

Rapport de projet

Microstructure des prix financiers

Binôme :

Peng-Wei CHEN & Diane BERANGER

18 avril 2020

Table des matières

1	Une première modélisation simplifiée	3
1.1	Estimation de la probabilité d'un prix négatif	4
1.1.1	Méthode de Monte-Carlo naïve	4
1.1.2	Méthode d'échantillonnage d'importance	5
1.1.3	Méthode de Splitting/MCMC	6
1.1.4	Résultats :	7
1.2	Estimation de quantiles du prix final	7
2	Prise en compte d'autres propriétés des données	9
	haute fréquence	9
2.1	Estimation de la probabilité d'un prix négatif	10
2.1.1	Méthode de Monte-Carlo naïve	10
2.1.2	Méthode d'échantillonnage d'importance	10
2.1.3	Résultats :	10
2.2	Estimation de quantiles du prix final	11
3	Une autre modélisation du phénomène de retour vers la moyenne	12
3.1	Estimation de la probabilité d'un prix négatif	12
3.1.1	Méthode de Monte-Carlo naïve	13
3.1.2	Méthode de sélection-mutation	13
3.1.3	Résultats :	14
3.2	Estimation de quantiles du prix final	15

4	Limite macroscopique	15
4.1	Estimation de la probabilité d'un prix négatif	16
4.1.1	Par méthode de Monte-Carlo sur le prix macroscopique discrétisé	16
4.1.2	Par méthode des particules en interaction sur le prix macroscopique discrétisé	18
4.2	Estimation de quantiles	19

Introduction

Lorsque l'on modélise les variations du prix *mid* P_t d'un actif sur des durées courtes (quelques heures, par exemple), il est nécessaire de prendre en compte la nature discrète des prix. En effet, les prix d'achat *ask* ou de vente *bid* ne varient que par incréments dont l'amplitude est multiple de la *tick size* (un *cent* sur la plupart des marchés actions, par exemple). Le meilleur *bid* et le meilleur *ask* (et donc également le prix *mid*, défini comme la moyenne de ces derniers) varient donc de manière discrète. On va ainsi modéliser le prix P_t comme un processus à temps continu et à valeurs discrètes :

$$P_t = P_0 + \sum_{n=0}^{N_t} J_n$$

où $(J_n)_{n \geq 1}$ est la suite des incréments des prix à valeurs dans l'ensemble $E = \{+1, -1, \dots, +m, -m\}$, $(T_n)_{n \geq 1}$ la suite des instants des sauts, et N_t le nombre d'incrément dans $[0, t]$:

$$N_t = \sum_{i=1}^{+\infty} \mathbb{1}_{T_i \leq t}$$

Le principal enjeu de ce projet consistera à estimer la probabilité q que le prix devienne négatif sur l'intervalle considéré.

1 Une première modélisation simplifiée

Dans un premier temps, on suppose les incréments indépendants identiquement distribués, et indépendants des instants des sauts. On postule de plus que les variables $S_n = T_{n+1} - T_n$ (représentant les intervalles de temps entre deux sauts consécutifs) sont indépendantes identiquement distribuées suivant $\mathcal{E}(\lambda)$.

Ceci revient à modéliser le prix P_t comme un processus de Poisson composé de paramètres (λ, ν) où ν est la loi des incréments J_n .

Comme valeurs numériques pour cette partie, on prendra :

- $\frac{1}{\lambda} = 300$ s
- $P_0 = 35$
- $T = 4$ h
- Pour les incréments, deux cas :
 - $m=1$: $\mathbb{P}(J_n = 1) = \mathbb{P}(J_n = -1) = \frac{1}{2}$ (J_n est une variable de Rademacher)
 - $m=3$: $\mathbb{P}(J_n = \pm 1) = \frac{1}{4}$, $\mathbb{P}(J_n = \pm 2) = \frac{1}{6}$ et $\mathbb{P}(J_n = \pm 3) = \frac{1}{12}$

Si on simule la trajectoire, on constate sans surprise que le prix s'écarte beaucoup plus de sa valeur initiale P_0 pour $m = 3$ que pour $m = 1$, les amplitudes de saut étant plus élevées.

Processus de Poisson composé d'intensité 0.00333 à horizon fixe T=14400 s

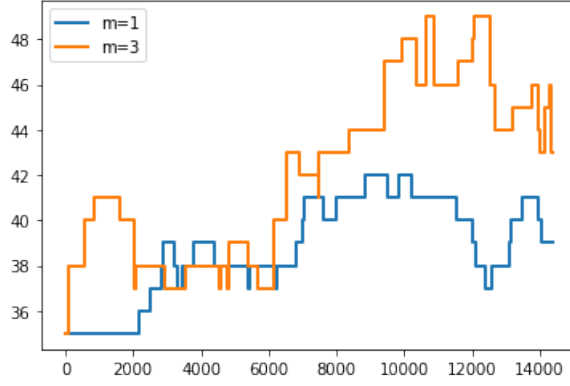


FIGURE 1 – Simulation de la trajectoire $(P_t)_{t \leq T}$

1.1 Estimation de la probabilité d'un prix négatif

Nous avons implémenté trois méthodes différentes pour estimer la probabilité q que le prix devienne négatif pour un $t \in [0, T]$.

1.1.1 Méthode de Monte-Carlo naïve

On simule M réalisations d'un processus de Poisson composé. Le nombre de sauts dans $[0, T]$ est simulé selon une loi de Poisson de paramètre λT . On n'a pas besoin de connaître les instants des sauts pour déterminer si la prix négatif a lieu pour cette réalisation ou non ; on simule donc uniquement les amplitudes des sauts (selon une loi ν_m , où $m = 1$ ou $m = 3$).

Codée de manière naïve, cette méthode n'est pas assez efficace pour permettre d'estimer la probabilité q pour $m = 1$. En revanche, elle fonctionne pour $m = 3$, qui donne une probabilité de prix négatif plus élevée (de l'ordre de 10^{-3} , contre 10^{-7} lorsque $m = 1$).

Nous avons implémenté deux versions de cet estimateur, le premier calculé de manière "intuitive" avec des tableaux numpy, la deuxième accélérée grâce à la librairie Numba, qui traduit le code python en code machine optimisé.

La première version n'est pas assez rapide pour permettre de simuler suffisamment de trajectoires pour estimer la probabilité q avec une précision satisfaisante. En revanche, elle est utile pour vérifier la correction de tous nos autres algorithmes, que l'on test avec des valeurs de P_0 plus faibles (donc des probabilités de prix négatif q plus élevées que notre méthode de Monte-Carlo naïve peut estimer).

Avec la version accélérée, nous obtenons un ordre de grandeur de la probabilité q pour $m = 1$: $q \sim 10^{-7}$. Ceci sera utile pour choisir le nombre de niveaux dans la méthode de Splitting (1.1.3).

Pour $m = 3$, nous obtenons une estimation de la probabilité (voir tableau de résultats en 1.1.4).

Intervalle de confiance :

On a un intervalle de confiance asymptotique à 95% pour q , qui s'obtient par le théorème central limite et le lemme de Slutsky :

$$\left[\hat{q}_M + \frac{1,96 \sqrt{\hat{q}_M(1 - \hat{q}_M)}}{\sqrt{M}}, \hat{q}_M - \frac{1,96 \sqrt{\hat{q}_M(1 - \hat{q}_M)}}{\sqrt{M}} \right]$$

où M est le nombre de trajectoires simulées.

1.1.2 Méthode d'échantillonnage d'importance

On procède à un changement de probabilité par transformation de Esscher.

Le changement de probabilité s'exprime avec :

$$L_{T,f}(P_T) = \exp \left[P_T(f) - \int_{\mathbb{R}} (e^{f(y)} - 1) \lambda T \nu(dy) \right] = \exp \left[\sum_{k=1}^{N_T} f(J_k) - \lambda T (\mathbb{E}[e^{f(J_1)}] - 1) \right]$$

Sous la loi \mathbb{P}_f de densité $L_{T,f}(P_T) \mathbb{P}$ sur la tribu engendrée par $(P_t)_{0 \leq t \leq T}$, $(P_t)_{0 \leq t \leq T}$ est toujours un processus de Poisson composé mais de caractéristiques :

$$\lambda^f = \lambda \mathbb{E}[\exp(f(J_1))], \quad \nu^f(dy) = \frac{\exp(f(y))}{\mathbb{E}[\exp(f(J_1))]} \nu(dy).$$

On choisit $f : y \rightarrow c \times y$, où c est choisi de manière à ce que $\mathbb{E}_{\mathbb{Q}}[P_T] \simeq 0$ (dans l'implémentation, on choisira un tel c par dichotomie en faisant varier les valeurs de c entre -100 et 100). Dans ces conditions,

$$\lambda^c = \lambda \mathbb{E}[\exp(c J_1)], \quad \nu^c(dy) = \frac{\exp(c y)}{\mathbb{E}[\exp(c J_1)]} \nu(dy).$$

On voit qu'avec $c \leq 0$, on peut diminuer la probabilité d'un saut positif et donc favoriser les trajectoires basses.

Pour $m = 1$, J_1 est une variable de Rademacher donc :

$$\lambda^c = \lambda \cosh(c), \quad \mathbb{P}_c(J_1 = +1) = \frac{\exp(c)}{2 \cosh(c)}, \quad \mathbb{P}_c(J_1 = -1) = \frac{\exp(-c)}{2 \cosh(c)}$$

Pour $m = 3$,

$$\left\{ \begin{array}{l} \lambda^c = \lambda \left[\frac{\cosh(c)}{2} + \frac{\cosh(2c)}{3} + \frac{\cosh(3c)}{6} \right] \\ \mathbb{P}_c(J_1 = \pm 1) = \frac{\exp(\pm c)/4}{\cosh(c)/2 + \cosh(2c)/3 + \cosh(3c)/6} \\ \mathbb{P}_c(J_1 = \pm 2) = \frac{\exp(\pm 2c)/6}{\cosh(c)/2 + \cosh(2c)/3 + \cosh(3c)/6} \\ \mathbb{P}_c(J_1 = \pm 3) = \frac{\exp(\pm 3c)/12}{\cosh(c)/2 + \cosh(2c)/3 + \cosh(3c)/6} \end{array} \right.$$

Dans les deux cas,

$$q = \mathbb{P}(\exists t \in [0, T] | P_t < 0) = \mathbb{E}[\mathbf{1}_{\exists t \in [0, T] | P_t < 0}] = \mathbb{E}_c \left[\frac{\mathbf{1}_{\exists t \in [0, T] | P_t < 0}}{L_{T,c}} \right]$$

On échantillonne les processus de Poisson composés avec les nouvelles caractéristiques, puis on estime l'espérance par Monte-Carlo en utilisant l'égalité ci-dessus.

Intervalle de confiance : On calcule un intervalle de confiance asymptotique pour q grâce au théorème central limite pour l'échantillonnage d'importance, avec $\hat{\sigma}_M^{IS}$ l'écart-type empirique de l'estimateur :

$$\left[\hat{p}_{n_M}^{IS} \pm 1.96 \frac{\hat{\sigma}_M^{IS}}{\sqrt{M}} \right]$$

$$\text{où } (\hat{\sigma}_M^{IS})^2 = \frac{1}{M} \sum_{k=1}^M \mathbf{1}_{\tilde{N}_T^{(k)} > x/\alpha} \exp \left[2 \lambda T (\mathbb{E}[e^{c J_1}] - 1) - 2c (P_T - P_0) \right] - (\hat{p}_{n_M}^{IS})^2.$$

Choix de c : On ajuste l'hyperparamètre c de manière à ce que l'espérance du prix final échantillonné sous la probabilité \mathbb{P} soit proche de 0. On obtient :

- Pour $m = 1$: $c = -0,68$
- Pour $m = 3$: $c = -0.21$

1.1.3 Méthode de Splitting/MCMC

On utilise une méthode de Splitting couplée à une simulation des processus de Poisson par chaînes de Markov.

On note $(X_t)_{t \geq 0}$ le processus de Poisson composé de paramètres λ, ν_m , et $\Phi(X_t) = \inf_{t \in [0, T]} X_t$. On choisit une liste de K niveaux (a_0, \dots, a_K) tels que $a_K = 0 < a_{K-1} < \dots < a_0 = +\infty$

La probabilité recherchée vaut :

$$q = \mathbb{P}(\Phi(X_t) < 0) = \prod_{k=1}^K \mathbb{P}(\Phi(X_t) < a_k | \Phi(X_t) < a_{k-1})$$

On estime ces probabilités conditionnelles par MCMC : pour chacune des M étapes, on supprime chaque saut avec une probabilité p , puis on rajoute les sauts obtenus suivant un processus de Poisson de paramètres $(\lambda(1-p), \nu_m)$.

On choisit a_k comme étant le quantile de niveau α de la loi de $\Phi(X_t)$ conditionnellement à $\{\Phi(X_t) < a_{k-1}\}$. Pour le choix du seuil α , on utilise la valeur approximative de q : si on veut K niveaux, on choisit α légèrement inférieur à $q^{1/(K+1)}$. On ajuste K (ou α) pour minimiser le rapport entre l'écart-type empirique de l'estimateur et sa valeur.

Optimisation des paramètres α (seuil) et p :

Nous avons visualisé l'évolution de la variance de l'estimateur \hat{p}_n par Splitting/MCMC en fonction de ces deux paramètres, à nombre total de simulations de trajectoires fixé.

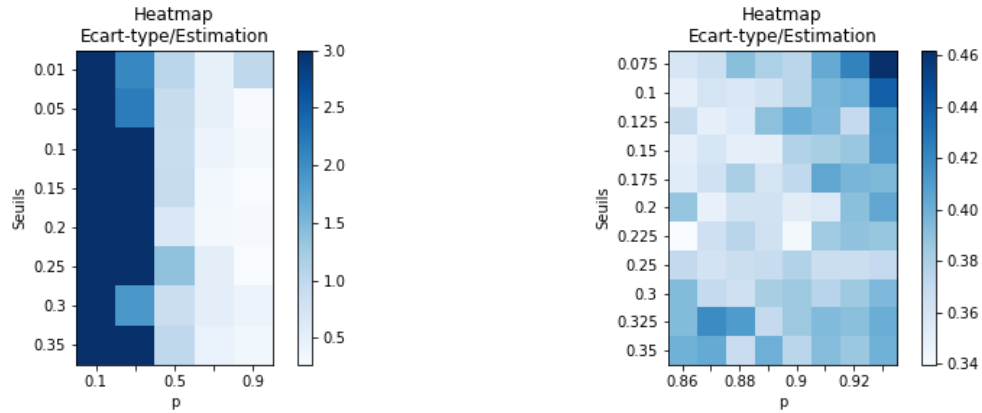


FIGURE 2 – HeatMap des écarts-types/probabilité estimée de \hat{p}_n en fonction des hyperparamètres

On constate que pour p , des valeurs d'environ 0.7 – 0.8 donnent un rapport écart-type sur valeur de l'estimateur faible (tandis que la qualité de l'estimateur diminue fortement pour des valeurs de p inférieures à 0.4). D'après les HeatMaps, le seuil α n'a qu'une influence très limitée sur la qualité de l'estimateur, on peut donc le fixer arbitrairement de manière à avoir environ 4 niveaux de Splitting, donc en prenant $\alpha \simeq q^{1/5}$

1.1.4 Résultats :

Sauf mention contraire, toutes les probabilités dans ce rapport sont estimées avec $M = 10^7$ simulations de trajectoires.

Méthode	\hat{q}_M	Ecart-type empirique de l'estimateur/ rayon de l'intervalle de confiance à 95%
Monte-Carlo	6.10^{-7}	$R_{IC} = 4.8 \cdot 10^{-7}$
Splitting-MCMC	$3.52 \cdot 10^{-7}$	$\sigma_{empirique} = 1.5 \cdot 10^{-8}$
Echantillonnage d'importance	$3.52 \cdot 10^{-7}$	$R_{IC} = 9. \cdot 10^{-10}$
Monte-Carlo ($M = 10^9$)	$3.61. \cdot 10^{-7}$	$R_{IC} = 3.7 \cdot 10^{-8}$

TABLE 1 – Probabilité de prix négatif pour $m = 1$

Méthode	\hat{q}_M	Ecart-type empirique de l'estimateur/ rayon de l'intervalle de confiance à 95%
Monte-Carlo	$4.35 \cdot 10^{-3}$	$R_{IC} = 4. \cdot 10^{-5}$
Splitting-MCMC	$4.348 \cdot 10^{-3}$	$\sigma_{empirique} = 7. \cdot 10^{-6}$
Echantillonnage d'importance	$4.317 \cdot 10^{-3}$	$R_{IC} = 7. \cdot 10^{-6}$
Monte-Carlo ($M = 10^9$)	$4.332 \cdot 10^{-3}$	$R_{IC} = 4. \cdot 10^{-6}$

TABLE 2 – Probabilité de prix négatif pour $m = 3$

1.2 Estimation de quantiles du prix final

Nous avons implémenté trois méthodes, qui diffèrent seulement par la manière d'obtenir M réalisations du prix final :

- Dans la première, on simule le prix final de manière "naïve" (nombre d'incrément suivant une loi de Poisson, puis simulation des incréments que l'on somme pour obtenir le prix final)
- Dans la deuxième (quantile ergodique), on simule une première trajectoire, que l'on fait ensuite "évoluer" (comme pour le calcul des niveaux de Splitting en question 1.1.3), en gardant pour chaque trajectoire de la chaîne de Markov son prix final. La seule différence avec la première méthode est la manière de simuler les trajectoires. On choisit $p = 0.5$ pour le "coloriage".

Dans ces deux cas, une fois l'échantillon de prix final construit, on estime le quantile en prenant la $[M\alpha]$ -ième valeur de l'échantillon trié :

$$\inf_x \{x; \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{P_T^i > x} \leq 1 - \alpha\} = P_T^{(\lceil n\alpha \rceil)} \rightarrow Q(\alpha)$$

- Dans la troisième méthode, on estime le quantile par échantillonnage d'importance, en utilisant la propriété :

$$\inf_x \{x; \frac{1}{n} \sum_{i=1}^n \frac{p(P_T^i)}{q(P_T^i)} \mathbf{1}_{P_T^i > x} \leq 1 - \alpha\} = \inf_x \{x; \frac{1}{n} \sum_{i=1}^n \frac{1}{L_{T,f}(P_T^i)} \mathbf{1}_{P_T^i > x} \leq 1 - \alpha\} \rightarrow Q(\alpha)$$

Intervalles de confiance : On peut calculer des intervalles de confiance asymptotiques à 95% pour les quantiles. En effet, en notant $Q(u)$ le quantile de niveau u et F_n la fonction de répartition empirique de la variable P_T :

$$\forall u \in]0, 1[, \sqrt{n}(F_n(Q(u)) - u) \Rightarrow \mathcal{N}(0, \alpha(1 - \alpha))$$

Donc $[Q_n(\alpha_n^-), Q_n(\alpha_n^+)]$ est un intervalle asymptotique à 95% pour $Q(\alpha)$, avec :

$$\alpha_n^- = \alpha - 1.96 \frac{\sqrt{\alpha(1 - \alpha)}}{\sqrt{n}}, \quad \alpha_n^+ = \alpha + 1.96 \frac{\sqrt{\alpha(1 - \alpha)}}{\sqrt{n}}$$

Les résultats obtenus sont reportés ci-dessous :

Niveau	Méthode	Quantile	Intervalle de confiance
10^{-6}	Monte-Carlo	2	[1, 2]
	Ergodique	1	[-1, 2]
	Echantillonnage d'importance	2	[0, 2]
10^{-5}	Monte-Carlo	5	[5, 5]
	Ergodique	5	[5, 5]
	Echantillonnage d'importance	5	[5, 5]
10^{-4}	Monte-Carlo	9	[9, 9]
	Ergodique	9	[9, 9]
	Echantillonnage d'importance	9	[9, 9]
$1 - 10^{-4}$	Monte-Carlo	61	[61, 61]
	Ergodique	61	[61, 61]
	Echantillonnage d'importance	61	[61, 61]
$1 - 10^{-5}$	Monte-Carlo	65	[65, 65]
	Ergodique	65	[65, 65]
	Echantillonnage d'importance	65	[65, 65]
$1 - 10^{-6}$	Monte-Carlo	68	[67, 70]
	Splitting-MCMC	68	[67, 69]
	Echantillonnage d'importance	68	[68, 70]

TABLE 3 – Q1 : Quantiles de niveau 10^{-4} , 10^{-5} , 10^{-6} à gauche et à droite pour $m = 1$

Niveau	Méthode	Quantile	Intervalle de confiance
10^{-6}	Monte-Carlo	-29	[-31, -26]
	Splitting-MCMC	-28	[-29, -26]
	Echantillonnage d'importance	-26	[-30, -25]
10^{-5}	Monte-Carlo	-20	[-21, -20]
	Splitting-MCMC	-20	[-21, -20]
	Echantillonnage d'importance	-20	[-20, -19]
10^{-4}	Monte-Carlo	-13	[-13, -13]
	Splitting-MCMC	-13	[-13, -13]
	Echantillonnage d'importance	-12	[-13, -12]
$1 - 10^{-4}$	Monte-Carlo	83	[83, 83]
	Splitting-MCMC	83	[83, 83]
	Echantillonnage d'importance	83	[83, 83]
$1 - 10^{-5}$	Monte-Carlo	90	[90, 91]
	Splitting-MCMC	90	[89, 91]
	Echantillonnage d'importance	90	[90, 91]
$1 - 10^{-6}$	Monte-Carlo	97	[95, 99]
	Splitting-MCMC	97	[95, 100]
	Echantillonnage d'importance	97	[95, 99]

TABLE 4 – Q1. Quantiles de niveau 10^{-4} , 10^{-5} , 10^{-6} à gauche et à droite pour $m = 3$

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

En plus de leur nature discrète et du caractère isotrope de leurs variations, on souhaite également modéliser une autre caractéristique des prix à haute fréquence, qui est la **tendance à osciller autour d'une valeur intermédiaire**. Dans cette partie, on va modéliser cette composante oscillante du prix. La nouvelle modélisation du prix *mid* est :

$$\begin{cases} P_t = P_t^{(1)} + P_t^{(2)} \\ P_t^{(1)} = P_0/2 + \sum_{i=1}^{N_t^{(1)}} J_n^{(1)} \\ P_t^{(2)} = P_0/2 + \sum_{i=1}^{N_t^{(2)}} J_n^{(1)} \end{cases}$$

où $N_t^{(1)}$ et $N_t^{(2)}$ sont des processus de Poisson de paramètres respectifs $\lambda_1 = 1/660$ et $\lambda_2 = 1/110$, où les $(J_n^{(1)})_{n \geq 1}$ sont distribués comme en 1, où $J_1^{(2)}$ est une variable de Rademacher et : $\forall n \geq 1, J_{n+1}^{(2)} = -J_n^{(2)}$. On garde $P_0 = 35$ et $T = 4h$.

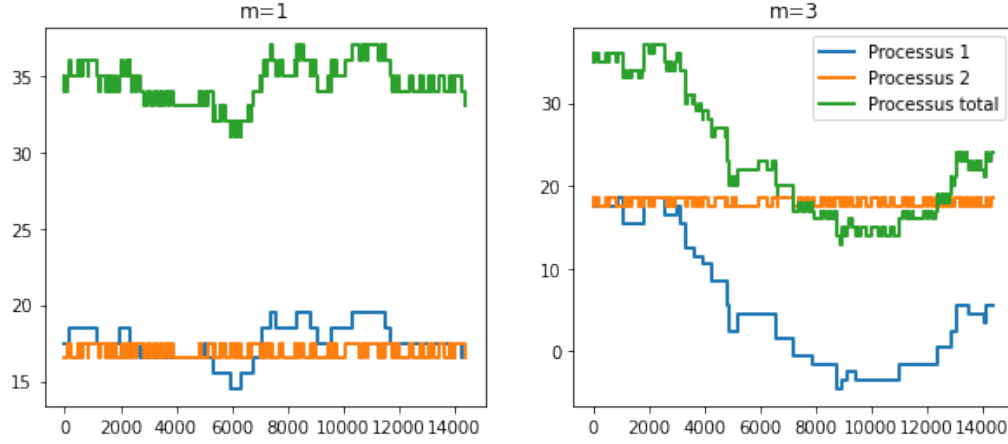


FIGURE 3 – Simulation de la trajectoire $(P_t)_{t \leq T}$

2.1 Estimation de la probabilité d'un prix négatif

2.1.1 Méthode de Monte-Carlo naïve

On reprend la méthode de la question 1.1.1. en l'adaptant pour prendre en compte le processus $P_t^{(2)}$.

La probabilité pour $m = 1$ est trop faible pour que l'on parvienne à l'estimer avec cette méthode.

2.1.2 Méthode d'échantillonnage d'importance

Il est toujours possible d'utiliser la méthode d'échantillonnage d'importance présentée en question 1.1.2. Il faut cette fois-ci réaliser un changement de loi pour $P_T^{(1)}$, et garder identique la loi de $P_T^{(2)}$. Ceci est possible car **les deux processus sont indépendants**.

Comme en 1.1.2, on a un intervalle de confiance asymptotique à 95%.

Choix de c : On choisit ici encore une fonction d'importance de la forme $x \rightarrow cx$, où $c \leq 0$ est ajusté de manière à que $\mathbb{P}_c[P_T < 0] \simeq 0.5$

- Pour $m = 1$: $c = -1.3$
- Pour $m = 3$: $c = -0.4$

2.1.3 Résultats :

Méthode	\hat{q}_M	Ecart-type empirique de l'estimateur/ rayon de l'intervalle de confiance à 95%
Echantillonnage d'importance	$1.488 \cdot 10^{-12}$	$R_{IC} = 6.5 \cdot 10^{-15}$

TABLE 5 – Probabilité de prix négatif pour $m = 1$

Méthode	\hat{q}_M	Ecart-type empirique de l'estimateur/ rayon de l'intervalle de confiance à 95%
Monte-Carlo	$5.05 \cdot 10^{-5}$	$R_{IC} = 4.4 \cdot 10^{-6}$
Echantillonnage d'importance	$5.10 \cdot 10^{-5}$	$R_{IC} = 1.2 \cdot 10^{-7}$

TABLE 6 – Probabilité de prix négatif pour $m = 3$

2.2 Estimation de quantiles du prix final

On a implémenté uniquement la méthode naïve du calcul de quantile empirique, qui renvoie la $\lceil n\alpha \rceil$ -ième valeur du tableau des P_T trié de longueur n .

Niveau	Méthode	Quantile	L'intervalle de confiance
10^{-6}	Monte-Carlo	12	[11, 12]
	Echantillonnage d'importance	12	[11, 12]
10^{-5}	Monte-Carlo	14	[14, 15]
	Echantillonnage d'importance	14	[14, 15]
10^{-4}	Monte-Carlo	17	[17, 17]
	Echantillonnage d'importance	17	[17, 17]
$1 - 10^{-4}$	Monte-Carlo	53	[53, 53]
	Echantillonnage d'importance	53	[53, 53]
$1 - 10^{-5}$	Monte-Carlo	56	[55, 56]
	Echantillonnage d'importance	56	[55, 56]
$1 - 10^{-6}$	Monte-Carlo	58	[57, 59]
	Echantillonnage d'importance	58	[58, 59]

TABLE 7 – Quantiles gauches et droits pour $m = 1$

Niveau	Méthode	Quantile	L'intervalle de confiance
10^{-6}	Monte-Carlo	-8	[-10, -7]
	Echantillonnage d'importance	-8	[-10, -7]
10^{-5}	Monte-Carlo	-3	[-3, -3]
	Echantillonnage d'importance	-3	[-4, -3]
10^{-4}	Monte-Carlo	2	[2, 2]
	Echantillonnage d'importance	2	[2, 2]
$1 - 10^{-4}$	Monte-Carlo	68	[68, 68]
	Echantillonnage d'importance	68	[68, 68]
$1 - 10^{-5}$	Monte-Carlo	73	[72, 73]
	Echantillonnage d'importance	73	[73, 74]
$1 - 10^{-6}$	Monte-Carlo	77	[76, 78]
	Echantillonnage d'importance	78	[77, 80]

TABLE 8 – Quantiles gauches et droits pour $m = 3$

Ces quantiles sont nettement plus proches de P_0 que ceux de la question 1, ce qui est cohérent avec le fait qu'on modélise la tendance au retour vers la moyenne.

3 Une autre modélisation du phénomène de retour vers la moyenne

On peut également modéliser les oscillations du prix autour du prix moyen en considérant cette fois comme loi des sauts :

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

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

— $(\hat{J}_n)_{n \geq 1}$ est une chaîne de Markov telle que $\hat{J}_1 \sim -1 + 2\mathcal{B}(1/2)$ et dont la matrice de transition est :

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

avec $\alpha_+, \alpha_- \in [-1, 1]$.

— Les $(\xi_n)_{n \geq 1}$ sont à valeurs dans $\{1, \dots, m\}$. Comme dans les questions précédentes, si $m=1$, $\xi_n = 1$;
et si $m = 3$, $\mathbb{P}(\xi_n = 1) = \frac{1}{2}$, $\mathbb{P}(\xi_n = 2) = \frac{1}{3}$, $\mathbb{P}(\xi_n = 3) = \frac{1}{6}$.

Il faut que $\alpha_+ < 0$ et $\alpha_- < 0$ pour modéliser le retour vers la moyenne : il est plus probable que le prix baisse lorsqu'il vient de monter que le contraire, et vice-versa.

On fixe la valeur estimée pour α dans l'article : $\alpha = -0.875$

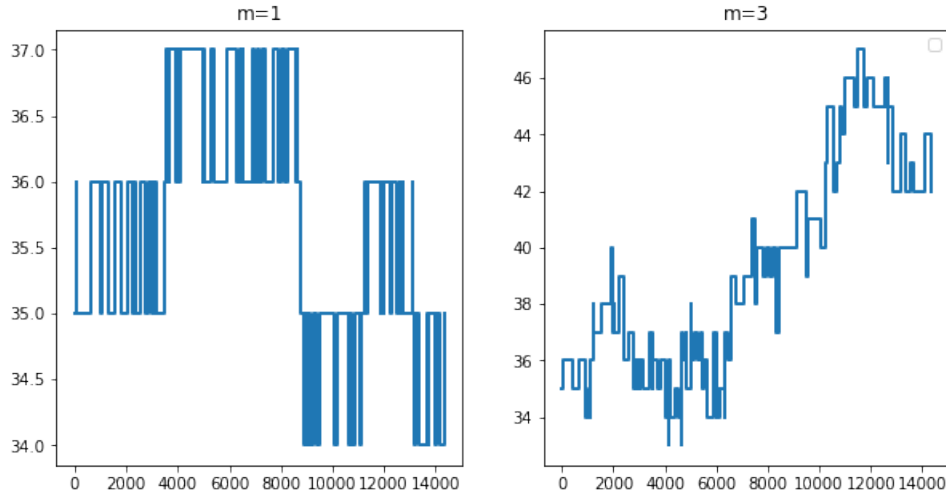


FIGURE 4 – Simulation de la trajectoire $(P_t)_{t \leq T}$

3.1 Estimation de la probabilité d'un prix négatif

On implémente deux méthodes :

- Une Monte-Carlo naïve
- Une méthode des particules en interaction (sélection-mutation) pour simuler la trajectoire jusqu'à N sauts.

3.1.1 Méthode de Monte-Carlo naïve

Cette première méthode permet de tester la correction de la méthode suivante avec des valeurs de P_0 plus faibles, mais elle n'est pas assez performante pour être utilisée pour estimer la probabilité de prix négatif pour $P_0 = 35$.

3.1.2 Méthode de sélection-mutation

Description de la méthode des particules en interaction appliquée au problème :

Avec cette nouvelle modélisation où on observe les variations du prix jusqu'à un **nombre de sauts N fixé**, on peut simuler les trajectoires du prix comme des chaînes de Markov, grâce à la méthode des particules en interaction.

En effet, on cherche à calculer $q = \mathbb{E}[f(P_0, \dots, P_N)]$, où $f(p_0, \dots, p_N) = \mathbf{1}_{\min(p_0, \dots, p_N) < 0}$

La suite des couples (prix, signe du saut) $(P_n, \hat{J}_n)_{n \in \mathbb{N}}$ est définie par :

- $\hat{J}_1 \sim -1 + 2\mathcal{B}(1/2)$
- $P_1 = 35 + \hat{J}_1$
- $\forall n \geq 1, \hat{J}_{n+1} = \hat{J}_n \times \delta_n$, où les δ_n sont iid suivant $-1 + 2\mathcal{B}\left(\frac{1+\alpha}{2}\right)$ (valent 1 avec probabilité $\frac{1+\alpha}{2}$, -1 sinon)
- $X_{n+1} = X_n + \xi_n \times \hat{J}_{n+1}$

Les amplitudes de saut ξ_n et les "changements de signe de saut" δ_n sont bien indépendantes identiquement distribuées, et indépendantes de (P_0, \hat{J}_0) . On a bien une récurrence aléatoire, donc $(P_n, \hat{J}_n)_{n \in \mathbb{N}}$ est une chaîne de Markov.

La probabilité d'avoir un prix négatif dans les N premiers sauts est donnée par :

$$q = \mathbb{E}\left[f((P_0, \hat{J}_0), \dots, (P_N, \hat{J}_N))\right]$$

où $f((p_0, \hat{j}_0), \dots, (p_N, \hat{j}_N)) = \mathbf{1}_{\min(p_0, \dots, p_N) < 0}$

On simule donc une trajectoire de la chaîne de Markov $(p_n, \hat{J}_n)_{n \in \mathbb{N}}$ jusqu'à $n = N$.

On choisit comme fonction de poids G une fonction qui favorise les trajectoires où le prix P_n est bas : $G_k(P_{0:k}) = \exp(-\mu P_k)$. Il va falloir ajuster le paramètre $\mu \in \mathbb{R}^+$.

Une autre possibilité aurait été de choisir une fonction G qui favorise les trajectoires croissantes, de la forme $G_k(P_{0:k}) = \exp(-\mu(P_k - P_{k-1}))$. Dans notre cas c'est une très mauvaise idée, car une trajectoire qui vient de descendre aura tendance à remonter ensuite. On sélectionnerait donc les trajectoires qui ont le plus de chance de monter au prochain saut, ce qui ne nous aiderait pas à atteindre un prix négatif.

Dans l'implémentation, il est inutile de simuler les intervalles de temps entre les sauts : seuls les valeurs des incréments sont important.

Optimisation de μ :

Le graphe ci-dessous représente l'évolution du rapport de l'écart-type empirique par la valeur moyenne de l'estimateur q_M :

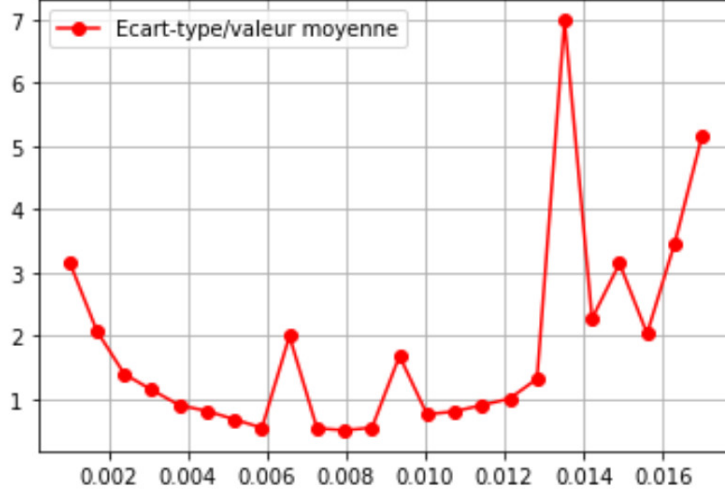


FIGURE 5 – Recherche de la valeur optimale de μ pour $m = 3$

Même si la courbe n'est pas tout à fait régulière, on observe tout de même une tendance qui permet de choisir $\mu = 0,008$.

Intervalle de confiance : On sait qu'il existe un théorème central limite pour l'estimateur par sélection-mutation, mais on ne connaît pas la variance asymptotique. On calcule donc une variance empirique, et un intervalle de confiance asymptotique pour q à 95% est :

$$\left[\frac{1}{L} \sum_{l=1}^L q_l \pm Stud_{0.975, L-1}^{-1} \frac{\sigma_L}{\sqrt{L}} \right]$$

où \hat{q}_l est l'estimateur de q obtenu au l -ième lancement de l'algorithme, $Stud_{0.975, L-1}^{-1}$ est le quantile de niveau 0.975 de la loi de Student à $L - 1$ degrés de liberté, et σ_L est l'écart-type empirique corrigé des estimateurs :

$$\sigma_L^2 = \frac{1}{L} \sum_{l=1}^L \left[\hat{q}_l - \frac{1}{L} \sum_{i=1}^L \hat{q}_i \right]^2$$

3.1.3 Résultats :

Pour $m = 1$, la probabilité de prix négatif est extrêmement faible (inférieur à 10^{-27}), il a été impossible de l'estimer. Voici les résultats pour $m = 3$:

Méthode	\hat{q}_M	Ecart-type empirique de l'estimateur/ rayon de l'intervalle de confiance à 95%
Monte-Carlo	$4.19 \cdot 10^{-5}$	$R_{IC} = 4.0 \cdot 10^{-6}$
Sélection-mutation	$4.41 \cdot 10^{-5}$	$R_{IC} = 3.8 \cdot 10^{-7}$

TABLE 9 – Probabilité de prix négatif pour $m = 3$

3.2 Estimation de quantiles du prix final

On a implémenté une méthode de Monte-Carlo naïve, qui suffit pour estimer les quantiles.

Niveau	Quantile	Intervalle de confiance
$1 \cdot 10^{-4}$	25	[25, 25]
$1 \cdot 10^{-5}$	23	[23, 23]

TABLE 10 – Quantiles gauches pour $m = 1$

Niveau	Quantile	Intervalle de confiance
10^{-4}	3	[3, 3]
10^{-5}	-2	[-3, -2]

TABLE 11 – Quantiles gauches pour $m = 3$

Ces quantiles sont plus proches de P_0 que dans la question 2 : on a donc modélisé une tendance au retour vers la moyenne plus prononcée.

4 Limite macroscopique

Dans cette question, on étudie les limites du modèle de prix décrit dans la question 3. sur des échelles de temps plus longues. Les nouvelles données sont :

- $P_0 = 1000$
- $T = 172800000$ s (250 jours de trading, 8h par jour)
- Comme dans la question 1, $\lambda = 1/300$.

Contrairement à la question 3., on ne considère plus une trajectoire à nombre de sauts fixé N , mais plutôt à horizon temporel fixé $T = 172800000$. Les autres hypothèses restent inchangées.

D'après l'article *Semi Markov model for market microstructure* (pp. 15-16), le processus $(P_t)_{0 \leq t \leq T}$ peut être modélisé par un mouvement brownien pour T suffisamment grand (la nature discrète des prix n'est plus à prendre en compte à l'échelle macroscopique).

Plus précisément, on pose $\Delta P_t = P_t - P_0$ pour $t \in [0, T]$, et $\Delta P_t^{(T)} = \frac{\Delta P_{tT}}{\sqrt{T}}$ pour $t \in [0, 1]$.

D'après la proposition 3.1 de l'article, il y a convergence en loi du processus normalisé $\Delta P^{(T)}$ vers un processus brownien :

$$\Delta P^{(T)} \Rightarrow \sigma_\infty W$$

où W est un processus brownien standard et $\sigma_\infty^2 = \lambda \left[\text{Var}[\xi_n] + (\mathbb{E}[\xi_n])^2 \frac{1+\alpha}{1-\alpha} \right]$

- Pour $m = 1$, $\sigma_{\infty,1}^2 = \lambda \times \left(0 + 1 \times \frac{1+\alpha}{1-\alpha} \right) = \lambda \frac{1+\alpha}{1-\alpha}$
- Pour $m = 3$, $\sigma_{\infty,3}^2 = \lambda \times \left(\frac{10}{3} + \frac{25}{9} \times \frac{1+\alpha}{1-\alpha} \right)$

Comme on se place dans le cas où $T \rightarrow +\infty$, on peut simuler le processus brownien standard pour obtenir une simulations du processus des prix :

$$P_t = P_0 + \Delta P_t = P_0 + \sqrt{T} \Delta P_{t/T}^{(T)} \simeq P_0 + \sqrt{T} \sigma_\infty W_{t/T}$$

Le mouvement brownien standard W est défini par :

$$\begin{cases} W_0 = 0 \\ \forall s \leq t, W_s - W_t \sim \mathcal{N}(0, t - s) \\ \forall n, \forall t_i, 0 \leq t_0 \leq \dots \leq t_n, (W_{t_n} - W_{t_{n-1}}, \dots, W_{t_1} - W_{t_0}) \text{ sont indépendantes.} \end{cases}$$

4.1 Estimation de la probabilité d'un prix négatif

4.1.1 Par méthode de Monte-Carlo sur le prix macroscopique discrétisé

Pour estimer la probabilité d'un prix négatif, il reste à discrétiser le mouvement brownien.

On fixe $N \in \mathbb{N}$ et on discrétise l'intervalle $[0, T]$ en $\{\frac{kT}{N} \mid 0 \leq k \leq N\}$

Nous faisons l'hypothèse qu'il est raisonnable d'approximer la probabilité que le prix prenne au moins une valeur négative sur l'intervalle $[0, T]$ par la probabilité que les prix prenne au moins une valeur négative sur cet ensemble de N points équidistants de l'intervalle :

$$\mathbb{P}(\exists t \in [0, T] \mid P_t < 0) \simeq \mathbb{P}(\exists k \in \{1, \dots, N\} \mid P_{kT/N} < 0)$$

$$= \mathbb{P}(\exists k \in \{1, \dots, N\} \mid P_0 + \sqrt{T} \sigma_\infty W_{k/N} < 0)$$

$$= \mathbb{P}\left(\exists k \in \{1, \dots, N\} \mid P_0 + \sqrt{T} \sigma_\infty \sqrt{1/N} \sum_{i=1}^k X_i < 0\right)$$

où $(X_1, \dots, X_N) \sim \mathcal{N}\left((0, \dots, 0), I_N\right)$

Pour s'assurer de la validité de ce modèle, on a simulé d'abord la trajectoire avec la même modélisation que dans la question 3, si ce n'est qu'ici l'horizon temporel T est fixé au lieu du nombre N de sauts :

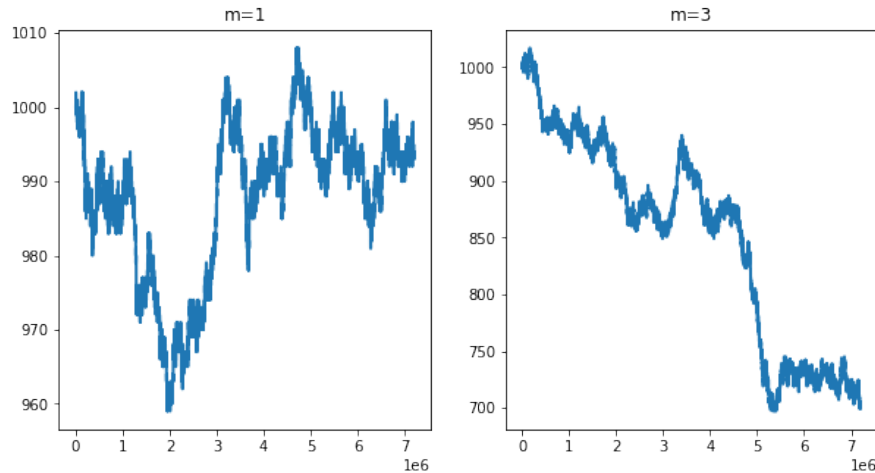


FIGURE 6 – Simulation de la trajectoire dans la modélisation de la question 3 à horizon fixé T

Dans un deuxième temps, on simule la trajectoire par une somme de gaussiennes indépendantes dans le cadre de la limite macroscopique du processus :

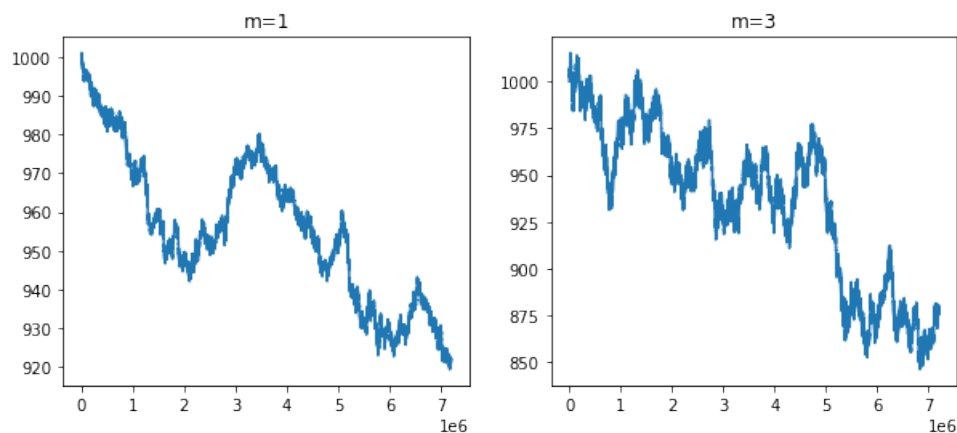


FIGURE 7 – Simulation de la trajectoire discrétisée $(P_{kT/N})_{0 \leq k \leq N}$ dans la limite brownienne

Les allures des trajectoires sont tout à fait comparables, ce qui semble valider l'approximation du processus des prix par sa limite comme décrite ci-dessus. Pour avoir une deuxième confirmation, nous avons réalisé l'histogramme des tirages des prix finaux sous la modélisation "non macroscopique" (celle de la question 3) :

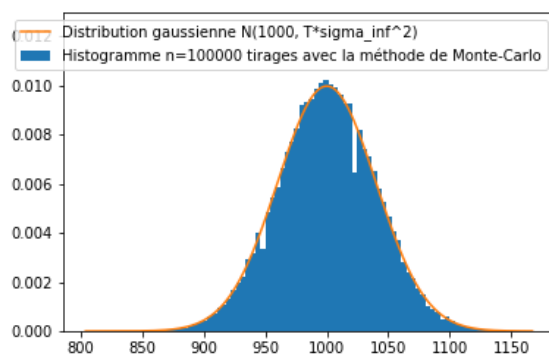


FIGURE 8 – Histogramme des prix finaux pour $m=1$

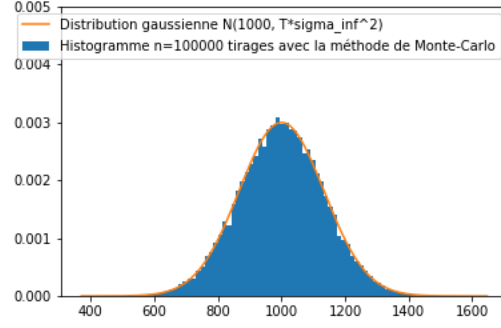


FIGURE 9 – Histogramme des prix finaux pour $m=3$

On observe que l'histogramme des prix finaux converge vers une distribution gaussienne, ce qui confirme la validité de l'approximation par la limite brownienne. On simule donc les trajectoires en se plaçant à la limite macroscopique. Cependant la probabilité q est trop faible pour être estimée simplement par Monte-Carlo, d'où la nécessité d'utiliser une méthode de mutation/sélection.

4.1.2 Par méthode des particules en interaction sur le prix macroscopique discrétisé

Dans la modélisation macroscopique, le prix discrétisé $(P_{kT/N})_{0 \leq k \leq N}$ est une chaîne de Markov (c'est une récurrence aléatoire) :

$$P_{kT/N} = P_0 + \sqrt{T/N}\sigma_\infty \sum_{i=1}^k X_i$$

On note $Y_k = \frac{P_{kT/N}}{\sqrt{T/N}\sigma_\infty}$ et on va simuler $Y_k = \frac{P_0}{\sqrt{T/N}\sigma_\infty} + \sum_{i=1}^k X_i$

En effet, $\mathbb{P}(\exists k \in \{1, \dots, N\} | Y_k < 0) = \mathbb{P}(\exists k \in \{1, \dots, N\} | P_{kT/N} < 0) \simeq q$

On peut utiliser sur la chaîne de markov Y une méthode de mutation/sélection. Comme dans la question 3.1, on choisit comme fonction de poids G une fonction qui favorise les trajectoires où le prix q est bas : $G_k(Y_{0:k}) = \exp(-\mu Y_k)$, où $\mu \in \mathbb{R}^+$.

Pour $m = 1$, la probabilité est extrêmement faible et nous ne parvenons pas à l'estimer avec cette méthode.

Pour $m = 3$, on choisit $\mu = 0.03$ expérimentalement, pour avoir une moyenne des prix finaux proche de 0, et on trouve une estimation de q .

L'intervalle de confiance asymptotique à 95% est obtenu comme décrit en 3.1.2.

On obtient pour $m = 3$: $\hat{q}_M = (1.1 \pm 0.8)10^{-14}$

4.2 Estimation de quantiles

Pour estimer les quantiles, on peut soit utiliser les méthodes de la question 3.2, soit se placer dans l'approximation de la limite brownienne.

Dans ce deuxième cas, on utilise le fait que :

$$P_T \simeq P_0 + \sqrt{T} \sigma_\infty W_1$$

où $W_1 \sim \mathcal{N}(0, 1)$

Ainsi, $\alpha = \mathbb{P}(P_T \leq Q_T(\alpha)) \simeq \mathbb{P}(W_1 \leq \frac{Q_T(\alpha) - P_0}{\sqrt{T} \sigma_\infty})$, où $Q_T(\alpha)$ dénote le quantile de niveau α de P_T .

En notant F^{-1} l'inverse de la fonction de répartition de la loi normale standard,

$$\frac{Q_T(\alpha) - P_0}{\sqrt{T} \sigma_\infty} \simeq F^{-1}(\alpha)$$

Donc :

$$Q_T(\alpha) \simeq P_0 + \sqrt{T} \sigma_\infty F^{-1}(\alpha)$$