

Introduction

Pourquoi des modèles et des simulations aléatoires ?

Les probabilités sont nées de l'étude des jeux de hasard et de leurs propriétés remarquables : il s'agissait de construire un modèle mathématique rendant compte, par exemple, de la convergence des fréquences empiriques de pile vers $1/2$ au cours d'un jeu équitabile de pile ou face, ainsi que d'apparitions arbitrairement longues de séries de pile, et de phénomènes de fluctuations autour de la moyenne apparemment universels quel que soit le jeu de hasard considéré. Pour modéliser les jeux et calculer les gains moyens, les risques de grandes pertes, etc., il a suffi de savoir définir des lois de probabilité sur des ensembles finis. Par contre, pour analyser les phénomènes de fluctuations, mais aussi pour modéliser les erreurs de mesure, il a fallu construire des lois de probabilités sur des espaces de dimension finie de type \mathbb{R}^d . C'est encore insuffisant pour des besoins apparus au début du vingtième siècle : pour modéliser des quantités continues évoluant en temps continu (par exemple, pour modéliser l'historique en temps continu d'un cours boursier ou d'une température), il faut construire des probabilités sur des espaces fonctionnels de dimension infinie appropriés. On conçoit aisément qu'une définition mathématique rigoureuse de « tirer au hasard une fonction » pose quelques difficultés techniques, et qu'il en est de même pour réaliser sur ordinateur un tel tirage.

Le but de ce cours est de présenter :

- quelques exemples de construction de lois de probabilité sur l'espace des fonctions continues ou continues par morceaux ;
- quelques algorithmes de simulation correspondants, ainsi que des estimations théoriques des erreurs de simulation ;
- quelques applications qui motivent les deux points précédents.

Les applications que nous étudierons sont représentatives des situations où l'on fait appel aux modèles probabilistes. Nous insistons sur deux points importants mais parfois mal compris dans la littérature.

1) Quand utiliser des modèles aléatoires ?

La modélisation aléatoire n'a de sens que pour des situations **critiques** où des modélisations déterministes échouent. De telles situations se multiplient dans les sciences

fondamentales et en ingénierie : puisqu'il est devenu raisonnable de simuler sur ordinateur des systèmes à grand nombre de degrés de liberté, à dynamiques complexes, à petites échelles de temps, la description de tels systèmes ne peut plus être fondée sur une énumération d'équations décrivant les actions et réactions en chacun des éléments constitutifs de l'ensemble et paramétrées par des quantités parfaitement mesurables par les expérimentateurs. Voici quelques exemples génériques :

- On veut modéliser un phénomène régi par des lois physiques bien connues mais certains paramètres du modèle sont difficiles à mesurer avec de bonnes précisions, voire impossibles à mesurer : par exemple, c'est une des difficultés rencontrées par l'industrie pétrolière qui ne peut disposer que d'informations imparfaites et très partielles sur la nature géologique et l'emplacement des failles des sous-sols ; pour des raisons évidentes, le manque de données est aussi inhérent aux modélisations des systèmes vivants.
- On veut modéliser un phénomène régi par des lois physiques bien connues mais soumis à des forces nombreuses, répétées, hautement variables : par exemple, il est raisonnable d'adopter une modélisation aléatoire pour les forces exercées par les creux et les bosses d'une route sur les amortisseurs d'une voiture lorsqu'on veut étudier la fatigue mécanique de ces amortisseurs ou mettre au point des systèmes de contrôle destinés à améliorer le confort des conducteurs.
- On veut modéliser un phénomène régi par des lois physiques bien connues mais comportant un nombre de degrés de liberté trop grand pour qu'on puisse énumérer les réactions de chacune des composantes du système : par exemple, un ensemble de grande taille d'électrons et de noyaux qui collisionnent.
- On veut modéliser un phénomène régi par des lois physiques mal connues : c'est le cas des échanges d'énergie dans des fluides turbulents, et des interactions entre solvants et chaînes de polymères.
- On veut modéliser un phénomène qui n'est pas régi par des lois physiques et est soumis à un grand nombre de causes diverses : c'est le cas d'un prix d'actif financier (comment établir une loi pour le comportement psychologique des traders et la survenance d'événements déstabilisateurs pour le marché ?) ou encore d'un réseau informatique (comment établir une loi pour les flux de données échangées par les internautes ?).

De plus, la complexité de certains phénomènes conduit à coupler des descriptions micro-macro qui combinent des dynamiques aléatoires pour les petites échelles et des équations macroscopiques pour les grandes échelles : on traite ainsi de la fatigue de structures mécaniques à grand nombre de degrés de liberté soumises aux effets du vent, des grands assemblages d'atomes constitutifs d'un polymère plongé dans un fluide turbulent, etc.

En résumé, les paramètres et les géométries mal connus sont considérés comme aléatoires, les forces mal comprises ou les déformations rapides sont décrites par des bruits aléatoires, l'évolution en temps de quantités inaccessibles aux mesures sont décrites par des processus stochastiques, etc. On recherche alors des informations d'ordre **statistique** sur le système considéré : valeurs moyennes de vitesse, de position, d'énergie, de prix – ou bien probabilités d'occurrence d'événements critiques

(fracture de longueur trop grande, pertes financières importantes, grande proportion de population infectée par un virus) – ou encore densité d’occupation pour une particule représentative de tout un système, densité de l’intervalle de temps entre deux décharges neuronales.

Jusqu’à présent nous avons été vagues sur les objets probabilistes mis en jeu dans les modèles. Les quelques exemples ci-dessus illustrent que les modèles probabilistes sont souvent appliqués à des phénomènes évoluant au cours du temps. Il n’est donc pas étonnant que l’on ait recours aux **processus stochastiques**. Nous définirons plus loin la **loi d’un processus stochastiques** comme une probabilité sur l’espace fonctionnel de ses trajectoires, ce qui permet de donner un sens rigoureux à ce que nous appelions ci-dessus « tirer au hasard une fonction ». Nous verrons comment tirer au hasard sur ordinateur, non seulement des réalisations de variables réelles selon des lois de probabilité sur \mathbb{R}^d , mais aussi des trajectoires de certains processus selon des lois de probabilité sur un espace de trajectoires. Nous nous limiterons essentiellement aux processus de saut pur et aux solutions d’équations différentielles stochastiques qui permettent de traiter à ce jour la plupart des modèles probabilistes utilisés en pratique.

Deux exemples détaillés illustrent particulièrement bien tout ce qui précède.

Le premier exemple concerne la dynamique moléculaire. Le texte est extrait de la thèse de S. Park [21] dirigée par un grand spécialiste du domaine, K. Schulten :

‘A protein is an unbranched chain of amino acids. (...) When placed in water (or, for some proteins, in a membrane), a protein folds into a particular structure which is also uniquely determined by the amino acid sequence.(...) In principle, the dynamics of biomolecules such as proteins must be described by solving Schrödinger’s equation for every constituent particles (nuclei and electrons). However, this is impractical because of the large size of biomolecules. The Born–Oppenheimer approximation provides a more practical way to model large molecules. The basic idea is that since electrons are much lighter than nuclei, electrons can rapidly adjust to the motion of nuclei. (...) A model of molecular interactions in terms of nuclear coordinates (...) is called a *force field*. In principle, since proteins are supposed to fold by themselves, a good force field should be enough to obtain naive protein structures from simulations. But, whereas protein folding typically takes milliseconds or longer, today’s molecular dynamics simulation is limited to the nanosecond time scale. This is because fast vibrational motions of atoms allow only a small time step for the integration of the equation of the motion. (...) Steered Molecular Dynamics is an efficient method which permits us to focus on important degrees of freedom while minimizing computational cost (...) The *reaction coordinate* is basically the most important degree of freedom, with the others being considered as fluctuations. With all the other degrees of freedom averaged out, the motion along the reaction coordinate is often well approximated by a **diffusive Brownian motion on an effective potential**.’

Le second exemple concerne les phénomènes multi-échelles, en particulier en météorologie. Le texte a été écrit par C. Penland [22], NOAA Climate Diagnostics Center (Boulder, Colorado), grande spécialiste des modèles probabilistes en climatologie :

‘The need to numerical model the interaction between geophysical processes having different timescales has led many modelers to represent rapidly varying components of the system as stochastic forcing. The method these stochastic modelers use to do this are almost as numerous as the modelers themselves. Yet, there does exist a prescription for making the stochastic approximation in a systematic manner consistent with the multiscale dynamics.

Many of us are familiar with some of the Central Limit Theorem, which states how sums of weakly dependent quantities are approximately Gaussian distributed. There is another version that states the conditions under which a multiscale dynamical system may be approximated as depending on the realizations of a whitenoise process, that is, as a stochastic differential equation’.

(...)

‘The importance of stochastic differential equations to probabilistic forecasting is only now beginning to be appreciated in meteorology and climate research. The advances in the subject have already allowed us to make real-time predictions of climate systems much more complicated than the simple univariate, additive-noise case considered by Hassellman (1976) in his landmark paper. We hope, and expect, that climate researchers will increasingly value the great utility and intrinsic beauty of the theory.’

2) Quel est le but d’une simulation aléatoire ?

Ainsi que nous venons de le souligner, les probabilités permettent de modéliser des phénomènes complexes dont les états sont impossibles à décrire de manière précise à partir de données mesurées avec précision. Le cadre ne peut alors plus être classique. La modélisation probabiliste consiste à se donner des lois de probabilité sur les données du problème (paramètres, états possibles, transformations possibles de ces états, ...). Dès lors, pour que la démarche soit consistante, cela n’a guère de sens de simuler des états aléatoires particuliers. Bien au contraire, il faut chercher des informations sur les lois de probabilité de quantités significatives pour le système étudié. Par exemple, en reprenant certains des exemples déjà évoqués :

- On se donne une loi de probabilité pour décrire, dans un marché financier risqué-neutre, l’évolution des prix futurs d’une action, et on cherche à calculer la valeur moyenne du flux d’un contrat reposant sur cette action.
- On se donne une loi de probabilité pour décrire les réactions physico-chimiques le long d’un axone, et on cherche à déterminer la densité de la loi de passage du potentiel électrique le long de l’axone au niveau de décharge.

- On se donne une loi de probabilité pour décrire les volumes des paquets de bits circulant dans un réseau ainsi que leurs dates d'émission, et on cherche à estimer la probabilité de congestion rapide d'une configuration donnée du réseau.

Dans tous les cas il s'agit donc de calculer une loi de probabilité, ou une espérance, ou une probabilité d'occurrence. Ces informations peuvent parfois être obtenues numériquement à l'aide d'algorithmes déterministes : par exemple, les prix d'options peuvent s'écrire à l'aide de solutions d'équations aux dérivées partielles paraboliques ; si la dimension d'espace n'est pas trop grande, on peut utiliser efficacement des algorithmes de résolution d'équations aux dérivées partielles ; inutile alors de recourir aux méthodes numériques probabilistes. Il en est de même pour tous les cas où l'on veut approcher une quantité du type $\mathbb{E} \Psi(X_T)$, où Ψ est une fonction donnée et (X_t) un processus stochastique, et où la densité de X_T peut être obtenue en résolvant une équation aux dérivées partielles (c'est le cas des solutions d'équations différentielles stochastiques), à condition toutefois que la dimension de (X_t) soit au plus trois ou quatre.

Ce cours, quant à lui, est consacré à quelques algorithmes probabilistes. Nous verrons que ces algorithmes permettent, non seulement d'obtenir des informations sur des modèles probabilistes, mais aussi de résoudre des équations complètement déterministes, y compris des équations décrivant des phénomènes sans aucun aspect aléatoire, mais difficiles à résoudre par des méthodes déterministes. L'exemple le plus simple est le suivant : la fonction f étant intégrable dans l'hypercube $[0, 1]^d$, on veut approcher la valeur de l'intégrale

$$I := \int_{[0,1]^d} f(x_1, \dots, x_d) dx_1 \cdots dx_d.$$

Les méthodes de quadrature numérique consistent à discrétiser l'espace d'intégration pour obtenir un ensemble de points $(\alpha_1^\ell, \dots, \alpha_d^\ell)$ de l'hypercube $[0, 1]^d$ et à utiliser des poids ρ_ℓ pour, finalement, approcher I par la somme pondérée

$$\sum_{\ell=1}^L \rho_\ell f(\alpha_1^\ell, \dots, \alpha_d^\ell).$$

Comment choisir le nombre L de points de la somme ? La réponse dépend de la précision souhaitée, de la régularité des dérivées de la fonction f , et surtout de la dimension d : si, pour la précision souhaitée, la méthode de quadrature utilisée requiert M points pour une intégrale en dimension 1, il en faudra M^d en dimension d . On conçoit aisément que, lorsque d est 6 (dimension de l'espace des phases en mécanique) les calculs peuvent être prohibitifs. A fortiori, lorsque d est de l'ordre de plusieurs centaines ou plusieurs milliers (situation fréquente en physique statistique), le nombre de termes de la somme rend le calcul totalement inenvisageable.

Par contre, une approche probabiliste permet de développer une méthode numérique dont la complexité croît seulement linéairement en fonction de la dimension d

(et dont, par ailleurs, la précision dépend de la norme $L^2([0, 1]^d)$ de f plutôt que de la régularité de ses dérivées). En effet, remarquons que I admet la **représentation probabiliste** suivante :

$$I = \mathbb{E} f(U_1, \dots, U_d)$$

où les U_i sont des variables aléatoires indépendantes de loi uniforme sur l'intervalle $[0, 1]$. La Loi Forte des Grands Nombres assure donc que I peut être approchée par

$$\frac{1}{N} \sum_{\ell=1}^N f(U_1^{(\ell)}, \dots, U_d^{(\ell)}),$$

où les $U^{(\ell)}$ sont des variables aléatoires indépendantes de loi uniforme sur l'intervalle $[0, 1]$, et N est « assez grand ». On appelle *méthode de Monte-Carlo* cette technique d'approximation de la quantité déterministe I par une moyenne de réalisations indépendantes d'une variable aléatoire.

De manière générale, la convergence des algorithmes probabilistes repose sur la Loi Forte des Grands Nombres (ou sur une extension que nous n'étudierons pas dans ce cours : le théorème ergodique). La plupart des algorithmes que nous étudierons consistent, comme dans l'exemple précédent, à simuler un grand nombre de variables aléatoires indépendantes de loi commune la loi choisie pour modéliser, puis à moyenner sur les simulations réalisées. On obtient ainsi une approximation aléatoire de la quantité recherchée. Nous étudierons avec soin comment évaluer et contrôler cette erreur en fonction du nombre de simulations et d'autres paramètres numériques, notamment le paramètre de discrétisation en temps des équations différentielles stochastiques. Nous nous attacherons à estimer des vitesses de convergence appropriées au but poursuivi, à savoir l'approximation de statistiques particulières de certaines lois de probabilité. Citons à ce propos les recommandations du 'Theory and Modeling in Nanoscience Report of the May 10 11, 2002, Workshop' émanant du Basic Energy Sciences and Advanced Scientific Computing Advisory Committees to the Office of Science, U.S. Department of Energy :

'It was noted that upscaling (going to coarser length and time scales) typically results in the introduction of stochastic terms to reflect high-frequency motion at smaller scales ; thus, more rigorous methods for including stochasm and more effective methods for characterizing and solving stochastic differential equations are required. New developments in nonequilibrium statistical mechanics, both of classical and quantum systems, are needed. For example, the recent discovery by nonequilibrium molecular dynamics, confirmed by nonlinear response theory, of violations of the second law of thermodynamics for nanoscale systems indicates the importance of new theoretical developments focused on nanoscale systems. Quantifying errors in calculated properties was identified as a significant problem that consists of two parts characterization of uncertainty due to inaccuracies inherent in the calculation (such as limitations of a force field

in molecular dynamics or of a basis set in electronic structure calculations) and performing useful sensitivity analyses. The important aspect of nanoscale systems that makes them such a Theory Model Simulation (TMS) challenge is that the characterization of uncertainty will often have to be done in the absence of experimental data, since many of the property measurement experiments one would like to perform on nanoscale systems are impossible in many cases or unlikely to be done in cases where they are possible. Hence, the concept of self-validating TMS methods arises as a significant challenge in nanoscience. (...)

Some degree of stochasticity is intrinsic at atomistic scales ; hence, a Monte Carlo simulation will yield a fluctuation with statistics that must be carried to larger length scales. The numerical analysis of stochastic partial differential equations could enter here, the goal being a coarse solution with the right statistics. The theory of large deviation for stochastic PDEs allows the designing of sampling methods for so-called rare events. The long time integrations that are burdensome to the accuracy of deterministic models may actually be a benefit to the accuracy of statistical models. It is axiomatic to a mathematician that there are better ways than brute-force propagation of fine scales to attack a multiscale problem. A mathematician would argue that anyone requiring a billion degrees of freedom for a week of wall clock execution time is running the wrong algorithm. Indeed, the nanoscale community must learn to compute smarter and not just harder.'

Plan de l'ouvrage

Cette monographie est organisée de manière à remplir un double objectif :

- décrire les méthodes numériques probabilistes de base et analyser finement leurs vitesses de convergence par rapport aux paramètres de la simulation : le nombre de simulations indépendantes de la méthode de Monte-Carlo, et le pas de discrétisation en temps ;
- introduire de manière aussi auto-suffisante que possible les outils avancés de la théorie des probabilités et du calcul différentiel stochastique (inégalités de concentration, flots stochastiques, interprétation probabiliste d'équations aux dérivées partielles, etc.) pour établir des vitesses de convergence optimales.

Des exercices sont présentés au fil du texte, qu'ils peuvent préciser. Chaque chapitre se conclut par quelques problèmes qui peuvent apporter des compléments utiles, les plus difficiles étant marqués d'une étoile (★).

Le chapitre 1 traite de la convergence des méthodes de Monte-Carlo, et donc de la Loi Forte des Grands Nombres. Nous décrivons aussi des techniques de simulation des variables aléatoires, qui sont les briques de base des méthodes numériques probabilistes. Nous donnons une preuve de la Loi Forte des Grands Nombres en utilisant un objet mathématique essentiel dans toute la théorie moderne des probabilités et dans cet ouvrage, à savoir les familles de variables aléatoires dépendant du temps qui ont la propriété de *martingale*. Cette preuve permet ainsi au lecteur de se familiariser avec cet objet crucial pour la suite.

Le chapitre 2 traite de la vitesse de convergence des méthodes de Monte-Carlo. Son contenu diffère de la plupart des ouvrages introductifs au calcul des probabilités qui se limitent à la démonstration du Théorème Limite Central. Notre but étant de déterminer le nombre de simulations nécessaire pour atteindre une précision souhaitée avec un seuil de confiance prescrit, nous mettons l'accent sur des estimations *non asymptotiques* de la fonction de répartition de l'erreur normalisée d'une méthode de Monte-Carlo : nous insistons donc sur les théorèmes de Berry–Esseen et de Bikelis, ainsi que sur les inégalités de concentration pour les moyennes d'échantillons de variables aléatoires indépendantes. Nous abordons aussi la question très difficile en pratique de la réduction de la variance de l'erreur d'une méthode de Monte-Carlo.

Le chapitre 3 commence par décrire les enjeux pratiques et théoriques de la modélisation par des processus de Markov à trajectoires continues à droite avec limites à gauche (cadlag). Ceci amène à introduire le processus de Poisson en tant que processus

ponctuel sans mémoire. La suite du chapitre consiste en une étude assez poussée de ce processus, incluant divers points de vue sur sa simulation et son approximation. Ceci se révélera fondamental aussi bien pour la compréhension et la construction abstraites que pour la simulation efficace des processus de Markov avec sauts.

Le chapitre 4 est une étude assez complète des processus de Markov à espace d'état discret, et montre leurs liens avec les processus de Poisson et les chaînes de Markov à temps discret. L'étude est essentiellement trajectorielle, et intimement mêlée aux nécessités de la simulation. Cette approche montre l'existence du processus de Markov pour un générateur infinitésimal avec taux de saut uniformément borné. Ce cas donne lieu à des développements autour des algèbres d'opérateurs bornés, qui éclairent les liens entre générateur et semi-groupe, et fournissent un bon cadre pour les équations de Kolmogorov progressives et rétrogrades et la formule de Feynman-Kac.

Le chapitre 5 aborde les processus de Markov à valeurs dans un espace d'états « continu », typiquement un fermé de \mathbb{R}^d . Leur étude rigoureuse exigerait des outils avancés de la théorie de la mesure, mais on essaie plutôt de développer et de favoriser l'intuition, notamment par des constructions trajectorielles liées à la simulation.

On met rapidement en évidence les fortes similitudes entre de tels processus qui n'évoluent que par des sauts à des instants isolés et les processus de Markov à espace discret. On introduit ensuite une classe de processus de Markov évoluant de manière continue par morceaux, sans excès de rigueur. On établit les équations de Kolmogorov et la formule de Feynman-Kac. On applique ceci aux équations cinétiques linéaires (équations de transport), qui décrivent l'évolution de particules dans l'espace des phases position-vitesse en physique statistique, et à quelques extensions.

Le chapitre 6 est consacré à des schémas de discrétisation d'équations différentielles stochastiques et leur application à la résolution numérique probabiliste des équations aux dérivées partielles paraboliques. Nous commençons par rappeler quelques propriétés essentielles des intégrales stochastiques d'Itô par rapport au mouvement brownien et le théorème d'existence et d'unicité des solutions d'équations différentielles stochastiques à coefficients lipschitziens. Ensuite nous démontrons à l'aide d'outils probabilistes exclusivement des résultats d'existence, d'unicité et de régularité pour les solutions d'équations aux dérivées partielles paraboliques : pour ce faire, nous introduisons la notion de *flot stochastique* associé à une d'équation différentielle stochastique, et en établissons quelques propriétés. Nous démontrons ainsi un résultat optimal de vitesse de convergence pour les schémas de discrétisation considérés.

Le chapitre 7 revient de manière détaillée sur le thème de la réduction de la variance de l'erreur d'une méthode de Monte-Carlo, en se concentrant au contexte de la résolution numérique probabiliste des équations aux dérivées partielles paraboliques. Ce sujet demande des outils assez avancés de calcul différentiel stochastique, notamment le théorème de Girsanov, que nous commençons par rappeler et commenter.

Le chapitre 8 aborde quelques enjeux théoriques et pratiques des algorithmes stochastiques d'optimisation. Il s'agit d'outils efficaces pour des applications variées, dont le principe est à la base des techniques de réduction de variance du chapitre 7. Nous présentons un cadre idéalisé, et y établissons des propriétés de convergence sans excès de technicité grâce aux théories des systèmes dynamiques et des martingales.