comportement asymptotique, la convergence. Bien sûr, il faut toujours quand même, pour se fabriquer une intuition, comprendre qu’on est en train de parler de phénomène aléatoire.

Allons y.

On appelle espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ un espace Ω (un ensemble), muni d’une tribu¹ \mathcal{A} et d’une probabilité, qui est une fonction

$$\begin{aligned} \mathbb{P} : \mathcal{A} &\longrightarrow [0, 1] \\ A &\longmapsto \mathbb{P}(A) \end{aligned}$$

qui satisfait, $\mathbb{P}(\Omega) = 1$ et qui est σ -additive, ce qui signifie que si on prend une famille au plus dénombrable $(A_i, i \in I)$ d’éléments disjoints de \mathcal{A} alors

$$\mathbb{P}\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mathbb{P}(A_i).$$

Un petit mot sur les tribus: la tribu \mathcal{A} est l’ensemble sur laquelle est définie \mathbb{P} . Il ne s’agit pas de l’ensemble Ω lui même, (c’est-à-dire, l’ensemble “des singletons”) comme on pourrait le croire de prime abord, car cela fournirait un concept de probabilité beaucoup trop pauvre²: une tribu sur Ω est une partie de $\text{Powerset}(\Omega)$ stable par union dénombrable, passage au complémentaire, et contenant l’espace Ω tout entier.

Dans le cas où E est fini ou dénombrable, souvent, la tribu considérée est $\mathcal{E} = \text{Powerset}(E)$; lorsque E est non dénombrable, par exemple $E = [0, 1]$ ou \mathbb{R} , on considère la tribu engendrée par les ouverts, qui s’appelle alors la tribu borélienne (on va y revenir): “en discret”, on attribue souvent bien une masse à chaque partie de Ω , mais en continu, c’est un peu plus subtile.

On appelle espace mesurable (E, \mathcal{E}) , un ensemble E muni d’une tribu \mathcal{E} . Les éléments de la tribu \mathcal{A} sont appelés *événements*.

Une variable aléatoire X sur notre espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ est une fonction

$$\begin{aligned} X : \Omega &\longrightarrow E \\ a &\longmapsto X(a) \end{aligned}$$

telle que X soit mesurable de (Ω, \mathcal{A}) dans (E, \mathcal{E}) , ce qui signifie que

$$X^{-1}(B) \in \mathcal{A}$$

¹appelée aussi σ -algèbre dans la littérature

²les mesures à densité, par exemple, donnent une masse 0 à chaque singleton

pour tout $B \in \mathcal{E}$.

La loi de la variable aléatoire X est alors la mesure \mathbb{Q} (dite mesure image) définie sur (E, \mathcal{E}) par

$$\mathbb{Q}(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(\{i \in \Omega : X(i) \in B\}).$$

Pour aller plus vite, on note parfois \mathbb{P}_X pour la loi de la variable X (au lieu de \mathbb{Q}).

Par exemple, si $\Omega = \{1, 2, 3, 4, 5, 6\}$, muni de la tribu $\text{Powerset}(\Omega)$ et de la mesure \mathbb{P} caractérisée par $\mathbb{P}(\{i\}) = 1/6$. Si on convient avec Toto que si le dé tombe sur i , je lui verserai $X(i) = (i - 3)(i - 5)$ euros, alors X est une variable aléatoire.

a	1	2	3	4	5	6
$X(a)$	8	3	0	-1	0	3

Alors, la loi du gain de Toto est justement la loi de X :

- toto gagnera la somme b ssi $i \in X^{-1}(b)$,
- il gagnera une somme dans B ssi $i \in X^{-1}(B)$,

Ici donc, $\mathbb{P}_X(0) = \mathbb{P}(X^{-1}(0)) = \mathbb{P}(\{3, 5\}) = 2/6$. Toto gagne 0 ssi le dé tombe sur 3.

Pour que la loi de X soit bien définie, il faut pouvoir calculer

$$\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B))$$

pour tout $B \in \mathcal{E}$, et là, on voit que X doit avoir la propriété d'être une application mesurable: $X^{-1}(B)$ doit être un élément de la tribu de départ ³

Lorsqu'on s'intéresse aux mesures continues, il faut mettre davantage les mains dans le cambouis: par exemple, si on veut définir des mesures sur $[0, 1]$ (construction de la mesure de Lebesgue); il est aisé de démontrer qu'il n'existe pas de mesure sur $\text{Powerset}([0, 1])$ qui donnerait comme mesure d'un intervalle sa longueur. Bref, la mesure uniforme ne peut être étendue à tout $\text{Powerset}([0, 1])$... On préfère, plutôt que de renoncer à cette mesure, renoncer à mesurer tous les ensembles. On mesurera uniquement ceux qui sont boréliens⁴.

- Les variables aléatoires pour avoir une loi, doivent donc être non seulement être définies sur un espace de probabilité, mais également prendre leur valeur sur un espace mesurable (muni d'une tribu, sans quoi cela n'aurait pas de sens).

Exemple 1.1. Prenons $\Omega =$ ensemble des arbres à n noeuds enracinés planaires; $\mathcal{A} = \text{Powerset}(\Omega)$, \mathbb{P} la loi uniforme sur Ω ; ou $\mathbb{Q}(\{t\}) =$ poids proportionnel à $\prod_{u \text{ noeuds de } t} \lambda_{\text{degre}(u)}$ pour une suite $(\lambda_j, j \geq 0)$ préfixée. Toute fonction définie de Ω dans n'importe quel espace, muni de n'importe quel tribu, est une variable aléatoire Par exemple, la fonction H hauteur, ou l'application qui envoie un arbre sur son chemin de contour.

³ Qu'aurait-il pu se passer ? Imaginons un exemple concret pour construire une fonction non mesurable. Toto lance un dé à 6 faces, et son espace est toujours $\Omega = \{1, 2, 3, 4, 5, 6\}$. Mais il se trouve, que pour des raisons indépendantes de sa volonté, la tribu \mathcal{A} est la tribu engendrée par $\{1\}, \{2\}, \{3, 4\}, \{5\}$ et $\{6\}$. La mesure associée \mathbb{P} n'attribue un poids qu'aux éléments de la tribu, et en quelque sorte, pour le modèle, 3 et 4 sont indiscernables, et ne forment qu'un unique élément. Il se trouve qu'alors, la fonction X donnée plus haut n'est pas mesurable puisque $X^{-1}(0) = \{3, 5\}$ n'est pas un élément de la tribu. On ne peut donc pas connaître la loi de X , car justement on ne connaît pas la masse de "3" seul.

⁴on va en pratique au delà, sans aller jusque $\text{Powerset}([0, 1])$ en introduisant les ensembles Lebesgue mesurables

Exemple 1.2. $\Omega = \{0, 1\}^{\mathbb{N}}$, \mathcal{A} tribu engendrée par les parties de type $\{x_1\} \times \cdots \times \{x_k\} \times \{0, 1\}^{\mathbb{N}}$ (les cylindres) pour $1 \leq k \leq +\infty$, et les $x_i \in \{0, 1\}$, et avec une probabilité définie par

$$\mathbb{P}(\{x_1\} \times \cdots \times \{x_k\} \times \{0, 1\}^{\mathbb{N}}) = p^{\#\{1 \leq i \leq k : x_i = 1\}} (1 - p)^{\#\{1 \leq i \leq k : x_i = 0\}}$$

il s'agit bien de l'espace naturel sur lequel est défini une suite de tirage 0 ou 1 indépendant, ceux-ci ayant proba respectives $1 - p$ et p . Pour tout $x = (x_i, i \geq 1)$ de Ω , l'application $X_i : \Omega \rightarrow \mathbb{R}$ qui donne $x \mapsto x_i$ en quelque sorte le résultat du i ème lancer, est une variable aléatoire sur cet espace. L'application

$$N_n(x) = \#\{1 \leq i \leq n : x_i = 1\}$$

est une variable aléatoire (de loi Binomiale $(n, 1/2)$).

Exemple 1.3. Si $\Omega = [0, 1]$ muni de sa tribu borélienne, alors l'application $I : x \mapsto x$ (l'identité) et aussi $F(x) = 1/x$ sont des variables aléatoires à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (car puisqu'elles sont continues, elles sont mesurables). L'application $F(x) = (1/x, x, \cos(x^2 + \arctan(x)))$ est aussi une variable aléatoire à valeurs dans \mathbb{R}^3 munie de sa tribu borélienne.

Note 1 : Il n'y a pas de différence de nature entre la loi des v.a. et les lois initiales, celles sur les espaces de probabilité: Si $(\Omega, \mathcal{A}, \mathbb{P})$ est un espace de proba, alors l'identité $Id : x \rightarrow x$ est mesurable de (Ω, \mathcal{A}) à valeurs (Ω, \mathcal{A}) : c'est donc une variable aléatoire, de loi \mathbb{P} (facile à voir); c'est comme cela qu'on voit, que toute loi, est la loi d'une variable aléatoire.

Dans ce cours, on va beaucoup parler des

1.1 Espaces polonais

Un espace polonais est un espace métrique (E, d) , qui est complet et séparable. Complet signifie que les suites de Cauchy convergent, séparable, signifie qu'une partie dénombrable de E est dense dans E . Ici, il n'est pas possible d'expliquer pourquoi cette notion est la bonne (un début d'explication est donnée dans la section 6), ce qu'il faut retenir c'est que l'on est dans une situation favorable:

- on a une métrique d ; avec cela on peut définir des boules $B(x, r) = \{y, d(x, y) < r\}$; cela permet de définir des ouverts (les unions de boules ouvertes), et une tribu, celle engendrée par les ouverts qui s'appelle tribu borélienne;
- ensuite, l'espace n'est pas trop gros, puisqu'il a un sous ensemble dénombrable dense,
- et la convergence n'est pas trop compliquée, car l'espace est complet,
- on peut définir les mesures sur un espace métrique, comme on le fait sur \mathbb{R} .

Des espaces Polonais, en voici: \mathbb{R}^d muni des distances usuelles (car \mathbb{Q}^d y est dense et dénombrable), $C[0, 1]$ muni de la distance $d(f, g) = \|f - g\|_{\infty}$, car les polynômes à coefficients rationnels y sont denses, etc.

Définition 1.1. • On appelle espace polonais, un espace métrique (E, d) complet et séparable,
 • la tribu borélienne sur un espace métrique (E, d) est la tribu engendrée par les ouverts de E (on la note $\mathcal{B}(E)$).

Une probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ attribuera donc une masse à tout ensemble de la tribu, mais pas aux autres: il existe en général, des ensembles non mesurables.

Un principe général est que la théorie de la mesure est plus simple lorsque l'ensemble sur lequel on travaille est un espace Polonais muni de sa tribu borélienne.

2 Convergences

Qui dit converger, dit suites (ou fonctions). On s'intéresse donc ici typiquement à une suite de variables aléatoires $(X_n, n \geq 0)$ qui seront définies ou pas sur le même espace de probabilité selon les moments.

→ Le fait que les variables aléatoires sont définies ou pas sur le même espace n'est pas un détail; en fait, c'est au coeur de l'affaire.

Exemple 2.4. Parfois les variables aléatoires que l'on inspecte sont définies sur le même espace de probabilité par essence:

– lorsqu'on construit un ABR, on insère successivement des données (par exemple des nombres aléatoires) dans l'arbre, et on définit par la même une suite d'arbres aléatoires $(t_n, n \geq 0)$ qui sont bien définis sur le même espace de probabilité (celui où les nombres aléatoires sont définies).

– si on regarde la hauteur H_n d'un arbre uniforme binaire enraciné à n noeuds, eh bien, les variables aléatoires $(H_j, j \geq 0)$ n'apparaissent pas naturellement sur le même espace; chaque H_j est défini sur son espace propre (ça ne veut pas dire qu'on ne peut pas les définir sur le même espace; mais si on les construit sur le même espace, on doit discuter de la dépendance introduite, et donc clairement on a ajouté un a priori qui n'était pas présent au départ).

On va voir qu'il existe **2 principales notions de convergence**:

- On va parler de convergence de variables aléatoires, ce qui va revenir, comme on va le voir, à la convergence de fonctions pondérées. Ce type de convergence ne peut avoir lieu que si les variables sont définies sur le même espace (il y en a plusieurs avatars néanmoins : convergence presque sûre, en proba, dans L_p, \dots).
- Considérons les lois des variables X_n , à savoir les \mathbb{P}_{X_n} . La loi d'une variable est un objet d'un autre type que la variable elle-même: une mesure versus une fonction. Est-il possible que la suite $(\mathbb{P}_{X_n}, n \geq 0)$ converge? Oui ! Il s'agit de la convergence en loi... On va en parler. Et il va être possible que les lois convergent, que les variables X_n soient définies sur le même espace ou pas; il s'agit juste d'une convergence de loi, pas de variables.

Mais avant même de parler de tout cela, le fameux et utile lemme de Borel-Cantelli:

Lemme 2.2. [Borel-Cantelli] • Si $(A_n, n \geq 0)$ est une suite d'événements telle que $\sum_n \mathbb{P}(A_n) < +\infty$, alors

$$\mathbb{P}\left(\limsup_n A_n\right) = 0,$$

ce qui signifie qu'avec probabilité 1, un nombre fini d'événement A_n seulement est réalisé.

• Si la suite $(A_n, n \geq 0)$ est une suite d'événements indépendants, et si $\sum_n \mathbb{P}(A_n) = +\infty$, alors $\mathbb{P}(\limsup A_n) = 1$.

Exemple 2.5. On lance un dé juste avec n^2 faces à l'instant n ; on appelle A_n l'événement, le n ème lancer donne 1: Borel Cantelli dit que $\mathbb{P}(\limsup_n A_n)$, et donc p.s., le dé va tomber sur 1 qu'un nombre fini de fois; par contre, si le dé n'a que n faces, et que les lancers sont indépendants, cette fois, on aura un nombre infini de succès.

Note 2 : Il existe, pour aller plus loin, la loi 0-1 de Kolmogorov, qui dit que si $(A_i, i \geq 1)$ est une suite d'événements indépendants, alors tout événement A qui ne dépend pas d'un nombre fini de A_i (on dit qu'il est dans la tribu de queue) a probabilité 0 ou 1. Voir Billingsley [2, p50].

2.1 Convergence en probabilité, p.s., etc

Définition 2.3. Soit X, X_1, X_2, \dots une suite de variables aléatoires définies sur un même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeurs dans un espace polonais (E, d) (muni de sa tribu borélienne). On dit que la suite $(X_n, n \geq 0)$ converge en probabilité vers la variable aléatoire X (on note $X_n \xrightarrow{\text{proba}} X$) si pour tout $\varepsilon > 0$

$$\mathbb{P}(d(X_n, X) \geq \varepsilon) \xrightarrow{n \rightarrow +\infty} 0.$$

Exemple 2.6. Prenons une suite $(B_n, n \geq 1)$ de v.a. indépendante de Bernoulli $B(p)$ (pour le p de votre choix). On construit pour tout n , $X_n = \sum_{j=1}^n B_j/2^j$, et X la somme $X = \sum_{j \geq 1} B_j/2^j$. Autrement dit, en base 2, X possède des chiffres après la virgule aléatoires. On obtient X_n en ne conservant que les n premiers chiffres. On voit alors que X_n converge vers X puisqu'on a $|X_n - X| \leq 1/2^n$, de sorte que pour $\varepsilon > 0$ fixé, $\mathbb{P}(|X_n - X| > \varepsilon) = 0$ pour n assez grand. Ainsi la suite (X_n) converge en probabilité, et dans ce cas, la limite est aléatoire... c'est X .

Lemme 2.4. Soit X, X_1, X_2, \dots une suite de v.a. définies sur le même espace de probabilité à valeurs dans le même espace polonais (E, d) , telles que,

$$\lim_n \mathbb{E}(d(X_n, X)^2) = 0$$

alors

$$X_n \xrightarrow{\text{proba}} X.$$

Proof. Par Markov, on a:

$$\mathbb{P}(d(X_n, X) \geq \varepsilon) = \mathbb{P}(d(X_n, X)^2 \geq \varepsilon^2) \leq \frac{1}{\varepsilon^2} \mathbb{E}(d(X_n, X)^2) \xrightarrow{n \rightarrow +\infty} 0.$$

□

Définition 2.5. [Convergence presque sûre] Soit X, X_1, X_2, \dots une suite de variables aléatoires définies sur un même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeurs dans un espace polonais (E, d) . On dit que la suite $(X_n, n \geq 0)$ converge presque sûrement vers la variable aléatoire X (on note $X_n \xrightarrow[p.s.]{n} X$), si

$$\mathbb{P}\left(\lim_n X_n = X\right) = 1.$$

Sans exemple il est difficile d'imaginer en quoi la convergence p.s. diffère de la convergence en probabilité.

Exercice 2.1. Trouver un exemple de suite de va. qui converge en probabilité mais pas presque sûrement (non trivial).

2.2 Loi faible des grands nombres

Théorème 2.6. (loi faible des grands nombres) Soit $(X_n, n \geq 1)$ une suite de v.a. indépendantes et de même loi, et de moyenne m . On a

$$\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow[n]{proba.} \mathbb{E}(X_1) = m. \quad (2.1)$$

Exemple 2.7. Si les X_j sont des variables aléatoires de Bernoulli indépendantes $B(p)$, alors $\mathbb{E}(X_1) = p$. On obtient

$$\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{proba.} p.$$

Puisque \bar{X}_n est la proportion de Bernoulli qui valent 1, c'est la loi faible des grands nombres qui affirme que la proportion de *pile* dans une suite de pile ou face tend vers p . C'est donc la loi des grands nombres qui fait le lien entre la probabilité "abstraite" P d'un événement A , et son interprétation fréquentielle, c'est-à-dire, la proportion empirique asymptotique des événements A obtenus dans une suite d'expériences aléatoires.

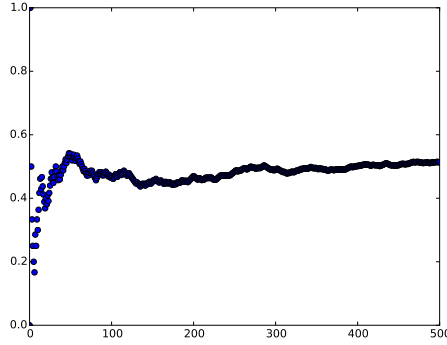


Figure 1: Traçage du graphe $n \mapsto p_n$, avec p_n proportion à l'instant n du nb de piles dans une suite de pile-face aléatoires indépendants, avec proba de pile valant $1/2$. On perçoit sur le dessin la convergence annoncée.

Proof. On donne une preuve sous la condition supplémentaire de l'existence d'une variance finie: On suppose donc que $\text{Var}(X_1) = \sigma^2$, et on note $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$. On a $\mathbb{E}(\bar{X}_n) = m$ et $\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$. Ainsi

$$\mathbb{E}((\bar{X}_n - M)^2) = \text{Var}(\bar{X}_n) \xrightarrow{n \rightarrow +\infty} 0.$$

Ce qui montre que $\bar{X}_n \xrightarrow{\text{proba}} m$ d'après le lemme précédent. □

Théorème 2.7. [Loi forte des grands nombres] Même hypothèse que la loi faible, mais conclusion plus forte:

$$\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow[n]{p.s.} \mathbb{E}(X_1) = m. \quad (2.2)$$

Proof. Il est facile, en passant le m à gauche, de supposer que les X_k ont moyenne 0.

On donne la preuve sous l'hypothèse additionnelle de l'existence de moment d'ordre 4: on note $S_n = X_1 + \dots + X_n$, et on développe S_n^4 ; comme on va calculer son espérance, tous les termes où un X_j apparaît avec un degré 1 va disparaître; reste

$$\mathbb{E}(S_n^4) = n\mathbb{E}(X_1^4) + \binom{n}{2} \mathbb{E}(X_1^2 X_2^2) = O(n^2);$$

Du coup, par Markov

$$\mathbb{P}(|S_n/n| \geq \varepsilon) = \mathbb{P}((S_n/n)^4 \geq \varepsilon^4) \leq \mathbb{E}(|S_n/n|^4)/\varepsilon^4 = O(1/(\varepsilon^4 n^2)).$$

En prenant maintenant $\varepsilon = n^{-1/8}$, on voit que

$$\mathbb{P}(|S_n/n| \geq n^{-1/8}) = O(1/n^{3/2})$$

et puisque cette affaire est sommable, d'après Borel-Cantelli on voit qu'un nombre fini d'événement du type $|S_n/n| \geq n^{-1/8}$ arrivent ps., ce qui implique que p.s., S_n/n converge vers 0. \square

2.3 Caractérisation des lois de probabilités sur \mathbb{R}

Tout d'abord, une définition valable plus généralement: avant de caractériser les lois, il nous faut une définition:

Définition 2.8. Soit μ et ν deux distributions définies sur le même espace mesurable (E, \mathcal{E}) . Ces deux distributions sont dites égales si

$$\int f d\mu = \int f d\nu$$

pour toute fonction $f : E \rightarrow \mathbb{R}$ continue bornée. Cela signifie que si X et Y ont pour loi respective μ et ν , $\mathbb{E}(f(X)) = \mathbb{E}(f(Y))$.

Notation : lorsque 2 variables X et Y ont la même loi, on dit qu'elles sont égales en loi, et on note $X \stackrel{(d)}{=} Y$ (parfois $\mathcal{L}(X) = \mathcal{L}(Y)$).

Pour savoir si deux mesures de probabilité sont égales, il y a des critères généraux, qui consistent à vérifier l'égalité $\mu(A) = \nu(A)$ pour des A parcourant les éléments d'une partie de la tribu (classe monotone, système de Dynkin (voir Billingsley [2, p36])).

→ Le plus souvent, on n'utilise pas ces critères ou définitions pour caractériser l'égalité des lois. On fait appel aux concepts suivant.

Soit μ une loi sur \mathbb{R} , et X une variable aléatoire de loi μ :

– on note F_μ la fonction de répartition de μ :

$$\begin{aligned} F_\mu : \mathbb{R} &\longrightarrow [0, 1] \\ x &\longmapsto \mu((-\infty, x]) = \mathbb{P}(X \leq x) \end{aligned}$$

– on dit que f_μ est une densité pour μ si pour tout fonction continue bornée g

$$\mathbb{E}(g(X)) = \int f_\mu(x)g(x)dx$$

– on dit que Φ_μ est la fonction caractéristique de μ (ou la transformée de Fourier) si

$$\begin{aligned} \Phi : \mathbb{R} &\longrightarrow \mathbb{C} \\ t &\longmapsto \mathbb{E}(e^{itX}) = \int_{-\infty}^{+\infty} e^{itx}d\mu(x) \end{aligned}$$

(elle est toujours définie, et même uniformément continue).

– Transformée de Mellin pour des variables à support dans \mathbb{R}^+

$$M_\mu(s) = \mathbb{E}(X^{s-1})$$

– Fonction génératrice de probabilité, pour les variables ne chargeant que \mathbb{N} :

$$G_\mu(t) = \sum_{k \geq 0} \mathbb{P}(X = k)t^k.$$

Comme série entière, on voit qu'elle possède un rayon de convergence $r \geq 1$.

– Moments: il s'agit de la suite $(m_k, k \geq 0)$ définie par $m_k = \mathbb{E}(X^k)$. Elle est définie pour $0 \leq k \leq K$, pour K pouvant être infini ou fini.

Théorème 2.9. – Deux mesures ν_1 et ν_2 de $\mathcal{M}(\mathbb{R})$:

- possédant la même fonction de répartition, sont égales,
- possédant la même densité, sont égales,
- dont les transformées de Fourier coïncident sur \mathbb{R} sont
- dont les fonctions génératrices coïncident sur un ouvert de \mathbb{R} sont égales contenant 0
- dont les transformées de Mellin coïncident sur un ouvert de \mathbb{R} contenant 1 sont égales
- possédant la même suite de moments sont égales, sous certaines conditions supplémentaires.

Bon, la transformée de Mellin c'est la transformée de Fourier à un changement de variable près, de $\log(X)$. On peut l'utiliser pour étudier la loi de produit de variables aléatoires, mais souvent dans les faits, on se ramène à la transformée de Fourier.

Pour les moments il faut aller plus loin: il y a tout un tas de lois qui n'ont pas de moment d'ordre 1. Donc il est clair qu'en toute généralité, que les moments ne caractérisent pas les mesures. Même si tous les moments existent, il existe des cas où les moments caractérisent la loi, ou pas; on a par exemple le critère de Carleman qui dit qu'il existe au plus une loi μ dont les moments sont les $(m_k, k \geq 0)$ si $\sum_k (m_{2k})^{-1/(2k)}$ diverge (ce qui inclut donc les lois à support compact), mais il existe des lois (pas délirantes) qui ne sont pas caractérisées par leur moment (Ref: Shohat-Tamarkin [18]).

2.4 Caractérisation de la convergence en loi

Définition 2.10. Soit $(\mu, \mu_0, \mu_1, \dots)$ une suite de mesures de probabilité définie sur un espace Polonais (E, \mathcal{E}) . On dit que la suite $(\mu_n, n \geq 0)$ converge faiblement vers μ , si μ est une loi de probabilité, et si pour toute fonction $f : E \rightarrow \mathbb{R}$ continue bornée.

$$\int f d\mu_n \rightarrow \int f d\mu.$$

En d'autres termes si, pour $X_n \sim \mu_n$ et $X \sim \mu$,

$$\mathbb{E}(f(X_n)) \rightarrow \mathbb{E}(f(X)).$$

On dit aussi que $(X_n, n \geq 0)$ converge en loi vers X , ou en distribution vers X , et on note $X_n \xrightarrow[n]{(d)} X$.

On ne peut s'empêcher de donner tout de suite une conséquence directe à la convergence en loi :

Corollaire 2.11. Supposons que $X_n \xrightarrow[n]{(d)} X$ pour des variables aléatoires à valeurs dans un espace polonais (E, \mathcal{E}) ; soit $f : (E, \mathcal{E}) \rightarrow (E', \mathcal{E}')$ une fonction continue de E dans un autre espace polonais E' (du coup, c'est une variable aléatoire sur l'espace de proba initial). On a

$$f(X_n) \xrightarrow[n]{(d)} f(X).$$

Si l'espace initial est \mathbb{R} , le corollaire, n'est pas très important; mais quand il s'agira de $C[0, 1]$, là, on verra toute la puissance de l'affaire: on pourra déduire de la convergence globale, la convergence de milliards de paramètres.

Il y a un théorème très général pour la convergence sur les espaces polonais, qui permet d'en comprendre quelques nuances;

Théorème 2.12. [le théorème porte-manteau] Soit $\mathbb{P}, \mathbb{P}_1, \mathbb{P}_2, \dots$ une suite de mesures de probabilité sur le même espace polonais (S, \mathcal{S}) . Les 5 assertions suivantes sont équivalentes :

(i) $\mathbb{P}_n \xrightarrow[n]{faible} \mathbb{P}$.

(ii) $\int f d\mathbb{P}_n \rightarrow \int f d\mathbb{P}$ pour toutes fonctions $f : S \rightarrow \mathbb{R}$ bornée, uniformément continue.

(iii) $\limsup_n \mathbb{P}_n(F) \leq \mathbb{P}(F)$ pour tout fermé F .

(iv) $\liminf_n \mathbb{P}_n(\mathcal{O}) \geq \mathbb{P}(\mathcal{O})$ pour tout ouvert \mathcal{O} .

(v) Pour tout $A \in \mathcal{S}$ tel que $\mathbb{P}(\partial A) = 0$, $\mathbb{P}_n(A) \rightarrow \mathbb{P}(A)$.

■ Notez que contrairement à ce qu'on est tenter de penser, la convergence en loi, n'est donc pas la convergence de la masse des ouverts, ou des fermés. C'est plus subtile.

■ Ici, on ne parle que de la convergence de mesures μ_n , pas des variables X_n : notez que l'on n'écrit pas $\mathbb{E}(f(X_n) - f(X)) \rightarrow 0$ ce qui sous entendrait qu'on travaille sur le même espace de probabilité... La convergence en loi n'est pas définie par une convergence fonctionnelle des X_n , mais uniquement

par une propriété de leur loi... et effectivement, elle n'a rien à voir avec la convergence des variables comme le montre l'exemple (a) suivant:

- Exemple 2.8.** (a) Si on prend des variables aléatoires $(B_n, n \geq 0)$ Bernoulli(1/2) indépendantes définies sur le même espace, alors B_n ne converge pas presque sûrement, ni en proba (facile à montrer, et à visualiser: pas de régularisation). Par contre $(B_n, n \geq 0)$ converge en loi, puisque la loi \mathbb{P}_{B_n} est constante.
- (b) Si on prend $\mathbb{P}_n = \delta_{1/n}$ la masse de Dirac en $1/n$, alors \mathbb{P}_n converge en loi vers δ_0 .
- (c) Si on prend $\mathbb{P}_n = (1 - 1/n)\delta_0 + (1/n)\delta_{e^{12n}}$ (proba $1 - 1/n$ en 0, et $1/n$ en e^{12n} , alors $\mathbb{P}_n \xrightarrow[n]{faible} \delta_0$; par contre les moments de \mathbb{P}_n tendent tous vers l'infini).

Dans les faits, à part dans des cas triviaux, on ne montre pas la convergence en loi par cette propriété-définition, mais par le truchement de critères reposant sur les caractérisations des lois exposés dans la section précédente, et d'un petit chouilla qu'il faut toujours ajouter:

Théorème 2.13. Soit $(\mu, \mu_1, \mu_2, \dots)$ une suite de loi de probabilité sur $\mathbb{R}, X, X_1, X_2, \dots$,

- Si $(F_n, n \geq 0)$ est la suite des fonctions de répartitions des lois $(\mu_n, n \geq 0)$ qui converge simplement sur c_F vers F , une fonction de répartition, où c_F est l'ensemble des points où F est continue, alors $\mu_n \xrightarrow[n]{faible} \mu$.
- Si la suite des transformée de Fourier $(\Phi_n, n \geq 0)$ converge simplement vers une fonction Φ qui est continue en 0, alors X_n converge en loi, et sa limite à pour fonction caractéristique Φ . [Théorème de continuité de Lévy].
- (cas discret, support \mathbb{N}) Si la suite de fonction génératrice $G_{(n)}$ converge simplement sur un intervalle contenant 0 vers la série génératrice d'une variable X , alors, $X_n \xrightarrow[n]{(d)} X$

- Il existe un critère similaire pour la transformée de Laplace $t \mapsto \mathbb{E}(e^{tX})$ (aussi appelée fonction génératrice des moments) qui, si elle existe dans un ouvert autour de 0, caractérise la mesure, et sa convergence simple, la convergence de mesure (Billingsley [2, Sec. 26]).

- Il existe aussi des critères pour la convergence de la transformée de Mellin qui caractérise aussi la convergence de la mesure, si cette convergence a lieu sur un domaine suffisant (on peut trouver des critères, en traduisant le tout en termes de transformée de Fourier ou Laplace: par exemple, sur la bande $\{s = 1 + it, t \in \mathbb{R}\}$ fait l'affaire, ou sur un intervalle de \mathbb{R} contenant 1). La limite, bien sûr, doit être une transformée de Mellin, pour faire l'affaire.

Exemple 2.9. Pourquoi le fait que F_n converge simplement, par exemple, n'implique pas la convergence en loi ? On a dit qu'il fallait que la limite soit en plus, une fonction de répartition (FR). La raison est que la masse peut partir à l'infini: supposons que $\mu_n = \delta_n$ la masse de Dirac en n ; alors la fonction de répartition de μ_n est $F_n(x) = 1_{x \geq n}$, et donc F_n converge simplement vers F définie par $F(x) = 0$ qui n'est pas une FR d'une mesure de proba sur \mathbb{R} , mais d'une mesure de masse totale 0.

Une conséquence de la définition de la convergence en loi est la suivante:

Théorème 2.14. Si $(X_n)_n$ est une suite de variables discrètes, prenant ses valeurs sur \mathbb{N} ,

$$X_n \xrightarrow{\text{loi}} X$$

si pour tout $k \in \mathbb{N}$, $\lim_n \mathbb{P}(X_n = k)$ existe, et si, en notant $p_k = \lim_n \mathbb{P}(X_n = k)$, on a $\sum_k p_k = 1$.

Exemple 2.10. (Convergence de la binomiale vers la loi de Poisson) Soit $(X_n, n \geq 1)$ une suite de v.a. de loi binomiale de paramètre $\mathcal{B}(n, \lambda/n)$ et X une v.a. de Poisson de paramètre λ . On a:

$$X_n \xrightarrow{\text{loi}} X.$$

En effet, pour $k \in \mathbb{N}$ fixé,

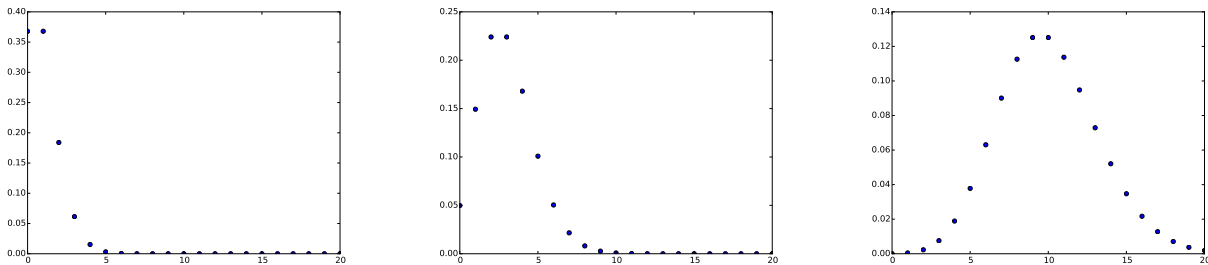


Figure 2: Distribution de Poisson de paramètres 1, 3 et 10

$$\mathbb{P}(X_n = k) = \frac{\lambda^k}{k!} \left(\frac{n!}{(n-k)!n^k} \right) \mathcal{B}\left(1 - \frac{\lambda}{n}\right)^{n-k} \rightarrow \frac{e^{-\lambda} \lambda^k}{k!}$$

car, le second terme du produit tend vers 1, le 3ème vers $e^{-\lambda}$.

Théorème 2.15. (Théorème de la limite centrale) Soit (X_n) une suite de v.a. réelles indépendantes et de même loi, d'espérance finie m et de variance $0 < \sigma^2 < +\infty$. Notons $S_n = X_1 + \dots + X_n$ la somme des n premières variables. On a

$$\frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow{(d)} \mathcal{N}(0, 1). \quad (2.3)$$

où $\mathcal{N}(0, 1)$ désigne une va gaussienne réduite centrée.

Autrement dit, puisque la fonction de répartition de la Gaussienne est lisse, le TCL est équivalent à la convergence simple (pour tout x)

$$\mathbb{P}\left(\frac{S_n - nm}{\sigma\sqrt{n}} \leq x\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$$

On dit aussi, théorème central limite, TCL, ou TLC. La fonction $t \mapsto \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)$ est appelée, courbe en cloche, courbe de Gauss, densité normale, ou densité Gaussienne.

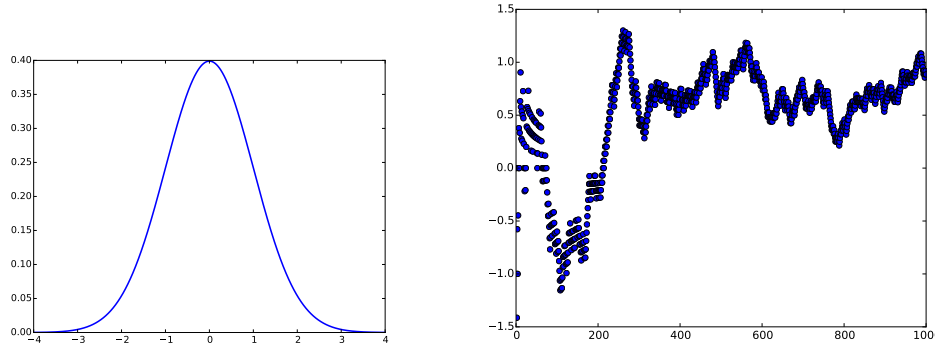


Figure 3: Une partie du graphe de la fonction $x \mapsto \exp(-x^2/2)$. Représentation de $n \mapsto \frac{S_n - np}{\sqrt{np(1-p)}}$.

Exemple 2.11. Si on prend des B_i Bernoulli(1/2) indépendantes, alors pour $S_n = B_1 + \dots + B_n$, le TCL dit que (pour $p = 1/2$ ici)

$$\mathbb{P}\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq y\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y \exp(-t^2/2) dt.$$

Si on trace le graphe $n \mapsto \frac{S_n - np}{\sqrt{np(1-p)}}$, alors on ne perçoit pas de régularisation, il n'y en a pas (voir 2ème figure de Fig. 3). C'est la fonction de répartition de $\frac{S_n - np}{\sqrt{np(1-p)}}$ qui converge; pour le voir, il faut calculer la loi de cette variable... éventuellement, on peut le voir en faisant des milliers de simulations de S_n pour approcher sa fonction de répartition empirique, mais en tout cas, pas avec une seule. ⁵

Remarque 2.16. • On peut relaxer à peu près toutes les hypothèses dans le théorème de la limite centrale;

- d'abord, le théorème de Lindeberg permet de traiter des variables non *i.i.d* (voir juste en dessous).
- on peut renoncer à l'indépendance des X_i , en la remplaçant par une structure de dépendance raisonnable (“processus⁶ faiblement dépendant”, incréments de chaînes de Markov pas trop méchantes, accroissement de martingales, etc).
- par contre, si on renonce à la variance, là, ça ne marche plus, en général. Si les variables sont *i.i.d.*, on regarde la queue de distribution des variables ajoutées; si elles sont régulières (de type $\mathbb{P}(X \geq x) \sim_{+\infty} c_1 x^{-1-\beta}$ et $\mathbb{P}(X \leq x) \sim_{-\infty} c_2 |x|^{-1-\beta}$ (pour $\beta \in (0, 2)$) alors, le théorème de Gnedenko-Kolmogorov dit que S_n/n^β converge en loi vers une variable stable (la limite n'est pas gaussienne si X n'a pas de second moment).

⁵enfin, pour être exact, puisqu'on peut toujours extraire d'une suite infinie, une infinité de suites infinies, on peut voir apparaître le TCL en utilisant qu'une seule trajectoire, mais il faut extraire ces suites, réarranger le tout, et compter. Bref, c'est pas direct

⁶famille de v.a ($X_j, j \in J$) définies sur le même espace de probabilité, indexée le plus souvent ici par $J = \mathbb{N}$, ou $[0, 1]$, \mathbb{R} ou \mathbb{R}^+

Théorème 2.17. [Lindeberg, 1920] Soit $(X_j, j \geq 1)$ une suite de variables indépendantes, centrées, de loi non nécessairement identiques, t.q. $\text{Var}(X_k) = \sigma_k^2$; supposons que pour tout $t > 0$,

$$\sum_{k=1}^n \mathbb{E} (X_k^2 1_{|X_k| \geq t s_n}) = o(s_n^2)$$

pour $s_n^2 = \sum_{k=1}^n \sigma_k^2$ alors

$$\frac{\sum_{k=1}^n X_k}{s_n} \xrightarrow[n]{(d)} \mathcal{N}(0, 1)$$

(Feller [6, p262]). L'hypothèse un peu complexe du théorème de Lindeberg signifie que la variance globale de $X_1 + \dots + X_n$ ne provient que marginalement de grand écart des variables à leur moyenne (en particulier, aucune variable X_i n'est de l'ordre de l'écart type global).

Théorème 2.18. [Berry-Esseen] Sous les hypothèses du TCL auxquelles on ajoute l'existence de moment d'ordre 3 pour les variables, c'est-à-dire, $\mathbb{E}(|X_i|^3) = \rho < +\infty$. Il existe une constante C indépendante de la loi des incréments t.q., pour tout n , tout $x \in \mathbb{R}$,

$$\left| \mathbb{P} \left(\frac{S_n - nm}{\sigma \sqrt{n}} \leq x \right) - \int_{-\infty}^x \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt \right| \leq \frac{C\rho}{\sigma^3 \sqrt{n}}.$$

(Feller [6, p.542] où on voit que $C = 3$ fait l'affaire.)

Comment observer le TCL sur des simulations ?

On prend n un peu grand, et on simule S_n , un grand nombre de fois aussi: on simule donc $S_n^{(1)}, \dots, S_n^{(N)}$. Ensuite, on représente ces données sous la forme d'un histogramme par exemple: on se donne une longueur ℓ , (on compte, pour tout k ,

$$N_k = \left\{ j \in [1, N] : \frac{S_n^{(j)} - Nm}{\sigma \sqrt{n}} \in [k\ell, (k+1)\ell] \right\}$$

et on trace N_k (on s'attend à trouver $N \int_{k\ell}^{(k+1)\ell} e^{-x^2/2} \sqrt{2\pi} dx$ points dans l'intervalle $[k\ell, (k+1)\ell]$; donc, si ℓ est pris assez petit, N et n assez grand, le graphe $k \mapsto N_k$ ressemblera à la courbe de Gauss.

Voici un théorème, qui permet une fois encore, de se fabriquer une intuition un peu plus précise sur ce que sont ces théorèmes de convergence:

Théorème 2.19. [Loi du log-itéré] Soit $(X_n, n \geq 0)$ une suite de v.a. i.i.d. de moyenne 0, de variance 1, et soit $S_n = X_1 + \dots + X_n$. On a

$$\limsup_n \frac{S_n}{\sqrt{2n \log(\log(n))}} \xrightarrow[n]{p.s.} 1$$

Sous les hypothèses du théorème $S_n/\sqrt{n} \xrightarrow[n]{(d)} \mathcal{N}(0, 1)$ (théorème central limite), et par la loi des grands nombres $S_n/n \xrightarrow[n]{p.s.} 0$. Il faut bien comprendre que la loi du log itéré parle de chacune des trajectoires pris dans son ensemble alors que $S_n/\sqrt{n} \xrightarrow[n]{(d)} \mathcal{N}(0, 1)$ parle de toutes les trajectoires coupées en n , pour voir qu'il n'y a pas de contradiction.

2.5 Un aperçu de quelques méthodes pour prouver la convergence

→ D'abord, on ne le dira jamais assez: pensez à la convergence de la fonction caractéristique; c'est ça qu'on utilise probablement le plus souvent en proba.

Le théorème suivant est utilisé pour montrer que les théorèmes locaux limites impliquent les théorèmes globaux.

Théorème 2.20. [Lemme de Scheffé] Soit $(\mu_n, n \geq 0)$ une suite de mesures de probabilité sur \mathbb{R}^d , telle que μ_n possède une densité f_n . Supposons que f_n converge simplement vers f (presque partout), avec $\int f = 1$, alors μ_n converge en loi vers la loi qui a f pour densité.

Bien sûr, très utilisé par ceux qui savent calculer des moments:

Théorème 2.21. [Convergence des moments] Soit $(X_n, n \geq 0)$ une suite de v.a. de loi $(\mu_n, n \geq 0)$ dont les moments existent, et convergent: pour tout $p \in \mathbb{N}$, $\mathbb{E}(X_n^p) \xrightarrow[n]{} m_p \in \mathbb{R}$. S'il existe une seule mesure μ possédant les moments $(m_p, p \geq 0)$, alors

$$\mu_n \xrightarrow[n]{(d)} \mu.$$

Voici un théorème très utilisé en combinatoire analytique

Théorème 2.22. [Théorème des quasi-puissances (Hwang)] Soit $(G_n, n \geq 0)$ une suite de fonctions génératrices de probabilité (de lois sur \mathbb{N}). Supposons que, uniformément, dans un voisinage de 1 dans \mathbb{C} , pour deux suites divergente β_n et κ_n ,

$$G_n(u) = A(u).B(u)^{\beta_n} (1 + O(\kappa_n^{-1}))$$

pour des fonctions analytiques A et B au voisinage de 1, tq. $A(1) = B(1) = 1$ et t.q. $v(B) := B''(1) + B'(1)(1 - B'(1)) \neq 0$, alors pour X_n une v.a. dont la fonction génératrice est G_n , on a

$$\mathbb{E}(X_n) = \beta_n B'(1) + A'(1) + O(1/\kappa_n), \tag{2.4}$$

$$\text{Var}(X_n) = \beta_n v(B) + v(A) + O(\kappa_n^{-1}) \tag{2.5}$$

et

$$\frac{X_n - \mathbb{E}(X_n)}{\sqrt{\text{Var}(X_n)}} \xrightarrow[n]{(d)} \mathcal{N}(0, 1).$$

Sous des hypothèses à peine plus restrictives un théorème limite locale existe, et une version, avec la transformée de Laplace au lieu de la fonction génératrice est également disponible (Flajolet-Sedgewick [8, Sec. XI]).

2.6 Théorème local limite

Les théorèmes locaux limites, sont des théorèmes “plus précis” que les théorèmes de convergence en loi classique (mais s’appliquent à une petite partie des modèles seulement): supposons que les X_i soient des variables discrètes avec support inclus dans \mathbb{Z} , et donc, c’est également le cas pour $S_n = \sum_{i=1}^n X_i$. Supposons que S_n satisfasse un théorème central limite:

$$\mathbb{P}(S_n \leq m_n + x\sqrt{n}) \rightarrow \int_{-\infty}^x e^{-t^2/2}/\sqrt{2\pi} dt \quad (2.6)$$

ou un autre théorème de ce type

$$\mathbb{P}(S_n \leq a_n + xb_n) \rightarrow \int_{-\infty}^x f(t) dt$$

où f est une densité. Alors, il est naturel de penser que

$$\sqrt{n}\mathbb{P}(S_n = \lfloor m_n + x\sqrt{n} \rfloor) \rightarrow e^{-x^2/2}/\sqrt{2\pi} \quad (2.7)$$

dans le 1er cas, (et plus généralement $b_n\mathbb{P}(S_n = \lfloor a_n + xb_n \rfloor) \rightarrow f(x)$).

Le facteur additionnel en \sqrt{n} devant la probabilité, provient du fait que

$$\mathbb{P}(S_n \leq m_n + x\sqrt{n}) = \int_{t \leq m_n + x\sqrt{n}} \mathbb{P}(S_n = \lfloor t \rfloor) dt \quad (2.8)$$

$$= \int_{-\infty}^x \sqrt{n}\mathbb{P}(S_n = \lfloor m_n + y\sqrt{n} \rfloor) dy \rightarrow \int_{-\infty}^x e^{-y^2/2}/\sqrt{2\pi} dy; \quad (2.9)$$

on a fait le changement de variable $t = m_n + y\sqrt{n}$ au passage.

Alors, en fait, si on regarde honnêtement les choses, on voit qu’on ne peut pas déduire de la convergence de l’intégrale, la convergence de l’argument, c’est-à-dire (2.7): un théorème local limite ne peut être déduit d’un théorème de convergence (on peut construire des contre-exemples si on veut, facilement).

C’est l’inverse qu’on fait habituellement: le lemme de Scheffé assure que quand la fonction $(y \rightarrow \sqrt{n}\mathbb{P}(S_n = \lfloor m_n + y\sqrt{n} \rfloor))$ converge vers $f(y)$ où f est une densité, alors $\frac{S_n - m_n}{\sqrt{n}}$ converge en loi vers la loi de densité f (il n’est pas dur de compléter les détails en utilisant notamment le Théorème 2.28). Or, quand on fait des calculs, il n’est pas rare que l’on puisse obtenir un théorème local limite immédiatement, plus facilement encore que la limite en loi. Voici l’énoncé classique du théorème central local limite:

Théorème 2.23. Soit $(X_n, n \geq 0)$ une suite de v.a. i.i.d. de moyenne 0, variance 1; supposons que le support de X_1 soit inclus dans $b + h\mathbb{Z}$ pour une paire de réels (a, h) et h maximal pour cette propriété. Alors

$$\sup_x \left| \frac{\sqrt{n}}{h} \mathbb{P} \left(\frac{X_1 + \dots + X_n}{\sqrt{n}} = x \right) - \frac{e^{-x^2/2}}{\sqrt{2\pi}} \right| \rightarrow 0$$

où le supremum est pris sur le plus petit réseau contenant les points du support de $\frac{X_1 + \dots + X_n}{\sqrt{n}}$ (c'est-à-dire $\sqrt{n}b + h\mathbb{Z}/\sqrt{n}$).

Feller II [6, p.517]

2.7 Méthodes de martingales

Le mot “martingale” a deux significations: tout d’abord, une stratégie (illusoire ou pas) pour (essayer) de gagner dans un jeu de hasard, ou un certain type de processus défini comme suit:

Définition 2.24. On appelle martingale, un processus $(M_n, n \geq 0)$ t.q. pour tout n , $\mathbb{E}(|M_n|)$ existe et pour lequel $\mathbb{E}(M_{n+1} | M_0, \dots, M_n) = M_n$ p.s..

Typiquement, une martingale⁷ c’est la fortune d’un joueur qui joue un jeu juste, avec des règles qui peuvent dépendre de sa fortune actuelle, ou de son passé, mais dont la moyenne de l’incrément à chaque instant vaut 0: son gain moyen au temps $n + 1$ sachant l’histoire de sa fortune (M_0, \dots, M_n) , vaut M_n

La théorie des martingales donne des théorèmes qui permettent de démontrer presque gratuitement la convergence p.s. de certaines suites sous des hypothèses vraiment faibles; on les utilise aussi pour dire des choses sur des limites d’objets combinatoires. Il existe aussi des notions de sur-martingales, ou de sous martingales, définie par $\mathbb{E}(M_{n+1} | M_0, \dots, M_n) \leq M_n$ et $\mathbb{E}(M_{n+1} | M_0, \dots, M_n) \geq M_n$ p.s. (les inégalités sont dans le bon sens).

Voici un théorème qui montre la puissance de l’affaire:

Théorème 2.25. [Doob] Soit $(M_n, n \geq 0)$ une sous martingale telle que $\sup_n \mathbb{E}(M_n^+) < +\infty$, alors (M_n) converge p.s. vers une v.a. M_∞ t.q. $\mathbb{E}(|M_\infty|) < +\infty$.

Note: on ne dit pas que ça converge dans L^1 , mais que la limite est dans L^1 . Nuance.

Exemple 2.12. Soit $(X_m, m \geq 1)$ une suite de v.a. i.i.d. de moyenne 1 (le cas < 1 est plus simple encore). Démontrons que $M_n = \prod_{j=1}^n X_j$ converge ps.

La suite $M_n = \prod_{j=1}^n X_j$ est une martingale, et $\mathbb{E}(M_n) = \mathbb{E}(|M_n|) = 1$ qui est bornée. Donc M_n converge p.s. et sa limite est intégrable. Presque sûrement, ce n’est donc pas $+\infty, \dots$. Que vaut cette limite ?

– Si dans notre modèle, $X_m = 1$ p.s. alors la limite est 1. C’est trivial.

– Sinon, pour un $\varepsilon > 0$, $\mathbb{P}(|X_1 - 1| \geq \varepsilon) > 0$. Montrons qu’alors, dans ce cas la limite est p.s. 0. Pour

⁷ plus généralement, on se donne une suite de tribus $(\mathcal{F}_n, n \geq 0)$ croissantes (“une filtration”), qui représente l’info disponible au temps n dans le système, et on demande à ce que M_n soit \mathcal{F}_n mesurable, et que $\mathbb{E}(M_{n+1} | \mathcal{F}_n) = M_n$ p.s..

voir cela, d'abord, on voit que M_∞ est solution du point fixe $M_\infty \stackrel{(d)}{=} M_\infty M'_\infty$ où les variables à droites sont 2 copies indépendantes de M_∞ (car $M_{2n+1} = (\prod_{j=1}^n X_{2j})(\prod_{j=1}^n X_{2j+1})$). Donc $\mathbb{E}(M_\infty)$ qui existe, est solution de $x = x^2$ et ne peut être que 0 ou 1 (vu qu'elle est finie). Or pour qu'un produit $\prod_{j \geq 1} X_j$ converge vers une constante $c > 0$, il faut que "le reste" $\prod_{j=N}^{+\infty} X_j$ converge vers 1 quand $N \rightarrow +\infty$; mais ce reste à la même tête que M_∞ , ça revient à dire ici que M_∞ vaut 1 ps. Ceci est incompatible avec l'égalité en loi $M_\infty \stackrel{(d)}{=} X_1 M_\infty$ car X_1 ne vaut pas 1 avec proba 1. Bref, si on met tout bout à bout, on voit que la moyenne de M_∞ doit être 0, et donc $M_n \xrightarrow[p.s.]{n} 0$.

2.8 Théorèmes de contractions

Il s'agit de théorème de points fixes pour les suite de v.a. (de lois, en fait) satisfaisant des équations du type

$$Y_n \stackrel{(d)}{=} \sum_{r=1}^{K_n} A_r(n) Y_{I_r(n)} + b_n, n \geq n_0$$

où les variables apparaissant à droite sont indépendantes, et typiquement les Y_n et b_n sont des vecteurs, et les A_r des matrices.

Idée: sous certaines hypothèses, si les $A_r(n)$ convergent dans un certain sens ainsi que b_n et ce, éventuellement après normalisation, alors, dans certains cas Y_n (éventuellement normalisé) va converger vers la solution de l'équation de point fixe

$$Y \stackrel{(d)}{=} \sum_{r=1}^K A_r Y_r + b_r.$$

Au delà de \mathbb{R}^d , des théorèmes du même type ont été démontrés pour des processus aléatoires (liés aux arbres, notamment). Ref: Travaux de Neininger de es dernières années, et Neininger et Sulzbach [16] pour les approches fonctionnelles.

2.9 Convergence des lois dans \mathbb{R}^d

Une fois encore, des théorèmes disent qu'il suffit de montrer la convergence de la transformée de Fourier multidimensionnelles définie par $\Phi(\lambda_1, \dots, \lambda_d) = \mathbb{E}(e^{i \sum_{k=1}^d \lambda_k X_k})$, utiliser le Lemme de scheffé, etc. En fait, il est facile de montrer, par exemple à l'aide des transformées de Fourier, que si $X_n = (X_n^{(1)}, \dots, X_n^{(d)})$ alors $X_n \xrightarrow[n]{(d)} X = (X^{(1)}, \dots, X^{(d)})$ sssi on a la convergence suivante pour les variables aléatoires réelles $(\sum_{j=1}^d \lambda_j X_n^{(j)}) \xrightarrow[n]{(d)} (\sum_{j=1}^d \lambda_j X^{(j)})$ pour tout vecteur $(\lambda_1, \dots, \lambda_d)$.

2.10 Comparaison des différents types de convergence

Bon, je voulais parler davantage de convergence dans L_p ; je le définis rapidement:

Définition 2.26. Soit (X, X_1, X_2, \dots) une suite de v.a. réelles définies sur le même espace; on dit que X_n converge vers X dans L_p , si pour tout n , $\mathbb{E}(|X_n|^p) < +\infty$ et si $\mathbb{E}(|X_n - X|^p) \rightarrow 0$.

On peut montrer que la convergence L_p implique la convergence L_q pour $1 \leq q < p$. La convergence dans L_p implique la convergence des moments $\leq p$ (attention, si p n'est pas entier, on a $\mathbb{E}(|X_n|^p) \rightarrow \mathbb{E}(|X|^p)$ seulement (avec les valeurs absolues, car sans, c'est pas super bien défini).

On a les implications suivantes:

Théorème 2.27. • $X_n \xrightarrow[n]{p.s.} X$ sur un espace de probabilité, alors $X_n \xrightarrow[n]{proba.} X$.

• $X_n \xrightarrow[n]{proba.} X \Rightarrow X_n \xrightarrow[n]{(d)} X$

• Pour des variables réelles: si $X_n \xrightarrow{L_p} X$, alors $X_n \xrightarrow[n]{proba.} X$

• si $X_n \xrightarrow[n]{proba.} X$, alors il existe une suite divergente (non aléatoire) $(k(n), n \geq 0)$ t.q. $X_{k(n)} \xrightarrow[n]{p.s.} X$ (converge ps. le long d'une sous suite).

Abus... Il est usuel de dire que si $X_n \xrightarrow[n]{(d)} X$ avec X déterministe (non aléatoire), alors $X_n \xrightarrow[n]{proba.} X$. C'est un abus de langage lorsque les X_n ne sont pas définis sur le même espace. On justifie néanmoins cet abus comme suit; lorsque qu'on se donne une suite de va. $(X_n, n \geq 0)$ définies sur des espaces différents, on peut toujours construire un espace de probabilité sur lequel existe une suite de variables aléatoires $(\tilde{X}_n, n \geq 0)$ et tq., pour tout n , $\tilde{X}_n \stackrel{(d)}{=} X_n$. Or, si on fait cela pour une suite $(X_n, n \geq 0)$, la suite $\tilde{X}_n \xrightarrow[n]{proba.} X$ quel que soit l'espace où on a choisi de représenter les X_i , bref, quel que soit la structure de dépendance qu'on leur impose; cela fonde l'abus. C'est un peu compliqué et probablement source de confusion(s) pour les débutants. Il serait profitable de dire qu'on converge en loi vers une constante.

Voici "un petit résultat" que l'on a toujours envie d'utiliser: il dit qu'une petite déformation d'une suite convergente en loi, converge toujours en loi:

Théorème 2.28. Si $((X_n, Y_n), n \geq 0)$ est une suite de paires de v.a. définies sur le même espace de proba à valeurs dans le même espace polonais (E, d) . Si $X_n \xrightarrow[n]{(d)} X$ et $d(X_n, Y_n) \xrightarrow[n]{proba.} 0$ alors $Y_n \xrightarrow[n]{(d)} X$.

(Billingsley [3, theo 3.1p 27]). Parfois appelé théorème de Slutsky.

Le théorème suivant est un théorème qui n'a l'air de rien de prime abord, mais qui probablement l'un des théorèmes les plus puissants dans ce domaine des probabilités.

Théorème 2.29. [Théorème de Skorokhod] Soit $(\mu_n, n \geq 0)$ une suite de lois de probabilité sur un espace polonais, t.q. $\mu_n \xrightarrow[n]{(d)} \mu$. Il existe un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ sur lequel on peut définir des variables aléatoires (X, X_1, X_2, \dots) t.q. $X \sim \mu$, $X_j \sim \mu_j$ pour tout j et t.q. $X_n \xrightarrow[n]{p.s.} X$.

Corollaire: si on se donne des variables aléatoires $(Y_n, n \geq 0)$ qui ne convergent pas p.s. (par exemple car elles ne sont même pas définies sur le même espace de probabilité) mais qui convergent en loi, alors, on peut fabriquer sur un autre espace de probabilité, des variables aléatoires $(X_n, n \geq 0)$ qui ont même loi une à une (“au niveau des marginales”, on dit) que les $(Y_n, n \geq 0)$ mais pour lesquelles on arrive à créer une structure de dépendance correcte entre les X_i de sorte que X_n converge p.s.

Remarque 2.30. On a dit que la convergence des moments de X_n vers des moments m_p qui caractérisent une mesure implique la convergence en loi de X_n vers la loi dont les moments sont les m_p . Par contre, on peut avoir convergence en loi $X_n \xrightarrow[n]{(d)} X$, sans avoir de convergence des moments, et y compris si X est caractérisé par ces moments (comme dans l'exemple 2.8(c)). En fait, la non convergence des moments n'indique rien du tout pour la convergence en loi.

Bon, on peut aller plus loin et ne pas avoir une vision trop négative de l'affaire: si $X_n \xrightarrow[n]{(d)} X$ et si pour un $\varepsilon > 0$ quelconque, $\sup_n \mathbb{E}(|X_n|^{p+\varepsilon}) < +\infty$, alors $\mathbb{E}(X_n^p) \rightarrow \mathbb{E}(X^p)$. En fait, il y a un critère un tout petit peu plus général encore: il s'agit de critère d'uniforme intégrabilité (voir Billingsley [3] p 31).

3 Convergence de structures combinatoires aléatoires

Ce qu'on va dire ici, est valable au delà des structures combinatoires, puisqu'on va s'entretenir de convergence en loi, et de la manière dont on prouve ces convergences. De nouveau, l'idée n'est pas de fouiller tous les détails, car non seulement on n'a pas la place ici, mais lorsque l'heure est venue de prouver la convergence de notre structure préférée, il faut se poser et ouvrir le(s) bon(s) livre(s), et suer sang et eau; l'idée ici est de comprendre le principe.

Tout d'abord, il existe en combinatoire, des palanqués de structures de différentes natures. Par exemple, on a des familles d'arbres, de mots, de permutations, de chemins particuliers, de graphes, de cartes, de polynômes, de partitions, des dissections du cercle, de chemins non intersectants, etc, etc. Établir un théorème de convergence ici, est forcément plus difficile que dans \mathbb{R} ou dans \mathbb{R}^d , puisqu'avant même de commencer, en général, il faut du temps pour exprimer ce que pourrait vouloir dire converger, et pour exprimer ce que pourrait être la limite; je parle ici de leur nature: qu'est-ce que la limite d'un arbre ? d'un chemin ? d'une quadrangulation ? d'une permutation ? Est-ce que cela a même un sens?

Par ailleurs, et c'est ici la même chose que dans \mathbb{R} , on va pouvoir converger en différent sens (ps, proba, ou en loi...).

Bon, prenons des exemples. Considérons, par exemple, ceux simulés et dessinés dans les premières pages du cours, ou les modèles suivants (tous ont été étudiés):

Exemple 3.13. a l'ensemble Ω_n des arbres binaires planaires enracinés à n noeuds internes muni de la loi uniforme,

b les cartes planaires enracinées biparties à n faces, avec un poids de Boltzmann $\prod_{f \text{ face}} \lambda_{\text{degre}(f \text{ face})}$ où $\lambda = (\lambda_j, j \geq 0)$ est une suite de paramètres

c Les tableaux carrés $n \times n$, dans lequel on place les nombres de 1 à n^2 , de sorte que ces nombres soient croissants sur les lignes et sur les colonnes (sous la loi uniforme) (Romik-Pittel).

d Les arbres non croisés dessinés sur un polygone régulier à n côtés (sous la loi uniforme)

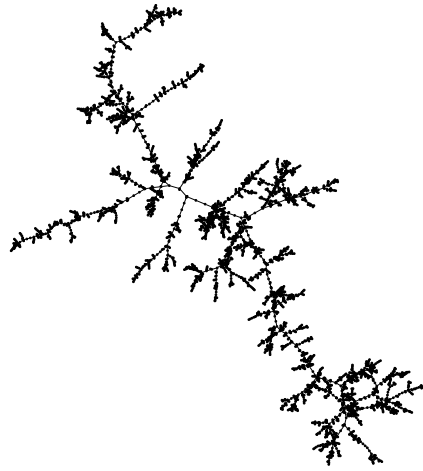


Figure 4: Arbre planaire enraciné uniforme parmi ceux à 8000 noeuds

On va donc chercher à dire des choses sur le comportement asymptotique de notre famille d'objets préférée. Pourquoi? Car on a besoin de comprendre à quoi ressemble un grand objet aléatoire typique... et on a remarqué, en faisant une simulation par exemple, que notre objet préféré semblait avoir un comportement asymptotique remarquable.

Les limites

Les “dessins ou objets limites” obtenus par simulation semblent, selon les cas, être asymptotiquement aléatoires ou déterministes. Lorsqu'on essaie de transformer les observations en théorème ou conjecture, la difficulté apparaît immédiatement: on a une espèce de convergence de l'image, soit, mais on n'a rien d'autre puisque la notion même de convergence n'a pas été définie! et en plus, dans la plupart des cas intéressants, la nature de la limite est différente en première apparence à celle des objets discrets desquels on est parti: la limite semble être “un objet continu souvent biscornu”, ou infini, en tout cas (si on ne normalise pas).

Lorsqu'on demande à un ordinateur de tracer une grande quadrangulation, par exemple, deux choses doivent être comprises:

- un programme particulier a été exécuté, et le rendu final dépend fortement de ce programme,
- une normalisation automatique a été opérée par le logiciel pour représenter l'objet dans une fenêtre: on ne sait pas ce que c'est gratuitement (voir simulations page 4).

Les questions principales qui surgissent normalement à l'esprit intrigué sont les suivantes:

1. que signifie qu'une suite d'objets combinatoires, normalisés ou non, converge ?
2. que signifie qu'une suite d'objets combinatoires **aléatoires**, normalisés ou non, converge ? (en loi ? en probabilité ?)

Supposons qu'un théorème a été écrit ...

3. que pourra-t-on déduire de cela ?
4. que ne pourra-t-on pas déduire de cette convergence?
et bien sûr,
5. Comment démontrer de tels résultats de convergence ?

Réponse rapide

1. Intrinsèquement rien du tout. “Converger” est une notion qui dépend de la topologie choisie. Il faut donc d’abord choisir une topologie avec laquelle on va travailler (on travaillera dans un espace polonais, sauf cas pathologique). Parfois, il peut arriver qu’aucune topologie n’ait été pensée pour décrire la convergence d’une famille d’objets donnée, ou alors que les topologies existantes ne soient pas satisfaisantes pour telle ou telle raison, parfois plusieurs topologies intéressantes existent déjà dans la littérature.
2. On peut dire rapidement, que si une bonne topologie existe pour décrire la convergence de structure non aléatoire, la notion de convergence en loi ou en proba / p.s. est définie automatiquement: c’est cette histoire de convergence faible sur un espace polonais.
3. De nouveau, cela dépend de la topologie choisie. Plus la topologie est fine, plus on peut déduire de choses. Si l’objet limite est non trivial, moralement, un nombre infini de sous produits sont donnés automatiquement (si on a trop écrasé notre suite discrète “en normalisant” tellement que la limite est réduite à un point, on ne peut rien déduire... sauf que la normalisation était trop forte). En toute généralité, on utilisera le Corollaire 2.11 pour déduire que toute fonction “raisonnable” de notre objet (tout paramètre) converge en loi vers le paramètre analogue de la limite.
4. De nouveau, la topologie et la normalisation choisie induisent le degré de finesse des déductions valides: par exemple, si on normalise une structure par \sqrt{n} , les détails qui sont à une échelle différente, disparaissent à la limite (par exemple, les arbres ternaires uniformes à n sommets, normalisés par \sqrt{n} convergent vers l’arbre continu d’Aldous, qui lui, est un arbre binaire).
5. bien sûr, il s’agit là de la question principale. Plusieurs problèmes de nature différente doivent être résolus simultanément: trouver un espace qui contient la famille d’objets dont on cherche à montrer la convergence, leur limite, définir une distance sur l’espace obtenu qui fait de cet espace un espace polonais; fabriquer des caractérisations des mesures sur cet espace, et des critères de convergence en distribution... et cela, bien sûr, en relation avec ce qu’on arrive à démontrer... Si on se place sur une nouvelle classe de limite, là, on peut dire qu’en général c’est une question difficile. Mais, dans la vie, rien de valeur n’est donné gratuitement.

3.1 Convergence des chemins

Supposons qu’on possède une famille de chemins aléatoires de longueur n et de loi μ_n , et ce pour $n \geq 0$; notons $S_n := (S_n(k), 0 \leq k \leq n)$ ce chemin pris au hasard selon μ_n .

Voici les 3 types de convergence en loi les plus souvent étudiés:

Convergence dans $C[0, 1]$

On plonge nos chemins dans $C[0, 1]$, en les normalisant comme suit:

$$s_n(x) = \frac{\overline{S_n}(nx)}{a_n}, \text{ pour } 0 \leq x \leq 1$$

où $\overline{S_n}$ est la fonction continue qui interpole linéairement par morceau S_n dans $[0, n]$:

$$\overline{S_n}(y) = S_n(\lfloor y \rfloor) + (y - \lfloor y \rfloor)(S_n(\lceil y \rceil) - S_n(\lfloor y \rfloor)),$$

et a_n est une normalisation spatiale. Ce qu'on a fait ? On la ramené simplement le temps sur $[0, 1]$, et si notre objet n'était pas continu de prime abord, on l'a forcé à l'être. Parfois une correction de la moyenne de type $\frac{\overline{S_n}(nx) - f(nx)}{a_n}$ est plutôt considérée.

Quand faire cela ? Quand on subodore que la limite est continue: on vient de plonger nos objets dans $C[0, 1]$, l'ensemble des fonctions continues sur $[0, 1]$ qui est un espace bien connu, et qui, muni de la métrique $d(f, g) = \|f - g\|_\infty$ est un espace polonais.

Convergence dans $D[0, 1]$

On plonge cette fois notre objet discret dans l'espace des fonctions càdlàg (continues à droite, limitées à gauche)

$$s_n(x) = \frac{S_n(\lfloor nx \rfloor)}{a_n}, \text{ pour } 0 \leq x \leq 1;$$

on ne peut pas démontrer une convergence dans $C[0, 1]$ car là, les objets discrets ne sont pas continus, et ne peuvent donc pas converger dans un espace qui ne les contient pas (même si la limite est continue). Ici, l'espace naturel pour contenir nos objets et l'espace $D[0, 1]$, espace des fonctions càdlàg. Cet espace est métrisable pour une distance sympathique à savoir

$$d(f, g) = \inf_{\lambda} \sup_x |\lambda(x) - x| + \sup_x |f(x) - g(\lambda(x))|$$

où λ est une fonction croissante valant 0 en 0 et 1 en 1 (“un changement de rythme dans le parcours de $[0, 1]$ qui permet, dans une certaine mesure, d'aligner les sauts importants de f et de g au prix d'une pénalisation). Pour une distance équivalente en terme topologique (voir détails dans [3, p.125], $D[0, 1]$ est un espace polonais). **On utilise cette topologie typiquement lorsque la limite n'est pas continue**, ou lorsque pour une raison technique on n'a pas interpolé le processus linéairement par morceaux, mais plutôt par une fonction en escalier... Enfin, j'abuse un peu ici, car une fois encore “la limite” n'est pas définie avant la topologie, mais bon, si on prend trop de précaution oratoire, on ne dit plus rien.

Convergence sans normalisation, convergence de suites

– Parfois on ne normalise pas du tout ! on cherche la limite au “processus S_n ”; cela revient sous la mesure produit, à chercher la limite en loi de

$$(S_n(k), 0 \leq k \leq K)$$

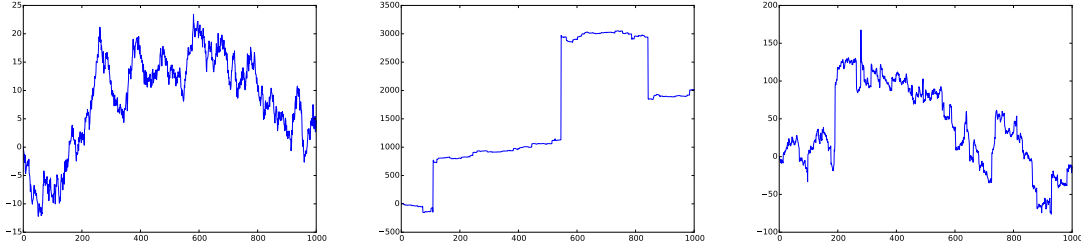


Figure 5: Simulation de 3 marches aléatoires: la première a des incréments Gaussien $N(0,1)$ i.i.d., la deuxième des incréments Cauchy (moment d'ordre 1 inexistant), la 3ème a pour incrément $\pm 1/U^{3/4}$ avec U uniforme sur $[0,1]$ (signe uniforme); pas de variance, mais une moyenne; les 3 marches ont 1000 pas.

pour tout K , typiquement, en tant que loi dans \mathbb{R}^{K+1} . On utilise alors le fameux théorème de Kolmogorov, qui dit que si la limite existe pour tout K , alors, S_n (complété par des 0 jusque l'infini pour obtenir un processus de longueur infinie) converge en loi, pour la topologie produit, vers un processus S dont les lois finies dimensionnelles sont les limites obtenues.

3.1.1 Exemple de critères de convergence dans $C[0,1]$:

D'abord, les lois dans $C[0,1]$ sont caractérisées par ce qu'on appelle les lois finies dimensionnelles: si on prend la loi μ d'un processus $X = (X_t)_{t \in [0,1]}$, la loi de μ est caractérisée par la loi des vecteurs $(X_{t_1}, \dots, X_{t_k})$ pour $k \geq 1$, $0 \leq t_1 < \dots < t_k \leq 1$ (on peut exprimer cela directement sur μ , en parlant des fonctions

$$\begin{aligned} \pi_{t_1, \dots, t_k} : C[0,1] &\longrightarrow \mathbb{R}^k \\ f &\longmapsto (f(t_1), \dots, f(t_k)) \end{aligned}$$

et des lois de $\mu \circ \pi_{t_1, \dots, t_k}^{-1}$, mais ça revient au même. On trouve cela facilement dans la littérature, à commencer chez Billingsley [3], d'ailleurs, comme le reste de cette partie).

- La convergence des lois finies dimensionnelles est une condition nécessaire pour la convergence en loi, mais non suffisante.
- On a besoin d'un critère dit de "tension": en gros, on veut s'assurer que notre distribution qui charge les fonctions continues, les charge toujours à la limite; or ceci n'étant pas garanti par la convergence simple. L'idée est de montrer que la masse ne sort pas de l'ensemble des fonctions continues à la limite; ça se fait, en montrant, qu'à ε , près la masse reste dans un compact de l'ensemble des fonctions continues (et ce pour tout $\varepsilon > 0$).

Toute la logique de l'affaire repose alors sur le théorème d'Arzelà-Ascoli qui dit qu'un ensemble A dans $C[0,1]$ est relativement compact (de fermeture compacte) si les fonctions de A sont uniformément bornées en 0 et équicontinues⁸ Si une suite de fonctions ne possède pas cette propriété, alors, elle peut tout à fait converger simplement, sans converger uniformément.

⁸ Ça dit que les fluctuations sont bornées uniformément: pour tout x , tout $\varepsilon > 0$, $\exists \eta > 0$ t.q. $|x - y| \leq \eta$, implique pour tout f dans A , $|f(x) - f(y)| \leq \varepsilon$.

Alors, les critères de tension sont plus compliqués que dans \mathbb{R} ; le plus général est le suivant: le module de continuité d'une fonction g de $C[0, 1]$ est

$$w_x(\delta) = \sup_{|s-t|<\delta} |g(s) - g(t)|.$$

Une condition suffisante pour l'équicontinuité sur un compact est

$$\limsup_{\delta \downarrow 0} \sup_{g \in H} w_g(\delta) = 0, \tag{3.10}$$

(uniforme équicontinuité et équicontinuité coïncident sur un compact).

Théorème 3.31. Soit $(\mathbb{P}_n, n \geq 1)$ une suite de probabilités sur $(C[0, 1], \mathcal{B}(C[0, 1]))$. Si les 2 conditions suivantes sont satisfaites, alors la suite de mesures de probabilité $(\mathbb{P}_n, n \geq 1)$ est tendue :

(i) Pour tout $\eta > 0$, il existe $a \in \mathbb{R}^+$ t.q.

$$\mathbb{P}_n(\{x \mid |x(0)| > a\}) \leq \eta \text{ for } n \geq 1.$$

(ii) Pour tout $\varepsilon > 0$ et $\eta > 0$, il existe $\delta \in]0, 1[$ et $n_0 \in \mathbb{N}$ telle que pour $n \geq n_0$,

$$\mathbb{P}_n(\{x \mid w_x(\delta) \geq \varepsilon\}) \leq \eta.$$

C'est cette caractérisation qu'il faut utiliser pour montrer le théorème de convergence des marches normalisées vers le mouvement Brownien (Donsker) en toute généralité (sans hypothèses additionnelles sur les moments).

En dehors de la recherche de résultats super généraux (de ce type), on n'utilise pas ce critère dans les applications, car en général on arrive à démontrer des meilleures bornes que cela, des bornes accessibles par les méthodes de moment, par exemple, voici un critère plus accessible aux méthodes calculatoires:

Théorème 3.32. Soit $(\mathbb{P}_n, n \geq 1)$ une suite de mesure de probabilités sur $C[0, 1]$. Soit $\mathbf{x}_n \sim \mathbb{P}_n$. S'il existe $\alpha > 1, \beta > 0$ et $\gamma > 0$ t.q. pour tout s, t dans $[0, 1]$ et tout n

$$\mathbb{E}(|\mathbf{x}_n(s) - \mathbf{x}_n(t)|^\alpha) \leq \beta |s - t|^{\gamma+1}$$

et si de plus la suite $(\mathbf{x}_n(0))$ est tendue^a alors $(\mathbb{P}_n, n \geq 1)$ est tendue.

^aPour tout $\varepsilon > 0$ il existe M t.q. pour tout n , $P(|\mathbf{x}_n(0)| \leq M) \leq 1 - \varepsilon$

(ref: Billingsley [3, Theorem 12.3]).

Théorème 3.33. [Donsker] Soit $(X_k, k \geq 1)$ une suite de v.a. indépendantes centrées, i.i.d. et de variance $0 < \sigma^2 < +\infty$. Alors pour

$$S(k) = X_1 + \dots + X_k$$

et \bar{S}_n obtenue par interpolation linéaire de S entre les points entiers, on a

$$\left(\frac{S(nx)}{\sigma\sqrt{n}} \right)_{0 \leq x \leq 1} \xrightarrow[n]{(d)} (B_t, 0 \leq t)$$

dans $(C[0, 1], \|\cdot\|_\infty)$ où B est appelé mouvement Brownien.

Les finies dimensionnelles du mouvement Brownien se déduisent immédiatement du théorème de la limite centrale: pour tout $0 = t_0 \leq t_1 \leq \dots \leq t_k \leq 1$

$$(B_{t_{i+1}} - B_{t_i}, 1 \leq i \leq k-1)$$

est une suite de v.a. indépendantes, et $B_t - B_s \stackrel{(d)}{=} \mathcal{N}(0, t-s)$. La tension, se montre facilement si on a des moments élevés, mais est dure à obtenir sous les hypothèses faibles du théorème de Donsker.

3.2 Comment procède-t-on en pratique ?

■ On considère le processus qui nous intéresse $(S(nx)/a_n, 0 \leq x \leq n)$ normalisé correctement (on divise par l'écart type, on retranche la moyenne; on doit connaître cela par ailleurs).

Ensuite, on calcule la limite en loi de $(S(nx_1)/a_n, \dots, S(nx_k)/a_n)$; par exemple, à l'aide d'un calcul de transformée de Fourier multidimensionnel, d'un théorème local limite, de l'utilisation d'un TCL, ou à l'aide d'un calcul de fonction génératrice, d'une transformée de Mellin, etc. Tous les coups sont permis.

■ Ensuite, le plus simple est de chercher une bonne description de $S(nt) - S(ns)$, puis on cherche les moments : on calcule $\mathbb{E}(|\frac{S(nt)-S(ns)}{a_n^k}|^k)$. Si tout se passe bien, pour p assez grand, on va réussir à montrer que $\mathbb{E}(|\frac{S(nt)-S(ns)}{a_n}|^p) \leq c|t-s|^{1+u}$ (pour $u > 0$). Typiquement, lorsqu'on montre la convergence vers des processus Brownien, il faut prendre p un peu plus grand que 2 pour que ça marche. Si ça ne marche pas, c'est que soit le processus limite n'est pas continu, soit que $S(nt) - S(ns)$ n'a pas de moments assez grands (par exemple, juste des moments d'ordre 2 mais pas plus), soit qu'on n'a pas de chance, et on n'arrive pas à calculer de moments; alors là, il faut regarder les livres (Billingsley [3] pour voir la preuve du théorème de Donsker dans toute sa généralité, ou trouver une autre approche).

3.2.1 Exemple de critères de convergence dans $D[0, 1]$:

Une fois encore, la convergence des finies dimensionnelles est une condition nécessaire non suffisante; un critère de tension est le suivant: On change la notion "de module de continuité"; on dit qu'une suite $t_0 = 0 < t_1 < \dots < t_k = 1$ (pour un k quelconque) est δ -éparse si $\inf |t_{i+1} - t_i| \geq \delta$; on considère alors le module

$$w'_f(\delta) = \inf_{(t_i) \delta\text{-éparse}} \max_i w_f[t_{i-1}, t_i]$$

Théorème 3.34. Un sous ensemble A de $D[0,1]$ est relativement compact s'il est borné ($\sup_{f \in A} \sup_{x \in [0,1]} |f(x)| < +\infty$) et si $\lim_{\delta \rightarrow 0} \sup_{f \in A} w'_f(\delta) = 0$

([3, Theo. 12.3 p 130]) un critère de tension consiste à remplacer dans le 2eme point du Théorème 3.31, le module w par w' . D'autres critères existent (voir [3, p.139 et suivantes]).

3.3 Convergence des arbres et des cartes

Commençons par les arbres. Les arbres sont plus compliqués que les chemins, mais:

- peuvent-être codés par des chemins (les contours par exemple).
- On peut également les voir comme des arbres réels (des espaces métriques, connexes, sans boucles).

On peut aussi, et c'est comme cela que le premier résultat de convergence dû à Aldous a été énoncé, les voir comme une partie de ℓ_1 (les suites de somme finie, oui). En effet, si en dehors de la racine, on choisit u_1, \dots, u_k des points dans un arbre, alors le sous arbre minimal contenant ces points est une sorte d'union de segments... mais ces segments si on les traces dans un espace euclidien de dimension suffisante pour que les segments ne s'intersectent pas, la distance sur l'arbre ne correspond pas à la distance $\|\cdot\|_2$, mais bien à la distance $\|\cdot\|_1$... Comme on doit faire cela pour $k = +\infty$, on peut faire coïncider la métrique de l'arbre et celle de l'arbre sous-jacent, si justement on prend la distance donnée par $\|\cdot\|_1$ dans ℓ_1 ...

Bref, derrière l'approche d'Aldous et les autres approches, il y a la volonté de plonger les objets discrets dans un espace susceptible de contenir également les limites, et bien entendu, cet espace doit être polonais si on veut éviter trop de complications. Chaque choix de topologie vient avec ses sous-produits, puisque une fois encore, quand $X_n \xrightarrow[n]{(d)} X$ alors $f(X_n) \xrightarrow[n]{(d)} f(X)$ pour f continue.

Convergence des arbres non normalisés

On peut s'intéresser à la convergence des arbres non normalisés, par exemple, le modèle des arbres binaires contenant n sommets sous la loi uniforme: on regarde alors l'arbre obtenu en ne conservant que les noeuds à distance $\leq r$ de la racine. On a besoin aussi dans ce cas de trouver un espace contenant les arbres finis initiaux et les arbres limites, mais dans ce cas, l'ensemble des arbres binaires finis ou infini offre un cadre simple pour discuter de ces questions de convergence; on munit alors l'ensemble des arbres d'une "topologie locale": 2 graphes (ou arbres) enracinés sont à distance $1/(n+1)$ si les deux arbres coïncident dans un voisinage à distance n de la racine (pour $n \geq 1$):

$$d(g, g') = \frac{1}{1 + \max\{r : B(g, r) \sim B(g', r)\}}$$

Comment procède-t-on ? Typiquement, on calcule à la main la loi du sous arbre $B(T_n, 1/k)$ de T_n (obtenu par troncation au niveau k), pour k fixé, et T_n l'arbre aléatoire sous la loi d'intérêt.

- On montre que pour k fixé, $B(T_n, 1/k)$ converge en loi lorsque $n \rightarrow +\infty$ vers une certaine loi μ_k qui ne charge donc, que les arbres dont les noeuds sont à distance $< k$ de la racine.
- Par construction, les lois $(\mu_k, k \geq 0)$ sont compatibles; on fait appel à un théorème dit "théorème

d'extension de Kolmogorov" qui assure alors, qu'il existe une loi μ sur les arbres infinis compatibles avec les μ_k (dans le sens ou un arbre sous μ , qu'on tronquerait à hauteur k coïnciderait avec μ_k),

■ On déduit de tout cela, la convergence des arbres sous la loi étudiée vers μ pour la topologie de la convergence locale.

3.4 Convergence des arbres normalisés

Deux topologies sont maintenant consacrées dans la plupart des cas pour décrire la convergence en loi des arbres; la première est liée à la convergence du contour de l'arbre, la deuxième consiste à prouver la convergence des arbres en tant qu'espace métrique pour une topologie sur l'espace des espaces métriques compacts (topologie de Gromov-Hausdorff).

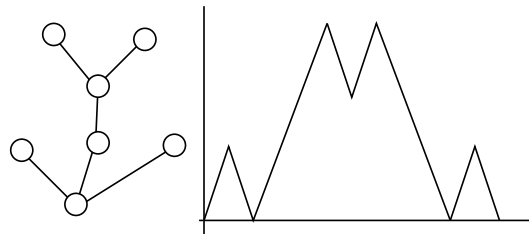
La première est plus fine que la seconde, mais puisqu'elle repose sur l'arbre en tant que structure ordonnée, elle ne s'applique pas bien aux structures non ordonnées naturellement, comme les arbres binaires enracinés non ordonnés (les arbres dit de Polya).

Toujours le même mantra:

Plongement des arbres discrets dans un espace polonais convenable.. qui contient leur limite.

Plongement dans un espace fonctionnel

Cette construction est due à JF Le Gall (dans ces aspects limites), elle repose sur l'idée suivante: un arbre discret peut-être codé de manière fidèle par un processus discret, son processus de contour. La fidélité: il s'agit, au delà, d'une bijection, de la représentation d'un arbre par une fonction qui est très proche topologiquement à l'arbre lui, même, puisque toutes les propriétés de l'arbre se retrouvent immédiatement sur la fonction: les noeuds, leur ordre, leurs distances relatives, etc.



Or, on vient justement, dans la section précédente de dire que on avait des outils pour traiter la convergence de fonctions aléatoires en loi. Il suffit de tout faire dans l'espace des fonctions $C[0, 1]$ muni de la topologie de la convergence uniforme et d'essayer de transférer notre question de convergence de l'arbre sur cet espace. On va donc normaliser le chemin de contour (le chemin de Dyck) pour qu'il puisse converger: on se met dans $C[0, 1]$:

Plus précisément, notons $C^+[0, 1]$ le sous ensemble de $C[0, 1]$ des fonctions $g : [0, 1] \rightarrow \mathbb{R}^+$, t.q. $g(0) = g(1) = 0$. On va associer un arbre à g , et c'est arbre va être construit topologiquement en identifiant des points de $[0, 1]$: pour g donné, on définit une relation d'équivalence sur $[0, 1]$ comme suit:

$$x \sim_g y \iff g(x) = g(y) = \check{g}(x, y),$$

où

$$\check{g}(x, y) = \inf\{g(u) \mid u \in [x \wedge y, x \vee y]\}.$$

Ainsi l'ensemble des classes $E_g = [0, 1] / \sim_g$ modulo \sim_g est considéré comme un arbre, et ses éléments sont les noeuds. Si on réfléchit en discret, on a ajouté à l'ensemble des noeuds discrets "les points" des arêtes. La surjection canonique F_g de $[0, 1]$ sur E_g

$$\begin{aligned} F_g : [0, 1] &\longrightarrow E_g \\ x &\longmapsto F_g(x) = \dot{x} := \{y \mid y \in [0, 1], x \sim_g y\} \end{aligned}$$

associe à chaque réel x la classe $\{y \mid y \in [0, 1], x \sim_g y\}$; l'ordre du parcours en profondeur se retrouve en prenant disant que $\dot{x} \leq \dot{y}$ si $\min \dot{x} \leq \min \dot{y}$. Bien sûr, la classe $\dot{0}$ est la racine de E_g .

La distance sur l'arbre se retrouve: soit x et y deux représentants de \dot{x} et de \dot{y} ; on définit la distance de \dot{x} à \dot{y} dans l'arbre comme

$$d_g(\dot{x}, \dot{y}) = d_{\mathcal{T}}(x, y) = g(x) + g(y) - 2\check{g}(x, y).$$

La fonction g joue le rôle du contour de E_g puisque

$$d_g(\dot{0}, \dot{x}) = g(x)$$

comme dans le cas discret. De là, on perçoit immédiatement qu'on peut définir sur le graphe de cette fonction g , la notion d'ancêtre, descendant, sous arbres, branches, etc. Par ailleurs, on peut aussi montrer à la main que E_g ne possède pas de cycle et est connexe. C'est un arbre, sans conteste.

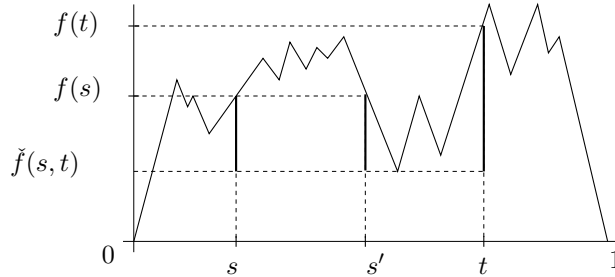


Figure 6: Graphe d'une fonction f de $C^+[0, 1]$. Dans cet exemple $s \equiv_f s'$ et la distance $d_f(s, t) = d_f(s', t) = f(s) + f(t) - 2\check{f}(s, t)$.

Soit $K = \{t_g := (E_g, d_g) \mid g \in C^+[0, 1]\}$ l'ensemble des arbres ainsi construits. On veut faire de cet espace un espace polonais (toujours la même obsession): ben, c'est facile, **on décrète** que la distance entre 2 arbres est donnée par la distance entre les fonctions qui les définissent: la distance dans K entre t_g et t_f est

$$d_K(t_g, t_f) = \|g - f\|_{\infty}.$$

Ca fait immédiatement de K un espace polonais, et bien sûr, la si la suite (f_n) converge vers f dans $C[0, 1]$, alors $t_{f_n} \rightarrow t_f$ dans K . De même la convergence faible est la même notion dans les 2 espaces $C^+[0, 1]$ et dans K : si $x_n \xrightarrow{(d)} x$ dans $C^+[0, 1]$ alors $t_{x_n} \xrightarrow{(d)} t_x$ dans K munie de sa tribu borélienne (héritée, une fois encore).

Pour comprendre la puissance de l'approche, il reste à observer plusieurs choses:

– la distance d_K est une distance entre des arbres. Est-ce une bonne distance? C'est une bonne question, en tout cas. Ben voui, c'est une bonne distance, car pour cette distance, 2 arbres t_g et t_f sont proches, disons à distance $< \varepsilon$, lorsque $\|f - g\|_\infty < \varepsilon$. Mais alors, lorsque c'est le cas, on voit que la classe de x respectivement dans les 2 arbres E_g et E_f , satisfont

$$|d_g(\dot{0}, \dot{x}) - d_f(\dot{0}, \dot{x})| \leq \varepsilon;$$

(les 2 noeuds ont la même distance à la racine à ε près. Et un petit calcul, ou dessin, montre que

$$\max_{u, v \in [0, 1]} |d_g(\dot{u}, \dot{v}) - d_f(\dot{u}, \dot{v})| \leq 3\varepsilon;$$

donc, en tant qu'espace métrique, les 2 arbres E_g et E_f sont proches.

Théorème 3.35. Soit \mathbf{p} une loi de probabilité sur \mathbb{N} de moyenne 1, et de variance $0 < \sigma^2 < +\infty$. Pour $n \rightarrow +\infty$ (dans l'ensemble des tailles possibles pour un arbre dont les degrés appartiennent à $\{k : p_k > 0\}$), le processus de contour $(C_k, 0 \leq k \leq 2n)$ d'un arbre de Galton-Watson de loi de reproduction \mathbf{p} conditionné à avoir n noeuds, normalisé comme suit:

$$\left(\frac{C_{2nt}}{\sqrt{n}} \right)_{0 \leq t \leq 1} \xrightarrow{(d)} \frac{2}{\sigma} (e_t)_{0 \leq t \leq 1}$$

où $(e_t, 0 \leq t \leq 1)$ est l'excursion brownienne normalisée. En conséquence, l'arbre $t_{c_n} \xrightarrow[n]{faible} t_{\frac{2}{\sigma}e}$ pour $c_n(\cdot) = \left(\frac{C_{2n\cdot}}{\sqrt{n}} \right)$ dans l'espace des probabilités sur K .

Pour cette topologie, il y a équivalence entre la convergence du contour et la convergence de l'arbre; les sous produits, à savoir ce qu'on peut déduire de cette convergence sont énormes: pour toute fonction(nelle) continue $F : C[0, 1] \rightarrow (E, \mathcal{E})$ dans un espace polonais quelconque, on e

$$F(c_n) \xrightarrow{(d)} F\left(\frac{2}{\sigma}(e_t)_{0 \leq t \leq 1}\right);$$

par exemple, si

$$F(f) = \left(\max(f), \int_0^1 f(t)dt, f(1/2) \right)$$

on voit qu'appliquée à notre modèle d'arbres, la convergence du contour implique la convergence jointe de la hauteur, de la longueur de cheminement, et de la hauteur du noeud qui est "au milieu" pour l'ordre lexicographique.

Autre notion de convergence: La topologie de Gromov-Hausdorff

Un espace métrique pointé est un triplet (X, d, x) où (X, d) est un espace métrique, et x un point distingué de X .

Deux tels espaces (X, d, x) et (X', d', x') sont dits isométriques s'il existe une isométrie (bijective donc) $\phi : X \rightarrow X'$ telle que $\phi(x) = x'$. La classe d'isométrie d'un espace métrique pointé (X, d, x) est l'ensemble des espaces métriques isométriques à (X, d, x) . Naturellement on peut considérer les arbres comme des espaces métriques pointés en leur racine, avec, bien sûr, comme métrique, la distance de graphe. Formellement, pour \mathbf{t} un arbre et $u, v \in \mathbf{t}$, la distance dans \mathbf{t} est donnée par

$$d_{\mathbf{t}}(u, v) = |u| + |v| - 2|u \wedge v|,$$

où bien sûr, $|u \wedge v|$ est l'ancêtre commun de u et v dont la hauteur est maximale. Deux arbres planaires sont égaux en tant qu'espaces métriques pointés, si on peut passer de l'un à l'autre par une suite d'échanges de sous-arbres enracinés en des frères.

Ce qu'on vient de faire sur les arbres, on peut le faire sur des graphes, des arbres, etc.

Ces dernières années, en particulier dans le cadre des cartes planaires, il a été imaginé qu'une bonne façon de décrire leur limite pouvait l'être au niveau de leur propriété métrique: une fois encore, l'obsession du chercheur de limite est de trouver un espace raisonnable, muni d'une topologie métrique non triviale, susceptible d'accueillir les cartes discrètes et leur limite. L'astuce imaginée a donc été de considérer ces objets comme des espaces métriques compacts; on perd un peu d'information topologique sur ces objets en faisant cela, par exemple, concernant l'ordre relatif des arêtes autour des sommets, mais on gagne un espace sur lequel travailler.

Voici ce qu'il en est. Il s'agit donc de travailler sur l'espace des classes d'isométries M d'espace métriques compacts, munie d'une distance, la fameuse distance de Gromov-Hausdorff (il ne s'agit pas tout à fait d'une distance sur les espaces compacts, car on va le voir, les espaces isométriques seraient à distance 0, donc, on travaille sur les classes d'isométries).

La distance de Gromov-Hausdorff entre classes d'isométries d'espaces pointés est définie par

$$d_{GH}((X, d, x), (X', d', x')) = \inf_{\phi, \phi'} \delta_H(\phi(X), \phi'(X')) \vee \delta(\phi(x), \phi'(x')), \quad (3.11)$$

l'infimum est pris sur les plongements isométriques ϕ, ϕ' de $(X, d), (X', d')$ dans un espace métrique commun (Z, δ) , et où δ_H est la distance de Hausdorff habituelle entre sous ensemble compacts de Z .

Une fois encore, dans les faits, on ne travaille pas directement avec cette topologie, car bien qu'elle ait toutes les bonnes propriétés (notre espace de travail est polonais), personne ne songe à vraiment étudier toutes les plongements isométriques simultanés de 2 espaces métriques dans le même espace. En fait, prosaïquement, on cherche plutôt une bonne paramétrisation des objets, sur lesquels on va travailler, et démontrer une convergence: par exemple, on va démontrer que le contour des arbres converge en loi. Cette convergence impliquera celle sous Gromov-Hausdorff. En fait, si on y songe un peu, la convergence de Gromov-Hausdorff est en quelque sorte, la plus faible des convergence raisonnable pour énoncer la convergence d'espace métrique.

Pour les cartes, on utilise donc les bijections qui envoient typiquement les cartes sur des paires d'arbres (ou des arbres étiquetés), et on montre que la convergence des arbres associés est suffisamment fidèle, pour entraîner la convergence de la métrique des cartes (c'est extrêmement difficile).

4 Quelques pointeurs bibliographiques

Voici des articles / livres / liens traitant de:

la convergence des marches aléatoires normalisées:

On trouve dans Billingsley [3] tout le matériel concernant la convergence des mesures sur \mathbb{R} , $C[0, 1]$ ou $D[0, 1]$. Une grande référence: Kallenberg [12] quand on veut le plus grand degré de généralité sur n'importe quel type de convergence (y compris, convergence de mesure aléatoire).

Sur ma page, un document, dans la partie "Cours" intitulé : "Cours M2 GRAZ : limit of random walks and of random trees." J'y détaille ces histoires de convergence dans $C[0, 1]$, la convergence du contour des arbres de Galton-Watson conditionnés par être de taille n , et énonce des sous-produits.

la convergence des arbres aléatoires normalisés:

Pour la convergence d'arbre, on trouve pas mal de ressources, à commencer par les articles de D.Aldous (Continuum random tree 1, 2, et 3) où le CRT est défini, le livre de J. Pitman [17], des cours JF Le Gall [14], G. Miermont [11]; pas mal de résultats récents, de N. Broutin, N. Curien, I. Kortchemski, L. Addario-Berry, T. Duquesne, JFM, ... ; un papier attrapant pas mal de modèles : G. Miermont- B. Haas [9]

Les arbres en $\log n$ ne convergent pas si on les normalise de manière homogène. Leur profil converge, leur hauteur, mais pas la structure globale, en tout cas, pas dans $C[0, 1]$ ni $D[0, 1]$.

la convergence des cartes aléatoires normalisés ou non

Beaucoup de résultats ces dernières années pour les cartes normalisées, à commencer par Ph. Chassaing - G. Schaeffer [4], JFM - Mokrjadem, G. Miermont, JF Le Gall, M. Albenque, E. Fusy, J. Bettinelli, N. Curien, ...

Non normalisées: O.Angel-O.Schramm, Ph.Chassaing-B.Durhuus, L Ménard, N. Curien, JF Le Gall, O. Bernardi, ...

Des résultats également pour les graphes planaires aléatoires (M. Drmota, O. Bernardi, E. Fusy, G. Collet, M. Noy,..)

5 Convergence de structures aléatoires p.s. ou en proba

Bon, alors, après avoir répété que souvent, pour les structures combinatoires, la bonne notion de convergence était plutôt celle de la convergence en loi, il se trouve qu'un certain nombre de cas particuliers viennent prouver que parfois, une convergence plus forte a lieu:

– Grossissement des marches simples (chemin de Dyck, ponts, méandres) à la Philippe Marchal [15]: Philippe a montré que l'on pouvait construire, pour chacun de ces modèles, une suite $(C_n, n \geq 0)$ de chemins aléatoires, t.q. pour tout n , C_n a la loi visée (uniforme dans la classe choisie), et t.q. C_n bien normalisé converge p.s. (avec une vitesse de convergence en $1/n^{1/4}$). Autrement dit, ces marches convergent en loi, mais on peut, sur un bon espace de probabilité, les relier par une bonne structure de dépendance qui entraîne leur limite en loi.

Il existe un procédé de grossissement pour les quadrangulations (Jérémy Bettinelli [1]) mais pour lequel, à ma connaissance, on ne sait pas si le processus induit converge p.s..

- Profil des arbres binaires de recherche, arbre m -aire: dans ces cas, on fait grossir l'arbre à l'infini, par l'insertion de données. Les arbres apparaissent alors naturellement sur le même espace de probabilité, et il se trouve que sur cette espace, le profil normalisé converge ps; cela arrive dans un sens très fort dans le cas des arbres m aires, puisque cela apparaît à plusieurs ordres (plusieurs travaux de B. Chauvin - N. Pouyanne, HK. Hwang, ..).
- Urnes: l'étude des urnes de Polya est diverse, et plusieurs travaux de la communauté bien connus, ont été discutés ces dernières années dans la communauté; dans de nombreux cas, au premier ordre, l'occupation de l'urne est déterministe (S. Janson [10], Ph. Flajolet, J. Gabarró et H. Pekari [7], B. Chauvin, N. Pouyanne, R. Sahnoun [5], M. Knappe, R. Neininger [13], ...

6 Appendix: Probability measure on a Polish set

Let \mathcal{S} be a set and ρ a distance on \mathcal{S} . For x in \mathcal{S} , $r \geq 0$, the ball $B(x, r)$ of center x and radius r , is the set

$$B(x, r) = \{y \in \mathcal{S}, \rho(x, y) < r\};$$

an open set of \mathcal{S} is any union of balls. The family of open subsets of a space forms the standard topology (a topology is a collection of subsets of a space E , stable by union, by finite intersection, containing E and the empty set).

A set equipped with a distance (\mathcal{S}, ρ) is called a **Polish space**, if it is separable and complete:

- separable: a countable subset D of S is dense (D is dense in S , if for any $s \in S$, there exists an element of D as close as wanted to s : for any $\varepsilon > 0$ there exists $d \in D$ such that $\rho(s, d) < \varepsilon$),
- complete: a space \mathcal{S} is complete if any Cauchy sequence of \mathcal{S} converges in \mathcal{S} (has a limit, and the limit is inside \mathcal{S}). A sequence (s_n) is a Cauchy sequence, if for any $\varepsilon > 0$, there exists N such that if $n \geq N$ and $m \geq N$ then $|x_n - x_m| \leq \varepsilon$.

Hence \mathbb{R} is a Polish space when equipped with the distance $d(x, y) = |x - y|$, since:

- \mathbb{Q} is countable and dense,
- The Cauchy sequences converge: a Cauchy sequence (x_n) is clearly bounded, then as some accumulations points. The set of accumulation points is necessarily of cardinality 1 (since accumulation points are at distance smaller than ε , for any ε , clearly).

The **Borelian** σ -field $\mathcal{B}(\mathcal{S})$ is the σ -field induced by the standard topology: this is the set of subsets of \mathcal{S} that are obtained by a (at most) countable sequence of unions, intersections, complementations, applied on the set of open sets of \mathcal{S} .

Hence, the open sets, closed sets, singleton, sets with a countable cardinality... are all Borelians sets. In fact, what is difficult is to find a non Borelian set (even on \mathbb{R} , it is not at all a simple exercise).

6.1 Measure on Polish space. Characterization

A function f defined on \mathcal{S} , taking its value in \mathbb{R} is continuous, if for any open set \mathcal{O} of \mathbb{R} , $f^{-1}(\mathcal{O})$ is an open set of \mathcal{S} . The function f is said to be measurable (or Borelian) if for any B in $\mathcal{B}(\mathbb{R})$, $f^{-1}(B) \in \mathcal{B}(\mathcal{S})$. It is easy to see, that a sufficient condition, is that $f^{-1}(\mathcal{O}) \in \mathcal{B}(\mathcal{S})$ for any open set \mathcal{O} of \mathbb{R} . In particular, the continuous functions from $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$ to \mathbb{R} are measurable.

The measure \mathbb{P} is sufficient to define a notion of integral for measurable non negative functions f :

$$\int f d\mathbb{P};$$

this is done in several steps : if $f = \mathbb{I}_A$ (is a indicator function of a set) where $A \in \mathcal{B}(\mathcal{S})$, set

$$\int_S \mathbb{I}_A(\omega) d\mathbb{P}(\omega) = \int_A d\mathbb{P}(\omega) = \mathbb{P}(A).$$

Prolong by linearity : let A_1, \dots, A_n a family of Borelian (disjoint or not) from $\mathcal{B}(\mathcal{S})$ and any real numbers $\lambda_1, \dots, \lambda_n$; for f defined by $f = \sum_{i=1}^n \lambda_i \mathbb{I}_{A_i}$, set

$$\int_S f(\omega) d\mathbb{P}(\omega) = \int \sum_{i=1}^n \lambda_i \mathbb{I}_{A_i}(\omega) d\mathbb{P}(\omega) = \sum_{i=1}^n \lambda_i \int_{A_i} d\mathbb{P}(\omega) = \sum_{i=1}^n \lambda_i \mathbb{P}(A_i).$$

This construction is valid, since it can be checked that the result does not depend on the representation of a given function f .

The notion of integral is then defined for the family of step functions (positive or not). Then one shows that one may build an increasing “function integral” and linear (that is, $f \longrightarrow \int f d\mathbb{P}$ is increasing and linear), on (first) the set of measurable non negative functions. For this, we use that any measurable function f , non negative is a limit of an increasing sequence of non negative scale functions f_n : for example,

$$f_n(x) = \begin{cases} [10^n f(x)]/10^n & \text{for } f(x) \leq n \\ n & \text{for } f(x) > n \end{cases}$$

which is the below approximation at 10^{-n} of $f(x)$ for $f(x) < n$. It is easy to check that f_n is measurable, is a step function, and that f_n is increasing.

One then sets
$$\int f d\mathbb{P} = \lim_n \int f_n d\mathbb{P}. \tag{6.12}$$

The integral of a positive bounded function f is always finite (since the measure is finite, $\mathbb{P}(\mathcal{S}) = 1$). The integral of a measurable function f is defined in the case where $\int f^+ d\mathbb{P}$ and $\int f^- d\mathbb{P}$ are finite, by

$$\int f d\mathbb{P} = \int f^+ d\mathbb{P} - \int f^- d\mathbb{P};$$

if only one of the integral $\int f^+ d\mathbb{P}$ or $\int f^- d\mathbb{P}$ diverges, we say accordingly that $\int f d\mathbb{P} = +\infty$ or $-\infty$. The words “one then sets” in (6.12) is a hidden theorem: one has to show that the limit does not depend on the chosen sequence f_n .

Lemme 6.36. Any probability measure on $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$ is regular, that is, for A in \mathcal{S} and $\varepsilon > 0$, there exists a closed set F and an open set \mathcal{O} such that $F \subset A \subset \mathcal{O}$ and $\mathbb{P}(\mathcal{O} - F) < \varepsilon$.

Proof. Let E be the set of subsets of A in \mathcal{S} satisfying the property stated in the Lemma. Let ρ denote the metric on \mathcal{S} ; $\rho(x, A)$ is the distance between x and the set A . If A is closed, take $F = A$ and $\mathcal{O} = \mathcal{O}_\delta = \{x, \rho(x, A) < \delta\}$ for δ chosen small enough (indeed, $\mathcal{O}_{1/n}$ is decreasing, and $\lim_n \mathbb{P}(\mathcal{O}_{1/n}) = \mathbb{P}(\lim \mathcal{O}_{1/n}) = \mathbb{P}(\overline{A}) = \mathbb{P}(F)$). Hence, every closed set belongs to E . Let us show that the family of sets E having this property is a σ -field (since it has already been shown that it contains the closed sets, it would contain the Borelian sets). Let A_n be a sequence of elements of E . Consider some open sets \mathcal{O}_n and closed set F_n such that

$$F_n \subset A_n \subset \mathcal{O}_n \text{ and } \mathbb{P}(\mathcal{O}_n - F_n) < \varepsilon/2^{n+1}.$$

Then, take $\mathcal{O} = \cup_n \mathcal{O}_n$ and $F = \cup_{n \leq n_0} F_n$ (n_0 is chosen such that $\mathbb{P}(\cup_n F_n - F) < \varepsilon/2$). One then has $A = \cup_n A_n \in E$. Since E is stable by taking the complementary, this ends the proof. \square

Lemma 6.37. Two measures on $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$ which coincide on the closed sets of \mathcal{S} (or on the open sets of \mathcal{S}) are equal.

Proof. Two measures \mathbb{P} and \mathbb{P}' on $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$ are equal if they are equal on all Borelian sets. The closed set generates the Borelian σ -field, but this does not provide a proof of the Lemma.

First, by taking the complementary, it suffices to prove this lemma for closed sets. Let us see why this is a consequence of the previous Lemma. Let A be a Borelian and F and \mathcal{O} as said in the previous Lemma. One has

$$\mathbb{P}(F) \leq \mathbb{P}(A) \leq \mathbb{P}(\mathcal{O}) = 1 - \mathbb{P}(\mathcal{O}^c).$$

Hence, to know the probability on the closed sets is sufficient to determine the probability of any Borelian. \square

To go faster, one can also use a Theorem about Dynkin systems, which states the following result. (A π -system of Dynkin is a set of subsets stable by finite intersection).

Lemma 6.38. If two measures on $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$ coincides on a π -system which generates $\mathcal{B}(\mathcal{S})$, then, they coincide on $\mathcal{B}(\mathcal{S})$.

(Billingsley, Probability and measure, Theorem 3.3). Open sets forms a π -system, and generate the Borelian σ -field.

Associated with \mathbb{P} , a notion of integral has been defined. If \mathbb{P} and \mathbb{P}' are two measures on $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$ such that for any measurable function f , one has $\int f d\mathbb{P} = \int f d\mathbb{P}'$ then $\mathbb{P} = \mathbb{P}'$ (that is, for all A in $\mathcal{B}(\mathcal{S})$, one has $\mathbb{P}(A) = \mathbb{P}'(A)$). This is clear, since it suffices to take $f = \mathbb{I}_A$ to check it. Which is less obvious is the following proposition:

Proposition 6.39. Let \mathbb{P} and \mathbb{P}' be two probability measures on $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$ such that $\int f d\mathbb{P} = \int f d\mathbb{P}'$ for all bounded continuous function f (with values in \mathbb{R}). Under these conditions, one has $\mathbb{P} = \mathbb{P}'$.

Proof. To prove this proposition, we will show that the two measures coincide on closed sets (and then, we use Lemma 6.37). For this, we approximate the function \mathbb{I}_F of a closed set F by bounded continuous functions. We make here the same work as on \mathbb{R} : let ψ defined on \mathbb{R} taking its values in \mathbb{R} defined by

$$\psi(t) = \begin{cases} 1 - t & \text{if } 0 \leq t \leq 1 \\ 0 & \text{if } t \geq 1 \end{cases}.$$

The function f_n defined by $f_n(x) = \psi(n\rho(x, F))$ is equal to 1 if $x \in F$, 0 if $\rho(x, F) > 1/n$ and is $[0,1]$ in the other case (it is uniformly continuous).

The sequence of functions f_n is decreasing and converge point-wise toward \mathbb{I}_F . Since all the f_n are dominated by the constant function $g = 1$ on \mathcal{S} (which is integrable), by the Theorem of Lebesgue (of dominated convergence) one has

$$\lim_n \int_{\mathcal{S}} f_n(x) d\mathbb{P}(x) = \int_{\mathcal{S}} \mathbb{I}_F(x) d\mathbb{P}(x) = \mathbb{P}(F).$$

Then, one gets $\mathbb{P}'(F) = \lim_n \int_{\mathcal{S}} f_n(x) d\mathbb{P}'(x)$. Hence $\mathbb{P}(F) = \mathbb{P}'(F)$. \square

6.2 Random variables with a values in a Polish space

Let $(\Omega, \mathcal{A}, \mathbb{P})$ a probability space, and $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$ a Polish space, equipped with its Borelian σ -field.

A map $X : \Omega \rightarrow \mathcal{S}$ is a random variable if X is measurable from (Ω, \mathcal{A}) to $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$ (that is, $X^{-1}(b) \in \mathcal{A}$ for any $b \in \mathcal{B}(\mathcal{S})$). The distribution of X , denoted by \mathbb{P}_X is the probability measure defined on $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$ by

$$\mathbb{P}_X(b) = \mathbb{P}(X^{-1}(b)) = \mathbb{P}(\{\omega \mid X(\omega) \in b\}).$$

Let f be a measurable function from $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$ with values in \mathbb{R} (equipped with $\mathcal{B}(\mathbb{R})$). The expectation of $f(X)$ is defined by

$$\mathbb{E}(f(X)) = \int f(x) d\mathbb{P}_X(x).$$

6.3 Convergence in distribution in a Polish space

Définition 6.40. Let $\mathbb{P}, \mathbb{P}_1, \mathbb{P}_2, \dots$ be a sequence of probability measure defined on the same Polish space $(\mathcal{S}, \mathcal{S})$. The sequence $(\mathbb{P}_n, n \geq 1)$ weakly converges to \mathbb{P} (we write $\mathbb{P}_n \xrightarrow[n]{faible} \mathbb{P}$), if for any bounded continuous function f (from \mathcal{S} to \mathbb{R}),

$$\int f d\mathbb{P}_n \longrightarrow \int f d\mathbb{P}. \tag{6.13}$$

A sequence of r.v. (X_n) converges in distribution to X if $\mathbb{P}_{X_n} \xrightarrow[n]{faible} \mathbb{P}_X$, in other words, if

$$\mathbb{E}(f(X_n)) = \int f(x) d\mathbb{P}_{X_n}(x) \longrightarrow \int f(x) d\mathbb{P}_X(x) = \mathbb{E}(f(X)),$$

for bounded continuous functions f . We note $X_n \xrightarrow[n]{d} X$.

We have

Proposition 6.41. Let $(X_n, n \geq 1)$ be a sequence of r.v. taking their values in a Polish space (S, \mathcal{S}) . If $X_n \xrightarrow[n]{faible} X$ then for any continuous function f from S to a Polish space (S', \mathcal{S}') , then $g(X_n) \xrightarrow[n]{d} g(X)$.

Proof. It suffices to show that for h bounded continuous, from S' to \mathbb{R} , we have $\mathbb{E}(h(g(X_n))) \rightarrow \mathbb{E}(h(g(X)))$. Since $h \circ g$ is bounded continuous from S to \mathbb{R} , it is indeed the case since $X_n \xrightarrow[n]{faible} X$. \square

References

- [1] J. Bettinelli. Increasing forests and quadrangulations via a bijective approach. J. Combin. Theory Ser. A, 122(0):107–125, 2014.
- [2] P. Billingsley. Probability and Measure. John Wiley and Sons, second edition, 1986.
- [3] P. Billingsley. Convergence of probability measures. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.
- [4] P. Chassaing and G. Schaeffer. Random Planar Lattices and Integrated SuperBrownian Excursion. Probability Theory and Related Fields, 128(2):161–212, 2004. 44 pages, 22 figures. Slides and extended abstract version are available at <http://www.loria.fr/~schaeffe/Pub/Diameter/> and <http://www.iecn.u-nancy.fr/~chassain/>.
- [5] B. Chauvin, N. Pouyanne, and R. Sahnoun. Limit distributions for large Pólya urns. Ann. Appl. Probab., 21(1):1–32, 2011.
- [6] W. Feller. An introduction to probability theory and its applications. Vol. II. Second edition. John Wiley & Sons Inc., New York, 1971.
- [7] P. Flajolet, J. Gabarró, and H. Pekari. Analytic urns. Ann. Probab., 33(3):1200–1233, 05 2005.
- [8] P. Flajolet and R. Sedgewick. Analytic Combinatorics. Cambridge University Press, New York, NY, USA, 1 edition, 2009.
- [9] B. Haas and G. Miermont. Scaling limits of markov branching trees with applications to galton–watson and random unordered trees. Ann. Probab., 40(6):2589–2666, 11 2012.
- [10] S. Janson. Functional limit theorems for multitype branching processes and generalized Pólya urns. Stochastic Process. Appl., 110(2):177–245, 2004.
- [11] G. M. JF Le Gall. Scaling limits of random trees and planar maps. Lecture notes for the Clay Mathematical Institute Summer School in Buzios, disponible sur la page de Grégory Miermont.
- [12] O. Kallenberg. Foundations of modern probability. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.
- [13] M. Knape and R. Neininger. Pólya urns via the contraction method. Combin. Probab. Comput., 23(6):1148–1186, 2014.
- [14] J.-F. Le Gall. Random trees and applications. Probab. Surveys, 2:245–311, 2005.

- [15] P. Marchal. Constructing a sequence of random walks strongly converging to Brownian motion. In C. Banderier and C. Krattenthaler, editors, Discrete Random Walks, DRW'03, volume DMTCS Proceedings vol. AC, Discrete Random Walks (DRW'03) of DMTCS Proceedings, pages 181–190, Paris, France, 2003. Discrete Mathematics and Theoretical Computer Science.
- [16] R. Neininger and H. Sulzbach. On a functional contraction method. Ann. Probab., 43(4):1777–1822, 07 2015.
- [17] J. Pitman. Combinatorial stochastic processes : Ecole d ete de probabilites de saint-flour xxxii - 2002 / jim pitman, jean picard. 03 2018.
- [18] J. Shohat, J. Tamarkin, and A. M. Society. The problem of moments. Mathematical surveys and monographs. American Mathematical Society, 1943.