

# Rapport MODAL MAP473D

Axel Benyamine, Julien Lambert

20 Mai 2023

## 1 Présentation du problème

Depuis une vingtaine d'années, le secteur bancaire adopte une approche globale de la gestion des risques, notamment le risque de crédit. Nous allons donc nous intéresser ici à l'évaluation du risque de crédit de plusieurs entreprises.

### 1.1 Modélisation

Nous allons modéliser la situation par un ensemble de  $N$  entreprises. Chacune de celles-ci aura une valeur  $S_t^i$  au cours du temps, modélisée par un brownien géométrique, c'est-à-dire

$$S_t^i = S_0^i e^{-\frac{1}{2}\sigma_i^2 t + \sigma_i W_t^i} \quad (1)$$

où  $\sigma_i$  représente la volatilité des rendements, et  $(W_t^i)_{t \geq 0}$  est un mouvement brownien.

On considèrera que l'entreprise  $i$  est en faillite lorsque sa valeur est inférieure à une valeur  $B_i$ , avec  $B_i \ll S_0^i$  de telle sorte que la probabilité de défaut soit un évènement rare.

Nous nous intéresserons à deux quantités : d'une part la probabilité qu'il y ait au moins  $k$  défauts, et d'autre part, les pertes engendrées par ces défauts. Nous modélisons de 3 manières ces quantités, par complexité croissante.

Dans la suite, on considèrera (sauf mention contraire) que  $N = 125$ ,  $B_i = B (= 36)$ ,  $S_0^i = 100$  pour tout  $1 \leq i \leq N$ , et on considèrera les 2 cas suivants :  $\sigma_i = 40\%$  et  $\sigma_i = 20\%$  pour  $1 \leq i \leq 25$ ,  $\sigma_i = 25\%$  pour  $26 \leq i \leq 50$ ,  $\sigma_i = 30\%$  pour  $51 \leq i \leq 75$ ,  $\sigma_i = 35\%$  pour  $76 \leq i \leq 100$ ,  $\sigma_i = 50\%$  pour  $101 \leq i \leq 125$ .

#### 1.1.1 Evolution indépendante

Nous allons d'abord nous intéresser au cas où l'on regarde la valeur des entreprise uniquement à la date  $T$  (1 an ici). On supposera de plus que les évolutions des mouvements browniens sont différents les uns des autres.

On peut alors introduire  $D_T$  les entreprises qui ont fait défaut, défini par

$$D_T = \{1 \leq i \leq N, S_T^i \leq B_i\} \quad (2)$$

Nous nous intéresserons donc à la loi de  $L = \#D_T$  le nombre de faillites à l'instant  $T$ . En particulier, nous chercherons à calculer les  $\mathbb{P}(L \geq k)$ .

On s'intéressera également à la perte engendrée par chacune des entreprises. Pour cela, on considèrera :

$$P_T = \sum_{i \in D_T} R_i S_T^i \quad (3)$$

où les  $R_i$  sont des variables aléatoires i.i.d. On cherchera à calculer les  $\mathbb{E}[P_T | L \geq k]$  avec  $R_i = 30\%$  et  $R_i$  suit une loi Beta de moyenne 30% et d'écart-type 15%.

Dans ce cas-là, on remarque que  $L = B_1 + \dots + B_N$  où  $B_i = \mathbb{1}_{\{S_T^i \leq B\}}$

En particulier, dans le cas où tous les  $\sigma_i$  sont égaux, on remarque que les  $(B_i)_i$  sont des variables de Bernoulli indépendantes de même paramètre  $p$ . Donc,  $L \sim \text{Bin}(N, p)$ . Le paramètre est déterminé par la probabilité qu'une seule entreprise fasse défaut, et est donc donné par

$$p = \mathbb{P}(S_T^i \leq B_i) = \mathbb{P}\left(W_T^i \leq \frac{\ln(B/S_0^i) + 0,5\sigma_i^2 t}{\sigma_i}\right) \quad (4)$$

avec  $W_T^i \sim \mathcal{N}(0, T)$

### 1.1.2 Evolution dépendante

Ensuite, nous nous intéresserons au cas où les valeurs des entreprises sont positivement corrélées. Ainsi, en suivant toujours l'équation (1), on suppose désormais que le vecteur  $(W_t^1, \dots, W_t^1)$  est un vecteur gaussien de moyenne nulle et de matrice de covariance

$$\Sigma_t = \begin{bmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \rho & \rho & 1 & \dots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \dots & 1 \end{bmatrix} t$$

avec  $\rho \in ]0, 1[$ .

On conserve dans ce cas les mêmes définitions ainsi que les mêmes quantités à estimer.

### 1.1.3 Extensions

Enfin, nous nous intéresserons au cas, plus réaliste, où on regarde la valeur des entreprises à différents intervalles de temps. Pour cela, on choisit plusieurs dates  $t_k$  croissantes (par exemple,  $t_k = \frac{k}{12}$  pour une fréquence par mois), et l'on regarde si l'entreprise fait faillite à une de ces dates  $t_k$ .

Ainsi, on pose :

$$\begin{aligned}\tau_i &= \inf\{t_k \text{ tq } S_{t_k}^i \leq B_i\}, \\ D'_T &= \{1 \leq i \leq N \mid \tau_i \leq T\}, \\ L'_T &= \#D'_T, \\ P'_T &= \sum_{i \in D'_T} R_i S_{\tau_i}^i.\end{aligned}$$

Conformément aux hypothèses du mouvement brownien, en posant  $W_t = (W_t^1, \dots, W_t^N)$ ,  $W_{t_{k+1}} - W_{t_k}$  est un vecteur gaussien de moyenne nulle et de matrice de covariance  $\Sigma_{t_{k+1}-t_k}$  et les  $W_{t_{k+1}} - W_{t_k}$  sont mutuellement indépendants.

Pour la suite, on choisira  $t_k = \frac{kT}{M}$  où  $M$  désigne le nombre de fois où l'on vérifie la valeur des entreprises. Ainsi, si on pose le vecteur ligne  $\Delta W = (W_{t_1}, W_{t_2} - W_{t_1}, \dots, W_{t_M} - W_{t_{M-1}})$  de taille  $MN$ , il s'agit d'un vecteur gaussien de moyenne nulle, et de matrice de covariance, la matrice par blocs (de taille  $MN \times MN$ ) :

$$\begin{bmatrix} \Sigma_{\frac{1}{M}} & 0 & \cdots & 0 \\ 0 & \Sigma_{\frac{1}{M}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \Sigma_{\frac{1}{M}} \end{bmatrix}$$

Ainsi,  $L$  peut être vu comme une fonction d'un vecteur gaussien centré.

## 2 Cas de l'évolution indépendante

Comme nous l'avons montré précédemment, le cas où les évolutions sont indépendantes, et ont la même volatilité, permet de faire des calculs facilitant l'estimation de notre probabilité précédente. Ainsi, à partir de  $p$ , on obtient

$$\mathbb{P}(L \geq k) = \sum_{j=k}^N \binom{N}{j} p^j (1-p)^{N-j} = f_k(p)$$

Ainsi, si on a un estimateur  $\hat{p}_n$  de  $p$  vérifiant  $\sqrt{n}(\hat{p}_n - p) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$ , alors par la méthode Delta, on obtient un estimateur de  $\mathbb{P}(L \geq k)$ , vérifiant  $\sqrt{n}(f_k(\hat{p}_n) - f_k(p)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2 f'_k(p)^2)$  Ici,

$$f'_k(p) = \sum_{j=k}^N \binom{N}{j} p^{j-1} (1-p)^{N-j-1} (j - Np)$$

Dans ce cas précis, nous avons pu donc estimer  $p$  par une méthode de Monte-Carlo par échantillonnage d'importance<sup>1</sup>, et ensuite propagé notre estimation des différentes probabilités par la méthode précédente. On obtient

<sup>1</sup>Section 4

alors l'intervalle de confiance à 95% suivant pour  $\mathbb{P}(L \geq k)$ :

$$I_n = \left[ f_k(\hat{p}_n) - \frac{1.96\sigma |f'_k(\hat{p}_n)|}{\sqrt{n}}, f_k(\hat{p}_n) + \frac{1.96\sigma |f'_k(\hat{p}_n)|}{\sqrt{n}} \right]$$

On peut remarquer ainsi que la qualité de cet estimateur par rapport aux autres dépend avant tout de la qualité de l'estimateur de  $p$ .

On constate aussi qu'étant donné la grande précision que l'on a obtenue avec la méthode de Monte-Carlo par échantillonnage d'importance, cette méthode est très précise pour calculer les différentes probabilités, même avec un grand nombre d'entreprises. Cependant, cette méthode est inutilisable, dès lors que les  $\sigma_i$  sont différents, ou que l'on introduit une corrélation entre les valeurs des différentes entreprises.

### 3 Monte-Carlo naïf

L'estimateur le plus naturel pour estimer  $\mathbb{E}[g(X)]$  est  $\hat{g}_n = \frac{1}{n} \sum_{k=1}^n g(X_k)$  où  $(X_k)_k$  sont i.i.d. suivant la loi de  $X$ , appelé estimateur de Monte-Carlo. Cet estimateur vérifie (sous condition de l'existence d'un moment d'ordre 2 de  $g(X)$ ) le TCL suivant :  $\sqrt{n}(\hat{g}_n - \mathbb{E}[g(X)]) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$  où  $\sigma^2 = \mathbb{V}(g(X))$ . Dans le cas de calcul d'une probabilité, on obtient le TCL suivant :  $\sqrt{n}(\hat{p}_n - p) \xrightarrow{\mathcal{L}} \mathcal{N}(0, p(1-p))$ . On peut ainsi en déduire un intervalle de confiance asymptotique à 95% :

$$I_n = \left[ \hat{p}_n - 1.96\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}, \hat{p}_n + 1.96\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}} \right]$$

L'erreur relative est donc

$$1.96\sqrt{\frac{1-\hat{p}_n}{n\hat{p}_n}}$$

Plus  $p$  est faible, donc plus l'évènement est rare, plus cette méthode est imprécise. En particulier, il faut en moyenne  $1/p$  simulations indépendantes de la loi de  $X$  avant que l'évènement en question se réalise, ce qui fait que l'estimation par méthode de Monte-Carlo naïve a tendance à estimer une probabilité nulle si  $n$  n'est pas suffisamment grand.

Nous avons implémenté cette méthode pour le calcul de  $\mathbb{P}(L \geq k)$  dans les différentes modélisations. Dans nos simulations, on peut voir que la méthode de Monte-Carlo naïf fonctionne très bien dès lors que l'on regarde  $\mathbb{P}(L \geq k)$  pour  $k$  assez petit. Au-delà d'un certain seuil de l'ordre de  $k = 10$ , l'estimation de  $\mathbb{P}(L \geq k)$  est nulle, et donc Monte-Carlo ne fonctionne pas, pour les raisons cités précédemment.

Malgré cet inconvénient, l'estimateur de Monte-Carlo naïf reste un algorithme simple et sûr pour estimer une probabilité, lorsque celle-ci est suffisamment grande.

## 4 Monte-Carlo par échantillonnage d'importance

Pour pallier aux défauts de la méthode de Monte-Carlo naïve lors de la simulation d'événements rares, une méthode peut être de changer la loi sous laquelle on simule  $X$ .

### 4.1 Présentation de la méthode

Si  $X$  suit une loi de densité  $f$  par rapport à la mesure  $\nu$ , alors  $\mathbb{E}[h(X)] = \int_{\Omega} h(x)f(x)d\nu$

On peut alors réécrire

$$\mathbb{E}_f[h(X)] = \int_{\Omega} h(x)\frac{f(x)}{g(x)}g(x)d\nu = \mathbb{E}_g[h(X)\frac{f(X)}{g(X)}]$$

pour  $g$  une densité par rapport à  $\nu$  telle que  $g(x) = 0 \implies f(x) = 0$  pour tout  $x \in \Omega$

On obtient alors un nouvel estimateur de  $E_f[h(X)]$  donné par  $\hat{h}_n = \frac{1}{n} \sum_{k=1}^n h(X_k) \frac{f(X_k)}{g(X_k)}$  où les  $(X_k)_k$  sont des variables aléatoires i.i.d. suivant la loi  $gd\nu$ . Cet estimateur est également un estimateur de Monte-Carlo, mais en simulant  $X$  sous une nouvelle loi. Il est donc sujet au même TCL que les estimateurs de Monte-Carlo naïfs :

$$\sqrt{n}(\hat{h}_n - \mathbb{E}_f[g(X)]) \xrightarrow{L} \mathcal{N}(0, \sigma^2)$$

où  $\sigma^2 = \mathbb{V}_g \left[ h(X) \frac{f(X)}{g(X)} \right]$ . L'intérêt de la méthode est donc de choisir  $g$  de telle manière à ce que  $\mathbb{V}_g \left[ h(X) \frac{f(X)}{g(X)} \right] < \mathbb{V}_f [h(X)]$ . Intuitivement, une manière de faire cela est de choisir  $g$  de telle sorte que du poids soit donné aux valeurs de  $X$  pertinentes pour l'évaluation de l'espérance de  $h(X)$ .

### 4.2 Décalage d'un vecteur gaussien

Comme nous l'avons montré dans la présentation de chaque modélisation,  $\mathbb{P}(L \geq k)$  peut être exprimé comme  $\mathbb{E}[h(X)]$  où  $X$  est un vecteur gaussien centré, éventuellement de dimension 1<sup>2</sup>.

Pour l'implémentation de la méthode de Monte-Carlo par échantillonnage d'importance, nous avons donc choisi à chaque fois de modifier la loi du vecteur gaussien de loi  $\mathcal{N}(0, A)$  en un vecteur gaussien de loi  $\mathcal{N}(\mu, A)$  où  $\mu$  est un vecteur que nous avons choisi pour chaque cas.

---

<sup>2</sup>cf. Section 2

Dans ce cas, on a

$$\begin{aligned}
\mathbb{E}_{(0,A)}[h(X)] &= \frac{1}{\sqrt{(2\pi)^n \det(A)}} \int_{\mathbb{R}^d} h(x) \exp\left(-\frac{1}{2}x^T A^{-1}x\right) dx \\
&= \frac{1}{\sqrt{(2\pi)^n \det(A)}} \int_{\mathbb{R}^d} h(x) \exp\left(-\mu^T A^{-1}x + \frac{1}{2}\mu^T A^{-1}\mu\right) \exp\left(-\frac{1}{2}(x-\mu)^T A^{-1}(x-\mu)\right) dx \\
&= \mathbb{E}_{(\mu,A)}\left[h(X) \exp\left(-\mu^T A^{-1}X + \frac{1}{2}\mu^T A^{-1}\mu\right)\right]
\end{aligned}$$

On obtient ainsi un nouvel estimateur des différents  $\mathbb{P}(L \geq k)$ .

### 4.3 Choix de $\mu$ pour les différentes modélisations

#### Probabilité de défaut d'une seule entreprise

Comme mentionné dans la section 2, nous avons estimé la probabilité de défaut d'une seule entreprise par une méthode de Monte-Carlo par échantillonnage d'importance. Nous avons  $p = \mathbb{P}\left(\mathcal{N}(0, T) \leq \frac{\ln(B/S_0^i) + 0,5\sigma_i^2 t}{\sigma_i}\right)$  Nous avons donc choisi  $\mu = \frac{\ln(B/S_0^i) + 0,5\sigma_i^2 t}{\sigma_i}$ . Cela nous a permis d'obtenir une estimation très précise de  $p$ , et donc ainsi des  $\mathbb{P}(L \geq k)$  par la suite. Comme dans le cours, le choix d'une valeur légèrement plus petite aurait pu permettre une variance encore plus faible, mais le résultat ici est déjà extrêmement satisfaisant comparé aux méthodes suivantes.

#### Evolution indépendante et dépendante

Dans les 2 cas où l'on ne regarde la faillite qu'à la date  $T$ , nous avons choisi de prendre  $\mu$  de telle sorte que le nombre moyen de faillites soit égal à  $k$ . On a

$$\mathbb{E}_{(\mu,A)}[L] = \sum_{k=1}^N \mathbb{P}_{(\mu,A)}(S_T^i \leq B)$$

Or,

$$\begin{aligned}
\mathbb{P}_{(\mu,A)}(S_T^i \leq B) &= \mathbb{P}\left(\mathcal{N}(\mu_i, T) \leq \frac{\ln(B/S_0^i) + 0,5\sigma_i^2 t}{\sigma_i}\right) \\
&= \mathbb{P}\left(\mathcal{N}(0, 1) \leq \frac{\ln(B/S_0^i) + 0,5\sigma_i^2 t}{\sigma_i\sqrt{T}} - \mu_i\right)
\end{aligned}$$

Donc en choisissant

$$\mu_i = \frac{\ln(B/S_0^i) + 0,5\sigma_i^2 t}{\sigma_i} - F^{-1}\left(\frac{k}{N}\right)$$

(où  $F$  est la fonction de répartition de la loi gaussienne centrée réduite), on obtient bien  $\mathbb{P}_{(\mu,A)}(S_T^i \leq B) = k/N$  et donc  $\mathbb{E}_{(\mu,A)}[L] = k$  En réalité, pour les cas  $k = 0$  et  $k = N$ , on a choisi respectivement  $F^{-1}(0.01)$  et  $F^{-1}(0.09)$  afin d'obtenir un vecteur  $\mu$  dont les composantes sont finies.

Dans nos simulations, le résultat de cette méthode est mitigé. En effet, il semble qu'en diminuant le nombre d'entreprises, cette méthode fonctionne très bien pour notre choix de changement de loi. Cependant, à mesure que  $N$  augmente, l'estimation baisse fortement en qualité. On peut expliquer cela par le fait que le terme correctif  $\exp(-\mu^T A^{-1} X + \frac{1}{2} \mu^T A^{-1} \mu)$  a une variance de plus en plus élevée lorsque la dimension augmente, ce qui rend les estimations mauvaises, bien qu'elle soient toujours meilleures que celles de Monte-Carlo naïf. On remarque également que des  $\sigma_i$  différents augmente la variance de l'estimateur et par conséquent diminuent la qualité de l'estimation. L'introduction d'une corrélation entre les différentes valeurs augmente également la qualité de l'estimation, la variance de cet estimateur est donc proportionnellement plus faible.

### Extension

Le choix du vecteur  $\mu$  dans le cas où l'on regarde la faillite à plusieurs dates paraissait moins naturel. Le fait que l'entreprise  $i$  soit en faillite à la date  $t_k$  s'exprime par la condition  $S_{t_k}^i \leq B_i$ , ce qui revient à  $W_{t_k}^i \leq K_{t_k}^i$  avec  $K_{t_k}^i = \frac{\ln(B/S_0^i) + 0,5\sigma_i^2 t_k}{\sigma_i}$ . Nous avons donc choisi  $\mu$  de telle sorte qu'en moyenne  $W_{t_k}^i$  soit égal à  $K_{t_k}^i$ .

Il semble que ce choix n'ait pas été bien judicieux car les estimations sont vraiment loin de celles de Monte-Carlo naïf, même lorsque celui-ci semble bien fonctionner, en particulier lorsque la fréquence de contrôle de la valeur des entreprises augmente. Il semble donc qu'il y ait à nouveau un effet de grande dimension faisant fortement augmenter la variance, et rendant ici notre choix de changement de loi inexploitable.

## 5 Méthodes de splitting

Les méthodes dites de splitting servent à calculer une probabilité de la forme  $\mathbb{P}(g(X) \geq a)$ . Pour cela, on écrit

$$\mathbb{P}(g(X) \geq a) = \mathbb{P}(g(X) \geq a_0) \prod_{i=0}^{J-1} \mathbb{P}(g(X) \geq a_{i+1} | g(X) \geq a_i)$$

avec  $a_0 < a_1 < \dots < a_J = a$

Il suffit alors d'estimer chaque  $\mathbb{P}(g(X) \geq a_{i+1} | g(X) \geq a_i)$ . Si les  $(a_i)_i$  sont bien choisis, ces événements ne sont plus rares, et sont donc plus simples à estimer correctement. Dans nos simulations, nous avons choisi de simuler les lois de  $X | g(X) \geq a_i$  grâce à une chaîne de Markov dont la loi de  $X$  est une loi réversible, en utilisant le théorème ergodique. Nous avons estimé  $\mathbb{P}(g(X) \geq a_{i+1} | g(X) \geq a_i)$  par l'estimateur

$$\frac{1}{n} \sum_{k=1}^n \mathbb{1}_{g(X_k) \geq a_i}$$

où  $(X_n)_n$  est une chaîne de Markov dont la loi invariante est la loi  $\mathcal{L}(X)|_{g(X) \geq a_i}$ . Cependant, cette méthode ne permet pas de calculer des intervalles de confiance asymptotiques car le TCL vérifié par cette estimateur ne permet pas de calculer en pratique la variance asymptotique.

## 5.1 Chaînes de Markov pour les vecteurs gaussiens

Pour construire nos chaînes de Markov, il nous a fallu construire un noyau de transition dans  $\mathbb{R}^d$  dont la mesure invariante est la loi du vecteur gaussien souhaité. Pour simuler les lois conditionnelles comme cela est nécessaire pour les estimations, nous avons ensuite utilisé ce noyau en n'acceptant la proposition du noyau que si  $g(X) \geq a_i$ . La loi invariante de cette nouvelle chaîne de Markov est alors bien  $\pi|_{g(X) \geq a_i}$ , où  $\pi$  est la loi invariante du premier noyau.

Pour simuler un vecteur gaussien centré de matrice de covariance  $A$ , nous avons choisi le noyau du processus suivant : à  $X$ , on associe  $aX + \sqrt{1-a^2}Y$  où  $Y \sim \mathcal{N}(0, A)$  est indépendant de  $X$ , et  $a \in ]0, 1[$  est un hyperparamètre. Ce noyau est réversible pour la loi  $\mathcal{N}(0, A)$ .

Dans les simulations de loi conditionnelle, le choix de  $a$  est important :  $a$  proche de 1 signifie que l'on change peu par rapport au vecteur initial et la convergence est plus lente, tandis que pour  $a$  proche de 0, on s'éloigne beaucoup de la valeur initiale, ce qui signifie que l'on a plus tendance à rejeter les propositions, la convergence est alors ralentie également.

## 5.2 Splitting sur $k$

Comme vu précédemment,  $L$  est une fonction du vecteur  $(W_T^1, \dots, W_T^N)$  dans les deux premières modélisations et ensuite est une fonction de  $\Delta W$  dans l'extension du problème. Ainsi, on peut choisir de simuler  $\mathbb{P}(L \geq k)$ , en décomposant

$$\mathbb{P}(L \geq k) = \prod_{i=0}^{k-1} \mathbb{P}(L \geq i+1 | L \geq i)$$

Nous avons implémenté cette méthode de 2 manières différentes : tout d'abord en prenant pour point de départ un état issu de la chaîne de Markov précédente, et ensuite en prenant comme point de départ un état fixe où les  $k$  premières entreprises étaient en défaut (à leur seuil de référence), mais pas les autres. Cette méthode n'a pas fonctionné pour  $k$  trop grand car les probabilités  $\mathbb{P}(L \geq i+1 | L \geq i)$  étaient trop faibles à partir de  $i = 9$ , et notre algorithme estimait ces probabilités comme étant nulles.

## 5.3 Splitting sur $B$

Nous avons donc choisi de changer de point de vue.

En posant  $g_k(x_1, \dots, x_N)$  qui à  $(x_1, \dots, x_N)$  associe la  $k$ -ième plus petite valeur parmi les vecteurs, et  $S^i = \min(S_{t_1}^i, \dots, S_{t_M}^i)$ . On peut alors réécrire  $\{L \geq k\}$  comme étant  $\{g_k(S_T) \leq B\}$  dans les deux premières modélisations et

$\{g_k(S^1, \dots, S^N) \leq B\}$  dans le cas de l'extension du problème. En réécrivant respectivement ces événements sous la forme  $\{S_k(W_T) \leq B\}$  et  $\{S'_k(\Delta W) \leq B\}$ , on peut alors réaliser la même méthode en décomposant suivant les niveaux de  $B$  :

$$\mathbb{P}(S_k(W_T) \leq B) = \mathbb{P}(S_k(W_T) \leq B_0) \prod_{i=0}^{J-1} \mathbb{P}(S_k(W_T) \leq B_{i+1} | S_k(W_T) \leq B_i)$$

avec  $B = B_J < \dots < B_1 < B_0$

Pour choisir les différents niveaux  $(B_i)_i$ , nous les avons choisi empiriquement de telle manière que chaque  $\mathbb{P}(S_k(W_T) \leq B_{i+1} | S_k(W_T) \leq B_i)$  vaille environ 0.1, en calculant approximativement un quantile grâce à une chaîne de Markov dont la loi invariante est la loi cible.

Cette méthode de splitting semble avoir donné des résultats concluants : les estimations sont plutôt précises (dans le cas où l'on connaît la valeur théorique). Pour  $n = 10^4$ , l'algorithme arrivait à approximer les quantiles, puis calculer les différentes probabilités jusqu'à  $k = 40$ , valeur pour laquelle la probabilité estimée est de l'ordre de  $10^{-49}$ . Les probabilités étant très faibles, nous avons choisi l'hyperparamètre  $a$  assez grand ( $a = 0.99$ ), afin de garantir un taux d'acceptation des propositions plutôt élevées. Enfin, on peut noter que le temps de calcul est rapide (comparé à la méthode suivante).

## 5.4 Evaluation des pertes à l'aide d'une chaîne de Markov

Pour évaluer l'espérance conditionnelle des pertes, nous l'avons tout d'abord estimé avec la même méthode que pour l'estimation des probabilités conditionnelles ci-dessus : en moyennant les pertes sur une chaîne de Markov. Bien que nous ne puissions à nouveau pas obtenir d'intervalles de confiance de manière explicite, les simulations semblent donner un résultat plutôt convaincant, montrant que  $\mathbb{E}[P_T | L \geq k]$  est plus ou moins linéaire en  $k$ .

## 6 Adaptative Multilevel Splitting

Enfin, nous avons utilisé une dernière méthode, la méthode de l'Adaptative Multilevel Splitting, et l'algorithme de la dernière particule. Celui-ci est un raffinement de la méthode de Splitting. En effet, plutôt que de choisir les niveaux à l'avance, l'algorithme adapte les niveaux au cours de la simulation.

Afin d'avoir une loi sans atomes, nous avons repris le point de vue en modifiant chaque probabilité comme étant  $\mathbb{P}(S_k(W_T) \leq B)$

L'algorithme est le suivant :

---

**Algorithm 1:** Algorithme de la dernière particule pour estimer  $\mathbb{P}(S(X) < B)$

---

**Entrée:** Nombre de particules  $n$

**Sortie :** Estimation de  $\mathbb{P}(S(X) < B)$

Générer  $n$  copies indépendantes  $(X_1, \dots, X_n)$  selon la loi de  $X$

**repeat**

    Calculer  $L = \min(S(X_1), \dots, S(X_n))$

**for**  $j = 1$  **to**  $n$  **do**

**if**  $S(X_j) = L$  **then**

            Remplacer  $X_j$  par  $X^*$  indépendant de  $(X_1, \dots, X_n)$  tel que

$X^* \sim \mathcal{L}(X) | S(X) < L$

**end**

**end**

**until**  $L < B$ ;

**return**  $(1 - \frac{1}{n})^{J_n}$  où  $J_n$  est le nombre d'itérations;

---

$J_n$  suit alors une loi de Poisson de paramètre  $-n \log p$  où  $p$  est la probabilité estimée. On obtient donc un estimateur  $\hat{p}_n$  sans biais vérifiant le TCL suivant :

$$\sqrt{n}(\hat{p}_n - p) \xrightarrow{\mathcal{L}} \mathcal{N}(0, -p^2 \log p)$$

On peut alors obtenir l'intervalle de confiance asymptotique à 95% suivant :

$$I_n = \left[ \hat{p}_n - \hat{p}_n \frac{1.96 \sqrt{-\log \hat{p}_n}}{\sqrt{n}}, \hat{p}_n + \hat{p}_n \frac{1.96 \sqrt{-\log \hat{p}_n}}{\sqrt{n}} \right]$$

\*

Dans les faits, il est compliqué de simuler la variable  $X^*$  de manière totalement indépendante des autres particules. Pour faire cela, nous avons simulé une chaîne de Markov par rejet, avec le noyau mentionné dans la section précédente en partant d'une des particules restantes, choisi au hasard. Plus cette chaîne est longue, plus l'indépendance entre  $X^*$  et les particules est importante. Nous avons choisi une longueur fixe  $l = 20$  afin de faire un compromis entre l'indépendance des variables, et le temps de calcul (qui était déjà assez long).

On peut d'abord remarquer d'un point de vue théorique que l'écart relatif de l'intervalle de confiance est asymptotiquement négligeable par rapport à celui de Monte-Carlo naïf. Cet algorithme a produit de bonnes estimations mais la contrepartie a été des temps de calcul assez longs. Le temps de calcul est proportionnel à  $J_n$ , et donc en moyenne, il est en  $\mathcal{O}(-n \log p)$ , ce qui signifie que l'algorithme tend à être plus long lorsque les probabilités sont petites. L'algorithme ne permettait pas de calculer les  $\mathbb{P}(L \geq k)$  en moins d'une heure au-delà de  $k = 10$ .

## 6.1 Estimation des pertes

L'algorithme de la dernière particule permet également de simuler une approximation de  $\mathcal{L}(X) | S(X) < B$ . En effet, à la fin de l'algorithme,  $X_1, \dots, X_n$  peu-

vent être assimilés à la réalisation de variables i.i.d. suivant  $\mathcal{L}(X)|S(X) < B$ . Ainsi, à partir de l'état final des particules, nous avons pu estimer les pertes, par une méthode de Monte-Carlo naïve, une fois l'échantillon sus-mentionné construit. Cela a fourni des résultats plutôt satisfaisants.

## 7 Influence des différents paramètres

Nous allons désormais nous intéresser à l'influence des différents paramètres sur les résultats de nos simulations.

### Influence de $B$

De manière générale, les probabilités sont croissantes avec  $B$ , ce qui se vérifie d'une part dans le calcul théorique et d'autre part sur les simulations réalisées, quelle que soit la méthode. Un seuil  $B$  plus élevé permet notamment d'avoir des résultats concluants pour des  $k$  plus grands pour les méthodes de Monte-Carlo (naïf et par échantillonnage d'importance). La valeur maximale passe de  $k = 10$  à  $k = 35$  lorsque  $B$  passe de 36 à 60. De plus, étant donné que l'augmentation des probabilités a tendance à diminuer l'erreur relative, celle-ci s'en trouve réduite dans nos simulations. Les temps de calcul ne sont quant à eux pas influencés (sauf pour la dernière particule, et le splitting).

### Influence de $\sigma_i$

Les variations de probabilité en ajoutant des  $\sigma_i$  différents sont relativement faibles. Cependant, cela a pour conséquences d'augmenter sensiblement l'erreur relative des simulations et l'augmentation drastique du temps de calcul impose de considérer moins de simulations, ce qui augmente d'autant plus l'erreur relative, et diminue le  $k$  maximal pour lequel on peut calculer  $\mathbb{P}(L \geq k)$ .

### Influence de $\rho$

L'introduction d'un facteur de corrélation, même faible ( $\rho = 0.1$ ) augmente drastiquement les probabilités considérées. Par exemple, pour  $k = 125$ ,  $\mathbb{P}(L \geq k)$  passe de  $10^{-250}$  à  $10^{-30}$  selon les estimations de Monte-Carlo naïf. Les algorithmes sont par conséquent beaucoup plus performants, notamment celui de Monte-Carlo par échantillonnage d'importance qui renvoie une courbe lisse pour  $10^4$  échantillons et  $\rho = 0.1$ . Comme attendu, les effets s'intensifient lorsque  $\rho$  augmente.

## 8 Conclusion

Ainsi, ce projet nous a permis d'implémenter plusieurs algorithmes de simulation d'événements rares dans un cadre concret. La méthode de Monte-Carlo naïve est insuffisante lorsque les probabilités sont trop faibles, ce qui nous a amenés à considérer de nouvelles méthodes. De manière générale, l'augmentation de la dimension des variables aléatoires simulées a toujours eu un impact négatif sur

la qualité de nos simulations. Les changements de paramètres conduisant à une augmentation des probabilités a toujours amélioré la qualité de nos algorithmes. On notera que le plus précis fût l'algorithme de la dernière particule, mais cela s'est payé par un temps de calcul bien plus élevé. De nombreux choix nous ont montré que les problématiques d'estimation d'évènements rares relèvent avant tout de l'ingénierie numérique par le choix pertinents d'hyperparamètres.