

Projets MODAL MAP 473D
Simulation Numérique Aléatoire (SNA)
autour des évènements rares
Promotion X2018

Equipe pédagogique :

S. De Marco, G. Fort, E. Gobet, A. Guyader,
I. Kortchemski, O. Saadi, A. Singh

Période : Février 2020 - Mai 2020

(Version du 17 février 2020)

Table des matières

| | | |
|----------|--|-----------|
| 1 | Modélisation d'épidémies | 4 |
| 2 | Risques extrêmes en finance de marché | 8 |
| 3 | Sûreté d'un réacteur nucléaire | 12 |
| 4 | Dynamique de population, extinction | 16 |
| 5 | Sûreté aérienne | 19 |
| 6 | Ruines en assurance | 23 |
| 7 | Barrages et risques d'inondation | 27 |
| 8 | Microstructure des prix financiers | 32 |

Objectifs. Il s'agit de réaliser un projet de modélisation et simulation sur un des thèmes applicatifs décrits ci-dessous, dans le contexte des événements rares.

Des références bibliographiques sont données dans le descriptif de chaque projet, ces documents sont accessibles sur le moodle du cours ou via la BCX. Prendre connaissance de ces références est important et fait partie du travail d'expérimentation et d'investigation du MODAL. Par cette démarche on attend des étudiants qu'ils :

- se familiarisent avec le thème proposé,
- en comprennent les enjeux de sorte à se poser les *bonnes questions*,
- développent des outils stochastiques de simulation pour mieux appréhender les événements rares (quantile, distribution conditionnelle à l'évènement, statistique dans de tels événements, scénario rare typique...). Ils auront en particulier à adapter à leur contexte les méthodes abordés en cours ou en TP.

Cela implique de l'autonomie et des initiatives.

Il est demandé de tester différentes approches, de les *comparer quantitativement et qualitativement*, avec un regard critique.

Encadrement : une permanence de suivi de projets est assurée chaque vendredi après-midi, de 13h30 à 15h30.

Livrables : un rapport écrit, avec les codes de simulation en annexe.

Outils informatiques : il est demandé de choisir **Python** comme langage de développement, nous recommandons de le faire sous forme de notebooks (fichier `.ipynb`). Toutefois, on pourra préférer exécuter des codes Python de fichier `.py` (écrit via un IDE, Spyder par exemple) : en effet cette approche offre plus de flexibilité (moins de contraintes de mémoires, run sur des serveurs distants, parallélisme...).

Soutenance : elle est prévue **vendredi 22 mai 2020**, d'une durée de 45' (30' de présentation, 15' de questions).

Barème de notation du projet : le projet est noté sur 20 pts.

- (4pts) Capacités d'initiatives/autonomie face aux difficultés/problèmes soulevés par le projet.

- (4pts) Capacités à décomposer le problème en plusieurs étapes intermédiaires, avant de parvenir à une version aboutie (démarche d'expérimentation).
- (5pts) Capacités à mesurer les améliorations, les apports, comparaison numérique.
- (4pts) Rapport écrit.
- (3pts) Soutenance orale.

Les 3 premiers items sont principalement notés pendant les séances de TPs.

1 Modélisation d'épidémies

Pour des questions de santé publique, il est important de prévoir l'évolution d'une épidémie dans une population. Historiquement, les premiers modèles mathématiques d'épidémies remontent au 18ième siècle avec, en particulier, des travaux de D. Bernoulli sur l'évolution de la variole. Depuis le début du 20ième siècle, de nombreux modèles ont été imaginés afin de simuler la propagation des épidémies, comme par exemple, l'évolution de l'épidémie du virus Ebola en Afrique en 2014

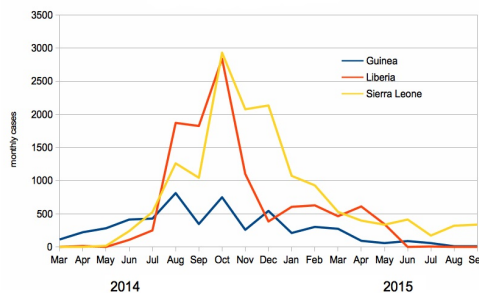


FIGURE 1 – Évolution de l'épidémie d'Ebola 2014 (source wikipedia)

ou encore la résurgence de la rougeole causée par la baisse du taux de vaccination. Plus récemment, on pense évidemment à modéliser l'évolution de l'épidémie de coronavirus ncov-2019 en Chine. . . On pourra consulter [4] pour un aperçu de différents modèles et de leurs applications.

Le but de ce projet est de regarder des modèles mathématiques classiques faisant partie de la classe des modèles d'épidémies dit *compartimentaux* c.f. [1]. Il s'agit de modèles où les individus sont classés selon des catégories : *sain*, *infecté*, *immunisé* . . . et où l'on précise les mécanismes de transitions entre les différents états selon des règles qui peuvent être déterministes ou probabilistes.

1.1 Modèle SIR

On considère une population de N individus. Chaque individu est dans un des 3 états suivants : *sain* (S), *infecté* (I), *immunisé/recovered* (R). On note par $S(t), I(t), R(t)$ le nombre d'individus au temps t dans chaque catégorie. Dans le modèle SIR, un individu ne peut être infecté qu'une fois puis devient immunisé par la suite (n'interagissant donc plus avec le reste de la population). Dans sa version *déterministe*, le modèle SIR est régi par le système d'équations différentielles :

$$\begin{cases} \frac{dS(t)}{dt} = -\frac{\beta I(t)S(t)}{N} \\ \frac{dI(t)}{dt} = \frac{\beta I(t)S(t)}{N} - \gamma I(t) \\ \frac{dR(t)}{dt} = \gamma I(t) \end{cases}$$

Ici, le paramètre β contrôle le taux d'infection et γ contrôle la durée des infections. On note $\mathcal{R}_0 = \beta/\gamma$ le *taux de reproduction de base* qui détermine l'évolution du système, c.f. [1, 4]. Le modèle ci-dessus est applicable dans le cas de grandes populations (N grand). Lorsque la population reste bornée, il est intéressant de considérer un modèle SIR *stochastique* qui tient compte de l'aléa des interactions entre individus. Pour cela, on considère la chaîne de Markov à temps continu $(S(t), I(t), R(t), t \geq 0)$, partant initialement de i_0 individus contaminés ($S(0) = N - i_0, I(0) = i_0, R(0) = 0$) et avec les taux de transitions suivants :

$$\begin{aligned} (S, I, R) &\mapsto (S - 1, I + 1, R) \quad \text{avec taux } \frac{\beta IS}{N}, \\ (S, I, R) &\mapsto (S, I - 1, R + 1) \quad \text{avec taux } \gamma I. \end{aligned}$$

L'épidémie se termine lorsque la chaîne de Markov atteint un état absorbant $I(t) = 0$. On notera $S(\infty)$, resp. $R(\infty)$ le nombre final d'individus épargnés (resp. contaminés).

Questions

1. Simuler numériquement le modèle SIR déterministe. Observer les différentes évolutions possibles de l'épidémie selon la valeur du taux de reproduction de base \mathcal{R}_0 . Quelle valeur critique apparaît pour \mathcal{R}_0 ?
2. Simuler le modèle SIR stochastique. On réfléchira en particulier à la manière de simuler la chaîne de Markov en temps continu (discrétisation ou construction à l'aide d'horloges exponentielles...).
 - Regarder l'influence des paramètres $I(0)$, β et γ sur la propagation de l'épidémie.
 - Dans quels cas le modèle SIR déterministe approxime-t-il bien le modèle SIR probabiliste ? Pour des résultats théoriques sur ce sujet, on pourra par exemple consulter le chap. 5 de [3].
3. On s'intéresse au nombre total $R(\infty)$ de personnes infectées au cours de l'épidémie. On cherchera à estimer la probabilité $\mathbb{P}(R(\infty) > N_0)$ pour différentes valeurs de N_0 :
 - Lorsque $\mathcal{R}_0 > 1$, on choisira $N_0 = N - o(N)$, c'est-à-dire que l'on regardera les événements (rares) ou presque toute la population est infectée. D'un point de vue pratique, on mettra en œuvre différentes approches (M.C., Importance sampling, IPS...) pour estimer $\mathbb{P}(R(\infty) > N_0)$ et l'on comparera leur efficacité respective. On pourra se référer à [5] pour la mise en œuvre de certaines méthodes. On comparera ensuite les résultats numériques avec les valeurs théoriques qui peuvent être obtenues explicitement dans le cadre de ce modèle c.f. Théorème 2.2 de [3].

- Dans le cas critique $\mathcal{R}_0 = 1$, comment faut-il choisir N_0 pour être encore dans le régime des événements rares ?
- Finalement, on pourra aussi s'intéresser au cas où \mathcal{R}_0 grandit logarithmiquement avec la taille de la population. D'un point de vue théorique, on pourra étudier la preuve (élégante) de [2] qui établit alors l'existence d'une distribution Poissonnienne pour $S(\infty) = N - R(\infty)$ et comparer ce résultat avec des estimations numériques.

Extensions envisageables.

On pourra regarder d'autres modèles d'épidémies, par exemple le modèle SIS où les individus redeviennent susceptibles à la fin de leur période d'infection ou bien encore des modèles compartimentaux plus complexes, par exemple avec de la démographie (naissances et morts dans la population).

1.2 Épidémie sur des graphes

Le modèle SIR défini précédemment est un modèle dit en champs moyen : tous les individus sont en contact avec tous les autres. Or il est naturel de vouloir tenir compte de la répartition géographique des individus. Pour cela, on représente maintenant l'épidémie sur un graphe $G = (V, E)$ où les sommets V représentent les individus et une arête de E relie deux individus si ils sont en contact direct et donc susceptibles de se transmettre une infection. On considère alors la dynamique markovienne sur $\{S, I, R\}^V$:

- Un site infecté (I) devient immunisé (R) à taux ρ .
- Un site sain (S) devient infecté (I) à un taux αk où k est le nombre de ses voisins qui sont infectés.

Questions

1. Pour quel graphe et pour quel choix de paramètres (ρ, α) retrouve-t-on le modèle SIR traité au paragraphe précédent ?
2. Dans le cas où $G = \mathbb{Z}$ et qu'un unique individu est initialement infecté, décrire la loi du nombre de personnes infectées ainsi que la durée totale de l'infection. On calculera en particulier les espérances de ces deux lois.
3. Dans le cas où G est un arbre régulier et que seule la racine de l'arbre est initialement infectée, montrer qu'une transition de phase apparaît en fonction de α, ρ et du degré des sommets de l'arbre. Dans le cas sous-critique (où l'épidémie s'éteint p.s.) on pourra regarder, dans le régime des événements rares, la loi conditionnelle du nombre de personnes

ayant été infectées à la distance n sachant que l'épidémie s'est propagée jusque là.

4. Dans le cas $G = \mathbb{Z}^2$. On étudiera par simulation l'existence d'une transition de phase en fonction des paramètres du modèle et on pourra estimer numériquement la taille et la durée de l'infection dans le régime sous-critique.

Références

- [1] Wikipedia. *Modèles compartimentaux en épidémiologie*. https://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology
- [2] Sellke, Th. (1982) On the asymptotic distribution of the size of a stochastic epidemic. *Technical reports No 17. Stanford University*. 1982. <https://statistics.stanford.edu/research/asymptotic-distribution-size-stochastic-epidemic>
- [3] Andersson H. et Britton T. (2000) Stochastic epidemic models and their statistical analysis. *Lecture Notes in Statistics, Vol 151. Springer-Verlag, New York*. <https://staff.math.su.se/tom.britton/newmain.pdf>
- [4] Hethcote, H. (2000) The mathematics of infectious diseases. *SIAM Review Vol 42. No 4*. <http://www.maths.usyd.edu.au/u/marym/populations/hethcote.pdf>
- [5] Cléménçon, S. et al. (2015). On computer-intensive simulation and estimation methods for rare-event analysis in epidemic models. *Stat. Med. Vol 34. No 28*. <https://hal.archives-ouvertes.fr/hal-00854458/document>

2 Risques extrêmes en finance de marché

Les produits dérivés sont des *produits financiers*, définis à partir d'un autre produit financier plus simple comme par exemple une action, un indice, une devise, une matière première ou un taux d'intérêt.

On trouve l'existence des produits dérivés à l'époque du pharaon Akhé-naton (XIV^e s. av. JC), pour garantir à l'avance l'achat et la vente de blé. Ce n'est que dans les années 1970 que les marchés financiers se sont vraiment organisés pour échanger de tels produits, les banques endossant par exemple les risques de fluctuation des taux de change à la place d'entreprises internationales leur ayant acheté un produit d'assurance ad hoc.

Depuis une quinzaine d'années, sous l'impulsion du Comité de Bâle, se développe une vision globale des risques au niveau bancaire. Cette vision globale a été adoptée dès 1998, avec la Value At Risk (VaR), indicateur de risque mesurant le seuil de *pertes potentielles* à un horizon donné et pour un quantile donné (typiquement 99% à horizon 10 jours, ou 99.9% ...). Cet indicateur réglementaire est relié au montant de fonds propres à immobiliser en regard des risques endossés, pour faire face à des pertes a priori rares. Ce projet se concentre sur l'évaluation des pertes extrêmes sous l'angle fréquence, sévérité et scénario les engendrant, en considérant en particulier des seuils à 99.99% voire plus.

● Portefeuilles d'options et modèles gaussiens

Le produit dérivé de base est le *Call* (option d'achat) donnant le droit (et non l'obligation) à son détenteur d'acheter à une date future T (*maturité*, de l'ordre de 6 mois à 1 an) une action de cours $(S_t)_{t \geq 0}$ à un prix d'exercice K fixé à l'avance. A la date T , cela conduit à un équivalent financier égal à $(S_T - K)_+$. De manière analogue, le *Put* (option de vente) donne un équivalent financier égal à $(K - S_T)_+$.

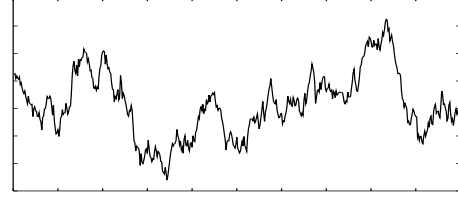
On supposera que le cours d'action est modélisé par un mouvement brownien géométrique

$$S_t = S_0 e^{-\frac{1}{2}\sigma^2 t + \sigma W_t}.$$

Précisons les notations :

- $W = (W_t)_{t \geq 0}$ est un mouvement brownien, c'est-à-dire un processus

continu à temps continu, partant de $W_0 = 0$, à accroissements indépendants et stationnaires, dont les accroissements $W_{t+h} - W_t$ sont de loi gaussienne $\mathcal{N}(0, h)$.



- Le rendement de S (variation relative) sur un court intervalle de temps $h \ll 1$ peut s'écrire $\frac{S_{t+h} - S_t}{S_t} \approx \sigma(W_{t+h} - W_t)$. Ainsi, σ mesure l'écart-type des rendements (appelé volatilité).

Dans ce modèle négligeant les taux d'intérêt, le prix $\text{Call}(t, S_t)$ de l'option d'achat à la date t avant maturité T lorsque le sous-jacent vaut S_t est donné par la formule de Black & Scholes :

$$\text{Call}(t, S_t) = S_t \mathcal{N}[d_+(T - t, S_t, K)] - K \mathcal{N}[d_-(T - t, S_t, K)]$$

avec $d_{\pm}(\tau, x, y) = \frac{1}{\sigma\sqrt{\tau}} \ln\left(\frac{x}{y}\right) \pm \frac{1}{2}\sigma\sqrt{\tau}$, où \mathcal{N} est la fonction de répartition de la loi normale, centrée réduite. Le prix $\text{Put}(t, S_t)$ du Put a une forme symétrique $\text{Put}(t, S_t) = K \mathcal{N}[d_+(T - t, K, S_t)] - S_t \mathcal{N}[d_-(T - t, K, S_t)]$.

On considère un portefeuille combinant des calls/puts sur I_0 actions différentes. La valeur V_t du portefeuille est donc

$$V_t = \sum_{i=1}^{I_0} (\alpha_i \text{Call}_i(t, S_t^i) + \beta_i \text{Put}_i(t, S_t^i)).$$

On prendra par exemple

1. $I_0 = 10$, $\alpha_i = -10$, $\beta_i = -5$, pour tout i ;
2. $I_0 = 10$, $\alpha_i = 10$, $\beta_i = 5$, pour tout i ;
3. $I_0 = 10$, $\alpha_i = -10$, $\beta_i = -5$ pour $i = \{1, \dots, 5\}$ et $\alpha_i = 10$, $\beta_i = 5$ pour $i = \{6, \dots, 10\}$.

On étudiera également le cas où le portefeuille est couvert (*hedgé*) initialement, c'est-à-dire

$$\begin{aligned} V_t^H &= \sum_{i=1}^{I_0} \alpha_i (\text{Call}_i(t, S_t^i) - \partial_{S^i} \text{Call}_i(0, S_0^i) (S_t^i - S_0^i)) \\ &\quad + \sum_{i=1}^{I_0} \beta_i (\text{Put}_i(t, S_t^i) - \partial_{S^i} \text{Put}_i(0, S_0^i) (S_t^i - S_0^i)), \end{aligned}$$

et il sera intéressant de comparer les risques de chaque portefeuille (avec ou sans couverture).

Evolution indépendante. Supposons d'abord que les mouvements browniens modélisant les actions sont indépendants.

▷ QUESTIONS À TRAITER :

1. évaluer numériquement les VaR et Conditional VaR (avec un seuil à 99.99% ou 99.999%),
2. évaluer la distribution conditionnellement à être au delà de la VaR,
3. influence de la composition du portefeuille, de savoir s'il est couvert ou pas.

Evolution dépendante. Supposons maintenant que les mouvements browniens sont corrélés positivement ($\rho \geq 0$) du fait de la relation linéaire $W_t^i = \sqrt{\rho}B_t^0 + \sqrt{1-\rho}B_t^i$, les $(B^i)_{0 \leq i \leq I_0}$ étant des mouvements browniens indépendants. B^0 peut s'interpréter comme un facteur commun à toutes les actions du marché.

▷ Reprendre les questions précédentes, tout en identifiant l'impact de la corrélation. On cherchera aussi à exhiber des trajectoires typiques de B^0 associées aux pertes extrêmes.

● Extension à un modèle avec sauts

A la lumière de la crise, les pertes extrêmes apparaissent lors de chocs dans le marché, chocs qui se traduisent par des sauts négatifs sur les cours boursiers (d'amplitude de -5% à -10% par exemple). Il est alors assez naturel de superposer au mouvement brownien B^0 un processus de Poisson composé avec des sauts négatifs, pour modéliser ces chocs. On abordera les questions précédentes dans ce nouveau modèle.

● Références bibliographiques

Sur les options et les modèles stochastiques en finance : [6], [7].

Sur les mesures de risque en finance : [2].

Sur le calcul de risques extrêmes par simulations : [3], [4], [5], [1].

Références

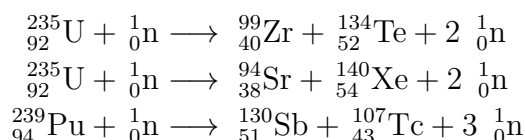
- [1] R. Carmona and S. Crépey. Particle methods for the estimation of credit portfolio loss distributions. *Int. J. Theor. Appl. Finance*, 13(4) :577–602, 2010.

- [2] H. Föllmer and A. Schied. *Stochastic finance*, volume 27 of *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 2002. An introduction in discrete time.
- [3] P. Glasserman, P. Heidelberger, and P. Shahabuddin. Variance reduction techniques for estimating value-at-risk. *Management Science*, 46 :1349–1364, 2000.
- [4] P. Glasserman, P. Heidelberger, and P. Shahabuddin. Portfolio value-at-risk with heavy-tailed risk factors. *Math. Finance*, 12(3) :239–269, 2002.
- [5] P. Glasserman. *Monte-Carlo methods in Financial Engineering*. Springer Verlag, New York, 2003.
- [6] D. Lamberton and B. Lapeyre. *Introduction au calcul stochastique en Finance*. Collection Mathématiques et Applications 9 - Ellipses, 1991.
- [7] P. Tankov and N. Touzi. *Calcul stochastique en finance*. Cours de l’Ecole Polytechnique, MAP552, 3A.

3 Sûreté d'un réacteur nucléaire

Le milieu du XX^e siècle a marqué le début du développement du nucléaire aux Etats-Unis, en URSS, en France . . . voir <http://nucleaire.cea.fr/fr/repere/historique.htm>. En dehors des applications militaires (bombe atomique, propulsion nucléaire navale), les applications civiles concernent surtout la production d'électricité via les centrales nucléaires. Depuis les années 1960, les technologies ont considérablement évolué, utilisant surtout le principe de fission alors que les projets à l'étude concernent la fusion (ITER par exemple).

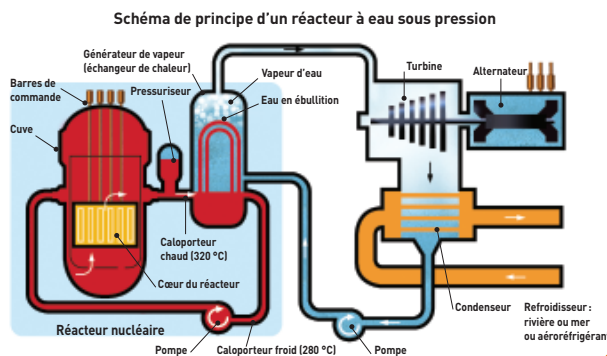
La fission nucléaire, découverte à la fin des années 30, est le phénomène de fragmentation de noyaux lourds (Uranium 235 ou 238, Plutonium 239) par capture de neutrons, donnant lieu à des nucléides plus légers et produisant une énergie importante. Par exemple



La fission produit de nouveaux neutrons qui peuvent entraîner une réaction en chaîne. L'énergie thermique dégagée est ensuite transformée (production de vapeur) en énergie mécanique ou électrique. Pour permettre à la réaction de continuer, il faut une quantité suffisante de combustible nucléaire dans le réacteur. A l'inverse, pour éviter un emballement de la réaction, des barres de commande peuvent être introduites dans le cœur du réacteur : ces barres constituées de matériaux spéciaux ont pour effet d'absorber les neutrons et ainsi d'arrêter la réaction.

L'accident de Three Mile Island en 1979 et la catastrophe de Tchernobyl en 1986 avaient déjà rappelé les risques du nucléaire. La catastrophe de Fukushima en 2011 a relancé le débat sur l'évaluation des risques extrêmes, dans un contexte où nos sociétés s'interrogent sur nos modèles de dépendance énergétique¹ et le développement d'énergies renouvelables.

L'accident de Three Mile Island en 1979 et la catastrophe de Tchernobyl en 1986 avaient déjà rappelé les risques du nucléaire. La catastrophe de Fukushima en 2011 a relancé le débat sur l'évaluation des risques extrêmes, dans un contexte où nos sociétés s'interrogent sur nos modèles de dépendance énergétique¹ et le développement d'énergies renouvelables.



Source : <http://www.cea.fr/content/download/4687/27890/file/livret-reacteur.pdf>

1. En France, la filière nucléaire fournit à elle seule 78% de la production électrique.

Dans une centrale nucléaire, le danger le plus important provient d'un défaut de refroidissement du cœur du réacteur (avec pour conséquence un risque de fusion libérant les éléments très radioactifs). La prise en compte de tous les facteurs est complexe² et nous étudions un modèle simplifié, cherchant à évaluer la capacité des barres de commande à absorber les neutrons (arrêt d'urgence de la réaction), dans un contexte où les réflecteurs entourant le cœur sont endommagés (fissure, dégradation des propriétés réfléchissantes).

● Modèle neutronique

Nous modélisons le déplacement d'un seul neutron. Pendant son parcours (modélisé dans cette partie dans le plan), il est soumis à de nombreux chocs qui se produisent de manière imprévisible : à chaque choc, sa vitesse saute d'une quantité tirée aléatoirement et indépendamment de son parcours précédent. Si $(\tau_i)_i$ est la suite des instants de chocs, sa vitesse (en km/s) est

$$V_t = V_{\tau_i^-} + Y_i, \quad \text{pour } t \in [\tau_i, \tau_{i+1})$$

où $(Y_i)_i$ est une suite de v.a.i.i.d. dans \mathbb{R}^2 modélisant les sauts de vitesse (chocs). La vitesse à la date t est donc :

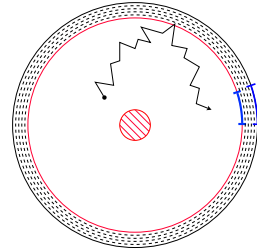
$$V_t = V_0 + \sum_{i=1}^{N_t} Y_i$$

où N est un processus de Poisson d'intensité λ grande. Les variables $(Y_i)_i$ et le processus N sont indépendants. La loi de Y est centrée et on pourra choisir une loi gaussienne sur chaque coordonnée, d'écart type 2000 km s^{-1} (neutron lent). Les ordres de grandeur réalistes pour λ sont de l'ordre 10^{12} s^{-1} à 10^{15} s^{-1} .

Si l'on ne tient pas compte des barres de commande et de l'enceinte du réacteur, la position du neutron au cours du temps évolue donc comme

$$X_t = X_0 + \int_0^t V_s ds.$$

C'est un modèle simple d'*équation de transport*, voir [3]. Le cœur du réacteur est composé d'une enceinte extérieure réfléchissante (qu'on suppose circulaire de diamètre $D = 5m$), et au centre d'une barre de commande (de diamètre d). Plusieurs situations peuvent se présenter.



Modèle de réacteur et déplacement de neutron

2. voir la conférence http://www.coriolis.polytechnique.fr/Confs/Hervieu_conf.pdf

- Lorsque le neutron touche la barre de commande, il est absorbé et s'arrête.
- Lorsque le neutron atteint l'enceinte réfléchissante, il est réfléchi vers l'intérieur.
- Une partie de l'enceinte est endommagée, sur une portion angulaire mesurant quelques mm. On peut imaginer que si le neutron atteint cette partie, il quitte le cœur du réacteur.

On cherche à évaluer la probabilité de la perte de contrôle du neutron décrit par l'évènement

$A = \{\text{le neutron atteint la zone endommagée avant d'être absorbée par la barre}\}.$

▷ QUESTIONS À TRAITER :

- Approximation asymptotique du modèle lorsque l'intensité est très grande et application. On pourra chercher à remplacer le modèle avec un modèle gaussien.
- Evaluer $\mathbb{P}(A)$ en fonction des positions et vitesses initiales.
- Déterminer la taille de la barre de commande pour maintenir une probabilité inférieure à 10^{-5} .
- Déterminer la distribution du temps d'absorption conditionnellement à A^c ou du temps de sortie conditionnellement à A .

● Réacteur à plusieurs barres de commande

Le cœur du réacteur contient maintenant 4 barres de commande disposées en carré. A quelle distance du centre disposer les barres pour garantir une efficacité maximale ? Discuter de l'influence des autres paramètres.

● Références bibliographiques

Sur le nucléaire : <http://www.cea.fr/content/download/4687/27890/file/livret-reacteur.pdf>.

Site de l'Autorité de Sûreté Nucléaire : <http://www.asn.fr/index.php>.

Sur les équations de transport : [3] ou [2].

Sur la neutronique : [1].

Sur les asymptotiques de diffusion : [4].

Références

- [1] J. Bussac and P. Reuss. *Traité de neutronique*. Hermann, Paris, 1978.
- [2] M. Cessenat, R. Dautray, G. Ledanois, P.L. Lions, E. Pardoux, and R. Sentis. *Méthodes probabilistes pour les équations de la physique*. Collection CEA, Eyrolles, 1989.
- [3] B. Lapeyre, E. Pardoux, and R. Sentis. *Methodes de Monte-Carlo pour les processus de transport et de diffusion*. Collection Mathématiques et Applications 29 - Springer Verlag, 1998.
- [4] G. Papanicolaou. Asymptotic analysis of transport processes. *Bulletin of the American Mathematical Society*, 81(2) :330–392, 1975.

4 Dynamique de population, extinction

La dynamique des populations étudie la répartition et le développement quantitatif de populations d'individus. Elle s'intéresse aux mécanismes d'auto-régulation des populations, au problème de l'extinction ou de croissance d'une population ou à l'existence d'un éventuel état stationnaire ou quasi-stationnaire. Dans ce projet, pour simplifier nous considérons une seule population d'individus semblables, n'étant pas en compétition pour les ressources, sans interaction entre eux (pas de proie ni prédateur), sans processus de migration additionnel.

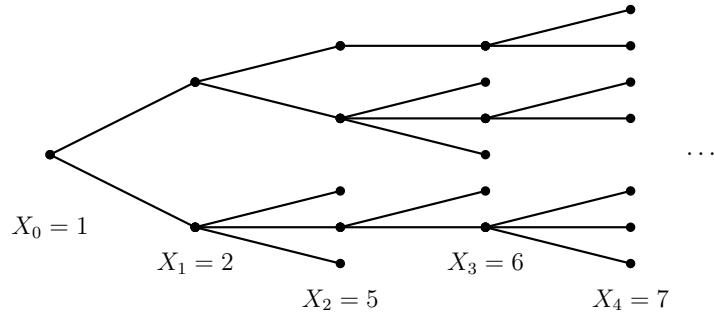


FIGURE 2 – Exemple d'arbre généalogique à partir d'un seul ancêtre

Un modèle simple et standard est le modèle de Bienaymé-Galton-Watson (BGW). On trouvera une analyse détaillée dans l'ouvrage [1]. Le modèle de BGW décrit l'évolution d'une population à travers l'évolution de la taille X_n de sa génération n . Avant de mourir, chaque individu k de la génération n donne naissance à $Y_{n,k}$ individus, de sorte que

$$X_{n+1} = \sum_{k=1}^{X_n} Y_{n,k}.$$

On suppose que les v.a $(Y_{n,k})_{n,k}$ sont indépendantes et de même loi ν . Le processus $(X_n)_{n \geq 0}$ est ainsi une chaîne de Markov à valeurs sur les entiers positifs. Pour que le problème ne soit pas dégénéré et un minimum intéressant, on supposera que $0 < \mathbb{P}(Y = 0) < 1$ et $\mathbb{P}(Y \geq 2) > 0$. La taille initiale X_0 de la population n'est pas forcément égale à 1.

Le point 0 est absorbant : si $X_n = 0$ pour un n , alors $X_p = 0$ pour tout $p \geq n$, autrement dit la population s'est éteinte. Le temps d'extinction est noté

$$T_E = \inf\{n \geq 0 : X_n = 0\}.$$

• Loi de reproduction géométrique

On suppose ici que $\nu = \mathcal{G}(m)$, c'est-à-dire

$$\mathbb{P}(Y = k) = \frac{1}{m+1} \left(\frac{m}{m+1}\right)^k, \quad k \geq 0.$$

Le paramètre $m > 0$ ci-dessus s'interprète comme la moyenne du nombre d'enfant d'un individu donné.

Le cas sous-critique : $m < 1$. Dans ce cas, *p.s.* la population s'éteint en temps fini. L'évènement de non-extinction $A_n = \{X_n > 0\}$ à la génération n devient rare si n est grand.

▷ QUESTIONS À TRAITER :

1. estimer numériquement $\mathbb{P}(A_n)$;
2. estimer la loi de X_n conditionnellement à la survie A_n ;
3. générer des scénarii typiques où la population n'est pas éteinte à la génération n .

L'influence jointe du nombre n de génération et de la taille initiale X_0 de la population sera aussi étudiée. Les impacts de m et de la distribution ν pourront être analysés. On pourra considérer également le problème des ressources consommées par la population, en évaluant $\mathbb{P}(\max_{1 \leq i \leq n} X_i > x)$ ou $\mathbb{P}(\sum_{1 \leq i \leq n} X_i > x)$ pour x grand.

Le cas sur-critique : $m > 1$. Dans ce cas, avec probabilité positive la population croît exponentiellement comme m^n . On s'intéressera au cas où la croissance est moins rapide, disons à la vitesse c^n où $c \in]1, m[$: c'est un évènement rare lorsque n est grand.

▷ QUESTIONS À TRAITER :

1. estimer numériquement $\mathbb{P}(0 < X_n \leq c^n)$;
2. quel scénario conduit à de telles croissances ?

• Extension à une loi de reproduction aléatoire

Maintenant on prend en compte le fait que la reproduction de la population dépend de son environnement, qui change aléatoirement à chaque génération. A la génération n , la loi du nombre d'enfants est encore de type géométrique avec un paramètre m_n , qui est lui-même une variable aléatoire prenant deux valeurs m_\heartsuit (environnement favorable) et m_\dagger (défavorable). Les v.a. $(m_n)_n$ sont supposées être i.i.d..

Dans ce contexte, a priori les comportements en temps long de la population peuvent être différents du cas précédent et il s'agit de mieux comprendre quel rôle a l'environnement dans les fluctuations anormales. On pourra traiter les questions précédentes dans ce nouveau contexte.

● Références bibliographiques

Sur les bases des modèles d'évolution de population : [1], [7].

Sur les applications : [4], [6], [5].

Sur des aspects plus avancés de modélisation : [8], [3], [2].

Références

- [1] K.B. Athreya and P.E. Ney. *Branching processes*. Springer-Verlag, New York, 1972.
- [2] V. Bansaye and J. Berestycki. Large deviations for branching processes in random environment. *Markov Process. Related Fields*, 15(4) :493–524, 2009.
- [3] J. Geiger, G. Kersting, and V. Vatutin. Limit theorems for subcritical branching processes in random environment. *Ann. Inst. H. Poincaré Probab. Statist.*, 39(4) :593–620, 2003.
- [4] V.A.A. Jansen and J. Yoshimura. Populations can persist in an environment consisting of sink habitats only. *Proc. Natl. Acad. Sci. USA, Ecology*, 95 :3696–3698, 1998.
- [5] D.A. Keith, H.R. Akçakaya, W. Thuiller, G.F. Midgley, R.G. Pearson, S.J. Phillips, H.M. Regan, M.B. Araújo, and T.G. Rebelo. Predicting extinction risks under climate change : coupling stochastic population models with dynamic bioclimatic habitat models. *Biol. Lett.*, 4 :560–563, 2008.
- [6] R. Lande, S. Engen, and B.E. Saether. *Stochastic population dynamics in ecology and conservation*. Oxford University Press, Oxford, 2003.
- [7] S. Méléard. *Modèles aléatoires en Ecologie et Evolution*. Cours de l'Ecole Polytechnique, MAP563, 3A.
- [8] O. Nerman. On the maximal generation size of a non-critical Galton-Watson process. *Scandinavian Journal of Statistics*, 4(3) :131–135, 1977.

5 Sûreté aérienne

D'après le code de navigabilité international réglementant les avions civils à turbine, la probabilité d'une défaillance conduisant à une catastrophe doit être inférieure à 10^{-9} par heure de vol. Les défaillances peuvent être d'origine multiples et nous ne les analysons pas de manière exhaustive dans ce projet.

Nous nous concentrons sur la probabilité de collision entre deux avions, que nous souhaitons être aussi très faible pour atteindre un objectif de sécurité raisonnable en adéquation avec la réglementation. Le cas contraire pourra être le point de départ pour renforcer la fiabilité des instruments de pilotage, modifier les plans de vols... Améliorer la sûreté aérienne et minimiser le risque de collision peut être réalisé à plusieurs niveaux.

- A un niveau global, il s'agit de répartir au mieux les vols entre les différents aéroports et entre les différents moments de la journée/semaine.
- Au niveau le plus fin, lorsque la collision est imminente, cela relève du sang-froid des pilotes et de la réactivité des tours de contrôles.
- A un niveau intermédiaire et pendant le vol, il s'agit d'orienter les avions sur des routes qui restent assez distantes les unes des autres. Néanmoins, les trajectoires des avions ne sont pas déterministes car elles sont soumises à des aléas météorologiques (vent), à des perturbations de pilotage, à des erreurs de mesure, qui hélas rendent la collision possible.

Pour évaluer ce dernier risque de collision et appréhender le type de scénario y conduisant, nous modélisons le déplacement aléatoire des avions.

Unités d'usage.

- Distance : 1 mile nautique (noté 1 nmi) = 1.852 km.
- Vitesse : 1 knot (noté 1 kt) = 1 nmi/h = 1.852 km/h.
Ainsi, une vitesse de 500 kt est égale à 926 km/h (c'est à dire une vitesse standard d'avion de ligne).
On a aussi $1 \text{ nmi/min} = 60 \text{ kt} = 111.12 \text{ km/h}$.

Typiquement, le déplacement d'un avion s'effectue selon un plan de vol, qui est constitué de points de passage (*waypoint*) à intervalle régulier (20 min). Pour simplifier, on considère uniquement des trajectoires planes. Ainsi, entre deux points de passage, la trajectoire *idéale* (non perturbée) de l'avion est rectiligne, à vitesse constante v : si on note $x_t = (x_{a,t}, x_{c,t})$ la position³ de l'avion au cours du temps dans un repère local centré sur le précédent point de passage, avec l'axe des abscisses vers le prochain point de passage, on a

3. $x_{a,}$ désigne la coordonnée *along-track* et $x_{c,}$ la coordonnée *cross-track*.

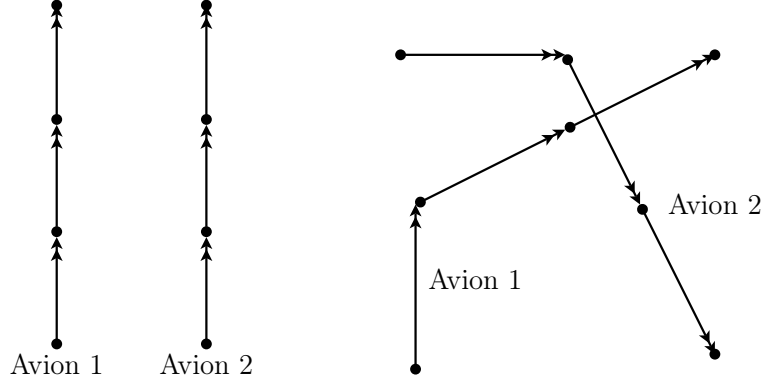


FIGURE 3 – Plan de vols parallèles et plan de vols se croisant. 3 points de passage (1 heure de vol).

$x_{a,t} = vt$ et $x_{c,t} = 0$. A partir de données de vol [5], considérons un modèle aléatoire $(x_t)_{t \geq 0}$ en posant

$$x_{a,t} = vt + M_{a,t}, \quad x_{c,t} = M_{c,t},$$

où $M_{a,\cdot}$ et $M_{c,\cdot}$ sont deux processus gaussiens continus, indépendants, de fonction de variance/covariance donnée par (pour $t < s$)

$$\begin{aligned} K_a(t, s) &= \text{Cov}(M_{a,t}, M_{a,s}) = r_a^2 t^2, \\ K_c(t, s) &= \text{Cov}(M_{c,t}, M_{c,s}) = \sigma_c^2 (1 - e^{-2\frac{r_c}{\sigma_c} vt}) e^{-\frac{r_c}{\sigma_c} v(s-t)}. \end{aligned}$$

Les paramètres $r_a = 0.25$ nmi/min, $r_c = 1/57$ et $\sigma_c = 1$ nmi sont communs à tous les avions civils à turbine. Ce qui peut changer d'un avion à un autre, c'est sa vitesse v et son plan de vol (la suite des points de passage).

Pour les besoins de simulations, les trajectoires ne seront pas considérées en temps continu, mais échantillonnées à intervalle régulier (par exemple 1 fois par minute) : c'est un paramètre de simulation à choisir convenablement. On s'attachera à détecter les *situations de conflit entre avions*, c'est-à-dire lorsque leur distance devient inférieure à 0.1 nmi, situation conduisant vraisemblablement à une collision imminente. La distance seuil pourra dépendre de la fréquence d'échantillonnage.

Quelques explications sur le modèle.

- *Composante along-track*. Lorsque le temps augmente, la coordonnée *along-track* est de moins en moins certaine (cumul des aléas de navigation) : la croissance quadratique de la variance peut s'interpréter comme

une incertitude sur la vitesse moyenne car $x_{a,t} \in [(v - 1.96r_a)t, (v + 1.96r_a)t]$ avec probabilité 95%.

- *Composante across-track*. Sur la coordonnée *cross-track*, le phénomène est différent et la variance de l'incertitude se stabilise quand t est grand. Cela s'interprète ainsi : le long de la composante *along-track*, le pilote et les instruments de bord cherchent à corriger les écarts à la trajectoire idéale (ici 0) induisant un phénomène de retour à 0. Ce modèle porte le nom de processus d'Ornstein-Uhlenbeck. Ce processus a un comportement ergodique, ce qui se voit en calculant la corrélation pour des temps lointains ($\lim_{s \rightarrow \infty} \text{Cor}(x_{c,t}, x_{c,s}) = 0$ avec vitesse exponentielle).

● Modèle à deux avions

On considère deux avions se déplaçant de manière indépendante, dans un même plan. Pour les tests, on considèrera différents plans de vol, comme par exemple celui de trajectoires parallèles, ou au contraire, celui des trajectoires qui se croisent (voir la figure 3). La distance entre deux points de passage correspond à un temps de vol prévu de 20 min. On supposera que chaque avion change de direction toutes les 20 minutes et se dirige vers son prochain point de passage.

▷ QUESTIONS À TRAITER :

1. Calcul numérique de la probabilité de conflit ; impact du plan de vol et des vitesses/distances entre avions ;
2. Distribution du temps de conflit, conditionnellement à l'occurrence du conflit ;
3. Distribution conditionnelle de la position de conflit.

On pourra dans un premier temps supposer aucune incertitude sur la composante *along-track* ($r_a = 0$), puis considérer le cas $r_a \neq 0$.

● Application à l'optimisation de plan de vols

On considère un plan de vol du type du second de la figure 3 : les avions se croisent pour rejoindre chacun un coin d'un carré. Cette fois, on cherchera à déterminer les trajectoires les plus courtes possibles (par exemple pour minimiser la consommation carburant, ou le temps de vol) sous contrainte de probabilité de collision faible. L'optimisation à réaliser concerne uniquement les 4 waypoints intermédiaires.

● Références bibliographiques

Sur les processus gaussiens : [3, Chap. 6 et 9].

Sur le calcul Monte Carlo des détections de conflit : [5].

Sur la modélisation des déplacements : [4], [1], [2], [5].

Le site web du projet européen HYBRIDGE : <http://hosted.nlr.nl/public/hosted-sites/hybridge/>

Références

- [1] W. Glover and J. Lygeros. A multi-aircraft model for conflict detection and resolution algorithm evaluation. *HYBRIDGE WP1 : Identification and modelling of uncertain hybrid systems*, 2004.
- [2] J. Hu, M. Prandini, and S. Sastry. Aircraft conflict prediction in the presence of a spatially correlated wind field. *IEEE Transactions on Intelligent Transportation Systems*, 6(3) :326–340, 2005.
- [3] E. Nelson. *Dynamical theories of Brownian motion*. Princeton University Press, Princeton, N.J., 1967.
- [4] R.A. Paielli and H. Erzberger. Conflict probability estimation for free flight. *Journal of Guidance, Control, and Dynamics*, 20(3) :588–596, 1997.
- [5] M. Prandini and O.J. Watkins. Probabilistic aircraft conflict detection. *HYBRIDGE WP3 : Reachability analysis for probabilistic hybrid systems*, 2005.

6 Ruines en assurance

• Contexte

L'assurance est un contrat de transfert de risques entre deux parties : l'assureur prend à son compte un risque (sinistres, dommages à la personne...) en échange d'une cotisation versée par l'assuré. Les sociétés d'assurance s'appuient sur un grand nombre d'assurés pour mutualiser les risques (effet loi des grands nombres) ; en 2013⁴, le chiffre d'affaires de l'assurance en France dépasse les 180 milliards d'euros. L'évaluation de la solvabilité d'une société d'assurance est une question cruciale et complexe ; en 2008, le groupe américain AIG a dû être renfloué par la Réserve Fédérale Américaine alors qu'il était au bord de la ruine, son capital ayant été largement déprécié suite à la crise des *sub-primes*. En Europe, la réforme réglementaire Solvabilité II⁵ impose de mieux mesurer les expositions aux risques, en évaluant par exemple les probabilités d'insolvabilité et de ruine. Afficher des excellents indicateurs de solvabilité sert aussi d'argument marketing auprès des clients.

Plus précisément, une compagnie d'assurance dispose d'un capital (fonds propres) ; elle perçoit régulièrement des cotisations de ses clients, et les indemnise en cas de sinistres. Cette compagnie propose en général des contrats de différents types liés aux différentes branches de son activité : assurance vie, complémentaire santé, assurance habitation, assurance automobile, assurance responsabilité civile, etc., avec des taux de cotisation qui dépendent de la *fréquence* des sinistres et de leur *sévérité*. Elle est ruinée lorsque les montants à rembourser sont supérieurs au capital de départ cumulé avec les cotisations perçues ; cela se produit en cas d'apparition de sinistres plus fréquents ou plus importants (catastrophes naturelles, épidémies...) ou si le capital initial et les cotisations sont insuffisants. Il est donc crucial de déterminer et/ou d'ajuster ces paramètres pour garantir la pérennité d'une compagnie d'assurance. Enfin, il est possible qu'une branche de l'assureur soit au bord de la ruine sans que l'assureur le soit globalement.



4. http://www.ffsa.fr/sites/jcms/p1_983486/1-assurance-francaise-demonstre-sa-solidite?cc=fn_7345

5. https://fr.wikipedia.org/wiki/Solvabilité_II

• Modélisation

Pour commencer, on considère une seule branche d'activité. La compagnie d'assurance a un capital de départ x , et on note R_t sa réserve à l'instant t . La compagnie d'assurance :

- perçoit des cotisations de ses clients, pour simplifier en continu et uniformément réparties sur l'année. On note p le taux de cotisation par unité de temps. Les recettes de la compagnie pendant un temps t sont donc pt .
- verse des indemnités à ses assurés sinistrés en fonction des dommages subis. On note $0 = T_0 < T_1 < T_2 < \dots$ les instants aléatoires des sinistres et $(Z_k)_{k \geq 1}$ les dommages. En posant $N_t \stackrel{\text{def}}{=} \sum_{k \geq 1} \mathbf{1}_{T_k \leq t}$ le nombre de sinistres entre l'instant 0 et t , la réserve à la date t est

$$R_t = x + pt - \sum_{k=1}^{N_t} Z_k.$$

▷ **Modèle, paramètres.** On modélise les montants remboursés $(Z_k)_{k \geq 1}$ par des lois Gamma i.i.d. de paramètres (α, β) , $\alpha > 0$ et $\beta > 0$. Les durées entre deux sinistres successifs $(\tau_k)_{k \geq 1}$ définies par $\tau_k \stackrel{\text{def}}{=} T_k - T_{k-1}$, sont supposées indépendantes entre elles et indépendantes des sinistres ; on suppose qu'elles sont distribuées suivant des variables exponentielles de paramètre λ , $\lambda > 0$.

▷ **Ruine.** La société est ruinée à l'horizon T si la réserve R_T est négative.

• Risque de ruine à l'horizon 1 an

L'objectif de cette partie est de calculer la probabilité de ruine $\mathbb{P}(R_T \leq 0)$ pour différentes valeurs des paramètres du modèle $(x, \alpha, \beta, \lambda, p)$ et pour $T = 1$ an.

1. Estimer la probabilité de ruine au bout d'un an pour les deux jeux de paramètres suivants $\{p = 14, \alpha = 2.5, \lambda = 2, \beta = 0.5, x = 100\}$ et $\{p = 14, \alpha = 3, \lambda = 1/24, \beta = 0.04, x = 250\}$. Proposer différentes méthodes de calcul.
2. Déterminer le capital initial que la compagnie doit avoir pour que sa probabilité de ruine à l'horizon annuel soit inférieure à 10^{-6} .
3. Estimer la distribution de R_T sachant qu'il y a eu ruine à l'instant T .

• Extensions

On traitera l'une des extensions suivantes.

▷ **Risque de ruine à horizon infini.** On veut quantifier si la compagnie d'assurance sera un jour ruinée. Si la société d'assurance est effectivement ruinée, la ruine interviendra à l'instant d'arrivée d'un sinistre : en conséquence, on s'intéresse au processus à temps discret de la somme déboursée échantillonnée aux instants d'arrivée des sinistres. La somme déboursée à l'instant d'arrivée du n -ième sinistre est égale à $X_1 + \dots + X_n$, où $X_k \stackrel{\text{def}}{=} Z_k - p\tau_k$. La probabilité de ruine à horizon infini est donc donnée par

$$\mathbb{P}(\sigma_x < \infty), \quad \text{où} \quad \sigma_x \stackrel{\text{def}}{=} \inf\{n \geq 1, \sum_{k=1}^n X_k > x\},$$

avec la convention $\inf \emptyset = +\infty$.

La difficulté dans la simulation en temps infini provient du fait qu'il faut simuler toute la trajectoire $(\sum_{k=1}^n X_k, n = 0, \dots, +\infty)$ pour déterminer si la réalisation est telle que $\sigma_x < +\infty$ ou $\sigma_x = +\infty$. Pour pallier à cette difficulté, on propose d'effectuer un changement de probabilité exponentiel décrit dans [3] et [2] pour lequel sous une nouvelle probabilité \mathbb{P}_θ , $\mathbb{P}_\theta(\sigma_x < \infty) = 1$.

Quelques résultats théoriques. On note ψ_X la fonction génératrice des cumulants de X . Le livre [1] et l'article [2] rappellent des nombreuses propriétés de la fonction génératrice des cumulants.

Si $\mathbb{E}[X] < 0$, il existe une unique valeur positive θ^* telle que $\psi_X(\theta^*) = 0$: on a alors

$$\mathbb{P}(\sigma_x < +\infty) = \mathbb{E}_{\theta^*} \left[e^{-\theta^* \sum_{k=1}^{\sigma_x} X_k} \right],$$

où \mathbb{E}_θ désigne l'espérance associée à la probabilité \mathbb{P}_θ . Voir [3].

Questions. On cherchera à

1. estimer la probabilité de ruine pour $\alpha = 1$, $\beta = 1$, $\lambda = 0.5$, $p = 1$ et différentes valeurs de x avec la méthode présentée ci-dessus.
2. démontrer certains des résultats énoncés dans cette section.

▷ **Compagnie diversifiée.** Dans cette section, on suppose que la compagnie étudiée propose des assurances habitations et des assurances automobiles, et qu'une partie des sinistres liés aux risques naturels touchent à la fois le secteur automobile et habitation. On ajoute des sinistres propres aux habitations N_t^H et des sinistres propres aux voitures N_t^A : le nombre de sinistres affectant la branche habitation jusqu'au temps t est donc $N_t^{RN} + N_t^H$, et le nombre des sinistres affectant la branche automobile est $N_t^{RN} + N_t^A$.

1. Estimer le risque de ruine au bout d'un an (on pourra notamment étudier la ruine de chaque branche et les effets de compensation entre branche).

2. Déterminer le capital initial dont doit disposer la compagnie pour que sa probabilité de ruine à horizon 1 an soit inférieure à 10^{-6} . Comment répartir un capital donné entre ses deux branches ?
3. Représenter la distribution de la réserve sachant qu'il y a eu ruine.

● Références bibliographiques

Théorie de la ruine : [1].

Changement de probabilité exponentiel : [3][2].

Références

- [1] S. Asmussen, and H. Albrecher *Ruin probabilities*, volume 14 of *Advanced Series on Statistical Science and Applied Probability*. World Scientific, Singapore, second edition, 2010.
- [2] J.H. Blanchet, K. Leder, and P.W. Glynn. Efficient simulation of light-tailed sums : an old-folk song sung to a faster new tune. . . . In *Monte-Carlo and quasi-Monte-Carlo methods 2008*, pages 227–248. Springer, Berlin, 2009.
- [3] P. Glasserman. *Monte-Carlo methods in Financial Engineering*. Springer Verlag, New York, 2003.

7 Barrages et risques d'inondation

Le risque d'inondation est classé premier risque naturel en France par le Ministère de l'Ecologie et du Développement durable⁶. Il concerne 13.300 communes sur le territoire français, dont 300 grandes agglomérations, pour un total de 5 à 6 millions de personnes concernées, directement ou indirectement.⁷ Dans le mécanisme de contrôle des risques d'inondation, un rôle est joué par les bassins artificiels, qui ont une fonction de régulation des crues (en plus de la production d'énergie hydroélectrique). En même temps, lorsqu'un bassin artificiel est alimenté par un volume d'eau trop important, pour éviter que le barrage qui le retient ne cède, on est obligé d'ouvrir les vannes d'évacuation de façon importante, ce qui entraîne un fort débit d'eau dans la vallée sous-jacente et donc un risque d'inondation est toujours présent. Il existe en France plusieurs lacs artificiels de grande taille, le lac de Serre-Ponçon en tête, avec ses $1.2 \cdot 10^9 m^3$ d'eau retenue.



7.1 Modélisation



Vidange du barrage, ouverture des vannes.

Dans une région montagneuse, deux vallées sont occupées par deux lacs artificiels créés par deux barrages B^1 et B^2 , dont on modélise leur volume contenu respectif $\{X_t^i, t \geq 0\}$, ($i = 1, 2$) au cours du temps $t \geq 0$. Chaque lac est alimenté par une rivière principale et plusieurs torrents, qui apportent un volume cumulé A_t^i ($i = 1, 2$) au temps t . En même temps, chaque bassin se vide à un taux instantané $r_i(X_t^i)$ qui est une fonction du volume d'eau X_t^i stocké dans le bassin (en cas de crue, le lac de Serre-Ponçon peut évacuer jusqu'à $1.24 \cdot 10^5 m^3$ par heure). L'équation qui gouverne l'évolution temporelle du contenu de chaque réservoir (unité= $10^5 m^3$) s'écrit donc

$$X_t^i = x_0^i + A_t^i - \int_0^t r_i(X_s^i) ds, \quad (1)$$

6. http://catalogue.prim.net/43__dppr-livretrisqmajeurs-v7.pdf

7. voir la conférence http://www.coriolis.polytechnique.fr/Confs/Leleu_poster.pdf

où x_0^i est le contenu à l'instant initial. On fait l'hypothèse de débit proportionnel en le volume du barrage : $r_i(x) = r_i x$.

▷ **Volumes surcritiques.** Chaque barrage $i = 1, 2$ est construit pour supporter un volume contenu maximal x_{crit}^i (la géométrie de chaque bassin artificiel étant fixée, le niveau critique pourrait aussi bien être exprimé par la hauteur maximale de l'eau à contact du barrage). Au-delà de ce niveau critique, on est obligé d'ouvrir les évacuateurs du barrage à tel point que cela entraîne un risque d'inondation pour la vallée sous-jacente. Pour réduire le plus possible ce risque, ce niveau critique x_{crit}^i est défini comme le quantile $Q_{X_T^i}(\alpha)$ d'ordre $\alpha = 10^{-6}$ à un horizon T .

On souhaite estimer à l'aide de simulations la valeur de ce niveau critique, et évaluer ensuite s'il est envisageable, du point de vue financier et technique, de construire un barrage qui puisse tolérer un tel niveau de remplissage. On cherche aussi à quantifier les risques d'inondation dus à un des deux barrages ou aux deux.

▷ **Modélisation de l'apport d'eau.** On suppose que l'apport d'eau $\{A_t^i, t \geq 0\}$ est essentiellement déterminé par des chutes de pluie intenses et de courte durée (on néglige les ruissellements continus). Les modèles que l'on peut utiliser pour les précipitations atmosphériques sont typiquement fondés sur l'usage de processus stochastiques à sauts : on décrit l'arrivée d'un épisode de pluie par un processus de Poisson (le début de l'événement correspondant à un instant de saut du processus), la durée de l'épisode par une variable aléatoire positive, et on décrit aussi une structure fine de l'événement entre son début et sa fin (voir [2]). On pourra retenir ici cette modélisation aléatoire de l'arrivée des pluies : on fait donc l'hypothèse que, pour chaque bassin, le débit entrant est modélisé par un processus de Poisson composé

$$A_t^i = \sum_{n=1}^{N_t^i} U_n^i, \quad i = 1, 2 \quad (2)$$

où pour tout $i = 1, 2$, $N^i = \{N_t^i, t \geq 0\}$ est un processus de Poisson d'intensité $\lambda_i > 0$, et $\{U_n^i, n \geq 1\}$ sont des variables aléatoires indépendantes et identiquement distribuées, et indépendantes de N^i . Pour la loi de U_1^1 et U_1^2 , une combinaison de lois exponentielles permettra de modéliser séparément des pluies de grande et petite intensité :

$$\nu(u) = b\delta_1 e^{-\delta_1 u} \mathbf{1}_{\{u > 0\}} + (1 - b)\delta_2 e^{-\delta_2 u} \mathbf{1}_{\{u > 0\}}, \quad 0 < b < 1.$$

Pour modéliser une différence importante entre les chutes de pluie intenses et les faibles, on pourra prendre δ_2/δ_1 de l'ordre de 10 et $\delta_2 = 0.7$; on pourra

choisir λ de façon à exprimer le déclenchement de plusieurs dizaines à une centaine de pluies par an en moyenne ; enfin, le taux de vidange annuel r_i sera de l'ordre de l'unité.

7.2 Objectifs.

▷ **Etude d'un seul barrage.** On étudie tout d'abord les barrages séparément. Les questions à aborder sont les suivantes.

- Q1. (a) Estimer le niveau critique x_{crit}^i pour chaque bassin $i = 1, 2$, à un horizon $T = 1$ an. Déterminer la distribution du volume d'eau dans chaque bassin lorsqu'il y a un dépassement du niveau critique en T .
- (b) Pour comprendre si l'évaluation des risques que l'on vient d'effectuer à un horizon fixé est pertinente, on reconsidère le critère de définition du niveau de remplissage critique, en se basant maintenant sur le volume maximum $\max_{t_j} X_{t_j}$ atteint par le bassin tout au long de l'année. Estimer par simulation le nouveau volume critique, ainsi que la distribution de dépassement de ce niveau au cours de l'année. Comparer avec les résultats précédents.

▷ **Prise en compte des deux barrages, inondation en aval.** Les chaînes de montagne en amont des deux bassins étant suffisamment distantes, on fait l'hypothèse que les précipitations qui alimentent les deux lacs sont indépendantes : les processus de Poisson composés définis par $\{N^i, U_n^i, n \geq 0\}$ pour $i = 1, 2$, sont indépendants.

Les deux vallées occupées par les bassins, raides et inhabitées, se rejoignent à l'embouchure d'une vallée plus large qui est occupée par des villages. Au point de jonction O entre les deux vallées, une rivière est formée par l'eau venant des deux bassins, qui atteint le point O après avoir parcouru la première vallée en un temps T_1 , et la deuxième en un temps T_2 . On supposera que ces retards d'arrivée d'eau en O sont fixes, et indépendants du volume d'eau qui se déplace. Au point de jonction des vallées, le débit observé au temps t est donc

$$D_t = r_1 X_{t-T_1}^1 + r_2 X_{t-T_2}^2.$$

La rivière qui coule dans la vallée et qui est alimentée par les deux barrages, peut tolérer un débit limite, au-delà duquel la rivière déborde et entraîne des inondations des zones habitables. On traitera les questions suivantes.

- Q2. (a) Quelle est la probabilité de dépasser le débit de seuil dans la vallée, soit $\mathbb{P}(D_T \geq \text{seuil})$?

- (b) Quelle est la probabilité de dépassement du débit de seuil lorsque l'un des deux bassins est rempli au-dessus du niveau critique ?
- (c) A partir de l'observation du débit instantané au point de jonction O , on souhaite remonter à de l'information sur l'état de remplissage des bassins. Estimer le volume d'eau dans chaque bassin lorsqu'on observe un niveau limite de débit à la jonction.

7.3 Extensions

On cherchera à aborder l'un des points suivants.

- **Dépendance** : On souhaite relaxer l'hypothèse d'indépendance des précipitations en amont des deux bassins (i.e. l'indépendance des deux processus de comptage des pluies). Proposer une modélisation qui tienne compte d'une dépendance entre les phénomènes atmosphériques. On reprendra l'analyse précédente dans ce cadre.
- **D'autres sources d'aléa** : La vallée habitée est alimentée en eau pas seulement par les deux bassins, mais aussi par l'arrivée de pluies. Prendre en compte ce phénomène dans la modélisation et l'analyse du risque d'inondation.
- **Une analyse en temps long** : Proposer une analyse du système (1)-(2) en temps long ($t \rightarrow \infty$), en étudiant notamment la possibilité d'obtenir des formules explicites pour la loi de X_t^i dans ce régime, et donc pour ses quantiles. Comparer avec les simulations de la question 7.

Références

- [1] S. Asmussen. *Applied probability and queues*, volume 51 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 2003. Stochastic Modelling and Applied Probability.
- [2] P. Cowpertwait, V. Isham, and C. Onof. Point process models of rainfall : developments for fine-scale structure. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 463(2086) :2569–2587, 2007.
- [3] P. Glynn. Importance sampling for Monte Carlo estimation of quantiles. Proceedings of the Second International Workshop on Mathematical Methods in Stochastic Simulation and Experimental Design, available at <http://www.stanford.edu/~glynn/papers/1996/G96.pdf>, 1996.

- [4] S. Sambou. Modèle statistique des hauteurs de pluies journalières en zone sahélienne : exemple du bassin amont du fleuve sénégal. *Hydrological Sciences – Journal des Sciences Hydrologiques*, 49(1) :115–129, 2004.

8 Microstructure des prix financiers

Les propriétés des séries temporelles du prix d'un actif coté sur un marché boursier dépendent fortement de la fréquence d'observation. Si pour des observations à basse fréquences (échelle de temps : jour, mois) les trajectoires sont bien décrites par des processus de diffusion⁸, Figure 4 (a), pour des fréquences plus élevées (échelle de temps : heure, minute, ou en dessous), le prix manifeste sa nature discrète et est représenté par des trajectoires discontinues, constantes par morceaux et qui varient sur une grille, Figure 4 (b). L'amplitude minimale de variation du prix à chaque sauts (0.01 dans la Figure 4) est appelée le *tick*. Cette nature discrète est intrinsèquement liée à l'existence d'un carnet d'ordre, Figure 5. A chaque instant, un certain nombre d'unités de l'actif (le *volume*) est disponible sur le marché à des prix de vente (ask) ou d'achat (bid) fixés : lorsque l'exécution d'une ordre consomme tout le volume disponible au meilleur ask (ou meilleur bid), le meilleur prix à disposition se déplace sur la valeur à coté.

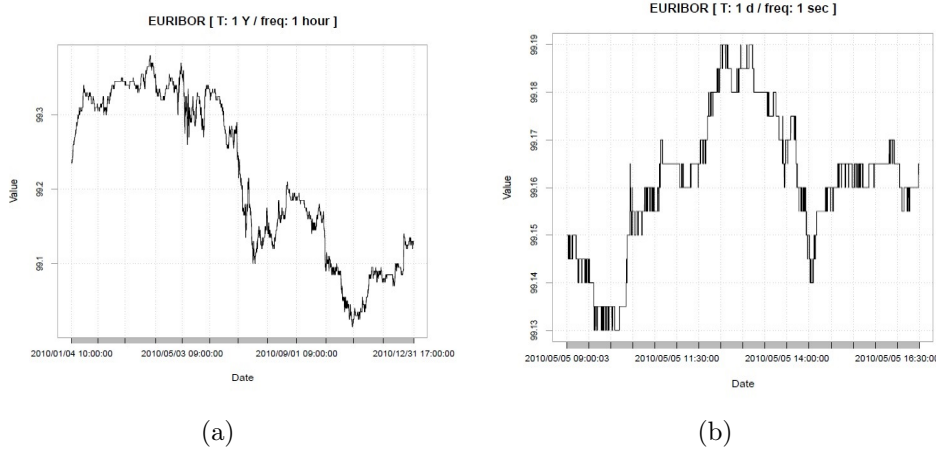


FIGURE 4 – Trajectoires du prix de l'Euribor future pour différentes fréquences d'observation. Pham (2013), présentation au NUS-UTokyo Workshop on Quantitative Finance.

Suivant une approche de modélisation typique, on décrit ce prix qui varie sur grille discrète par le biais d'une processus ponctuel $(T_n, J_n)_{n \geq 1}$. La suite croissante $(T_n)_n$ représente les instants de variations (sauts d'un ou plusieurs ticks) du prix. Les v.a. J_n sont à valeurs dans l'ensemble fini

$$E = \{+1, -1, \dots, +m, -m\} \subset \mathbb{Z} \setminus \{0\}.$$

8. Par exemple, un mouvement Brownien : mouvement continu d'une particule soumise à des chocs aléatoires gaussiens.

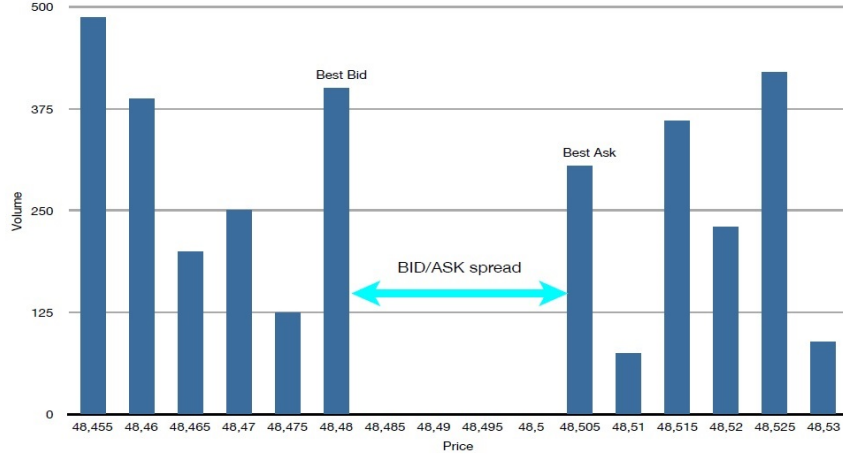


FIGURE 5 – Instantané d'un carnet d'ordre limite (*limit order book*).

et représentent les incréments (positifs ou négatifs) du prix. Le prix de l'actif est donné par le processus

$$P_t = P_0 + \sum_{k=1}^{N_t} J_n \quad (3)$$

où N_t est le nombre d'événements avant t ,

$$N_t = \sum_{n \geq 1} 1_{T_n \leq t}$$

La valeur courante P_t représente la dernière cotation du prix *mid*, soit la moyenne entre le meilleur ask et le meilleur bid.

Une modélisation simplifiée. Dans le cadre d'une première modélisation simplifiée, on suppose que la suite des incréments $(J_n)_n$ est i.i.d. (absence de dépendance entre les incréments du prix). On suppose de plus que les instants de sauts $(T_n)_n$ sont indépendants des $(J_n)_n$, et que la suite des intervalles entre sauts $S_n = T_{n+1} - T_n$ est i.i.d. de loi exponentielle de paramètre λ .

Q1. On souhaite estimer la probabilité que dans le modèle ainsi défini les trajectoires du prix prennent des valeurs négatives,

$$\mathbb{P} \left(\inf_{t \leq T} P_t < 0 \right)$$

pour $P_0 = 35$ et $T = 4h$. On prendra une valeur de λ telle que l'attente moyenne entre un saut et le suivant soit autour de 300s. On

considérera les cas $m = 1$ et $m = 3$ avec loi des sauts symétrique : $\mathbb{P}(J_n = 1) = \mathbb{P}(J_n = -1) = 1/2$ pour $m = 1$ et $\mathbb{P}(J_n = i) = \frac{1}{2}p_{|i|}$ pour $i \in E$ et $m = 3$, où $p_1 = 1/2$, $p_2 = 1/3$, $p_3 = 1/6$.

Pour ces mêmes valeurs des paramètres, on veut estimer les quantiles à 10^{-4} , à gauche et à droite, pour la distribution de P_T – ensuite les quantiles plus extrêmes, à 10^{-5} ou 10^{-6} .

Prise en compte d'autres propriétés des données haute fréquence.

On souhaite qu'un modèle de prix à haute fréquence soit capable de reproduire, en plus de la nature discrète des prix, les faits stylisés suivants :

- *Retour vers la moyenne* : les variations du prix sur des courts intervalles de temps sont fortement anti-corrélées. Cela détermine un effet appelé retour vers la moyenne : le prix a tendance à osciller autour d'une valeur intermédiaire.
- *Clusters de volatilité* : on observe une alternance de périodes avec activité de trading intense (volatilité du prix plus élevée) et de périodes plus calmes (volatilité plus faible).

► Afin de rajouter une plus forte composante de retour vers la moyenne, on peut modéliser le prix mid par une superposition de processus

$$P_t = P_t^{(1)} + P_t^{(2)},$$

où $P_t^{(i)} = P_0/2 + \sum_{n=1}^{N_t^{(i)}} J_n^{(i)}$ pour $i = 1, 2$, toutes les variables aléatoires indexées par $i = 1$ et $i = 2$ étant indépendantes. On considérera :

(1) les mêmes hypothèses qu'avant sur les distributions de $(N_t^{(1)})_{t \geq 0}$ et $(J_n^{(1)})_{n \geq 1}$, en changeant la durée moyenne $1/\lambda_1$ entre les sauts de 300s à 660s.

(2) une suite de sauts $J_{n+1}^{(2)} = -J_n^{(2)}$, à partir de $J_1^{(2)}$ avec $\mathbb{P}(J_1^{(2)} = 1) = \mathbb{P}(J_1^{(2)} = -1) = 1/2$. Ce choix modélise la composante du prix soumise à oscillation. On prendra pour $N_t^{(2)}$ un processus de Poisson simple, de durée moyenne entre sauts $1/\lambda_2 = 110s$.

Q2. Reprendre la question précédente, afin d'estimer l'impact d'une telle modélisation.

► Une autre approche à la modélisation du phénomène de retour vers la moyenne consiste à s'affranchir de l'hypothèse d'indépendance entre les sauts. On décompose à ce fin la variation du prix à chaque saut en

$$J_n = \hat{J}_n \xi_n$$

où $\hat{J}_n = \text{sign}(J_n)$ et $\xi_n = |J_n|$. On supposera maintenant que

- Les variations du prix ne sont pas indépendantes entre elles, mais on suppose que ce n'est que le signe d'un incrément de prix (et non sa valeur absolue) qui influence les incréments futures ; de plus, le signe de l'incrément à l'instant T_n ne dépend que de celui en T_{n-1} , et pas des précédents. Mathématiquement, cela se traduit en prenant pour \hat{J}_n une chaîne de Markov, de matrice de transition

$$\hat{Q} = \begin{pmatrix} \frac{1+\alpha_+}{2} & \frac{1-\alpha_+}{2} \\ \frac{1-\alpha_-}{2} & \frac{1+\alpha_-}{2} \end{pmatrix}$$

où $\alpha_+, \alpha_- \in [-1, 1)$.

- Les $(\xi_n)_n$ sont i.i.d. à valeurs dans $\{1, \dots, m\}$, indépendants des $(\hat{J}_n)_n$, de loi $\mathbb{P}(\xi_n = i) = p_i$, $i = 1, \dots, m$. On modélise de cette manière des amplitudes de saut qui sont indépendantes entre elles, et indépendantes de la direction du saut (à la hausse ou à la baisse). Cette propriété traduit une indépendance des volumes échangés dans le carnet d'ordre à chaque événement de trading, et leur indépendance par rapport au type de transaction (vente ou achat).

Q. Quel signe faudra-t-il choisir pour les paramètres α_{\pm} afin de modéliser le phénomène de retour vers la moyenne ?

Pour les $(J_n)_n$ ainsi définis, on modélise le prix P_t toujours par l'équation (3), sous une hypothèse d'indépendance entre les suites $(J_n)_n$ et $(T_n)_n$.

Q3. Des estimations à partir des données sur l'Euribor future [1] ont montré qu'il est réaliste de considérer le cas symétrique $\alpha_+ = \alpha_- = \alpha$. On fixera pour α la valeur estimée dans l'article [1], Section 2.

Simuler des trajectoires du prix P_t , observer empiriquement sur les simulations que l'on a $\mathbb{P}(\hat{J}_n \hat{J}_{n+1} = 1) \approx (1 + \alpha)/2$. Estimer dans ce cas également la probabilité que le prix P_t prenne des valeurs négatives au cours de $N = 100$ variations, ainsi que les quantiles à 10^{-4} et 10^{-5} de la valeur finale du prix.

Limite macroscopique. On souhaite étudier le comportement des prix engendré par le modèle considéré dans la question précédente sur des échelles de temps plus longues.

En supposant une durée de $8h$ pour chaque jour et 250 jours de trading par an, horizon $T = 1$ an, estimer la probabilité que le prix P_t prenne des valeurs négatives, ainsi que les quantiles à 10^{-6} de la valeur finale du prix, à partir de la valeur $P_0 = 1000$ (typique pour un indice boursier) et à un horizon $T = 1$ an.

Références

- [1] P. Fodra and H. Pham. *Semi-markov model for market microstructure*. *Applied Mathematical Finance, Preprint arXiv* <http://arxiv.org/abs/1305.0105>, 22(3) :261–295, 2015.