

RISQUE DE DÉFAUT D'ENTREPRISE

Rapport final - MAP473D

23 mai 2021

Yassine ETTACHFINI
Outhmane ITRO



TABLE DES MATIÈRES

Introduction	3
1 Modèle structurel	4
2 Cas de débiteurs indépendants	5
2.1 Distribution de la variable L	5
Estimation du paramètre p	5
Simulation du vecteur $W = (W_1, \dots, W_N)$	8
Exponential tilting	13
2.2 Influence du seuil B et de σ	18
2.3 Perte associée au défaut	20
2.4 Commentaires	22
3 Cas de débiteurs dépendants	24
3.1 Calcul des probabilités $P(L = j)$	24
3.2 Estimation par échantillonnage d'importance	25
Influence du seuil B et de la corrélation ρ	26
Perte associée au défaut	27
3.3 Exponential Tilting - (Copule gaussienne)	28
3.4 Commentaires et conclusions	31
4 Défaut d'entreprise pour $\tau \leq T$	32
4.1 Ce qui change	32
4.2 Interacting Particle System	32
Principe	32
Description de la méthode IPS	33
4.3 Résultats	34

INTRODUCTION

Le risque de crédit est le risque qu'un emprunteur ne rembourse pas tout ou partie de son crédit aux échéances prévues par le contrat signé entre lui et l'organisme prêteur (généralement une banque). La maîtrise du risque de crédit est au coeur du métier du banquier car il détermine la rentabilité des opérations effectuées. Si l'établissement financier sous-évalue ce risque, le montant prêté et les intérêts dus ne seront pas perçus et s'inscriront en perte. L'évaluation du risque de crédit passe par une bonne connaissance du client et, dans le cas d'une entreprise, par une bonne évaluation de son projet et du secteur dans lequel elle exerce son activité. L'évaluation de ces risques est aussi d'importance majeure pour la construction de produits dérivés et leur pricing. Les CDO (Collateralized Debt Obligations) par exemple sont des dérivés de crédits composés de titres représentatifs (entre 120 et 250 actifs) de portefeuilles de créances bancaires ou d'autres instruments financiers. Ces produits sont généralement divisés en tranches selon le degré de risque pris par l'investisseur en faisant l'acquisition. L'évaluation des taux de retour des tranches et de leurs risques demande une bonne connaissance et une bonne modélisation des actifs et du risque de défaut de certains débiteurs dont les emprunts forment le CDO. Selon le niveau de risque de ces tranches, on lui attribue une note (un rating) :

Rating	Description	
AAA	Highest Credit Quality	Investment Grade*
AA	Very High Credit Quality	
A	High Credit Quality	
BBB	Good Credit Quality	
BB	Speculative	Non-Investment Grade**
B	Highly Speculative	
CCC	Substantial Credit Risk	
CC	Very High Levels of Credit Risk	
C	Near Default	
RD	Restricted Default	
D	Default	

FIGURE 1 : Modèle de rating des tranches d'un CDO

Nous nous intéressons à l'estimation et la quantification du risque de défaut de crédit d'entreprises et d'émetteur dans un portefeuille d'actifs. La difficulté étant d'estimer correctement différentes métriques de risques quand l'événement de défaut d'un ou de plusieurs entreprises devient très rare. Nous commencerons pour cela par étudier le cas de débiteurs (ou entreprises) indépendants puis le cas où nos actifs seraient corrélés et nous nous intéresserons à l'influence de cette corrélation ainsi que d'autres paramètres (volatilité des actifs, seuil de défaut ...) sur les risques de défaut de crédit. Nous nous aiderons de plusieurs méthodes de réduction de variance et de simulation d'événements rares (Importance sampling, Interaction Particle System , MCMC ...) et compareront les performances de ces différentes méthodes dans l'évaluation de certains indicateurs numériques qui nous permettront de quantifier le risque de défaut de nos actifs.

1

MODÈLE STRUCTUREL

On considère un ensemble de $N = 125$ entreprises dont les valeurs économiques sont modélisées par des variables aléatoires qui peuvent évoluer au cours du temps. On supposera, notamment, que la valeur de l'entreprise i est modélisée par un mouvement brownien géométrique :

$$S_t^i = S_0 \exp\left(-\frac{1}{2}\sigma_i^2 t + \sigma_i W_t^i\right)$$

pour $1 \leq i \leq N$, où les $(W_t^i)_{t \geq 0}$ des mouvements brownien et où les σ_i sont les écarts-types des rendements et représentent donc les volatilités respectives des actifs.



FIGURE 2 : Exemple d'évolution temporelle des valeurs des entreprises

L'entreprise i est considérée en défaut si sa valeur à un instant précis est inférieure à une valeur B_i fixée, typiquement très inférieure à S_0^i . Nous cherchons à évaluer la distribution du nombre de faillites à une date T future fixée (disons $T = 1$ an). Notons par D_T l'ensemble des indices des entreprises en défaut à la date T , alors :

$$D_T = \{1 \leq i \leq N : S_T^i \leq B_i\}$$

On considère la variable aléatoire L qui donne le nombre d'entreprises qui font défaut dans notre modèle. L s'écrit donc comme :

$$L = \#D_T = \sum_{i=1}^N \mathbb{1}_{\{S_T^i \leq B_i\}}$$

L peut être vu comme une perte, et pour évaluer le risque de nos actifs on s'intéressera notamment à la distribution de cette variable aléatoire. Il est à noter que l'estimation de $\mathbb{P}(L \geq k)$ devient plus délicate pour des valeurs de k de plus en plus importante.

Pour le reste de notre analyse, dans le cas où cela n'est pas spécifié, nous prendrons : $B_i = B$ pour un certain $B > 0$ qui sera précisé, $S_0^i = S_0 = 100$ et $\sigma_i = \sigma = 40\%$.

2

CAS DE DÉBITEURS INDÉPENDANTS

Dans cette première partie, on suppose que les mouvements browniens modélisant les valeurs des entreprises sont indépendants.

2.1 DISTRIBUTION DE LA VARIABLE L

Sous ces hypothèses, nous n'avons pas besoin de tout le processus historique pour décider du défaut de l'entreprise et il nous suffit d'évaluer la valeur de l'actif à maturité, soit au temps final $T = 1$. On peut donc obtenir des simulations de la valeur finale d'une entreprise i en simulant une seule variable $W_i \sim \mathcal{N}(0, 1)$. On obtient donc les valeurs de toutes les entreprises au temps T en simulant un vecteur gaussien (W_1, \dots, W_N) de loi $\mathcal{N}(0_N, \mathbb{I}_N)$.

En reprenant les notations du modèle, peut voir que dans ce contexte d'étude, les variables $(\mathbb{1}_{\{S_T^i \leq B\}})_{1 \leq i \leq N}$ sont des variables de Bernoulli indépendantes et de même paramètre $p = \mathbb{P}(S_T^i \leq B) = \mathbb{P}(Z \leq M_{(B, \sigma)})$ avec $M_{(B, \sigma)} = \frac{1}{\sigma} \log \left(\frac{B}{S_0 e^{-\frac{1}{2}\sigma^2}} \right)$ et $Z \sim \mathcal{N}(0, 1)$.

On en déduit que **L suit une loi binomiale de paramètre $N = 125$ et $p = \mathbb{P}(Z \leq M_{(B, \sigma)})$.**

Pour déterminer la distribution de la variable aléatoire L , on utilisera différents procédés selon que l'on s'accorde ou non un accès direct au paramètre p :

- Estimation du paramètre p par méthode de Monte Carlo ou échantillonnage d'importance pour pouvoir en déduire $\mathbb{P}(L \geq k)$ pour tout k .
- Estimation directe par simulation du vecteur (W_1, \dots, W_N) par encore une fois, Monte Carlo naïf ou échantillonnage d'importance.
- Estimation de la distribution en simulant L par Monte Carlo naïf ou Exponential tilting en supposant p connue.

• ESTIMATION DU PARAMÈTRE p

Estimation par Monte Carlo naïf :

D'après ce qui précède :

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

Connaître le paramètre $p = \mathbb{P}_{\mathcal{N}(0,1)}(Z \leq M_{(B,\sigma)})$ nous permet donc de déterminer toutes les probabilités $\mathbb{P}(L \geq k)$. Puisque $M_{(B,\sigma)}$ peut être arbitrairement petit selon les valeurs de B et σ , nous devons nous assurer de pouvoir l'estimer au mieux quelque soient les valeurs choisies pour ces paramètres. La première méthode d'estimation que nous entreprenons est une méthode de Monte Carlo naïve en utilisant une suite $(X_j)_{1 \leq j \leq M}$ de v.a.r normale centrée réduite avec M suffisamment grand s'appuyant sur l'estimateur suivant :

$$\hat{p}_m^{MC} = \frac{1}{M} \sum_{j=1}^M \mathbb{1}_{X_j \leq B}$$

On obtient l'estimation suivante pour $M = 10^5$:

$$I_c^{95\%}(p) = [0.0105 \pm 0.0006]$$

Estimation par échantillonnage d'importance :

Nous estimons également p par d'échantillonnage d'importance par changement de moyenne pour la variable Z . On utilise la relation suivante :

$$p = \mathbb{P}_{\mathcal{N}(0,1)}(Z \leq M_{(B,\sigma)}) = \mathbb{E}_{\mathcal{N}(\lambda,1)} \left(\frac{\mathcal{N}(0,1)(Z)}{\mathcal{N}(\lambda,1)(Z)} \mathbb{1}_{Z \leq M_{(B,\sigma)}} \right) = \mathbb{E}_{\mathcal{N}(\lambda,1)} \left(\exp\left(\frac{\lambda^2}{2} - \lambda Z\right) \mathbb{1}_{Z \leq M_{(B,\sigma)}} \right)$$

On choisit λ de sorte à minimiser la variance asymptotique de notre estimateur :

$$\mathbb{V}(\lambda) = \mathbb{V}_{\mathcal{N}(\lambda,1)} \left(\exp\left(\frac{\lambda^2}{2} - \lambda Z\right) \mathbb{1}_{Z \leq M_{(B,\sigma)}} \right)$$

Et l'on montre que :

$$\mathbb{V}(\lambda) = \mathbb{E}_{\mathcal{N}(0,1)} \left(\exp\left(\frac{\lambda^2}{2} - \lambda Z\right) \mathbb{1}_{Z \leq M_{(B,\sigma)}} \right)^2 - \left(\mathbb{P}(Z \leq M_{(B,\sigma)}) \right)^2$$

On en déduit que réduire la variance $\mathbb{V}(\lambda)$ revient à réduire un moment d'ordre 1. Il s'agit d'une fonction deux fois dérivable convexe, le point λ^* qui minimise donc cette variance vérifie $\mathbb{V}'(\lambda^*) = 0$. On trouve que $\mathbb{V}'(0) \geq \mathbb{V}'(M_{(B,\sigma)}) \geq 0$ et on en déduit que le changement de moyenne $\lambda = M_{(B,\sigma)}$ permet d'avoir une variance asymptotique inférieure à celle de la loi de base d'une gaussienne centrée réduite. Sous cette nouvelle loi on simule $M = 10^5$ variables $(Z_i)_{i \geq 1}$ en s'appuyant sur l'estimateur :

$$\hat{p}_m^{IS} = \frac{1}{M} \sum_{i=1}^M \mathbb{1}_{Z_i \leq B} \exp \left(\frac{M_{(B,\sigma)}^2}{2} - M_{(B,\sigma)} Z_i \right)$$

On obtient l'estimation suivante pour $M = 10^5$:

$$I_c^{95\%}(p) = [0.0093 \pm 9.10^{-5}]$$

Calcul des probabilités $P(L \geq k)$:

On utilise les différentes estimations obtenues pour p pour calculer les probabilités $P(L \geq k)$ pour tout $1 \leq k \leq N$

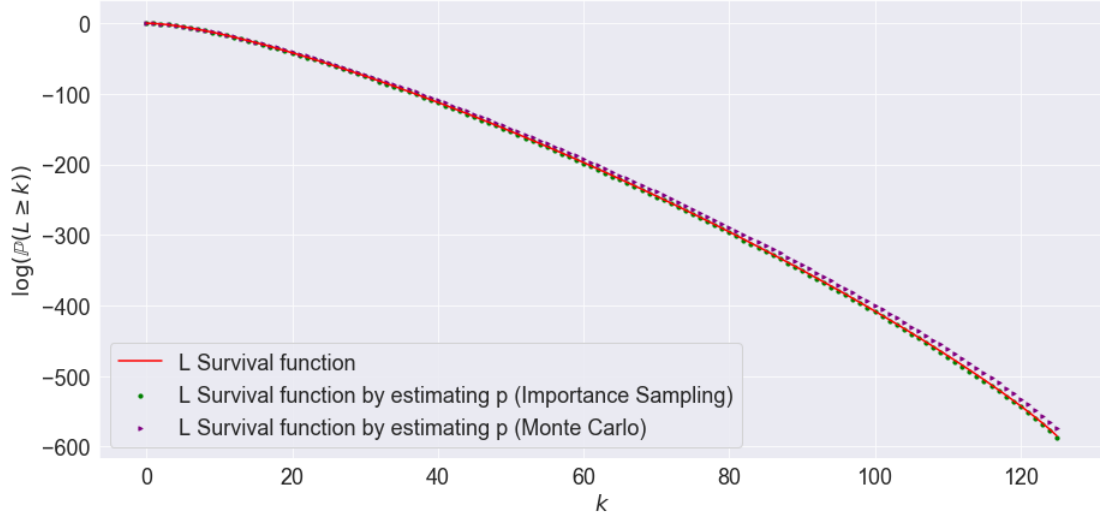


FIGURE 3 : $\log(\mathbb{P}(L \geq k))$ en fonction de k avec différentes estimations de p , $N = 125$ et $M = 10^5$

Pour continuer, il nous reste à définir des intervalles de confiance pour chaque probabilité estimée de la sorte en partant de ceux retrouvés pour le paramètre p . On s'assure pour cela du caractère monotone de $P_k(p) = \sum_{i=k}^N p^i (1-p)^{(N-i)} \binom{N}{i}$ (du moins pour des petites valeurs de p). P_k est différentiable par rapport à p et on a :

$$\forall p \text{ et } 1 \leq k \leq N, \quad P'_k(p) = \sum_{i=k}^N i p^i (1-p)^{(N-i-1)} \binom{N}{i} \left(\frac{N}{i} + \frac{1}{p} - 2 \right)$$

Or dans notre cas, $p \ll \frac{1}{2}$ donc $\forall k \geq 1$, la fonction P_k est une fonction croissante (même strictement croissante) de p . On peut donc calculer les intervalles de confiance à 95%. Nous en donnons quelques exemples pour certaines valeurs de k , s'appuyant sur les valeurs de p calculées avec échantillonnage d'importance :

k	2	4	10	25
$P(L \geq k)$	0.31	0.027	2.10^{-07}	5.10^{-26}
$I_c^{95\%}$	[0.30 , 0.33]	[0.025 , 0.031]	[$1.9.10^{-07}$, $3.5.10^{-07}$]	[2.10^{-26} , 11.10^{-26}]

TABLE 1 : Intervalles de confiance à 95% calculés avec la probabilité p estimée par échantillonnage d'importance

- SIMULATION DU VECTEUR $W = (W_1, \dots, W_N)$

Estimation par Monte Carlo naïf :

Dans cette partie, nous simulons le vecteur gaussien W de taille N et nous calculons, pour chaque vecteur simulé, le nombre de composantes inférieures à $M_{(B,\sigma)}$. On notera ces variables $(L_j)_{1 \leq j \leq M}$ et on évalue les probabilités $\mathbb{P}(L \geq k)$ à l'aide de l'estimateur :

$$\hat{p}_k = \frac{1}{M} \sum_{j=1}^M \mathbb{1}_{L_j \geq k}$$

Cette méthode échoue malheureusement très rapidement, et n'arrive presque pas à voir l'occurrence de plus de 8/9 défauts. On s'oriente donc comme ce qui précède vers une méthode d'échantillonnage d'importance dans le but de rendre plus fréquent le défaut d'un nombre plus important d'entreprises.

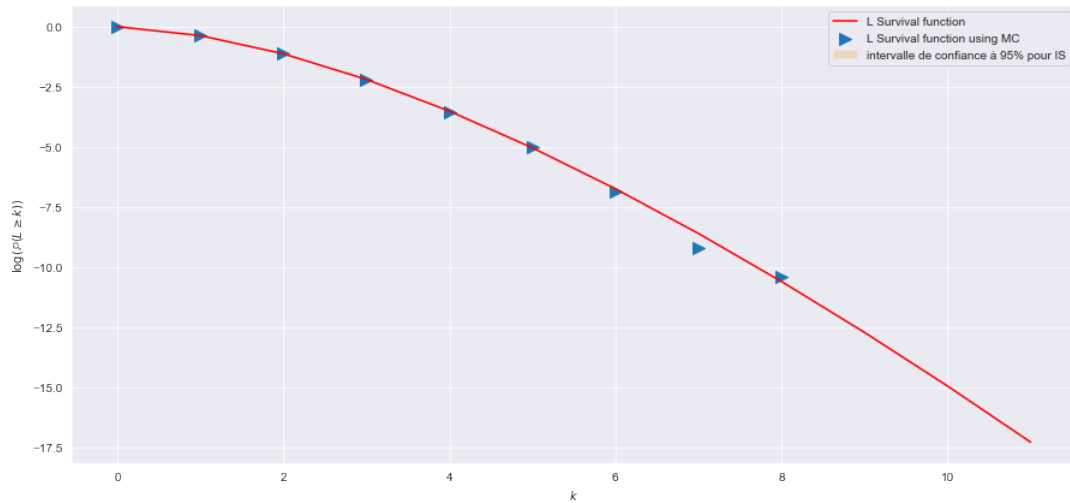


FIGURE 4 : $\log(\mathbb{P}(L \geq k))$ en fonction de k par MC naïf pour $M = 10^5$ réalisations et $N = 125$

k	2	4	6	8
$P(L \geq k)$	0.320	0.029	0.00097	1.10^{-5}
$I_c^{95\%}$	[0.31 , 0.33]	[0.02 , 0.03]	[0.0007, 0.0012]	[0.0 , 3.10^{-5}]

TABLE 2 : Estimation par méthode de MC naïf pour $M = 10^5$ réalisations

Estimation par échantillonnage d'importance :

Il est beaucoup plus difficile dans ce cas de déterminer un changement de probabilité qui nous garantisse une réduction suffisante de la variance, voire même juste une réduction de celle-

ci dans certains cas. S'ajoute à cela les risques d'erreurs de précision dans nos calculs avec le terme exponentiel nouvellement introduit.

Première tentative : Intuitivement, nous pensons que plus les moyennes des composantes du vecteur seront négatives, plus nous aurons de chances de tomber sur des nombres plus importants de défaut. Ceci nous conduit à notre première approche : pour tout k , nous changeons la moyenne nulle par le vecteur $\Lambda = (\lambda, \dots, \lambda) = \lambda u$, où λ dépend de k . On a dans ce cas :

$$\mathbb{P}_{\mathcal{N}(0_n, I_n)}(L \geq k) = \mathbb{E}_{\mathcal{N}(\Lambda, I_n)} \left(\exp \left(\sum_{i=1}^N W_i \lambda + \frac{1}{2} N \lambda^2 \right) \mathbb{1}_{L \geq k} \right)$$

étant donné que L est, ici, une fonction mesurable de notre vecteur gaussien W .

Le choix du paramètre λ dépend de la probabilité que l'on cherche à estimer et donc dépend de k . Plus k est grand plus nous serons enclins à prendre des valeurs négatives plus grandes en valeur absolue de notre moyenne.

Le premier changement possible est donc de faire varier λ en fonction de k de sorte à le pousser vers des valeurs plus négatives plus k est grand. Le graphe suivant résume ce qu'on obtient pour, par exemple, $\lambda(k) = (\frac{k}{N+1} - 1)^{-1}$ et $B = 36$ pour $N = 10$:

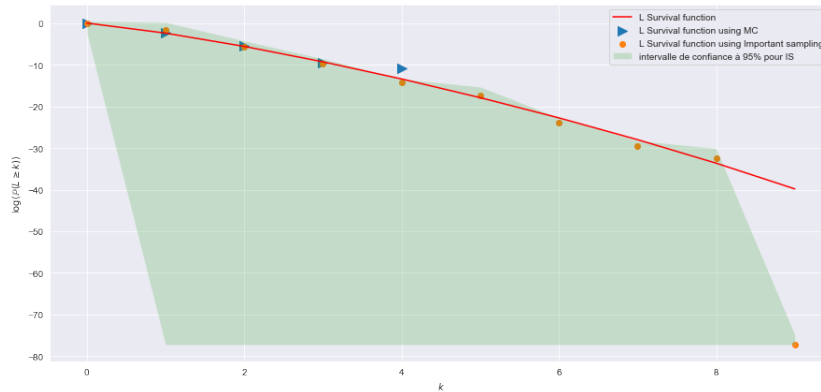


FIGURE 5 : $\log(\mathbb{P}(L \geq k))$ en fonction de k pour $N = 10$ et $M = 10^5$

On arrive avec cette première approche intuitive à estimer $P(L \geq 8) = 7.10^{-15}$ (la probabilité réelle étant $2,44.10^{-15}$) avec l'intervalle de confiance à 95% suivant :

$$I_c^{95\%} = [0.0, 8.10^{-14}]$$

On remarque cependant que l'on estime cette fois ci très mal $P(L \geq k)$ pour des valeurs plus petites de k . La qualité de nos intervalles de confiances pour certaines de ces valeurs est inférieure à celle obtenue par méthode de Monte Carlo naïve.

Deuxième tentative : Nous souhaitons garder la qualité manifestement supérieure de notre estimateur de MC naïve pour des petites valeurs de k . Nous avons jusque-là pris des shifts de moyenne dans l'espace $\text{Vect}(u)$ où $u = (1, \dots, 1)^\top$. On essaie donc dans cette seconde approche

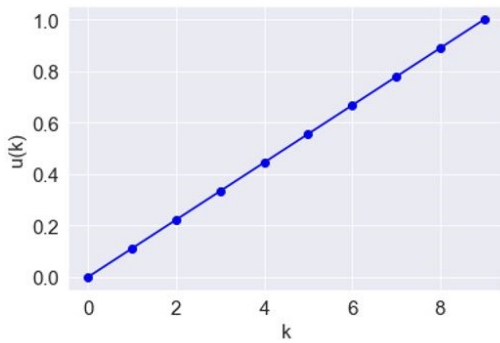
de prendre des vecteurs u de profils différents et qui permettraient de créer un compromis en variant différemment les moyennes des différentes composantes du vecteur, notamment en cherchant à garder certaines moyennes de composantes proches de leurs valeurs initiale ($= 0$). On propose par exemple :

$$u = u_1 = \left(0, \frac{1}{N-1}, \dots, \frac{j-1}{N-1}, \dots, 1\right)^\top$$

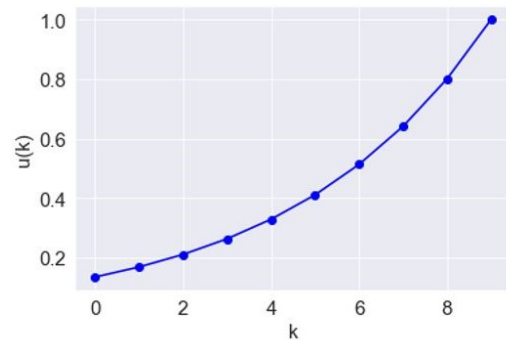
ou encore :

$$u = u_2 = \left(\exp\left(-\alpha \frac{N-1}{N}\right), \dots, \exp\left(-\alpha \frac{N-j}{N}\right), \dots, 1\right)^\top$$

où α est un nouveau paramètre que l'on introduit, que l'on peut modifier mais que l'on choisira pour simplifier égal à $\alpha = 2$.



(a) Profil de u_1



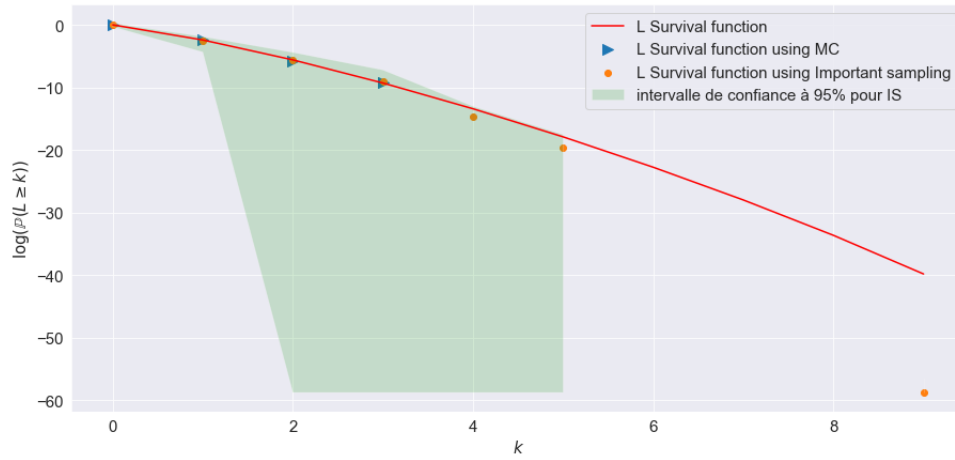
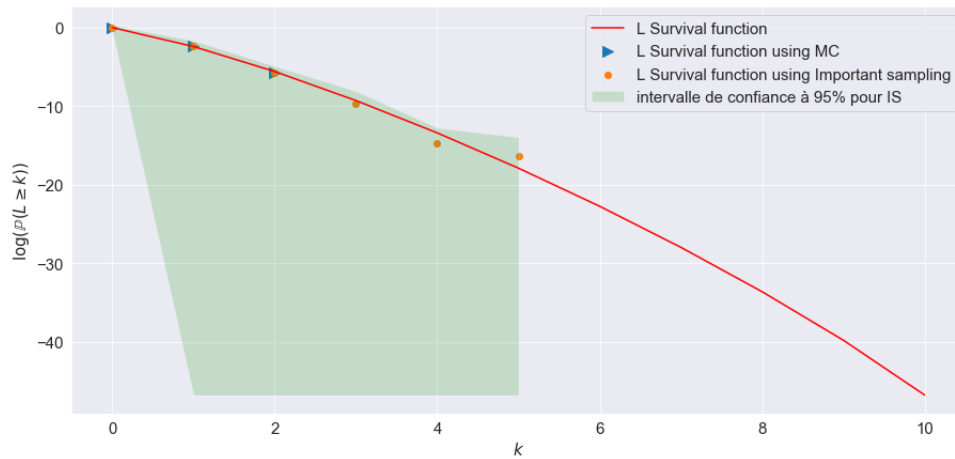
(b) Profil de u_2

FIGURE 6 : Différents profils de u

Pour ces nouveaux vecteurs de la forme $u = (u(1), \dots, u(N))$ et pour $\Lambda = \lambda u$, nous obtenons l'identité suivante :

$$\mathbb{P}_{\mathcal{N}(0_n, I_n)}(L \geq k) = \mathbb{E}_{\mathcal{N}(\Lambda, I_n)} \left(\exp \left(\sum_{i=1}^N u(i) W_i \lambda + \frac{1}{2} \sum_{i=1}^N u(i)^2 \right) \mathbb{1}_{L \geq k} \right)$$

Les figures 8 et 7 ci-dessous illustrent nos résultats de simulations en utilisant respectivement le vecteur u_1 et le vecteur u_2 :

FIGURE 7 : $\log(\mathbb{P}(L \geq k))$ en fonction de k pour $N = 10$, $u = u_2$ et $M = 10^5$ FIGURE 8 : $\log(\mathbb{P}(L \geq k))$ en fonction de k pour $N = 10$, $u = u_1$ et $M = 10^5$

Le tableau 3 résume les résultats de l'échantillonnage d'importance pour les différents profil u que nous avons utilisés :

k	2	4	6
p_k^{MC}	0.0038	-	-
$I_c^{95\%}$	[0.0034 , 0.0043]	-	-
$p_k^{IS}(u_0)$	0.003	7.10^{-7}	1.10^{-9}
$I_c^{95\%}$	[0.0 , 0.007]	[0.0 , 2.10^{-6}]	[0.0, 1.10^{-8}]
$p_k^{IS}(u_1)$	0.032	3.10^{-7}	8.10^{-10}
$I_c^{95\%}$	[0.0 , 0.007]	[0.0 , 3.10^{-6}]	[0.0, 8.10^{-9}]
$p_k^{IS}(u_2)$	0.320	4.10^{-7}	-
$I_c^{95\%}$	[0.31 , 0.33]	[0.0 , 2.10^{-6}]	-
$\mathbb{P}(L \geq k)$	0.0036	$1, 5.10^{-6}$	$1, 3.10^{-10}$

TABLE 3 : Résultats d'échantillonnage d'importance vectoriel pour $N = 10$ et $M = 10^5$ réalisations

Il reste cependant très difficile d'avoir des résultats satisfaisants même pour des dimensions aussi petites que $N = 10$. Cela peut s'expliquer en partie par le fait que notre choix de λ est arbitraire et ne suit que l'intuition de devoir être de plus en plus négatif pour des k plus grand. La prochaine partie illustre comment nous avons voulu de, manière adaptative, trouver le meilleur choix de λ (qui minimise la variance de notre estimateur) sur les espaces $\text{Vect}(u)$ en s'inspirant de la méthode de Lelong et Jourdan explicitée ci dessous.

Vers une méthode adaptative :

Pour un certain vecteur u pris comme ci-dessus et pour une valeur de k fixée, la variance de notre estimateur est égale à :

$$\text{Var}(\lambda) = N^{-1} \left(V(\lambda) - (\mathbb{E}[\mathbf{1}_{L \geq k}])^2 \right)$$

avec :

$$V(\lambda) = \mathbb{E} \left[\mathbf{1}_{L \geq k} e^{-\langle \lambda u, X \rangle + \|\lambda u\|^2/2} \right]$$

On cherche le paramètre λ qui minimise $V(\lambda)$. $V(\lambda)$ est généralement inconnu et incalculable de manière exacte en pratique. Cette fonction est de classe \mathcal{C}^2 et convexe, et afin d'approcher le paramètre optimal sur l'espace $\text{Vect}(u)$, on utilise une version bruitée de la méthode de Newton avec projection. On définit donc P_u l'opérateur de projection orthogonale sur l'espace $\text{Vect}(u)$ et $(\lambda_n)_{n \geq 1}$ une suite de scalaire vérifiant :

$$\lambda_{k+1} = P_u \left(\lambda_k - (H_M(\lambda_k))^{-1} G_M(\lambda_k) \right) \quad (1)$$

où $H_M(\lambda)$ et $G_M(\lambda)$ sont respectivement les approximations Monte Carlo de $H(\lambda)$, la hessienne de $V(\lambda)$ et de $G(\lambda)$, sa dérivée. Ces approximations sont calculées à partir des mêmes M

réalisations indépendantes X_1, \dots, X_M des variables aléatoires suivant la loi $\mathcal{N}_N(0, I_N)$:

$$G_M(\lambda) = \frac{1}{M} \sum_{i=1}^M \mathbb{1}_{L \geq k}(X_i) (\lambda - X_i) e^{-\lambda u \cdot X_k + \|\lambda u\|^2/2},$$

$$H_M(\theta) = \frac{1}{M} \sum_{i=1}^M \mathbb{1}_{L \geq k}(X_i) (I_N + (\theta - X_k)(\theta - X_k)') e^{-\lambda u \cdot X_k + \|\lambda u\|^2/2}.$$

En pratique, on se fixe un seuil $\varepsilon > 0$ et on itère la relation 1 tant que $\|G_M(\lambda)\| > \varepsilon$.

Malheureusement, cette méthode s'est avérée très instable et ce, même pour de petites valeurs de N (5, 10 par exemple), la fonction n'arrivait à calculer les probabilités au-delà de $k = 2$. (Voir fonction **LelongJourdan_k**).

• EXPONENTIAL TILTING

Dans cette partie, nous supposons que le paramètre p peut être exactement déterminé ou calculé et nous pouvons ainsi étudier notre problème sous le prisme des copules (gaussiennes).

Sous ces conditions, on suppose que chaque entreprise dispose de son propre seuil de défaut B_i et de sa propre volatilité σ_i .

Chaque entreprise a donc dans ce cadre une probabilité $p_i = \mathbb{P}(Z \leq M_{(B_i, \sigma_i)})$ que l'on suppose déterminée et connue (ou du moins calculable exactement par une relation mathématique fixe). L'étude du nombre de défauts peut donc être faite de manière équivalente en étudiant la variable aléatoire L comme somme de lois de Bernoulli indépendantes $L = \sum_{i=1}^N Y_i$ de paramètres $p = (p_1, \dots, p_N)$ sous la loi que l'on note \mathcal{L}_p . On peut donc réaliser à nouveau un changement de probabilité vers une loi \mathcal{L}_q en changeant les paramètres des Bernoullis par $q = (q_1, \dots, q_N)$ pris de sorte à minimiser la variance de l'estimateur observé.

$$\mathbb{P}_p(L \geq k) = \mathbb{E}_q \left[\mathbb{1}_{L \geq k} \prod_{i=1}^N \left(\frac{p_i}{q_i} \right)^{Y_i} \left(\frac{1-p_i}{1-q_i} \right)^{1-Y_i} \right]$$

Ici encore, au lieu de changer arbitrairement les paramètres, on se limite à une classe de changements défini par :

$$q_i(\theta) = \frac{p_i e^\theta}{1 - p_i + p_i e^\theta}$$

On remarque que pour $\theta > 0$, $q_i(\theta) > p_i$. On augmente donc la probabilité de défaut d'une entreprise en prenant des valeurs positives de θ . On observe également que les probabilités de défaut correspondent pour $\theta = 0$.

Ce choix de changement de probabilité connu sous le nom de "Exponential Tilting" ou "Exponential Twisting" nous permet d'avoir une expression assez simple du rapport de vraisemblance. En effet, on montre que :

$$\prod_{i=1}^N \left(\frac{p_i}{q_i} \right)^{Y_i} \left(\frac{1-p_i}{1-q_i} \right)^{1-Y_i} = \exp(-\theta L + \psi(\theta))$$

où

$$\psi(\theta) = \log \mathbb{E}[e^{\theta L}] = \sum_{i=1}^N \log(1 + p_i(e^\theta - 1))$$

Il ne reste donc plus qu'à chercher le bon paramètre θ : celui qui minimise la variance de l'estimateur, dans notre cas la minimisation du moment d'ordre 2 suffit.

Pour simplifier les notations on notera $\mathbb{E}_{q(\theta)}[f(Y_1, \dots, Y_N)] = \mathbb{E}_\theta[f(Y_1, \dots, Y_N)]$, on observe donc que le moment d'ordre 2 est majoré par :

$$\mathbb{E}_\theta[\mathbf{1}_{L \geq k} e^{-2\theta L + 2\psi(\theta)}] \leq e^{-2\theta k + 2\psi(\theta)}$$

On peut donc essayer de minimiser au mieux ce moment d'ordre 2 en minimisant sa borne supérieure et donc le terme $\psi(\theta) - \theta k$. Il s'agit d'une fonction de classe \mathcal{C}^2 fortement convexe sur \mathbb{R} , de plus ψ passe par l'origine. On en déduit donc qu'un choix raisonnable de θ^* pourrait être :

Si $k > \psi'(0)$:

$$\theta_k^* = \{\text{solution unique de } \psi'(\theta) = k\}$$

Si $k \leq \psi'(0)$:

$$\theta_k^* = 0$$

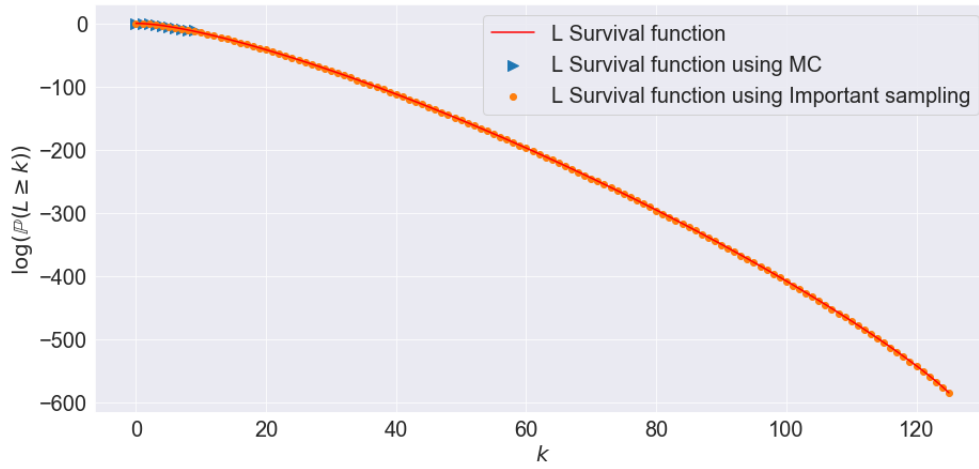
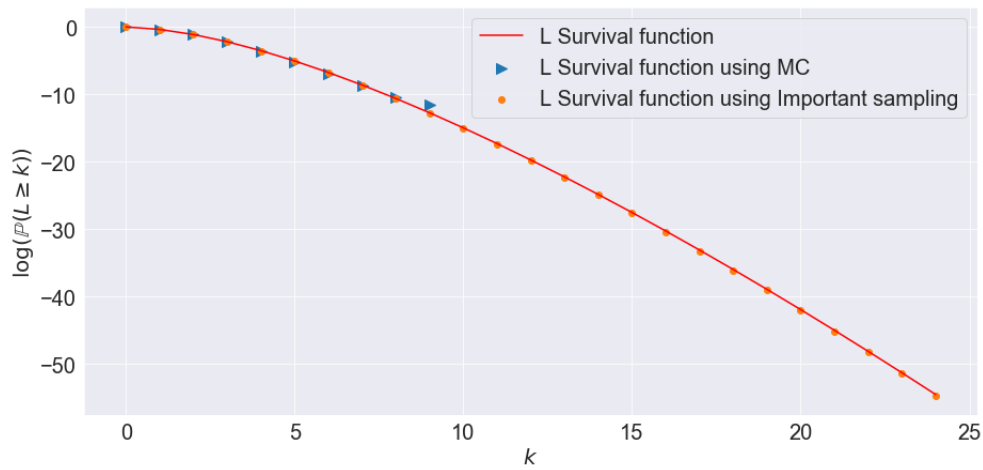
On obtient à partir de ce changement de probabilité la propriété suivante :

$$\mathbb{E}_\theta[L] = \psi'(\theta)$$

ce qui, pour le changement θ_k^* , reviendrait à changer la moyenne de L vers $\max(k, \psi'(0))$ sous la nouvelle loi.

Cas où $\sigma_i = \sigma$ et $B_i = B$

Dans cette configuration précise, les probabilités p_i sont prises toutes égales, on note p leur valeur commune. On peut donc calculer exactement le paramètre $\theta_k^* = \log\left(\frac{k-1}{N-k+1} \frac{1-p}{p}\right)$ On obtient les résultats suivants :

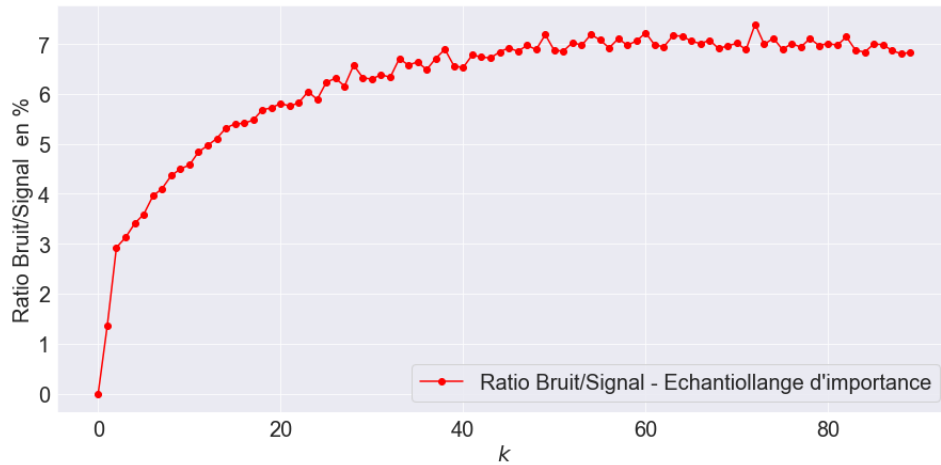
(a) $k \in \{0, \dots, 125\}$ (b) $k \in \{0, \dots, 25\}$ FIGURE 9 : $\log(\mathbb{P}(L \geq k))$ en fonction de k pour $N = 125$ et $M = 10^5$

On a donc une estimation de $\mathbb{P}(L \geq k)$ pour toutes les valeurs de k , on en donne quelques unes avec leurs intervalles de confiances pour référence, ainsi que le ratio bruit/signal pour certaines de ces estimations dans le tableau 4.

k	Estimation de $\mathbb{P}(L \geq k)$	Rayon I_c^{ET}	Rayon I_c^{MC}
3	0.322	0.0011	0.0007
5	0.0055	7.10^{-5}	3.10^{-5}
10	3.10^{-7}	4.10^{-9}	-
25	9.10^{-26}	4.10^{-27}	-
40	$1, 5.10^{-49}$	3.10^{-51}	-
80	1.10^{-131}	1.10^{-133}	-

TABLE 4 : Résultats de l'Exponential Twisting pour $N = 125$ et $M = 10^5$ s

La figure suivante donne une meilleure vision du rapport bruit/signal pour k variant de 0 jusqu'à 90.

FIGURE 10 : Ratio bruit/signal de l'estimation des $\mathbb{P}(L \geq k)$ en fonction de k pour $N = 125$ et $M = 10^5$

Cas où $\sigma_i \in I$

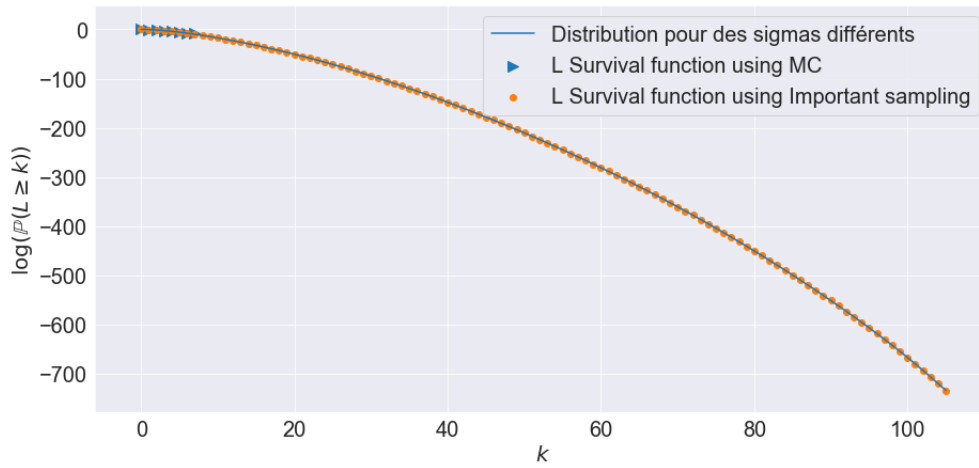
On se place dans le cas où le portefeuille est formé de d'entreprises de volatilités différentes prenant leurs valeurs dans $I = \{0.2, 0.25, 0.3, 0.35, 0.5\}$. On regroupe les différentes entreprises suivant leur volatilité dans des groupes homogènes. On note la probabilité de défaut d'une entreprise dans chaque groupe homogène par p_j avec $j \in \{1, 2, 3, 4, 5\}$.

Pour trouver le paramètre θ optimal, on utilise une fonction qui par méthode de Newton donne une valeur approché du point de minimisation. En effet, on construit une suite θ_n qui vérifie :

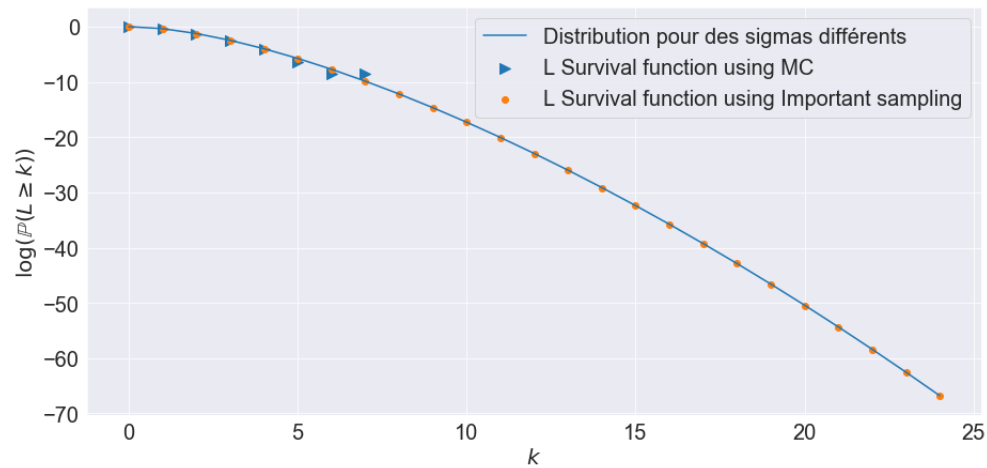
$$\theta_{n+1} = \theta_n - h(\psi''(t))^{-1}\psi'(t)$$

où h est un pas de descente que l'on prend assez petit pour obtenir une convergence satisfaisante de l'algorithme.

On obtient alors les résultats suivants :



(a) $k \in \{0, \dots, 125\}$



(b) $k \in \{0, \dots, 25\}$

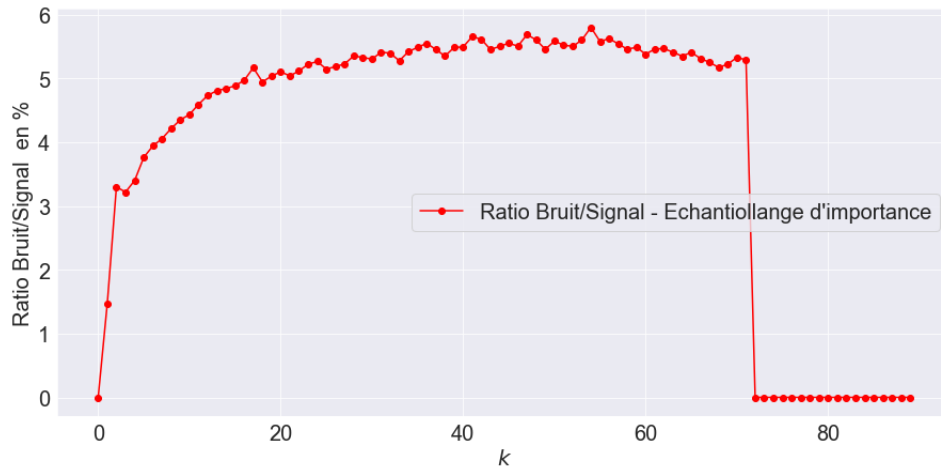
FIGURE 11 : $\log(\mathbb{P}(L \geq k))$ en fonction de k pour $N = 125$ et $M = 10^5$

Des exemples de résultats de calcul avec accompagnés de leurs différents intervalles de confiance sont donnés dans le tableau suivant :

k	Estimation de $\mathbb{P}(L \geq k)$	Rayon I_c^{ET}	Rayon I_c^{MC}
3	0.076	0.0024	0.0052
5	0.003	0.00011	0.001
10	3.10^{-8}	$1, 3.10^{-9}$	-
25	1.10^{-31}	6.10^{-33}	-
40	7.10^{-65}	3.10^{-66}	-

TABLE 5 : Résultats de l'Exponential Twisting pour $N = 125$ et $M = 10^5$

On peut également observer le ratio bruit/signal sur la figure suivante :

FIGURE 12 : Ratio bruit/signal de l'estimation des $\mathbb{P}(L \geq k)$ en fonction de k pour $N = 125$ et $M = 10^5$

2.2 INFLUENCE DU SEUIL B ET DE σ

Après avoir étudié et implémenté divers moyens d'obtenir la distribution de L , il est temps d'évaluer dans ce paragraphe l'influence du seuil de défaut B et de la volatilité σ sur celle-ci. On commence par tracer la distribution de L pour différentes valeurs de B :

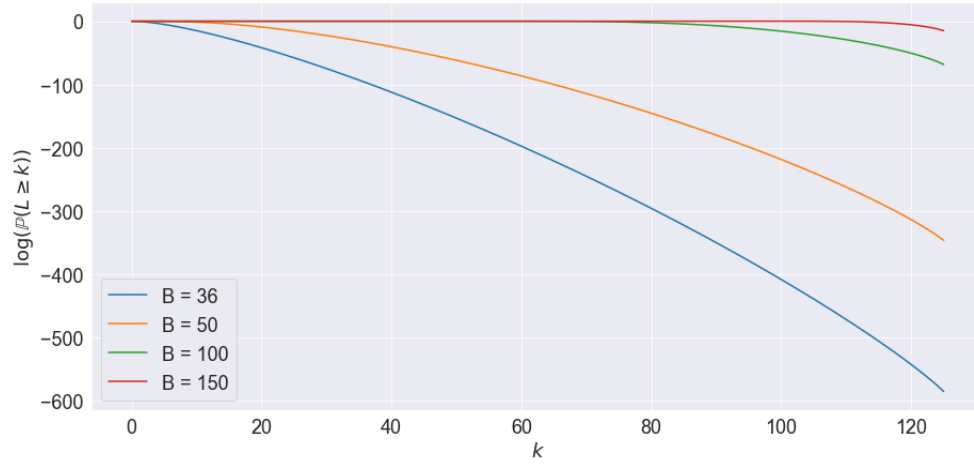


FIGURE 13 : $\log(\mathbb{P}(L \geq k))$ en fonction de k pour $N = 125$ pour différentes valeurs de B - (Pas de simulations)

Comme nous aurions pu conjecturer, plus le seuil de défaut croît, plus les probabilités de voir plus de k défauts à maturité croissent pour presque tout k . Nous avons attesté du comportement monotone des $\mathbb{P}(L \geq k)$ comme fonctions de $\mathbb{P}(Z \leq M_{(B,\sigma)})$ et donc comme fonctions de B et de σ . Or nous savons que :

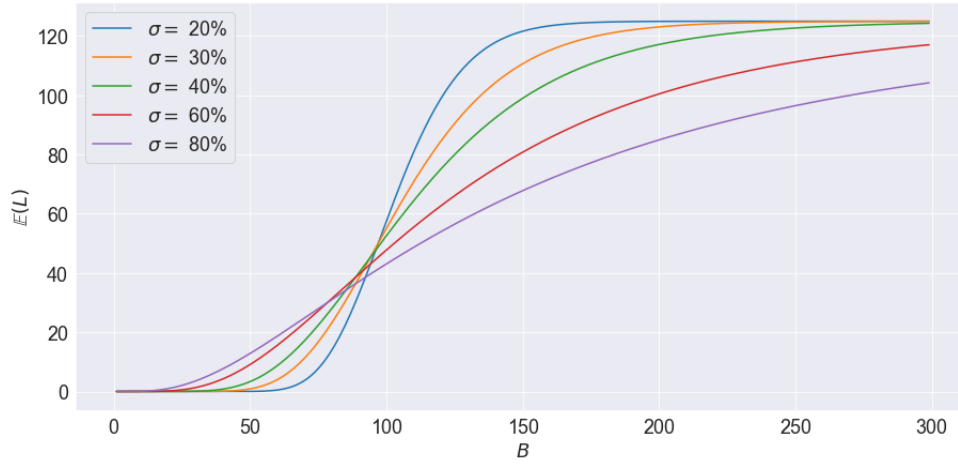
$$\mathbb{P}(Z \leq M_{(B,\sigma)}) = \Phi\left(\frac{1}{\sigma} \log\left(\frac{B}{S_0}\right) + \frac{\sigma}{2}\right)$$

Il s'agit là d'une fonction croissante de B , on en déduit que pour presque tous les k , $\mathbb{P}(L \geq k)$ croît avec B ou du moins pour un domaine de B où $M_{(B,\sigma)}$ reste négatif.

Pour continuer notre analyse qualitative de l'effet du changement du seuil de défaut et de la volatilité sur la distribution. On observe que :

$$\mathbb{E}(L) = Np = N\mathbb{P}(W \leq W_0) = N\Phi(W_0) = N\Phi\left(\frac{1}{\sigma} \log\left(\frac{B}{S_0}\right) + \frac{\sigma}{2}\right)$$

Il s'agit donc d'une fonction strictement croissante en B , le comportement de cette espérance lors d'une variation de σ dépend quant à lui de la valeur de B . On obtient donc numériquement les courbes d'espérances suivantes, qui attestent bien de ces observations :

FIGURE 14 : $\mathbb{E}(L)$ en fonction de B pour $N = 125$ (pas de simulations)

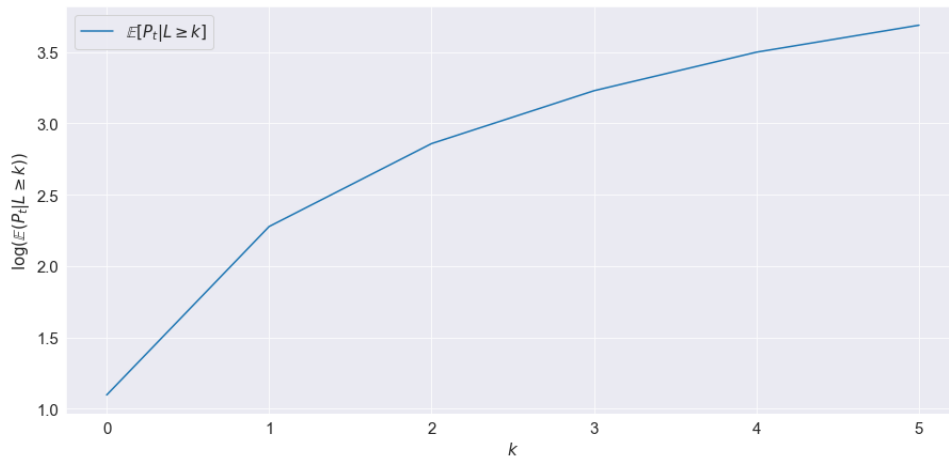
2.3 PERTE ASSOCIÉE AU DÉFAUT

On définit la perte associée au défaut à l'instant T par

$$P_T = \sum_{i \in D_T} R_i S_T^i$$

et on cherche à évaluer la quantité $\mathbb{E}(P_T | L \geq k)$ pour tout k .

Dans un premier temps on utilise une estimation par Monte Carlo naïve pour déterminer les limites de cette méthode :

FIGURE 15 : $\log(\mathbb{E}(P_t | L \geq k))$ en fonction de k pour $N = 125$ et $M = 10^5$

On échoue ici aussi à estimer cette espérance à partir de $k = 6$, cette méthode de Monte Carlo permet cependant d'avoir les valeurs suivantes :

k	2	3	4	5
$E(Pt L \geq k)$	17.43	25.25	33.082	39.957
Rayon de $I_c^{95\%}$	0.024	0.012	0.0043	0.0011

TABLE 6 : Estimation de la perte associé au défaut par méthode de MC naïf pour $N = 125$ et $M = 10^5$ réalisations

On essaie cependant de tirer profit de notre cas particulier et de l'indépendance et nos entreprises pour diminuer la complexité de notre problème d'estimation. On commence par développer l'expression de la perte associé au défaut à l'instant T . On a :

$$\begin{aligned}
f(k) &= \mathbb{E}(P_T | L \geq k) \\
&= \mathbb{E}\left(\sum_{i=1}^N S_T^i \mathbf{1}_{\{S_T^i \leq B\}} | L \geq k\right) \\
&= (\mathbb{P}(L \geq k))^{-1} \sum_{i=1}^N \mathbb{E}\left(S_T^1 \mathbf{1}_{\{S_T^1 \leq B\}} \mathbf{1}_{\{L \geq k\}}\right) \\
&= (\mathbb{P}(L \geq k))^{-1} \sum_{i=1}^N \mathbb{E}\left(S_T^1 \mathbf{1}_{\{S_T^1 \leq B\}} \mathbf{1}_{\{\sum_{j=2}^N \mathbf{1}_{\{S_T^j \leq B\}} \geq k\}}\right) \\
&= N \frac{\mathbb{P}(\sum_{j=2}^N \mathbf{1}_{\{S_T^j \leq B\}} \geq k)}{\mathbb{P}(L \geq k)} \mathbb{E}\left(S_T^1 \mathbf{1}_{\{S_T^1 \leq B\}}\right)
\end{aligned}$$

et nous pouvons observer que :

$$\mathbb{E}\left(S_T^1 \mathbf{1}_{\{S_T^1 \leq B\}}\right) = S_0 \mathbb{P}(Z \leq M_{(B,\sigma)} - \sigma)$$

et donc $\forall 1 \leq k \leq N$:

$$f(k) = N S_0 \frac{\mathbb{P}(\sum_{j=2}^N \mathbf{1}_{\{S_T^j \leq B\}} \geq k)}{\mathbb{P}(L \geq k)} \mathbb{P}(Z \leq M_{(B,\sigma)} - \sigma)$$

Il nous faut donc pour cela estimer les quantités $\mathbb{P}(\sum_{j=2}^N \mathbf{1}_{\{S_T^j \leq B\}} \geq k)$, $\mathbb{P}(L \geq k)$ et $\mathbb{P}(Z \leq M_{(B,\sigma)} - \sigma)$.

Les deux premières quantités étant déjà étudiées plus haut, il s'agit de l'estimation de la distribution du nombre de défaut pour deux portefeuilles, le premier contenant 124 actifs homogènes, et l'autre en contenant 125. On peut donc ramener notre problème à l'estimation de deux probabilités :

$$\mathbb{P}(Z \leq M_{(B,\sigma)} - \sigma) \text{ et } \mathbb{P}(Z \leq M_{(B,\sigma)})$$

Un calcul exact de la perte pour tout k et donc possible. On le compare aux premiers points estimés par méthode de Monte Carlo :

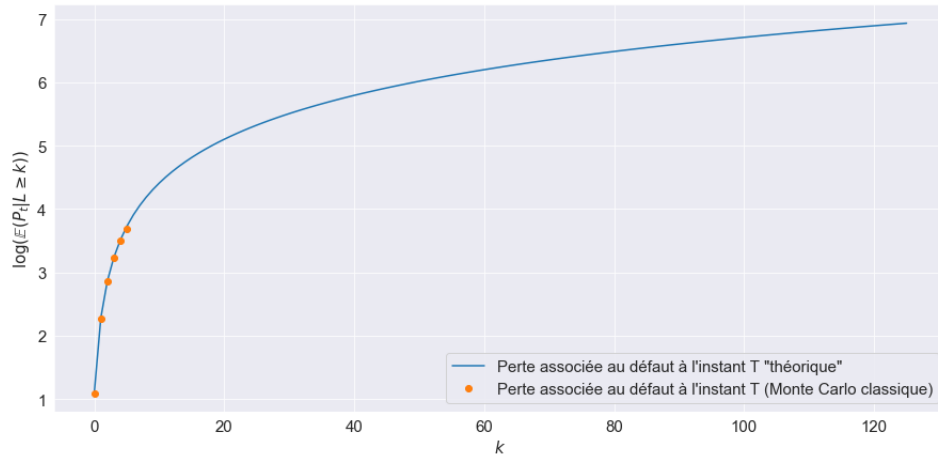


FIGURE 16 : Évolution de $\mathbb{E}(P_T | L \geq k)$ en fonction de k - $N = 125$ et $M = 10^5$

2.4 COMMENTAIRES

L'estimation de la distribution et de la perte des portefeuilles de manière directe s'est avérée très délicate et difficile à réaliser de manière satisfaisante. Heureusement, certaines hypothèses de ce modèle permettent de diminuer la complexité du problème et le ramener à des cas plus simples à analyser.

Nous avons ainsi pu estimer la distribution du nombre de défaut à maturité pour deux portefeuilles différents. Un avec une seule classe d'actifs de même volatilité et un second avec cinq classes différentes qui se distinguent des autres par la volatilité des valeurs de leurs entreprises :

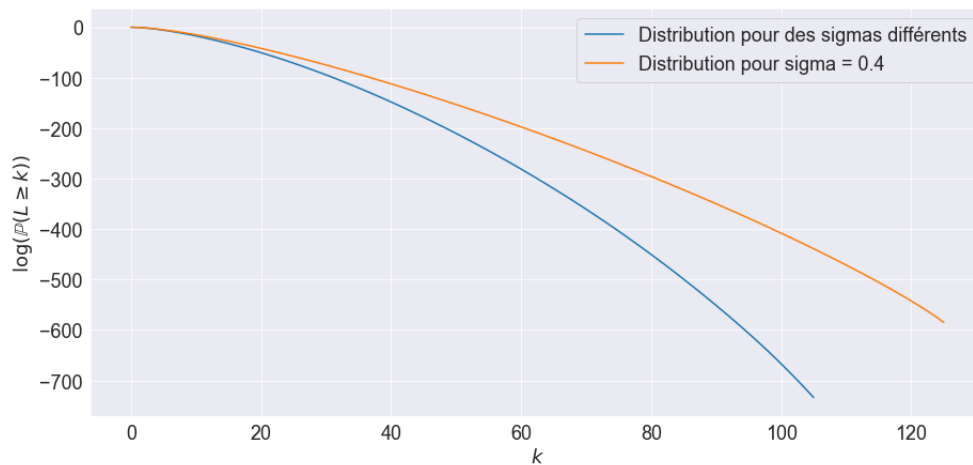


FIGURE 17 : Evolution de $\log(\mathbb{P}(L \geq k))$ en fonction de k

On observe que le second portefeuille est plus "sûr". Bien qu'il contienne une classe d'actifs plus volatiles et donc plus risqués que le premier, la diversification de celui là fait que la sûreté relative des autres classes ait compensé les risques encourues par la possession de cette classe d'actifs hautement volatils. Ceci atteste bien de la nécessité de la diversification, d'abord car cela augmente les chances d'avoir un portefeuille avec certains groupes d'actifs pouvant être considérés indépendants, mais aussi, étant donné qu'il est difficile d'estimer à tout moment les volatilités des actifs, cela pourrait garantir une minimisation du risque encouru lors de l'investissement. La perte associée au défaut à l'instant T permet aussi de voir pourquoi le second portefeuille est plus "sûr".

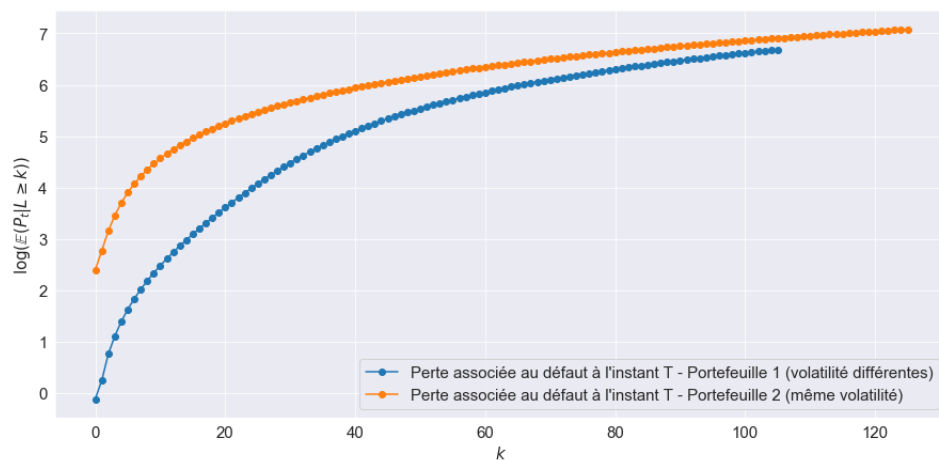


FIGURE 18 : Evolution de $\log(\mathbb{P}(L \geq k))$ en fonction de k (calcul)

Nous nous intéresserons dans la prochaine partie à des portefeuilles d'actifs dépendants et positivement corrélés. Nous pouvons dès lors émettre l'hypothèse que sous ce nouveau modèle, la distribution de la variable L favorisera en terme de probabilité les situations où un nombre plus importants d'entreprises tombent en défaut au détriment des événements où un nombre très restreints d'entre elles sont en faillite.

3

CAS DE DÉBITEURS DÉPENDANTS

3.1 CALCUL DES PROBABILITÉS $P(L = j)$

On suppose dans cette partie que les mouvements browniens modélisant les entreprises sont corrélés positivement. Une simplification nous conduit à considérer le vecteur aléatoire $W_t = (W_t^1, \dots, W_t^N)$ dont les composantes suivent une loi gaussienne, centrée réduite. On s'intéresse au nombre d'entreprises en défaut à l'horizon $t = 1$. Nous adoptons la notation simplifiée $W_1 = (W^1, \dots, W^N)$. La matrice de covariance de ce vecteur est ($t=1$) :

$$\Sigma = \begin{pmatrix} 1 & \rho & \dots & \dots & \rho \\ \rho & 1 & \rho & \ddots & \rho \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho & \dots & \rho & 1 & \rho \\ \rho & \dots & \dots & \rho & 1 \end{pmatrix}$$

Nous adoptons, dans cette partie, les notations suivantes pour tout $j \in [1, N]$:

- A_j , matrice diagonale de taille $N \times N$ telle que $(A_j)_{i,i} = \mathbb{1}_{i \leq j} - \mathbb{1}_{i > j}$ pour tout $i \in [1, N]$.
- $M_{(B,\sigma)}^{(j)}$, vecteur colonne de taille N tel que $(M_{(B,\sigma)}^{(j)})_i = M_{(B,\sigma)}(\mathbb{1}_{i \leq j} - \mathbb{1}_{i > j})$ pour tout $i \in [1, N]$.
- Z_j , le vecteur $A_j W_t$ qui est donc également un vecteur gaussien centré de matrice de covariance égale à $A_j \Sigma A_j^T$.

L'entreprise k est en défaut à $t = 1$ si et seulement si, d'après les remarques que nous avons émises dans la partie précédente, $W^k \leq M_{(B,\sigma)}$. On a donc $\forall j \in [1, N]$:

$$\begin{aligned} \mathbb{P}(L = j) &= \mathbb{P} \left(\bigcup_{\{i_1, \dots, i_N\} \in \mathfrak{S}_N} W^{i_1} \leq M_{(B,\sigma)}, \dots, W^{i_j} \leq M_{(B,\sigma)}, W^{i_{j+1}} > M_{(B,\sigma)}, \dots, W^{i_N} > M_{(B,\sigma)} \right) \\ &= \mathbb{P} \left(\bigcup_{\{i_1, \dots, i_N\} \in \mathfrak{S}_N} W^{i_1} \leq M_{(B,\sigma)}, \dots, W^{i_j} \leq M_{(B,\sigma)}, -W^{i_{j+1}} \leq -M_{(B,\sigma)}, \dots, W^{i_N} \leq -M_{(B,\sigma)} \right) \\ &= \mathbb{P}(A_j W_1 \leq M_{(B,\sigma)}^{(j)}) \binom{N}{j} \\ &= \mathbb{P}(Z_j \leq M_{(B,\sigma)}^{(j)}) \binom{N}{j} \end{aligned}$$

où \mathfrak{S}_N est le groupe des permutations de $[1, N]$.

3.2 ESTIMATION PAR ÉCHANTILLONNAGE D'IMPORTANCE

On procède au changement du vecteur de moyennes de W_1 (qui est donc le vecteur nul) par le vecteur $M_{(B,\sigma)}^N = (M_{(B,\sigma)}, \dots, M_{(B,\sigma)})$. On a ainsi :

$$\mathbb{P}_{\mathcal{N}(0_n, \Sigma)}(L \geq k) = \mathbb{E}_{\mathcal{N}(M_{(B,\sigma)}^N, \Sigma)} \left(\exp \left(\frac{1}{2} M_{(B,\sigma)}^N \Sigma^{-1} (M_{(B,\sigma)}^N - W_T) - \frac{1}{2} W_T \Sigma^{-1} W_0^N \right) \mathbb{1}_{L \geq k} \right)$$

A des fins de comparaison, nous avons également procédé au calcul des probabilité $P(L \geq k)$ par méthode de Monte Carlo naïf. Les résultats obtenus numériquement par les deux méthodes pour $\rho = 0.5$, $B = 36$, $\sigma = 0.4$, sont représentés sur la figure suivante :

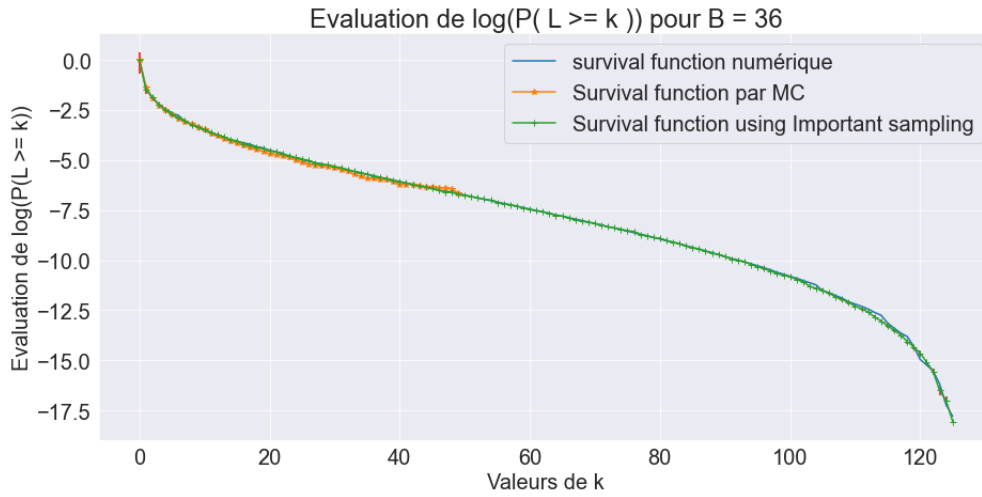


FIGURE 19 : Evolution de $\log(\mathbb{P}(L \geq k))$ - $N = 125$ et $M = 10^5$

On remarque directement que les probabilités de défauts sont beaucoup plus élevées que dans le cas où les variables aléatoires modélisant les valeurs des entreprises étaient indépendantes. La corrélation prise entre les entreprises étant assez élevée, avec $\rho)0.5$, ce résultat semble réaliste. Étant donné qu'il est réaliste de penser qu'une entreprise en défaut pourrait entraîner le défaut d'une autre à laquelle elle est liée.

Le tableau suivant donne quelques résultats relatives à certaines valeurs de k de l'estimation avec échantillonnage d'importance ainsi que les intervalles de confiance des valeurs respectives. On y voit clairement que les probabilités sont beaucoup plus grandes que dans le cas où les entreprises sont supposées indépendantes.

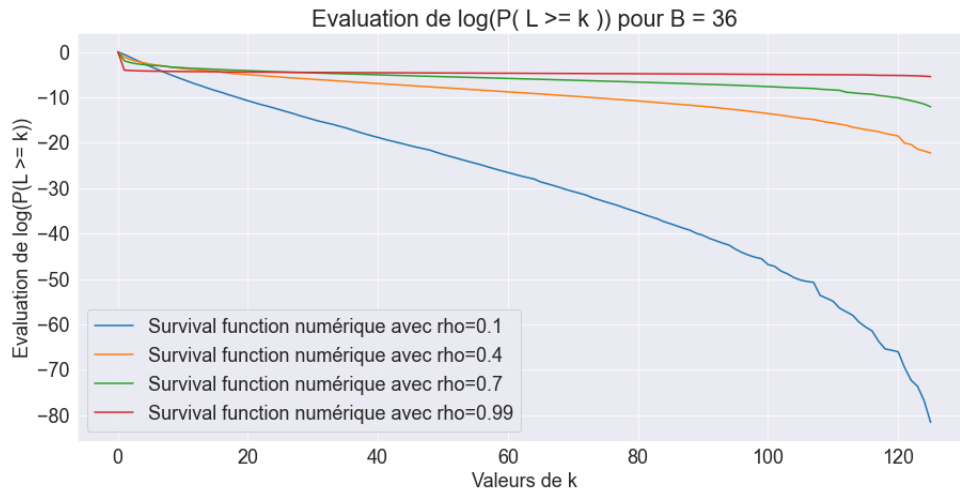
k	Estimation numérique de $\mathbb{P}(L \geq k)$	Estimation IS de $\mathbb{P}(L \geq k)$	Rayon I_c^{IS}	Rayon I_c^{MC}
5	0.0597	0.0665	0.0089	0.0096
10	0.0273	0.0315	0.0032	0.0063
30	0.0048	0.0047	$3.66 \cdot 10^{-4}$	0.0028
90	$5.13 \cdot 10^{-5}$	$5,49 \cdot 10^{-5}$	$5.35 \cdot 10^{-6}$	-

TABLE 7 : Résultats de d'estimation par Importance Sampling pour $N = 125$ et $M = 10^5$ s

• INFLUENCE DU SEUIL B ET DE LA CORRÉLATION ρ

Influence de ρ :

On s'intéresse à l'évolution de $\log(\mathbb{P}(L \geq k))$ sous l'influence de différentes valeurs de ρ . Nous calculons directement les probabilités correspondantes grâce à la relation précédemment établie $P(L = j) = \mathbb{P}(Z_j \leq M_{(B,\sigma)}^{(j)}) \binom{N}{j}$.

FIGURE 20 : Influence de différentes valeurs de ρ sur $\log(\mathbb{P}(L \geq k))$

Les différentes corrélations observées étant positives strictement, on observe bien le fait que plus la corrélation est grande, plus la probabilité d'obtenir des valeurs plus grandes pour le nombre de défaut à l'horizon final est grand. Encore une fois, cela est conforme avec l'intuition que plus la corrélation est grande, plus le défaut d'une unique entreprise entraînera le défaut d'un nombre d'entreprises qui y sont liées. Nous obtenons également des résultats similaires à ceux obtenus par ce calcul direct en utilisant la technique d'Important Sampling.

Influence du seuil de défaut B :

Il est évident que plus le seuil B est grand plus les entreprises ont plus de chance de tomber en défaut. Pour mieux observer cette influence, dans notre cas précis, nous estimons encore une

fois les différentes probabilités sous différents niveaux de seuil B , cette fois par Importance Sampling. La figure suivante représente les résultats que nous avons ainsi obtenu :

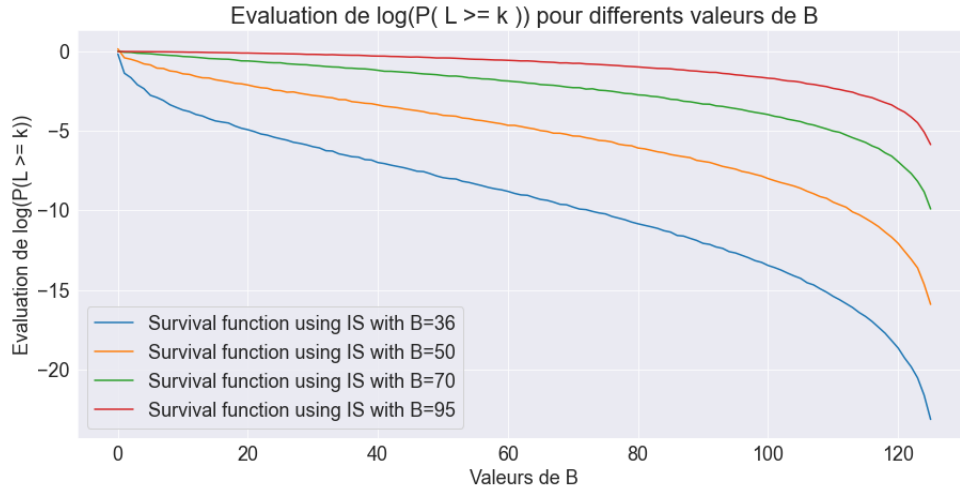


FIGURE 21 : Evolution de $\log(\mathbb{P}(L \geq k))$ pour différentes valeurs de B - $\rho = 0.5$ $N = 125$ et $M = 10^5$

On observe clairement que plus le seuil de défaut fixé est grand, plus les probabilités de défauts d'un nombre k d'entreprises sont grandes. Encore une fois, nous observons des probabilités beaucoup plus élevées que dans le cas indépendant, ceci est encore une fois expliqué par le fait que les événements de défauts des entreprises sont désormais liés. Nous avons également obtenus les mêmes résultats par le calcul direct des probabilités par les relations mathématiques établies précédemment.

• PERTE ASSOCIÉE AU DÉFAUT

Nous nous intéressons maintenant à la perte engendrée par un niveau de défaut L . On a :

$$\mathbb{E}(P_T | L \geq k) = \frac{\mathbb{E}(P_T \cap L \geq k)}{\mathbb{P}(L \geq k)}$$

Les probabilités $\mathbb{P}(L \geq k)$ sont calculées par Importance Sampling tandis que $\mathbb{E}(P_T | L \geq k)$ est obtenu grâce aux relations précédemment établies.

Dans un premier temps, nous utilisons le même changement de vecteur moyen que lors du début de cette partie. Nous conduisons également une estimation par Monte Carlo naïf. Toutefois, cette méthode ne nous a permis d'estimer la perte associée seulement à des valeurs de k très petites. Nous obtenons ainsi les résultats suivants :

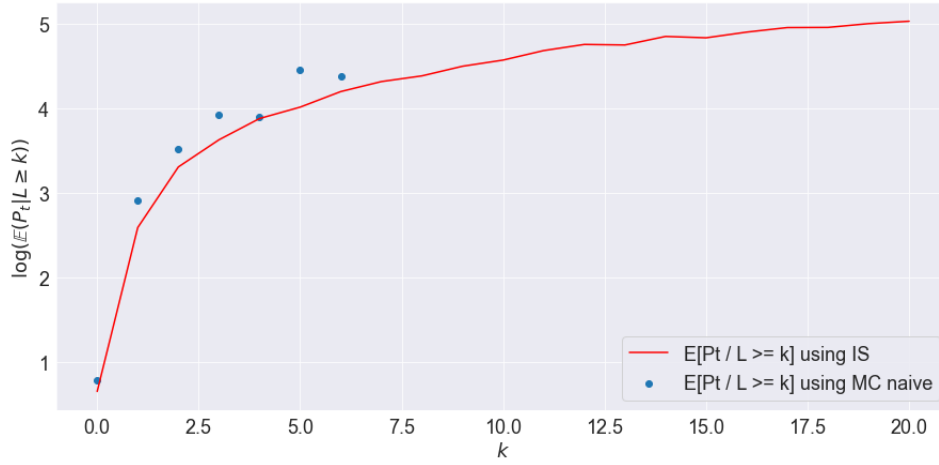


FIGURE 22 : Evolution de $\log(\mathbb{E}(P_T | L \geq k))$ en fonction de k - $\rho = 0.5$, $N = 125$ et $M = 10^5$

On essaie maintenant de voir si on peut faire mieux, donc on a essayé de faire, pour chaque k différent, un changement de moyenne qui minimise la variance asymptotique de l'estimateur, nous avons pu améliorer un peu nos résultats :

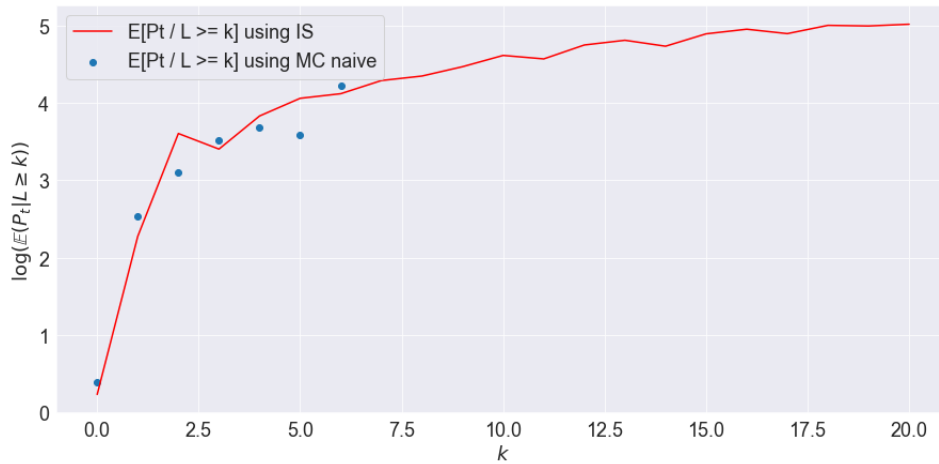


FIGURE 23 : Evolution de $\log(\mathbb{E}(P_T | L \geq k))$ en fonction de k - $\rho = 0.5$, $N = 125$ et $M = 10^5$

Ces résultats sont meilleurs parce qu'ils ont été construits de telle sorte à avoir des intervalles de confiance de largeur minimale.

3.3 EXPONENTIAL TILTING - (COPULE GAUSSIENNE)

Pour cette nouvelle partie, nous commençons par simplifier le problème en modélisant la dépendance entre les entreprises de manière à mieux comprendre le lien entre les différents

actifs.

Le vecteur (W_T^1, \dots, W_T^N) que l'on notera pour simplifier, encore une fois, (W^1, \dots, W^N) suit une loi de normale centrée de matrice de covariance Σ . On peut donc le simuler de manière équivalente à l'aide de Z_0 une variable scalaire de loi normale centrée réduite et d'un vecteur gaussien indépendant (Z_1, \dots, Z_N) de loi $\mathcal{N}(O, \mathbb{I}_N)$. On peut considérer que :

$$\forall i \in [1, N] \quad W_i = \sqrt{\rho} Z_0 + \sqrt{1 - \rho} Z_i$$

conditionnellement à $\{Z = z\}$ les actifs sont à nouveau indépendants et nous pouvons utiliser la méthode d'Exponential Twisting utilisé pour le cas indépendant. On moyennera ensuite sur z pour retrouver l'espérance de la fonction recherchée. On résume donc l'algorithme qui permet de créer une valeur de notre estimateur de la probabilité $\mathbb{P}(L \geq k)$ pour une valeur fixé de $z \in \mathbb{R}$:

1. Générer $Z \sim \mathcal{N}(0, 1)$
2. Calculer les probabilités de défaut conditionnel $p_i(Z)$ pour toute entreprise et si $\sum_{i=1}^N p_i(Z) \geq k$, prendre $\theta_k^*(Z) = 0$. Sinon poser $\theta_k^*(Z)$ comme l'unique solution de :

$$\frac{\partial}{\partial \theta} \psi(\theta, Z) = k$$

avec :

$$\psi(\theta, z) = \sum_{i=1}^N \log(1 + p_k(z)(e^\theta - 1))$$

3. Générer $L = \sum_{i=1}^N Y_i$ avec Y_i une bernoulli de loi :

$$q_i(\theta_k^*(Z), Z) = \frac{p_k(Z) e^{\theta_k^*(Z)}}{1 - p_i(Z) + p_i(Z) e^{\theta_k^*(Z)}}$$

4. Retourner l'estimateur :

$$\mathbb{1}_{L \geq k} \exp(-\theta_k^*(Z)L + \psi(\theta_k^*(Z), Z))$$

En suivant ce procédé, nous estimons la fonction de survie du nombre de défaut à maturité pour $N = 10$ mais avec un ratio bruit/signal assez important pour les plus grandes valeurs de k .

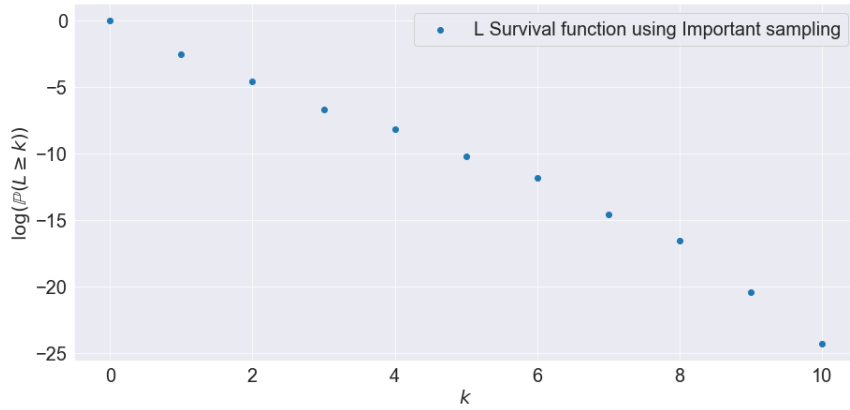


FIGURE 24 : Evolution de $\log(\mathbb{P}(L \geq k))$ en fonction de k - $\rho = 0.2$, $N = 10$ et $M = 10^4$



FIGURE 25 : Evolution de du ratio bruit/signal de l'estimation de la fonction de survie en tout point en fonction de k - $\rho = 0.2$, $N = 10$ et $M = 10^4$

k	4	5	6	10
Estim $P(L \geq k)$	0.0016	3.10^{-5}	2.10^{-6}	9.10^{-12}
$I_c^{95\%}$	[0.0012 , 0.0021]	[9.10^{-6} , 3.10^{-5}]	[1.10^{-6} , 3.10^{-6}]	[1.10^{-12} , 2.10^{-11}]

TABLE 8 : Estimation de la perte associé au défaut par méthode de MC naïf pour $\rho = 0.2$, $N = 10$ et $M = 10^4$ réalisations

Finalement on se propose d'observer l'effet de la corrélation sur la distribution du nombre de défaut. On trace pour cela les distributions pour différentes valeurs de $\rho \in \{0.2, 0.3, 0.5, 0.7\}$:

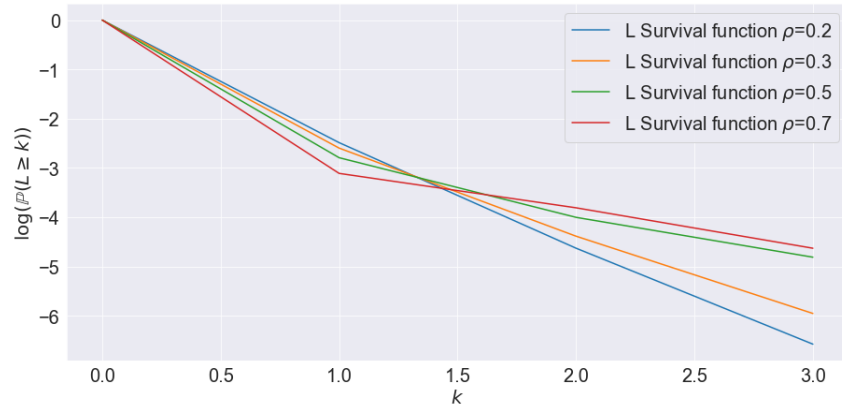


FIGURE 26 : Evolution de $\log(\mathbb{E}(P_T | L \geq k))$ en fonction de k pour différentes valeurs de ρ pour $M = 10^4$ et $N = 10$

3.4 COMMENTAIRES ET CONCLUSIONS

Comme nous l'avions intuité, plus la corrélation est importante moins la probabilité de voir un nombre petit est importante. On pourrait l'apparenter à un effet de groupe, où plus la corrélation est importante plus les scénarios favorisés sont un défaut collectif (ou en grand nombre des entreprises) soit un aucune voir très peu.

4

DÉFAUT D'ENTREPRISE POUR $\tau \leq T$

4.1 CE QUI CHANGE

Dans cette dernière partie, on s'intéresse au défaut des entreprises non plus à l'instant final T mais tout au long du processus, et donc sur tout l'intervalle de temps $[0, T]$. On discrétise cet espace de sorte à prendre exactement douze instants $(t_k = \frac{k}{12})_{k \in \{1, \dots, 12\}}$ où seront calculés les valeurs des entreprises. On considère qu'une entreprise est en défaut si sa valeur descend au dessous de son seuil B_i à un instant $t_i \leq T$.

On pose alors :

$$\tau_i = \inf\{t_k \text{ tq } S_{t_k}^i \geq B_i\}$$

$$D'_T = \{i \in \{1, \dots, N\} \text{ tq } \tau_i \geq T\}$$

et :

$$L' = D'_T$$

De manière équivalente, on peut se restreindre à l'étude du processus $(W_{t_k}^1, \dots, W_{t_k}^N)_{k \leq 0}$ qui est une chaîne de Markov de loi initiale $\delta_{(0, \dots, 0)}$. Le noyau de transition de la chaîne est un noyau gaussien qui dépend des hypothèses faites sur le modèle (indépendants, dépendants positivement comme à la partie précédente ...).

Notre but était dans cette partie d'arriver à estimer l'espérance de fonctions mesurables de cette chaîne de Markov après 12 pas. Nous avons commencé pour cela à explorer la possibilité d'utiliser une méthode de système de particule en interactions.

4.2 INTERACTING PARTICLE SYSTEM

• PRINCIPE

La méthode IPS est une méthode de simulation d'événements rares qui consistent à faire évoluer un nombre M de processus (copies identiques et indépendants et qui dans notre cas sont des chaînes de Markov) et de les faire passer un processus de mutation/sélection à chaque étape d'évolution de la chaîne. La probabilité de survie d'une de ces copies est pondérée par des fonctions G_p qui dépendent notamment de l'étape d'évolution du processus. Ses fonctions de potentiels sont pris de telle sorte à favoriser l'occurrence d'un événement que l'on désire voir se produire à une étape ou à la fin du processus.

• DESCRIPTION DE LA MÉTHODE IPS

En notant $X_{0:n} = (X_0, \dots, X_n)$, supposons que l'on cherche à estimer

$$\mathbb{E}[f(X_{0:n})]$$

pour une fonction f donnée. La méthode de Monte-Carlo standard (ou naïve) consiste à simuler M copies indépendantes de la trajectoire $X_{0:n}$ notées $X_{0:n}^{(1)}, \dots, X_{0:n}^{(M)}$ et à faire la moyenne empirique des valeurs de $f(X_{0:n})$ obtenues

$$\mathbb{E}[f(X_{0:n})] \approx \frac{1}{M} \sum_{m=1}^M f(X_{0:n}^{(m)}).$$

Dans la méthode IPS, on va aussi simuler M trajectoires et faire une moyenne empirique, mais les trajectoires simulées ne seront plus indépendantes.

1) **(Initialisation)** On simule M v.a. indépendantes de même loi que X_0 , notées :

$$X_0^{(0,1)}, \dots, X_0^{(0,M)}$$

2) **(Sélection, itération 1)** On va faire, parmi ces M simulations, une sélection en associant à chaque tirage un poids proportionnel à $G_0(X_0^{(0,m)})$ et en tirant avec remise M points en tenant compte du poids de chacun.

La fonction potentiel G_0 est choisie par l'utilisateur de façon à biaiser les tirages dans une direction pertinente pour le problème à résoudre. A l'issue de cette sélection, nous disposons de M nouvelles valeurs initiales, notées $X_0^{(1,m)}$, $1 \leq m \leq M$.

3) **(Mutation)** Pour chacune de ces valeurs $X_0^{(1,m)}$, et de façon indépendante, on simule $X_1^{(1,m)}$ selon la dynamique de la chaîne de Markov définie par le noyau Q . On dispose donc de M paires $(X_0^{(1,m)}, X_1^{(1,m)})$.

4) **(Sélection, itération 2)** Comme lors de la première sélection, nous n'allons pas forcément conserver toutes ces paires, mais plutôt faire M tirages avec remise, tenant compte d'une pondération associée à chaque paire. Cette pondération nécessite de choisir une fonction potentiel

$$(x_0, x_1) \mapsto G_1(x_0, x_1)$$

.

A l'issue de cette étape de sélection, nous disposons de M nouvelles paires notées $(X_0^{(2,m)}, X_1^{(2,m)})$.

5) **(Mutation)** Pour chacun de ces vecteurs, et de façon indépendante, on simule $X_2^{(2,m)}$ selon la transition de la chaîne de Markov définie par le noyau Q . On obtient ainsi M triplets $(X_0^{(2,m)}, X_1^{(2,m)}, X_2^{(2,m)})$, $1 \leq m \leq M$.

6) On recommence ...

7) (**Sélection finale, itération n**) A l'issue de cette étape de sélection, basée sur la pondération donnée par $(x_0, \dots, x_{n-1}) \mapsto G_{n-1}(x_0, \dots, x_{n-1})$, on dispose de M vecteurs de longueur n , et notés $(X_0^{(n,m)}, \dots, X_{n-1}^{(n,m)})$ pour $1 \leq m \leq M$.

8) (**Mutation finale**) Pour chacun et de façon indépendante, on simule la valeur suivante de la chaîne de Markov selon la dynamique imposée par le noyau Q . On obtient M vecteurs de longueur $n+1$, ou trajectoires, notés $X_{0:n}^{(n,m)} = (X_0^{(n,m)}, \dots, X_n^{(n,m)})$ pour $1 \leq m \leq M$.

Le prix à payer pour cette torsion de l'aléa est l'incorporation, dans la moyenne empirique, d'un ratio de vraisemblance :

$$\mathbb{E}[f(X_{0:n})] \approx \left(\frac{1}{M} \sum_{m=1}^M \frac{f(X_{0:n}^{(n,m)})}{\prod_{p=0}^{n-1} G_p(X_{0:p}^{(n,m)})} \right) \times \left(\prod_{p=0}^{n-1} \frac{1}{M} \sum_{m=1}^M G_p(X_{0:p}^{(p,m)}) \right)$$

Proposition de potentiels L'application de cette méthode nécessite donc principalement de faire le bon choix de potentiels pour favoriser l'évolution des processus dans la direction de faillite d'un nombre d'entreprise. Pour cela nous avons fait le choix de quelques formules pour le potentiels G_p . Etant donné une trajectoire à l'étape p $(W_0^{(p,m)}, \dots, W_p^{(p,m)})$ on définit les potentiels suivants :

Favoriser la décroissance de la moyenne

$$G_p(W_0^{(p,m)}, \dots, W_p^{(p,m)}) = \exp(\alpha_1 (M(W_p^{(p,m)}) - M(W_{p-1}^{(p,m)})))$$

où α_1 est un paramètre négatif à choisir et $M(W)$ la moyenne de composante de W . **Favoriser les trajectoire de moyenne basse**

$$G_p(W_0^{(p,m)}, \dots, W_p^{(p,m)}) = \exp(\beta_1 (M(W_p^{(p,m)})))$$

où β_1 est un paramètre négatif à choisir. **Favoriser la décroissance du maximum**

$$G_p(W_0^{(p,m)}, \dots, W_p^{(p,m)}) = \exp(\alpha_2 (Max(W_p^{(p,m)}) - Max(W_{p-1}^{(p,m)})))$$

où α_2 est un paramètre négatif à choisir et $Max(W)$ la composante maximale du vecteur W .

Favoriser les trajectoire de maximum bas

$$G_p(W_0^{(p,m)}, \dots, W_p^{(p,m)}) = \exp(\beta_2 (Max(W_p^{(p,m)})))$$

où β_2 est un paramètre négatif à choisir.

4.3 RÉSULTATS

Nous avons réussi à implémenter la méthode IPS pour estimer les probabilités $P(L \geq k)$ pour certaines valeurs de k avec $n \leq 12$. (**Voir notebook jupyter CreditDefault.ipynb**)