MASTER DE MATHÉMATIQUES M1 - SORBONNE UNIVERSITÉ
SUJETS T.E.R. - 2023/2024

FRÉDÉRIC KLOPP
RESPONSABLE DES T.E.R.

Date: 5 avril 2024.

Encadrant(e)(s)	Bureau	Courriel	Intitulé du ou des sujets	Étudiant(e)(s)
Anne-Laure Basdevant	16-26-125	basdevant@lpsm.paris	Graphe d'Erdös Rényi	Baptiste Le Duigou
Motasem Alrahabi Gérard Biau	15-25-219	motasem.alrahabi@sorbonne-universite.fr	Extraction automatique de mots clés à partir de textes	Nacef Ben Mansour
Jérémie Bouttier	15-16-116A	jeremie.bouttier@sorbonne-universite.fr	Orientations pfaffiennes des graphes	Thomas De Myttenaere
Dario Cordero-Erausquin	16-26-418	dario.cordero@imj-prg.fr	The Johnson-Lindenstrauss flattening Lemma	Xiaotang Sun
Nguyen Viet Dang	16-26-523	dang@imj-prg.fr	Du théorème de Gauss Bonnet aux classes de Chern.	Benjamin Mutenfu
Adrien Deloro	15-25-519	adrien.deloro@imj-prg.fr	Cohérence de l' arithmétique de Peano	Hugo Galland
Cyril Demarche	16-26-513	cyril.demarche@imj-prg.fr	Courbes algébriques et arithmétique	Tianyang Wang
			Autour des théorèmes de transcendance de Baker	Jihyuk Seo
Antoine Ducros	16-26-507	antoine.ducros@imj-prg.fr	Introduction au langage des catégories	Thomas Moore
Javier Fresán	15-25-430	javier.fresan@imj-prg.fr	Quotients de factorielles et monodromie de fonctions hypergéométriques	
Omer Friedland	16-26-408	omer.friedland@imj-prg.fr	Autour du théorème de Dvoretzky	Tomasz Kostrzewa
David García-Zelada	16-26-123	david.garcia-zelada@sorbonne-universite.fr	Notions catégoriques en topologie et géométrie algébrique	Louis Burton
Frédéric Klopp	16-26-412	frederic.klopp@imj-prg.fr	Théorie de Floquet	Sameo Herry
Camille Laurent	16-26-312	laurent@ann.jussieu.fr	EDP dispersives et géométrie	Adil Ihtajja
Julien Marché	16-26 518	julien.marche@imj-prg.fr	Isomorphismes de De Rham	Maxime Debois
Jean-Pierre Marco	15-25-517	jean-pierre.marco@imj-prg.fr	Géométrie et topologie de l' information	Alexandre Bestandji
Sébastien Martineau	16-26 210	sebastien.martineau@sorbonne-universite.fr	Une initiation pas à pas à la théorie des catégories avec un regard tourné vers les applications	Garance Henrion
			Mathématiques et crise environnementale	Charlie Marsily
Frédéric Naud	16-26 525	frederic.naud@imj-prg.fr	Groupes fuchsiens arithmétiques	Lucas Vaudron
Frédéric Paugam	15-25-518	frederic.paugam@imj-prg.fr	Les schémas spectraux comme treillis annelés	Tianjian Tan
Nastassia Pouradier Duteil	16-26-331	nastassia.pouradierduteil@ljll.math.upmc.fr	Etude du flocking dans des modèles de type Cucker-Smale	Jean Leroy
Marco Robalo	15-25-512	marco.robalo@imj-prg.fr	Catégories et Homologie	Hongxiao Bu
Stéphane Robin	15-25-221	stephane.robin@sorbonne-universite.fr	Estimation du squelette d' un modèle graphique	Malo Ruellan
			Segmentation / classification dans un processus de Poisson	Roxane Cellier
			Classification non supervisée d'espèces par SDM	Louisa Djerbroun
Etienne Roquain	15-16-213	etienne.roquain@upmc.fr	Interpolation et sur-apprentissage en estimation non-paramétrique	Aidan Buttimore
			Autour de la prédiction conformelle	Théo René
Maxime Sangnier	15-25-224	maxime.sangnier@upmc.fr	Apprentissage supervisé large échelle	Tiziano Fassina Yassine Ghariani
Benoît Stroh	15-25-413	benoit.stroh@imj-prg.fr		
Jean-Baptiste Teyssier	15-25-512	jean-baptiste.teyssier@imj-prg.fr	Groupe fondamental en géométrie algébrique	Pierre Nouhaud

Sujet TER : Graphe d'Erdös Rényi

proposé par Anne-Laure Basdevant

En 1959, Erdös et Rényi ont introduit un modèle de graphes aléatoires permettant de modéliser de grands réseaux. Etant donné un ensemble de sommets, $\{1, \ldots, n\}$, et un paramètre $p \in [0, 1]$, le graphe $\mathcal{G}(n, p)$ est obtenu en plaçant de manière i.i.d. une arête entre chaque couple de sommets $i \neq j$ avec probabilité p. Ce modèle, assez ancien maintenant, a été beaucoup étudié mais reste encore aujourd'hui un sujet de recherche actif. Dans ce TER, nous nous s'intéresserons à décrire différentes propriétés de connectivité de ce graphe, comme par exemple les suivantes :

- Le graphe est-il connexe ?
- Quelle est la taille de la plus grande composante connexe du graphe ?
- Le graphe possède-t-il des cycles ?
- Quel est le diamètre du graphe ?

Les réponses à ces questions dépendent de la valeur du paramètre p. Nous montrerons en particulier qu'une transition de phase apparait lorsque l'on fait varier p avec n et nous décrirons les valeurs critiques pour p en fonction de n.

Bibliographie

1. Van der Hofstad R. Random Graphs and Complex Networks. Cambridge University Press ; 2016.
2. N. Curien. Notes de cours "Random graphs". https ://www.imo.universite-paris-saclay.fr/ nicolas.curien

Extraction automatique de mots clés à partir de textes

Motasem Alrahabi / Gérard Biau

L'extraction de mots-clés demeure un processus crucial dans la sélection des termes les plus importants et représentatifs d'un document ou d'un corpus. Cette technique offre une variété d'applications, que ce soit pour condenser le contenu d'un texte, améliorer les performances de recherche d'informations, ou encore analyser des tendances émergentes. Les approches courantes d'extraction de mots-clés englobent la fréquence des termes (comme BM25, TF-IDF...), les graphes de mots, les modèles d'apprentissage automatique, ainsi que les grands modèles de langage. Dans ce contexte, le présent travail explore différentes perspectives à aborder dans le cadre de ce TER. Après une première phase d'état des lieux des travaux récents sur l'extraction des mots-clés, l'objectif est d'approfondir les expérimentations avec diverses bibliothèques existantes (Yake, KeyBERT, Textrank...) et les modèles pré-entraînés (Mistral, GPT...), en vue de les comparer rigoureusement. Cette comparaison permettra d'identifier les choix techniques les plus pertinents pour l'identification et l'évaluation des résultats. L'étudiant(e) sera alors en mesure d'appliquer les solutions retenues sur des corpus diversifiés, englobant des livres, publications, blogs, etc. Cette diversité de corpus s'avère cruciale pour tester la robustesse des méthodes sélectionnées dans des contextes variés, tout en évaluant leur adaptabilité à une gamme étendue de contenus.

Bibliographie

Campos, R., Mangaravite, V., Pasquali, A., Jorge, A. M., Nunes, C., & Jatowt, A. (2018). A Text Feature-Based Automatic Keyword Extraction Method for Single Documents. In Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC).

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL).

Mihalcea, R., & Tarau, P. (2004). TextRank: Bringing Order into Texts. In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP).

Sharma, P., & Yingbo Li. (2019). "Self-Supervised Contextual Keyword and Keyphrase Retrieval with Self-Labelling" Preprints. https://doi.org/10.20944/preprints201908.0073.v1

Orientations pfaffiennes des graphes

Sujet de TER proposé par Jérémie Bouttier (IMJ-PRG)

Année 2023-2024

Étant donné un graphe, un *couplage parfait* est une partie C de l'ensemble des arêtes telle que tout sommet du graphe est incident à une et une seule arête dans C. Les couplages parfaits sont des objets fondamentaux de la combinatoire, liés notamment au classique «lemme des mariages», mais ils jouent également un rôle en physique statistique dans le cadre des «modèles de dimères» [1].

Compter le nombre de couplages parfaits d'un graphe donné est, en général, un problème difficile (il est «#P-complet» au sens de la théorie de la complexité). Dans le cas des graphes plans, Kasteleyn a donné en 1961 un algorithme de comptage [2], qui consiste à orienter les arêtes du graphe selon une certaine règle, et construire ainsi une matrice d'adjacence signée dont le pfaffien (ou le déterminant dans le cas biparti) donne le nombre de couplages parfaits. Cet algorithme est efficace car tant l'orientation que le pfaffien peuvent être calculés en temps polynomial en la taille du graphe.

Au-delà du cas plan, la méthode de Kasteleyn s'applique généralement aux graphes possédant des orientations dites *pfaffiennes*. Il est a priori difficile de déterminer si un graphe donné possède une orientation pfaffienne, cependant en 1999 un algorithme en temps polynomial a été donné dans le cas biparti par Robertson, Seymour et Thomas [3]. Dans le cas de ce TER, nous proposons d'étudier les concepts de théorie des graphes sous-tendant cet algorithme, et leurs ramifications. Un point de départ possible est l'article de revue [4].

Références

[1] Richard Kenyon et Andrei Okounkov. "What is ... a dimer ?" English. In : *Notices of the American Mathematical Society* 52.3 (2005), p. 342-343. URL : `https://www.ams.org/notices/200503/what-is.pdf`.

[2] P.W. Kasteleyn. "The statistics of dimers on a lattice : I. The number of dimer arrangements on a quadratic lattice". In : *Physica* 27.12 (1961), p. 1209-1225. DOI : `http://dx.doi.org/10.1016/0031-8914(61)90063-5`.

[3] Neil Robertson, P. D. Seymour et Robin Thomas. "Permanents, Pfaffian orientations, and even directed circuits". English. In : *Annals of Mathematics. Second Series* 150.3 (1999), p. 929-975. DOI : `10.2307/121059`.

[4] Robin Thomas. "A survey of Pfaffian orientations of graphs". English. In : *Proceedings of the international congress of mathematicians (ICM), Madrid, Spain, August 22–30, 2006. Volume III : Invited lectures.* Zürich : European Mathematical Society (EMS), 2006, p. 963-984. ISBN : 978-3-03719-022-7. URL : `http://people.math.gatech.edu/~thomas/PAP/pfafsurv.pdf`.

The Johnson-Lindenstrauss flattening Lemma

Dario Cordero-Erausquin

January 2024

In data science, it is common to analyse a set of N data using algorithms based on linear algebra on $\mathbb{R}^N$. When N is large, the question of reducing the dimension, without loosing too much information on the data, becomes crucial. The Johnson-Lindenstrauss Lemma, established for other purposes actually, asserts surprisingly that one can reduce the dimension up to $\log(N)$.

The precise statement is as follows. Assume we are given N vectors $v_1, \dots, v_N$ is an Euclidean space $(E, \|\cdot\|)$. Then, for a given $\epsilon \in (0,1)$, there exists a linear map u on E with rank

$$c(\epsilon)\log(N)$$

where $c(\epsilon)$ depends only on ϵ, such that, for any $i, j \leq N$

$$(1-\epsilon)\|v_i - v_j\| \leq \|u(v_i) - u(v_j)\| \leq (1+\epsilon)\|v_i - v_j\|.$$

The only known ways to obtain the map u is at random. Two descriptions will be constructed : one with random $\pm$ signs (Bernouilli variables), one using independent Gaussian variables.

For the Bernouilli case, relying on a concentration inequality by Talagrand, we will follow the "sujet de concours" *MP Centrale-Supelec 2018*. For the gaussian case, we will use the book of Boucheron-Lugosi-Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence* (Oxford, 2016) and the book of G. Pisier, *The volume of convex bodies and Banach space geometry* (Cambridge, 1989) for the proof of Gaussian concentration.

Du théorème de Gauss Bonnet aux classes de Chern.

1 Du théorème de Gauss Bonnet aux classes de Chern.

L'objectif du TER est d'abord de se familiariser suffisamment avec les notion de connexion, courbure et du calcul différentiel extérieur. Le but est de comprendre la preuve du théorème de Gauss-Bonnet dans le cas des surfaces suivant [2, 1].

Dans un second temps, si le temps le permet, nous allons étudier la définition et des exemples des classes de Chern et comprendre l'énoncé du théorème de Chern-Weil suivant [1, 3, 4].

Nous allons faire un effort pour comprendre le point de vue original de Chern lui-même dans le TER.

References

[1] Chern, Shiing-shen, Complex manifolds without potential theory. No. 15. Princeton: van Nostrand, 1967.

[2] Hélein, Frédéric. "Systèmes différentiels extérieurs." (2009): 90.

[3] Ma, Xiaonan, and George Marinescu. Holomorphic Morse inequalities and Bergman kernels. Vol. 254. Springer Science, 2007.

[4] Zhang, Weiping. Lectures on Chern-Weil theory and Witten deformations. Vol. 4. World Scientific, 2001.

Cohérence de l'arithmétique de Peano

Sujet proposé par Adrien Deloro

Printemps 2024

C'est un sujet de logique mathématique, mais avec un regard vers la théorie de la démonstration. Soit PA l'arithmétique élémentaire « de Peano », i.e. la théorie du premier ordre décrivant l'addition et la multiplication des entiers avec *schéma de récurrence* sur toutes les formules.

En 1931 [Göd31], Kurt Gödel démontrait :

— que PA n'est pas complète,

— qu'il existe un énoncé Coh(PA) formalisant la cohérence de PA,

— que PA $\nvDash$ Coh(PA).

Dès 1936 [Gen36], Gentzen avait une démonstration de Coh(PA). Naturellement il ne le démontrait ni dans PA (exclu d'après Gödel) ni dans ZF (trivial) mais au moyen d'une remarquable analyse de preuve, s'appuyant sur une récurrence transfinie contrôlée par l'ordinal ε_0.

On propose d'étudier les concepts en jeu et la preuve de Gentzen, notamment via [Horská].

Références

[Gen36] Gerhard Gentzen. « Die Widerspruchsfreiheit der reinen Zahlentheorie ». In : *Math. Ann.* 112.1 (1936), p. 493-565 (cf. p. 1).

[Göd31] Kurt Gödel. « Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I ». In : *Monatsh. Math. Phys.* 38.1 (1931), p. 173-198 (cf. p. 1).

[Horská] Anna Horská. *Where is the Gödel-point hiding : Gentzen's consistency proof of 1936 and his representation of constructive ordinals.* SpringerBriefs in Philosophy. Springer, Cham, 2014, p. x+77 (cf. p. 1).

Courbes algébriques et arithmétique

Encadrant : Cyril Demarche
cyril.demarche@imj-prg.fr

Si K est un corps, et $f(x, y) \in K[x, y]$ un polynôme non constant, alors l'équation $f(x, y) = 0$ défini une courbe algébrique dans le plan affine sur K. Dans un premier temps, on étudiera les courbes algébriques sur un corps quelconque K, et on démontrera l'équivalence entre la catégorie des courbes projectives lisses irréductibles sur K et celle des extensions de corps de K de degré de transcendance 1. Pour cela, on se familiarisera d'abord avec le vocabulaire et les outils de la géométrie algébrique classique.

Dans un second temps, on s'intéressera au choix à l'une des deux questions suivantes sur l'arithmétique des courbes algébriques :

— les courbes elliptiques, et le théorème de Hasse affirmant que si E est une courbe elliptique sur un corps fini K à q éléments, alors $|\#E(K) - (q+1)| \leq 2\sqrt{q}$. On pourra préciser ce résultat en démontrant le cas particulier des conjectures de Weil pour les courbes elliptiques.

— le théorème de Riemann-Roch pour les courbes, et son application à la preuve de Bombieri-Stepanov des bornes de Hasse-Weil : si C est une courbe de genre g sur un corps fini K à q éléments, alors $|\#E(K) - (q+1)| \leq 2g\sqrt{q}$.

Références

[1] J. F. Dat, *Cours introductif de M2 : Variétés algébriques*, disponible sur `https://webusers.imj-prg.fr/~jean-francois.dat/enseignement/VA/VA.pdf`

[2] R. Harsthorne, *Algebraic geometry*. Graduate Texts in Mathematics. 52. New York-Heidelberg-Berlin : Springer-Verlag, 496 p. (1977).

[3] J. Silverman, *The arithmetic of elliptic curves*. Graduate Texts in Mathematics 106, 513 p. (2009).

[4] T. Tao, *The Bombieri-Stepanov proof of the Hasse-Weil bound*, disponible sur `https://terrytao.wordpress.com/2014/05/02/the-bombieri-stepanov-proof-of-the-hasse-weil-bound`

Autour des théorèmes de transcendance de Baker

Encadrant : Cyril Demarche
cyril.demarche@imj-prg.fr

Un nombre complexe est dit algébrique s'il est racine d'un polynôme non nul à coefficients rationnels. Dans le cas contraire, on dit qu'il est transcendant.

En 1882, Lindemann (suivi par Weierstrass) démontre que si α est algébrique non nul, alors e^α est transcendant, prouvant ainsi que π est transcendant. En 1934, Gelfond et Schneider démontrent la généralisation suivante de Lindemann : si α est algébrique distinct de 0 et 1, et $\beta \notin \mathbf{Q}$ est algébrique, alors α^β est transcendant.

L'un des théorèmes de Baker (1968) généralise ces résultats et affirme que si $\alpha_1, \ldots, \alpha_n, \beta_0, \beta_1, \ldots, \beta_n$ sont algébriques non nuls, alors $e^{\beta_0}\alpha_1^{\beta_1} \ldots \alpha_n^{\beta_n}$ est transcendant.

On se propose d'étudier en détail la preuve de ce théorème. Si le temps le permet, on pourra ensuite s'intéresser à des applications de ce théorème, notamment à la résolution de certaines équations diophantiennes, ou à la détermination de tous les corps de nombres de petit nombre de classes.

Références

[1] A. Baker, *Transcendental number theory*, Cambridge Mathematical Library (2nd ed.), Cambridge University Press, 1990.

[2] J. P. Serre, *Travaux de Baker (Exposé 368)*, Séminaire Bourbaki. Vol. 1969/70 : Exposés 364–381, Lecture Notes in Mathematics, vol. 180, pp. 73–86 (1971).

Projet de TER : *Introduction au langage des catégories*

Antoine Ducros
Bureau 16-26-507
antoine.ducros@imj-prg.fr

Le langage des catégories décrit de manière systématique et unifiée un grand nombre de situations formelles que l'on rencontre dans des domaines très variés des mathématiques. Son premier avantage est qu'il permet d'éviter de refaire des dizaines de fois exactement le même raisonnement dans des contextes différents, et aide, dans une une théorie donnée, à bien distinguer l'*abstract non sense* des questions plus profondes spécifiques à la théorie en question ; mais au-delà de ça, il s'avère indispensable à un grand nombre de constructions en algèbre homologique (catégories dérivées et plus récemment ∞-catégories, K-théorie...). Pour ces deux raisons, ce langage est devenu un outil majeur en topologie algébrique et en géométrie algébrique à la Grothendieck.

Le but de ce TER est de s'initier aux catégories. Il s'agira d'en découvrir les concepts de base : catégories, foncteurs, transformations naturelles, équivalences de catégories, foncteurs représentables, lemme de Yoneda, produits et coproduits, limites et colimites, adjonction...et d'essayer à chaque fois autant que faire se pourra de les illustrer par des exemples mathématiques variés : constructions de limites et colimites dans différentes catégories concrètes, mise en évidence que de nombreux foncteurs classiques sont des adjoints, etc. Si le temps le permet, on pourra ensuite aborder des questions plus avancées, comme par exemple les catégories abéliennes, les foncteurs dérivés et l'exemple des groupes de cohomologie (avec l'approche originelle, celle de Grothendieck dans le *Tohoku*).

Références :

(1) A. Ducros, *Introduction à la théorie des schémas* (chapitre I), polycopié de cours de M2 disponible à l'adresse
https://webusers.imj-prg.fr/~antoine.ducros/Cours-schemas.pdf

(2) A. Grothendieck, *Sur quelques points d'algèbre homologique*, Tohoku Math. J. **9** (2) : 119-221 (1957).

(3) E. Riehl, *Category theory in context*, ouvrage disponible en ligne avec l'accord de l'éditeur à l'adresse
https://math.jhu.edu/~eriehl/context.pdf

QUOTIENTS DE FACTORIELLES ET MONODROMIE DE FONCTIONS HYPERGÉOMÉTRIQUES

par

Javier Fresán

Dans ses travaux sur les nombres premiers, Tchebychev utilise implicitement que

$$u_n = \frac{(30n)!n!}{(15n)!(10n)!(6n)!}$$

est un nombre *entier* pour tout $n \geq 0$. C'est surprenant car u_n n'a pas d'interprétation combinatoire évidente, contrairement à par exemple $\frac{(2n)!}{n!n!} = \binom{2n}{n}$. En 2007, Rodríguez Villegas a établi un lien entre l'intégralité de ces quotients de factorielles et l'algébricité de leurs séries génératrices.

Théorème. — *Soit $\gamma = (\gamma_\nu)_{\nu \geq 1}$ une suite d'entiers presque tous nuls tels que $\sum_{\nu \geq 1} \nu\gamma_\nu = 0$ et $\sum_{\nu \geq 1} \gamma_\nu = -1$. Alors le quotient de factorielles $u_n = \prod_{\nu \geq 1} (\nu n)!^{\gamma_\nu}$ est entier pour tout $n \geq 0$ si et seulement si la série entière $u = \sum_{n \geq 0} u_n \lambda^n$ est algébrique sur $\mathbf{Q}(\lambda)$, c'est-à-dire s'il existe un polynôme $F \in \mathbf{Z}[x, y]$ tel que $F(\lambda, u(\lambda)) = 0$.*

Par exemple, pour $u = \sum_{n \geq 0} \frac{(2n)!}{n!n!} \lambda^n = 1/\sqrt{1 - 4\lambda}$, on peut choisir $F(x, y) = (1 - 4x)y^2 - 1$. Mais pour les factorielles de Tchebychev, le plus petit polynôme F est de degré 483.840 en y !

Dans un premier temps, nous étudierons la démonstration de ce théorème en complétant tous les détails qui manquent dans l'esquisse dans [**1**] ; elle repose sur les propriétés de la monodromie des solutions de l'équation différentielle de type hypergéométrique satisfaite par la série génératrice u, et ce sera le prétexte pour les aborder. En fonction du temps qui reste et des goûts des élèves, on pourra ensuite s'orienter vers l'un des sujets suivants :

— les fonctions hypergéométriques à monodromie finie suivant Beukers-Heckman [**2**] ;

— l'interprétation géométrique de l'intégralité en termes de polytopes [**3**] ;

— une approche par la théorie analytique des nombres due à Soundararajan [**4**].

Références

[1] F. Rodríguez Villegas, *Integral ratios of factorials and algebraic hypergeometric functions*, Oberwolfach Reports **2** (2005), no. 3, 1813–1816.

[2] F. Beukers and G. Heckman, *Monodromy for the hypergeometric function ${}_nF_{n-1}$*, Invent. math. **95** (1989), 325–354.

[3] F. Rodríguez Villegas, *Mixed Hodge numbers and factorial ratios*, arXiv :1907.02722.

[4] K. Soundararajan, *Integral factorial ratios*, Duke Math. J. **171** (2022), no. 3, 633–672.

Javier Fresán • *E-mail :* `javier.fresan@imj-prg.fr`

AUTOUR DU THÉORÈME DE DVORETZKY

OMER FRIEDLAND

Quelle est la forme des ensembles convexes en grande dimension? Le théorème de Dvoretzky nous assure que si n est assez grand, alors tout convexe de $\mathbb{R}^n$ a des sections de basse dimension qui sont presque sphériques. Le but de ce sujet de TER est de donner une introduction au théorème de Dvoretzky. Sans doute c'est le résultat le plus important de la théorie locale des espaces de Banach. En 1961, Aryeh Dvoretzky résout positivement une conjecture formulée en 1956 par Grothendieck en établissant le théorème fondamental suivant:

Theorem 1. *Soit n un entier, et $\varepsilon > 0$. Il existe un entier $N = N(n, \varepsilon)$ tel que si X est un espace normé de dimension N, il existe une application linéaire $T : \ell_2^n \to X$ telle que $\|T\|\|T^{-1}\| \leq (1 + \varepsilon)$ (où bien sûr T^{-1} est définie sur l'image de T).*

En 1971, Vitali Milman a donné une nouvelle preuve de ce théorème, par une méthode probabiliste, en utilisant la concentration de mesure sur la sphère. Ce fut l'un des points de départ du développement de l'analyse géométrique asymptotique. La preuve de Milman donne une estimation fine en fonction de n:

$$N(n, \varepsilon) \leq \exp(c(\varepsilon)n).$$

Cette dépendance par rapport à ε a été étudiée par Y. Gordon, G. Schechtman, et d'autres, mais la bonne dépendance reste toujours ouverte. Dans ce TER, on étudiera ces différentes preuves en s'appuyant, en particulier, sur les ouvrages [1, 2].

References

[1] Gilles Pisier, *The volume of convex bodies and Banach space geometry*, Cambridge Tracts in Mathematics, vol. 94, Cambridge University Press, Cambridge, 1989.

[2] Gideon Schechtman, *Euclidean sections of convex bodies*, Asymptotic geometric analysis, Fields Inst. Commun., vol. 68, Springer, New York, 2013, pp. 271–288.

Notions catégoriques en topologie et géométrie algébrique

Le langage catégorique est devenu un outil incontournable qui apparaît naturellement dans l'étude de divers objets mathématiques. Trois des domaines qui ont joué un rôle particulier dans le développement de la théorie des catégories sont la géométrie algébrique, l'algèbre homologique et la théorie de l'homotopie avec, bien sûr, beaucoup d'intersections entre ces domaines. Dans ce TER, nous explorerons les notions catégoriques nécessaires pour aborder ces trois domaines, en gardant toujours à l'esprit les motivations géométriques et algébriques qui ont données naissance à ces notions.

Références

[1] Antoine Ducros. *Introduction à la théorie des schémas.* `https://webusers.imj-prg.fr/~antoine.ducros/Cours-schemas.pdf`

[2] S. I. Gelfand et Yu. I. Manin. *Homological algebra* (Vol. 38). Springer Science & Business Media (1994).

[3] Emily Riehl. *Category theory in context.* Dover Publications (2017).

[4] Joseph J. Rotman. *An introduction to homological algebra* (Vol. 2). New York : Springer (2009).

[5] Pierre Schapira. *Categories and Homological Algebra.* Lecture notes. `https://perso.imj-prg.fr/pierre-schapira/wp-content/uploads/schapira-pub/lectnotes/HomAl.pdf`

David García-Zelada

Théorie de Floquet

Frédéric Klopp

Coordonnées de l'encadrant : Frédéric Klopp
bureau 16-26-412,
frederic.klopp@imj-prg.fr

Le but de ce sujet de TER est de donner une introduction à la théorie de Floquet des opérateurs différentiels elliptiques. L'étude suivra le fil de l'article [1]. On étudiera de nombreux aspects de la théorie des opérateurs différentiels, aspects spectraux, aspects lié à l'analyse des EDPs. On y verra aussi de nombreuses applications, en particulier à la physique du solide.

Références

[1] Peter Kuchment. An overview of periodic elliptic operators. *Bull. Amer. Math. Soc. (N.S.)*, 53(3):343–414, 2016.

EDP dispersives et géométrie

Camille Laurent*†

Les Équations au dérivées aux dérivées partielles dispersives sont actuellement le sujet de nombreuses études. Le sujet est très vaste, en particulier puisque de nombreuses équations de la physique mathématique ont ce type de propriétés: équation de Schödinger, équation de Korteweg de Vries... La définition d'une EDP dispersive n'est pas complèment rigide, mais l'idée générale est que les solutions se "dispersent" dans l'espace ce qui induit une sorte de régularisation et parfois une décroissance. La géométrie de l'espace considéré a ainsi une grande importance. Nous proposons dans ce TER d'explorer les possibles liens entre cette géométrie et ces propriété analytiques.

Adil Ihtajja a déjà étudié les espaces de Strichartz dans un précd́ent stage. Nous pourrions dans un premier temps revoir ces notions dans d'autres cadres et ensuite considérer les espaces de Bourgain qui sont utiles pour des géométries bornées. La référence générale [1] pourrait être un bon point de départ ou [2].

Par la suite, selon les goûts et l'avancement de l'étudiant, il serait possible de se diriger plus vers les raffinements analytiques liés [3] ou géométriques comme [4] qui est une extension sur la sphère.

References

[1] Terence Tao. *Nonlinear dispersive equations*, volume 106 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, 2006. Local and global analysis.

[2] Jean Ginibre. Le problème de Cauchy pour des EDP semi-linéaires périodiques en variables d'espace (d'après Bourgain). *Astérisque*, (237):Exp. No. 796, 4, 163–187, 1996. Séminaire Bourbaki, Vol. 1994/95.

[3] Jean Bourgain. Fourier transform restriction phenomena for certain lattice subsets and applications to nonlinear evolution equations, Part I. *Geom. Funct. Anal.*, 3:107–156, 1993.

[4] Nicolas Burq, Patrick Gérard, and Nikolay Tzvetkov. Bilinear eigenfunction estimates and the nonlinear Schrödinger equation on surfaces. *Invent. math*, 159:187–223, 2005.

*CNRS, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France

†UPMC Univ Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France, email: laurent@ann.jussieu.fr

Isomorphismes de De Rham
Sujet de TER

par Maxime Debois, encadré par Julien Marché

Le but du TER est d'introduire l'homologie et la cohomologie singulière des variétés et d'établir l'isomorphisme de De Rham. En préalable, il faudra développer la cohomologie de De Rham jusqu'à la démonstration de la dualité de Poincaré et la théorie de l'intersection. Ensuite, on étudiera la théorie singulière jusqu'à prouver l'isomorphisme de De Rham. On pourra (par exemple) appliquer cela pour comprendre la topologie des groupes de Lie.

Références:

1) Topology and Geometry, G. Bredon.
2) Differential forms in algebraic topology, Bott et Tu.

Géométrie et topologie de l'information

Jean-Pierre Marco

La théorie de l'information, fondée par C. Shannon dans les années 1950, s'est progressivement développée après de spectaculaires applications au problème de la communication de données (codes, turbocodes) et a donné lieu récemment à une analyse mêlant probabilités, statistiques et géométrie différentielle (métrique de Fischer par exemple), très utilisée dans toutes les méthodes de traitement de données. En parallèle et plus récemment, D. Bennequin et ses collaborateurs ont développé une approche topologique du problème basée sur des notions de géométrie algébrique. Le but de ce stage est de donner une introduction à la théorie de Shannon et de permettre d'acquérir ou de consolider les connaissances requises pour une toute première approche des problèmes géométriques et topologiques associés –seule une connaissance préalable de la théorie élémentaire des probabilités et quelques éléments de géométrie différentielle seront nécessaires. On s'attachera aussi à détailler quelques exemples d'applications récentes dans des domaines techniques.

Références

C. Shannon. La théorie mathématique de l'information (Cassini)

A. Chambert-Loir. Information theory (Springer)

N. Ay, J. Jost, H. Vân Lê, L. Schwaschhöfer. Information Geometry (Springer)

S. Amari, H. Nagaoka. Methods of Information Geometry (AMS, Oxford University Press)

D. Bennequin, P. Baudot. The homological nature of entropy (Entropy 2015)

Une initiation pas à pas à la théorie des catégories avec un regard tourné vers les applications

Sébastien Martineau

smartineau@lpsm.paris

En mathématiques, nous sommes habitués à considérer des objets et des morphismes entre ces objets. Il peut s'agir d'espaces vectoriels, de groupes, d'ensembles, d'espaces topologiques, etc. [1] Pour démontrer des résultats, il s'agit alors d'inspecter en détail nos objets d'étude. La théorie des catégories incarne un tout autre point de vue, moderne et fécond : pour bien comprendre un objet, on ne l'inspecte pas en détail ; on étudie plutôt comment il interagit avec tous les objets possibles (on regarde la collection de tous les morphismes partant de cet objet ou allant vers cet objet).

La théorie des catégories se prête naturellement à une présentation en termes algébriques, abstraits et généraux, avec un regard tourné vers la géométrie algébrique, la topologie algébrique ou la logique. Dans ce TER, on propose une approche assez différente. Au lieu de sauter rapidement vers des abstractions générales, on procèdera de façon progressive. On inspectera en détail un certain nombre de situations concrètes et la généralité de notre propos s'étendra graduellement. En particulier, les catégories elles-mêmes ne seront pas étudiées dès le début du TER. Par ailleurs, les constructions seront présentées avec un regard tourné vers des applications extra-mathématiques, telles que la gestion de ressources ou les bases de données. Aucun prérequis n'est demandé concernant ces domaines d'application potentiels.

Références :

— *Applied Category Theory Course*, de John Baez.
— *Seven Sketches in Compositionality: An Invitation to Applied Category Theory*, de Brendan Fong et David Spivak.
— *La Théorie des Ensembles selon les Shadoks*, d'Alain Prouté.
— Discussions régulières avec Sébastien Martineau.

1. Pour les ensembles, les morphismes sont les fonctions. Pour les espaces topologiques, ce sont les fonctions continues.

Mathématiques et crise environnementale

Sébastien Martineau
smartineau@lpsm.paris

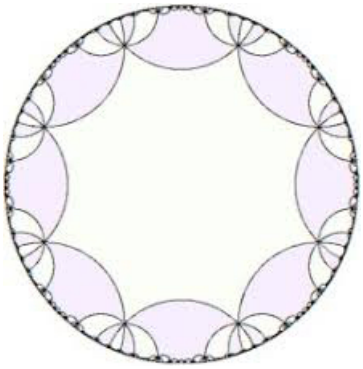
La situation actuelle de la biosphère est préoccupante, avec des extinctions massives, une pollution à grande échelle, ainsi qu'une évolution climatique rapide. Il s'agit dans ce TER de se renseigner sur cette situation et d'en aborder un ou plusieurs aspects de façon mathématique.

Plus précisément, la première moitié du semestre sera dédiée à rassembler informations et compréhension concernant la situation actuelle. Durant cette phase, le but sera d'aller vers ce qu'on considérera comme essentiel, sans se préoccuper de voir les choses à tout prix sous un angle mathématique.

Durant la seconde moitié, on sélectionnera, parmi les aspects rencontrés jusqu'alors, un ou plusieurs points qui se prêtent à une étude mathématique de niveau M1 et mènera une telle étude. Cette procédure en deux temps a pour but de vraiment garder en tête le sujet d'étude et d'y employer les mathématiques comme outils de compréhension.

Groupes Fuchsiens Arithmétiques
Sujet de stage M1, Sorbonne Université.
Proposé par Frédéric Naud.

Toute surface de Riemann de genre $g \geq 2$ s'uniformise comme quotient du disque unité $\mathbb{D}$ par un groupe discret d'automorphismes de $\mathbb{D}$. Ces automorphismes sont des homographies et des isométries par rapport à la distance hyperbolique. La surface quotient associée est donc naturellement munie d'une structure métrique, dite hyperbolique. Dans ce TER, on étudiera les propriétés de tels groupes et en particulier ceux que l'on appelle "arithmétiques". Leur construction sera l'occasion d'apprendre un peu de théorie algébrique des nombres. L'objectif final de ce travail sera de comprendre le théorème de Takeuchi qui caractérise de tels groupes par les propriétés algébriques du spectre des longueurs de la surface associée.



Références:

- Svetlana Katok, *Fuchsian groups*, Chicago Lectures in Mathematics.
- Pierre Samuel, *Théorie algébrique des nombres*, Hermann.
- Kisao Takeuchi, *A characterization of arithmetic Fuchsian groups*, J. Math. Soc. Japan, 27(4):600612, 1975.

Les schémas spectraux comme treillis annelés

Proposition de mémoire

Frédéric Paugam

frederic.paugam at imj-prg.fr

25 janvier 2024

Joyal à proposé une définition du spectre d'un anneau comme treillis, ce qui permet de l'étudier dans un cadre constructif, dans lequel l'existence d'idéaux premiers ne peut être garantie, puisque l'axiome du choix n'est pas disponible [1].

On propose d'étudier la construction par Coquand, Lombardi et Schuster dans [CLS09] du faisceau d'anneaux sur ce spectre, et de la notion de schéma spectral constructif, en l'illustrant par l'exemple de l'espace projectif.

Prérequis :

— Un peu d'algèbre commutative.

— Un peu de théorie des treillis distributifs (à rappeler en cours de mémoire).

Références

[CLS09] Thierry Coquand, Henri Lombardi, and Peter Schuster. Spectral schemes as ringed lattices. *Ann. Math. Artif. Intell.*, 56(3-4) :339–360, 2009.

1. On pense par exemple au cas d'une implémentation dans un assistant de preuve basé sur une théorie constructive des types (agda cubique), qui a récemment été mise en oeuvre en se basant sur ces travaux.

Etude du flocking dans des modèles de type Cucker-Smale

TER proposé par Nastassia Pouradier Duteil
nastassia.pouradier_duteil@sorbonne-universite.fr

On considère un modèle de type Cucker-Smale, décrivant l'évolution d'une nuée de N oiseaux dont les positions sont notées $(x_i)_{i\in\{1,\cdots,N\}}$ et les vitesses $(v_i)_{i\in\{1,\cdots,N\}}$. Le modèle s'écrit

$$\begin{cases} \dfrac{d}{dt}x_i(t) = v_i(t), \quad \text{pour tout } i \in \{1,\cdots,N\}, \text{ pour tout } t \in \mathbb{R}_+ \\ \dfrac{d}{dt}v_i(t) = \dfrac{1}{N}\displaystyle\sum_{j=1}^{N} w_{ij}(t)\phi(\|x_i(t)-x_j(t)\|)(v_j(t)-v_i(t)), \end{cases} \tag{1}$$

complété par des données initiales $(x_i^0)_{i\in\{1,\cdots,N\}} \in (\mathbb{R}^d)^N$ et $(v_i^0)_{i\in\{1,\cdots,N\}} \in (\mathbb{R}^d)^N$. La fonction $\phi \in C(\mathbb{R}_+;\mathbb{R}_+)$ est considérée Lipschitz, et représente l'interaction spatiale entre deux individus. La matrice $(w_{ij})_{i,j\in\{1,\cdots,N\}} \in L^\infty(\mathbb{R}_+;\mathcal{M}_N(\mathbb{R}))$ contient les poids de communication, qui sont susceptibles d'évoluer en temps.

Introduit en 2007 par Cucker et Smale [1], ce modèle met en évidence la capacité du groupe à s'auto-organiser, c'est-à-dire à former une nuée cohérente dont les vitesses sont alignées les unes aux autres. On appelle ce comportement *flocking*.

Ce TER aura pour objectif d'étudier les conditions suffisantes et nécessaires sur la fonction d'interaction ϕ et sur les poids de communication w_{ij} pour la formation du flocking, par l'étude d'articles de recherche et à travers des simulations numériques. Le travail pourra s'organiser de la façon suivante:

1. Démontrer que le système (1) est bien posé.
2. Etudier et redémontrer le Théorème de Flocking pour un système *échangeable*, c'est-à-dire dans le cas où les poids de communication satisfont $w_{ij} \equiv 1$ pour tout $(i,j) \in \{1,\cdots,N\}^2$. On s'appuiera sur l'article [2].
3. Simuler numériquement le système dans le cas échangeable et retrouver la condition suffisante pour le flocking. Cette condition est-elle aussi nécessaire?
4. Etudier et redémontrer le Théorème de Flocking pour un système *non-échangeable*, avec des poids de communication positifs ou nuls. On s'appuiera sur l'article [3].
5. Simuler numériquement le système dans le cas non-échangeable et identifier des matrices de communication telles que le système converge (ou non) vers le flocking.
6. Pour aller plus loin, on étudiera également le cas où les poids de communication sont autorisés à être négatifs à l'aide d'explorations numériques. Observe-t-on des comportements collectifs différents du flocking? Que peut-on démontrer dans ce cas?

References

[1] F. Cucker and S. Smale, *Emergent Behavior in Flocks*, in IEEE Transactions on Automatic Control, vol. 52, no. 5, pp. 852-862, (2007)

[2] S.-Y. Ha, J.-G. Liu, *A simple proof of the Cucker-Smale flocking dynamics and mean-field limit*, Communications in Mathematical Sciences, Commun. Math. Sci. 7(2), 297-325, (2009)

[3] B. Bonnet, E. Flayac, *Consensus and flocking under communication failures for a class of Cucker–Smale systems*, Systems & Control Letters, Volume 152, 104930 (2021)

Estimation du squelette d’un modèle graphique

Les modèles graphiques constituent un cadre générique pour décrire les dépendances (conditionnelles) liant un ensemble de *p* variables. L’inférence d’un modèle graphique à partir de réalisation iid de vecteurs constitués de ces *p* variables, dite inférence de structure, pose un problème combinatoire difficile, du fait du nombre de graphes pouvant relier *p* noeuds.

Récemment, Duan & Dunson (2021) ont proposé de restreindre l’inférence du graphe à celle de son squelette qui prend la forme d’un arbre recouvrant de l’ensemble des variables. Ils proposent et analyse également les propriétés d’un estimateur bayésien de ce squelette.

L’objectif de ce TER est de lire et de synthétiser l’article de Duan & Dunson et de réfléchir à une alternative à l’algorithme d’inférence qu’ils proposent, plus économe en temps de calcul, en se fondant sur la structure particulière de l’ensemble des arbres recouvrants.

- Duan, L. L., & Dunson, D. B. (2021). Bayesian spanning tree: estimating the backbone of the dependence graph. *arXiv preprint arXiv:2106.16120*.

Segmentation / classification dans un processus de Poisson

Le processus de Poisson permet de décrire les apparitions d'événements ponctuels au cours du temps. L'intensité du processus contrôle le taux de survenu des événements et, dans beaucoup de cas, on peut supposer que ce taux varie au cours du temps.
La segmentation vise à estimer les dates auxquelles surviennent ces changements de taux (ainsi que l'intensité au sein des intervalles – appelés 'segments' – qui séparent ces dates – appelées ruptures). Cette estimation pose un problème d'optimisation spécifique qu'on peut résoudre au moyen d'un algorithme de programmation dynamique (Dion-Blanc & al, 2024).
Dans beaucoup de situations, on peut supposer que les changements de taux correspondent aux changements d'un comportement sous-jacent et que le nombre d'intensités différentes est en fait limité, ce qui revient alors à classer les segments selon le comportement associé : on parle alors de segmentation-classification.
L'objectif de ce TER est de concevoir et d'implémenter un algorithme d'inférence du modèle de segmentation-classification pour un processus de Poisson. On pourra également s'intéresser au problème du choix (simultané) du nombre de segments du nombre de comportements.

- Dion-Blanc, C., Lebarbier, E., & Robin, S. (2023). Multiple change-point detection for Poisson processes. *arXiv preprint arXiv:2302.09103*.

Classification non supervisée d'espèces par SDM

En écologie, les modèles de distribution d’espèces (= SDM pour *species distribution models*) servent à prédire l’abondance (i.e., le nombre d’individus) d’une espèce dans un milieu donné en fonction des caractéristiques de ce milieu. Le modèle le plus classique est la régression poissonnienne, mais plusieurs alternatives existent.
L’écologie des communautés s’intéresse plus particulièrement à la constitution des ensembles d’espèces (communautés) qui occupent un même milieu. Pour cela, il est intéressant de savoir si certaines espèces “répondent” de façon semblable aux caractéristiques environnementales. Une façon de répondre à cette question est de déterminer des groupes d’espèces dont les SDM respectifs reposent sur des paramètres proches, voire égaux.
L’objectif de ce TER est de concevoir et d’implémenter une méthode de classification non-supervisée (*clustering*) d’espèces. La classification obtenue pourra être confrontée, par exemple, aux traits caractérisants les différentes espèces ou à la phylogénie qui les relie.

Autour de la prédiction conformelle

Sujet proposé par Etienne Roquain, etienne.roquain@upmc.fr

Les méthodes actuelles de machine learning peuvent avoir des performances impressionnantes, mais font des erreurs. Ainsi, un travail est de quantifier l'incertitude liée à ces méthodes, et ce malgré le côté "boîte noire" de ces procédures et la complexité des données observées.

L'objectif de ce TER est de comprendre comment la prédiction conformelle permet d'apporter une solution à ce problème. Plus précisément, en s'appuyant sur Angelopoulos and Bates (2022), une feuille de route pourra être la suivante :

- Comprendre et prouver les théorèmes fondamentaux sur laquelle cette approche repose ;
- L'appliquer sur des modèles simples en régression et classification non paramétrique ;
- (Plus difficile) Aller vers des guaranties de couverture conditionnelles, par exemple Gibbs et al. (2023).

La rédaction du rapport se fera en langage LaTeX et une soutenance orale devra appuyer ce document.

Il sera important que l'étudiante ou l'étudiant en question soit à l'aise avec les statistiques de base. De plus, il est conseillé de suivre le cours de "statistiques avancées" en parallèle.

Références

Angelopoulos, A. N. and Bates, S. (2022). A gentle introduction to conformal prediction and distribution-free uncertainty quantification.

Gibbs, I., Cherian, J. J., and Candès, E. J. (2023). Conformal prediction with conditional guarantees.

Interpolation et sur-apprentissage en estimation non-paramétrique

Sujet proposé par Etienne Roquain, `etienne.roquain@upmc.fr`

Un obstacle majeur des procédures d'apprentissage automatique est le sur-apprentissage ("over-fitting" en "machine learning") : si la procédure "colle" trop aux données d'apprentissage, alors elle aura de piètres performances sur d'autres jeux de données, pourtant générés selon la même loi.

Ces dernières années, les statisticiens portent un nouveau regard sur ce phénomène bien connu : il s'avère qu'il y a une façon de "coller" parfaitement aux données qui garantisse encore la validité de l'inférence statistique. Cela s'appelle l'"interpolation".

L'objectif de ce TER sera d'explorer ce nouveau phénomène dans le cadre de l'estimation non-paramétrique, en s'appuyant sur Belkin et al. (2018). Une feuille de route pourra être la suivante :

- Commencer par un travail bibliographique autour de l'interpolation dans les différents domaines de la statistique, cf par exemple Bartlett et al. (2021) ;
- Etudier profondément le travail de Belkin et al. (2018), par exemple avec des preuves détaillées en se restreignant à la classe des fonctions Lipschitziennes. Il faudra notamment percer le mystère de la Figure 1 ;
- Appliquer ce résultat théorique sur des données simulées et des données réelles (langage R de préférence) ;
- Si le candidat est très à l'aise, regarder le cas de l'interpolation dans le cas de la régression linéaire, cf Bartlett et al. (2021).

La rédaction du rapport se fera en langage LaTeX et une soutenance orale devra appuyer ce document.

Il sera important que l'étudiante ou l'étudiant en question soit à l'aise avec les statistiques de base. De plus, il est conseillé de suivre le cours de "statistiques avancées" en parallèle.

Références

Bartlett, P. L., Montanari, A., and Rakhlin, A. (2021). Deep learning : a statistical viewpoint.

Belkin, M., Rakhlin, A., and Tsybakov, A. B. (2018). Does data interpolation contradict statistical optimality ?

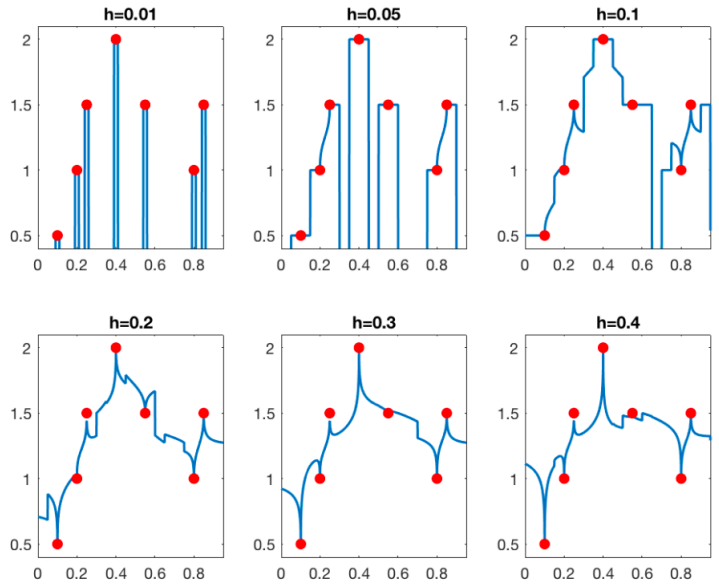


FIGURE 1 : Interpolation en régression non paramétrique (extraite de Belkin et al. (2018)).

Apprentissage supervisé large échelle

Sujet proposé par Maxime Sangnier

maxime.sangnier@sorbonne-universite.fr

En apprentissage statistique supervisé, on suppose observer des couples $(X_1, Y_1), \ldots, (X_n, Y_n)$ i.i.d. à valeurs dans $\mathbb{R}^p \times \mathbb{R}$, et on cherche généralement à estimer un paramètre inconnu $\beta^\star \in \mathbb{R}^p$, permettant d'exprimer, pour tout $x \in \mathbb{R}^p$, une caractéristique de la forme ${\beta^\star}^\top x$ de la loi $\mathcal{L}\left(Y_1 \mid X_1 = x\right)$. C'est le cas par exemple de la régression linéaire, dans laquelle $\mathbb{E}\left[Y_1 \mid X_1 = x\right] = {\beta^\star}^\top x$. Très souvent, la procédure d'estimation consiste alors en la résolution d'un problème d'optimisation de la forme :

$$\hat{\beta} \in \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \ell(Y_i, \beta^\top X_i),$$

où $\ell : \mathbb{R}^2 \to \mathbb{R}$ est une fonction de perte adéquate. Si l'on peut montrer que, sous certaines hypothèses de régularité, un tel estimateur $\hat{\beta}$ possède les propriétés asymptotiques attendues (convergence et normalité asymptotique lorsque $n \to \infty$), il n'est pas toujours raisonnable (en particulier il devient difficile sur le plan calculatoire) de chercher à traiter toujours plus d'observations. On a donc souvent un intérêt, étant donné une grande quantité n d'individus, à construire un jeu de données plus petit, permettant un calcul plus aisé de l'estimateur.

L'une des manières d'arriver à cette fin est le sous-échantillonnage, consistant à extraire aléatoirement m individus représentatifs du jeu de données, menant à un nouvel estimateur :

$$\hat{\beta}_S \in \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{m} \pi_{(i)} \ell(Y_{(i)}, \beta^\top X_{(i)}),$$

où $(X_{(1)}, Y_{(1)}), \ldots, (X_{(m)}, Y_{(m)})$ sont les individus sélectionnés et $\pi_{(i)}$ est un facteur de redressement adéquat.

Le but de ce TER est d'étudier les propriétés statistiques d'estimateurs de β de la forme de $\hat{\beta}_S$ (avec évenuellement différentes stratégies d'échantillonnnage). Cela fera appel à des résultats de normalité asymptotique des estimateurs, qui s'obtiennent par d'autres moyens que le Théorème Central Limite.

Après s'être familiarisé avec les principaux théorèmes asymptotiques et avec leurs applications en apprentissage statistique, on s'attachera à caractériser la variance asymptotique des estimateurs obtenus suivant les stratégies de sous-échantillonnage. Les résultats seront illustrés par une étude numérique utilisant le language Python.

Références

H.Y. Wang and J.K. Kim. Maximum sampled conditional likelihood for informative subsampling. *Journal of Machine Learning Research*, 23(332) :1–50, 2022.

H.Y. Wang, R. Zhu, and P. Ma. Optimal Subsampling for Large Sample Logistic Regression. *Journal of the American Statistical Association*, 113(522) :829–844, 2018.

Y. Yao and H. Wang. Optimal subsampling for softmax regression. *Statistical Papers*, 60(2) :585–599, 2019.

GROUPE FONDAMENTAL EN GÉOMÉTRIE ALGÉBRIQUE.

Encadrant: Jean-Baptiste Teyssier
jean-baptiste.teyssier@imj-prg.fr

Le but de ce TER sera de comprendre comment définir le groupe fondamental pour les courbes algébriques [Sza09], le lien avec la théorie de Galois et son application au théorème de Belyi montrant que le groupe de Galois absolu de $\mathbb{Q}$ agit fidèlement sur le groupe fondamental de $\mathbb{P} \setminus \{0, 1, \infty\}$.

Si le temps le permet, on pourra aussi chercher à comprendre la théorie abstraite des catégories galoisiennes et en quoi le π_1 des variétés algébriques en est une incarnation.

References

[Sza09] T. Szamuely, *Galois Groups and Fundamental Groups*, Cambridge Studies in Advanced Mathematics,, vol. 117, 2009.