

Mémoire rédigé sous la direction de Raphaël Cerf

PIERRE PETIT

1 Présentation du domaine de recherche

En mécanique classique du point ou du solide, on parvient généralement à décrire l'état d'un système en résolvant formellement, ou numériquement, l'équation différentielle régissant l'évolution complète de ce système. Par exemple, le mouvement de la Terre autour du soleil ou d'un électron émis par le tube cathodique d'un poste de télévision. Mais, si déjà le problème à trois corps (pour l'interaction gravitationnelle) n'est plus intégrable, comment espérer calculer, même numériquement, l'état d'un simple verre d'eau dont la description nécessite, *a priori*, de l'ordre de 10^{25} paramètres que sont les positions et vitesses de toutes les molécules à un instant donné ?

« The objective of statistical mechanics is to explain the macroscopic properties of matter on the basis of the behaviour of the atoms and molecules of which it is composed. »

Lanford

La mécanique statistique offre une réponse satisfaisante à l'étude de ces systèmes qui contiennent un grand nombre de degrés de liberté. Et l'idée fondamentale qui la sous-tend est la notion de loi des grands nombres : avec des hypothèses raisonnables, vérifiées en pratique, les mesures — ou observables —

effectuées sur un système contenant un grand nombre de particules dépendent essentiellement d'un petit nombre de paramètres macroscopiques, par exemple température et densité, qui sont, pour le coup, faciles à connaître.

1.1 Loi des grands nombres

La théorie des probabilités permet de formaliser cette idée. Donnons-nous une suite $(X_n)_{n \geq 1}$ de variables aléatoires réelles, indépendantes et identiquement distribuées (i.i.d.). Elles peuvent, par exemple, représenter les énergies de particules sans interaction. On définit les moyennes empiriques

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}$$

qui correspondent à l'énergie totale du système par particule. Alors, on montre que, si X_1 est intégrable et de moyenne m ,

$$\forall \varepsilon > 0 \quad \mathbb{P}(|\bar{X}_n - m| \geq \varepsilon) \xrightarrow[n \rightarrow \infty]{} 0 \quad (1)$$

résultat qui est connu sous le nom de *loi faible des grands nombres*. Autrement dit, bien que l'énergie de chacune des particules puisse varier — éventuellement beaucoup, du moment que $\mathbb{E}[|X_1|]$ reste fini —, l'énergie globale du système par particule est égale à $m \pm \varepsilon$ du moment que le nombre de particules est suffisamment grand.

Plusieurs questions se posent alors : peut-on tout d'abord donner une estimation de la vitesse de convergence ? Dans le cas présenté ci-dessus, quitte à supposer X_1 de carré intégrable et de variance σ^2 , l'inégalité de Markov permet d'écrire

$$\mathbb{P}(|\bar{X}_n - m| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \quad (2)$$

Peut-on faire mieux ?

Ensuite, peut-on généraliser le résultat ? À des variables non nécessairement i.i.d. ? Ce point est fondamental dans l'optique de décrire des systèmes physiques où les particules interagissent, autrement dit, ne sont pas indépendantes. À des variables non nécessairement réelles ? À valeurs dans un espace vectoriel quelconque ? Il est notamment naturel de s'intéresser aux mesures empiriques

$$L_n = \frac{1}{n}(\delta_{X_1} + \cdots + \delta_{X_n})$$

lorsqu'on veut procéder à un sondage. Or, si X_1 est à valeurs dans $\{\pm 1\}$, elles prennent déjà leurs valeurs dans l'espace $\mathcal{M}([0, 1])$ des mesures signées sur

$[0, 1]$, espace de dimension infinie. La loi forte des grands nombres¹ permet de montrer que

$$L_n \xrightarrow[n \rightarrow \infty]{} \mu$$

où μ est la loi de X_1 , la convergence ayant lieu au sens de la convergence faible des mesures. Mais, encore une fois, la vitesse de convergence seule permet d'estimer la taille d'un échantillon raisonnable pour que le sondage reflète la réalité.

La recherche de réponses à ces questions a donné naissance à deux approches, dès la première moitié du vingtième siècle : les inégalités de *concentration* et les *grandes déviations*. Toutes deux ont trouvé un véritable essor au début des années 1970.

1.2 Inégalités de concentration

L'approche concentration s'attache à donner de bonnes majorations, qui améliorent (2) tout en restant valables pour tout n . Remarquons tout de suite que (2) n'est pas très optimale. En effet, le théorème de la limite centrale et une majoration classique de la queue de la gaussienne montrent que

$$\mathbb{P}\left(\frac{\sqrt{n}}{\sigma}(\bar{X}_n - m) \geq \varepsilon\right) \xrightarrow[n \rightarrow \infty]{} \frac{1}{\sqrt{2\pi}} \int_{\varepsilon}^{+\infty} e^{-x^2/2} dx \leq \frac{1}{\sqrt{2\pi}} \frac{e^{-\varepsilon^2/2}}{\varepsilon}$$

On s'attend² donc à une décroissance, dites *exponentielle*, du type

$$\mathbb{P}(|\bar{X}_n - m| \geq \varepsilon) \approx C \exp\left(-n \frac{\varepsilon^2}{2\sigma^2}\right) \quad (3)$$

et on ne peut pas espérer mieux étant donné que, si X_1 est normale, une minoration de la queue de la gaussienne montre cette fois que

$$\mathbb{P}(|\bar{X}_n - m| \geq \varepsilon) \geq \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{n+1}{2}\varepsilon^2\right)$$

¹On peut énoncer un résultat plus précis que (1) : en fait,

$$\bar{X}_n \xrightarrow[n \rightarrow \infty]{ps} m$$

C'est la *loi forte des grands nombres*.

²On trouve dans [BKR61] une condition nécessaire et suffisante de décroissance exponentielle dans le cas i.i.d. portant sur les moments exponentiels.

Les inégalités les plus connues donnent des majorations pour des variables aléatoires indépendantes (non nécessairement de même loi) et bornées et utilisent des comparaisons avec les moments exponentiels. Donnons-nous désormais, pour comparer l'efficacité des résultats obtenus, une suite (X_n) de variables aléatoires indépendantes sur $[-1, 1]$ et de moyenne nulle. L'inégalité d'Hoeffding montre que

$$\mathbb{P}(\bar{X}_n \geq \varepsilon) \leq \exp\left(-n\frac{\varepsilon^2}{2}\right) \quad (4)$$

L'inégalité de Bennett améliore le résultat en faisant intervenir la variance moyenne des X_k

$$\sigma_n^2 = \frac{\text{var}(X_1) + \cdots + \text{var}(X_n)}{n}$$

et donne

$$\mathbb{P}(\bar{X}_n \geq \varepsilon) \leq \exp\left(-n\sigma_n^2 h\left(\frac{\varepsilon}{n\sigma_n^2}\right)\right) \quad (5)$$

où

$$h(x) = (1+x)\log(1+x) - x \geq \frac{x^2}{2(1+\frac{x}{3})}$$

Les travaux plus récents, sous l'impulsion de V. Milman dans les années 1970 puis développés par Talagrand à la fin des années 1990, ont une approche géométrique qui met en lumière le phénomène de *concentration de la mesure*, notamment dans les espaces de grandes dimensions. C'est là-dessus que porte l'exposé de maîtrise, sous la direction de J.-F. Le Gall. L'idée sous-jacente est de voir la concentration comme une propriété d'isopérimétrie de l'espace métrique mesuré (X, d, μ) .

Illustrons cette démarche sur la sphère \mathbb{S}^n munie de la mesure de probabilité uniforme σ^n . On montre que, pour tout borélien A , si B est une calotte sphérique de même mesure,

$$\forall \varepsilon > 0 \quad \sigma^n(A_\varepsilon) \geq \sigma^n(B_\varepsilon)$$

où C_ε est le ε -voisinage ouvert de C . On en déduit que, si $\sigma^n(A) \geq \frac{1}{2}$,

$$\forall \varepsilon > 0 \quad \sigma^n(A_\varepsilon) \geq 1 - \exp\left(-(n-1)\frac{\varepsilon^2}{2}\right)$$

Ainsi, au sens de la mesure, la majeure partie des points de \mathbb{S}^n se trouvent concentrés à une distance de A inférieure à $1/\sqrt{n}$. Voilà ce qui amène Gromov

(cf. [Gro99]) à définir le diamètre observable de X et

$$ObsDiam(\mathbb{S}^n) = O\left(\frac{1}{\sqrt{n}}\right)$$

On en déduit alors des résultats de concentration de fonctions F suffisamment régulières sur X — par exemple les fonctions 1-lipschitziennes ou à variation bornée — autour d'une médiane m_F .

Diverses méthodes, notamment l'usage de martingales, permettent d'évaluer la concentration d'espaces produits. Or, pour en revenir aux moyennes de variables aléatoires,

$$\bar{X}_n = F(X_1, \dots, X_n)$$

où F est une fonction 1-lipschitzienne pour la distance $d = d_1 + \dots + d_n$ sur l'espace produit. On récupère alors les inégalités de concentration évoquées plus haut, en affinant encore les constantes dans certains cas.

1.3 Grandes déviations

1.3.1 Introduction

La seconde approche, celles des grandes déviations, ne cherche plus des majorations valables pour tout n , mais s'attache à l'ordre de grandeur asymptotique de la décroissance. Plus précisément, guidé par (3), on s'attend à une convergence du type

$$\frac{1}{n} \log \mathbb{P}(|\bar{X}_n - m| \geq \varepsilon) \xrightarrow{n \rightarrow \infty} s(\varepsilon) \quad (6)$$

qu'on appelle *principe de grandes déviations*³.

Par exemple, dans le cas où (X_n) est une suite de variables aléatoires i.i.d. sur $\{\pm 1\}$ de moyenne nulle, on obtient⁴

$$s(\varepsilon) = -\frac{1}{2} \left[(1 - \varepsilon) \log(1 - \varepsilon) + (1 + \varepsilon) \log(1 + \varepsilon) \right]$$

³La définition précise est un peu moins restrictive, toutefois, et concerne une suite de mesures, ici la suite $(\mu_n)_{n \geq 1}$ des lois des \bar{X}_n .

⁴On notera qu'à ε petit, on retrouve l'ordre de grandeur trouvé dans les inégalités de concentration puisque

$$s(\varepsilon) \approx -\frac{\varepsilon^2}{2}$$

La fonction s ainsi obtenue, n'est autre que l'*entropie* définie par le physicien Boltzmann, dès les années 1870. Le second principe de la thermodynamique, énoncé par lui, assure qu'un système physique tend vers un état d'entropie maximale. En particulier, il est facile de voir que, quitte à supposer s suffisamment régulière⁵, si s ne s'annule qu'en 0, alors

$$\overline{X}_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} m$$

Notons que l'on obtient aussi, grâce à cette formalisation du second principe, des lois des grands nombres pour des conditionnements dégénérés. En reprenant l'exemple des variables sur $\{\pm 1\}$, conditionnellement à

$$\{\overline{X}_n \geq \alpha\}$$

l'entropie associée n'est autre que $s|_{[\alpha,1]}$ et ainsi, pour $\alpha > 0$,

$$\overline{X}_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}(\cdot|\overline{X}_n \geq \alpha)} \alpha$$

Cette approche est donc naturelle dans l'étude des états d'équilibre des systèmes qui ont un grand nombre de degrés de liberté. De plus, elle donne des résultats dans les deux directions de généralisation mentionnées plus haut, à savoir se passer de l'indépendance des variables et considérer des espaces vectoriels de dimension quelconque — ce qui est appréciable pour obtenir des résultats satisfaisants en physique.

Deux questions se posent alors : d'abord comment déterminer si une suite vérifie un principe de grandes déviations ? Et, le cas échéant, existe-t-il une expression simple de l'entropie associée ?

1.3.2 Cramér vs Sanov

Le cas des variables réelles a naturellement été le premier sujet d'étude. Il a donné lieu à deux approches.

⁵Il s'avère que, sous réserve d'existence, s est convexe, semi-continue supérieurement (s.c.s.) et négative. L'hypothèse qu'on rajoute est qu'elle est coercive, c'est-à-dire que les ensembles de niveau

$$\{s \geq \beta\}$$

sont compacts.

D'un côté, Cramér (1938) et Chernoff (1952) montrent un principe de grandes déviations (PGD) pour la suite (\bar{X}_n) dans le cas où les X_n sont i.i.d. De plus, l'entropie associée est donnée par l'opposée de la conjuguée de la *pression*

$$s(x) = -\sup_{\lambda \in \mathbb{R}} (\lambda x - p(\lambda)) = -p^*(\lambda)$$

où

$$\forall \lambda \in \mathbb{R} \quad p(\lambda) = \mathbb{E}[e^{\lambda X_1}]$$

Actuellement, on démontre ce résultat à l'aide d'un outil importé de la mécanique statistique par Lanford (1973) : la sous-additivité. En effet, on note que, si A est un intervalle de \mathbb{R} ,

$$\mathbb{P}(\bar{X}_{m+n} \in A) \geq \mathbb{P}(\bar{X}_m \in A; \bar{X}_n \in A) = \mathbb{P}(\bar{X}_m \in A)\mathbb{P}(\bar{X}_n \in A)$$

donc la suite

$$\left(\log \mathbb{P}(\bar{X}_n \in A) \right)_{n \geq 1}$$

est sous-additive et on en déduit que

$$\frac{1}{n} \log \mathbb{P}(\bar{X}_n \in A)$$

converge.

Depuis, Bahadur et Zabell [BZ79] ont généralisé le résultat à des variables à valeurs dans un espace vectoriel topologique localement convexe vérifiant quelques hypothèses de régularité. En particulier, ils obtiennent le PGD dans les espaces polonais. On peut aussi le montrer dans un espace localement convexe muni de sa topologie faible et dans un espace de Banach, modulo une hypothèse sur la loi de X_1 . Ce dernier résultat a entre autres pour conséquence le théorème de Schilder sur les grandes déviations du mouvement brownien $(\varepsilon W_t)_{t \in [0,1]}$.

De son côté, Sanov (1957) s'intéresse aux mesures empiriques

$$L_n = \frac{1}{n} (\delta_{X_1} + \dots + \delta_{X_n})$$

et montre pour elles un PGD. La preuve directe du théorème repose sur les propriétés de l'entropie de Shannon et on montre que le PGD est gouverné par l'entropie

$$h(\nu) = \int_{\mathbb{R}} \frac{d\nu}{d\mu} \log \frac{d\nu}{d\mu} d\mu$$

où μ est la loi de X_1 .

Les différentes généralisations de ces deux résultats permettent de montrer leur intrication. En effet, le théorème de Cramér dans le cadre des topologies faibles contient le théorème de Sanov. Dans l'autre sens, le théorème de Sanov redonne le théorème de Cramér dans le cas compact, via le *principe de contraction* qui assure la stabilité des PGD par une application continue.

1.3.3 Grandes déviations pour les champs de variables aléatoires dépendantes

On souhaite maintenant relaxer l'hypothèse d'indépendance des variables. Un lemme très utile, énoncé par Varadhan, montre que si $(\mu_n)_{n \geq 1}$ vérifie un PGD, alors $(e^{nf} \mu_n)_{n \geq 1}$ aussi dès que f est continue bornée. On peut donc perturber un peu une suite de variables i.i.d. et récupérer un PGD. C'est ainsi qu'on peut montrer un PGD dans le cas du modèle d'Ising standard, par exemple, défini par les mesures en volume fini $\Lambda \subset \mathbb{Z}^d$

$$\forall \sigma \in \{\pm 1\}^\Lambda \quad \mu_{\Lambda, \beta}^+(\sigma) = \frac{1}{Z_{\Lambda, \beta}} e^{-\beta H_\Lambda^+(\sigma)}$$

où

$$H_\Lambda^+ = -\frac{1}{2} \sum_{x \sim y} \sigma(x) \sigma(y) - \sum_{\substack{x \sim y \\ x \in \Lambda \\ y \notin \Lambda}} \sigma(x)$$

est le hamiltonien d'une configuration σ avec condition au bord $+1$. Ici, les variables aléatoires ne sont plus indexées par \mathbb{N}^* mais par \mathbb{Z}^d .

On montre alors que, si

$$M_n = \frac{1}{|\Lambda(n)|} \sum_{z \in \Lambda(n)} \sigma(z)$$

alors

$$\mu_{\Lambda(n), \beta}^+(M_n) \rightarrow m^*(\beta)$$

qui s'interprète comme la magnétisation spontanée à température $1/\beta$. On arrive ainsi à décrire le phénomène connu de l'aimant qui perd sa magnétisation lorsqu'il est chauffé.

Pour obtenir des résultats sur des mesures plus générales que celle du modèle d'Ising, les mesures de Gibbs, Ellis [Eli85] distingue trois « niveaux » de PGD :

Niveau 1 : la suite des moyennes empiriques

$$M_n = \frac{1}{|\Lambda(n)|} \sum_{z \in \Lambda(n)} \sigma(z)$$

satisfait un PGD sur X (résultat de type **Cramér**).

Niveau 2 : la suite des mesures empiriques

$$L_n = \frac{1}{|\Lambda(n)|} \sum_{z \in \Lambda(n)} \delta_{\sigma(z)}$$

satisfait un PGD sur $\mathcal{M}_1^+(X)$ (résultat de type **Sanov**).

Niveau 3 : la suite des mesures

$$T_n = \frac{1}{|\Lambda(n)|} \sum_{z \in \Lambda(n)} \delta_{\theta_{z\omega}}$$

satisfait un PGD sur $\mathcal{M}_1^+(X^{\mathbb{Z}^d})$.

Il s'avère que le niveau 3 est le bon cadre pour obtenir le PGD pour les mesures de Gibbs. Les niveaux 1 et 2 sont alors obtenus via le principe de contraction, dans les bons cas. C'est l'approche retenue dans les articles parus depuis les années 1970. L'article de Pfister [Pfi00] englobe les résultats obtenus et introduit la notion de mesure asymptotiquement découplée : grossièrement, la probabilité μ sur $X^{\mathbb{Z}^d}$ est asymptotiquement découplées si des événements dépendants de sites éloignés sont presque indépendants. Plus précisément, on se donne deux applications g et $c : \mathbb{N} \rightarrow [0, +\infty[$ telles que

$$\frac{g(n)}{n} \rightarrow 0 \quad \text{et} \quad \frac{c(n)}{|\Lambda(n)|} \rightarrow 0$$

On dit que μ est asymptotiquement découplée inférieurement de paramètres g et c si, pour toutes boîtes finies Λ_1 et Λ_2 , pour tous $A_1 \in \mathcal{B}(Y^{\Lambda_1})$ et $A_2 \in \mathcal{B}(Y^{\Lambda_2})$,

$$\left. \begin{array}{l} \Lambda_1 \subset \Lambda(n) \\ d(\Lambda_2, \Lambda(n)) > g(n) \end{array} \right\} \Rightarrow \mu(A_1 \times A_2) \geq e^{-c(n)} \mu(A_1) \mu(A_2)$$

Cette notion permet une généralisation naturelle du lemme sous-additif de

Lanford. Il prend alors la forme suivante

$$\begin{aligned} \exists M(\varepsilon, \delta) \quad \forall m \geq M \quad \exists N(m, \varepsilon, \delta) \quad \forall n \geq N \\ \frac{1}{|\Lambda(n)|} \log \mathbb{P}(m_{\Lambda(n)} \eta \in C) \geq (1 - \delta) \frac{1}{|\Lambda(m)|} \log \mathbb{P}(m_{\Lambda(m)} \eta \in C(\varepsilon)) \\ - \frac{c(m)}{|\Lambda(m)|} \end{aligned}$$

où η est une configuration de loi μ et

$$m_{\Lambda} \eta = \frac{1}{|\Lambda|} \sum_{z \in \Lambda} \eta(z)$$

On obtient alors un PGD de niveau 3 pour ces mesures asymptotiquement découplées.

Références

- [BKR61] L. E. BAUM, M. KATZ & R. R. READ – Exponential convergence rates for the law of large numbers, *Trans. Amer. Math. Soc.* **102**, p. 187 – 199 (1968).
- [BZ79] R. R. BAHADUR & S. L. ZABELL – Large Deviations of the Sample Mean in General Vector Spaces, *Ann. Prob.* **7**, no. 4, p. 587–621 (1979).
- [Cer07] R. CERF – *On Cramér’s Theory in Infinite Dimensions*, Panoramas et synthèses (2007).
- [Ell85] R. S. ELLIS – *Large Deviations and Statistical Mechanics*, Springer (1985).
- [Gro99] M. GROMOV – *Metric Structures for Riemannian and Non-Riemannian Spaces*, Birkhäuser (1999)
- [Lan73] O. E. LANFORD – *Entropy and Equilibrium States in Classical Statistical Mechanics*, Statistical Mechanics and Mathematical Problems, Lecture Notes in Physics, vol. 20, Springer (1973).
- [Pfi00] C.-E. PFISTER – *Thermodynamical Aspects of Classical Lattice Systems* (2000).
- [Rev68] P. RÉVÉSZ – *The Laws of Large Numbers*, Academic Press (1968).