



ELEMENTS FONDAMENTAUX DE LA METHODE MONTE CARLO

 Ibrahima Dione (Ph.D.) & Van Son Lai (Ph.D., CFA)

 Simulations Stochastiques et Applications en Finance

- Introduction au Concept de Précision
- Qualité des Résultats Obtenus par Simulation MC
- Amélioration de la Qualité des Simulations MC

Introduction au Concept de Précision



- ▷ La méthode Monte Carlo est une méthode de calcul numérique dont l'objectif est d'effectuer des calculs numériques des fonctionnelles de variables aléatoires.
- ▷ L'approche consiste à réaliser une série d'expériences et à faire la moyenne des valeurs issues de chacune d'elles.
- ▷ Soit une variable aléatoire X de fonction de densité de probabilité $f_X(x)$.
- ▷ On cherche à calculer $E[g(X)]$ (où $g(\cdot)$ est une fonction donnée), dont la valeur numérique est souvent difficile à obtenir.
- ▷ On l'estime alors par la technique Monte Carlo qui consiste à générer N échantillons indépendants X_i de la variable aléatoire X et à approximer l'espérance $h = E[g(X)]$ par la moyenne arithmétique

$$\hat{h} = \frac{1}{N} \sum_{i=1}^N g(X_i) \quad (1)$$

- ▷ Sachant que d'après la loi des grands nombres, nous avons la relation

$$P\left(\left|h - \hat{h}\right| < \varepsilon \sigma_h\right) \geq 1 - \frac{1}{N\varepsilon^2} \quad (2)$$

- ▷ Alors h appartient, avec une probabilité d'au moins $1 - \frac{1}{N\varepsilon^2}$, à l'intervalle

$$\left[\hat{h} - \varepsilon \sigma_h, \hat{h} + \varepsilon \sigma_h\right] \quad (3)$$

- ▷ Pour un niveau de confiance α fixé dans $[0, 1]$ tel que la probabilité que h soit dans l'intervalle en (3) est supérieure ou égale à $1 - \alpha$, érige un nombre minimum de simulations vérifiant

$$N > \frac{1}{\alpha \varepsilon^2} \quad (4)$$

Note: En terminologie statistique, cela veut dire que l'hypothèse que h se trouve dans l'intervalle en (3) est acceptable avec un niveau de confiance α (pour $N > \frac{1}{\alpha \varepsilon^2}$).

- ▷ Comme σ_h^2 n'est pas connue, on le remplace par son estimateur

$$\hat{\sigma}_h^2 = \frac{1}{N-1} \sum_{i=1}^N \left(g(X_i) - \hat{h}\right)^2 \quad (5)$$

- ▷ Le nombre de simulations nécessaires, calculé à partir de la loi des grands nombres peut être amélioré par le théorème centrale limite. Introduisons en effet la variable aléatoire Z

$$\begin{aligned} Z &= \frac{1}{\sigma_h \sqrt{N}} \sum_{i=1}^N (g(X_i) - h) \\ &= \frac{\sqrt{N}}{\sigma_h} \frac{1}{N} \sum_{i=1}^N (g(X_i) - h) = \frac{\sqrt{N}}{\sigma_h} (\hat{h} - h) \end{aligned}$$

où $h = E[g(X)]$, $E[Z] = 0$ et $\text{Var}[Z] = 1$.

- ▷ Cela implique alors l'égalité suivante

$$\begin{aligned} P(|h - \hat{h}| < \varepsilon \sigma_h) &= P\left(\left|\frac{\sqrt{N}}{\sigma_h} (h - \hat{h})\right| < \frac{\sqrt{N}}{\sigma_h} \varepsilon \sigma_h\right) \\ &= P(|Z| < \varepsilon \sqrt{N}) \end{aligned}$$

- ▷ Si le nombre de simulations N est élevé, la variable Z est approximée par une gaussienne de moyenne nulle et de variance unitaire. On a donc

$$\begin{aligned} P(|Z| < \varepsilon \sqrt{N}) &= P(-\varepsilon \sqrt{N} < Z < \varepsilon \sqrt{N}) \simeq \frac{1}{\sqrt{2\pi}} \int_{-\varepsilon \sqrt{N}}^{\varepsilon \sqrt{N}} e^{-\frac{x^2}{2}} dx \\ &\simeq \frac{2}{\sqrt{\pi}} \int_0^{\varepsilon \sqrt{\frac{N}{2}}} e^{-u^2} du \\ &\simeq \text{erf}\left(\varepsilon \sqrt{\frac{N}{2}}\right) \end{aligned}$$

▷ On peut obtenir le nombre de simulations en remarquant d'abord que

$$P\left(|Z| < \varepsilon\sqrt{N}\right) > 1 - \alpha = \operatorname{erf}(\beta), \text{ où } \beta = \operatorname{erf}^{-1}(1 - \alpha)$$

▷ Le nombre de simulations en fonction de l'intervalle de confiance est

$$N \geq 2 \left(\frac{\operatorname{erf}^{-1}(1 - \alpha)}{\varepsilon} \right)^2$$

Exemple:

Ainsi en prenant $\alpha = 0.05$, on a alors $\beta = \operatorname{erf}^{-1}(0.95) = 1.96$. Et pour $\varepsilon = 0.01$, on obtient le nombre approximatif de simulations suivant

$$\begin{aligned} N &\simeq \frac{2(1.96)^2}{10^{-4}} \\ &\simeq 76\,832 \end{aligned}$$

Ce nombre est largement inférieur à celui généré via l'équation (4) qui est 200 000. Ce qui est une nette amélioration du nombre de simulations.

Qualité des Résultats Obtenus par Simulation MC



- ▷ En général, la précision de l'approximation de σ_h^2 par sa valeur empirique en (5) est relativement faible et complexe.
- ▷ Toutefois si $g(X)$ est gaussien, l'approximation de σ_h^2 est assez satisfait et simple pourvu que le nombre de simulations N soit très élevé.
- ▷ En pratique l'hypothèse gaussienne sur $g(X)$ est difficilement acceptable.
- ▷ Nous avons par conséquent besoin d'une approche qui permet d'approximer le cas gaussien.

- ▷ Cette approche consiste à effectuer M calculs, chacun correspondant à N simulations MC, comme suit

$$\hat{h}_k = \frac{1}{N} \sum_{i=1}^N g(X_{k,i}), \quad k = 1, \dots, M \quad (6)$$

où $X_{k,i}$ est le résultat de la i^e simulation du k^e calcul. D'après le théorème central limite, quand N est élevé, \hat{h}_k peut être considéré comme gaussien.

Note: On peut appliquer la précision discutée précédemment à chacune des \hat{h}_k , $k = 1, \dots, M$. Cependant la valeur exacte de σ_h étant inconnue, on utilisera celle de $\hat{\sigma}_h$.

- ▷ Avec les M expériences réalisées, tous les résultats \hat{h}_k peuvent être considérés comme des variables aléatoires gaussiennes indépendantes de même statistique.
- ▷ L'estimateur h et sa variance sont estimés comme suit

$$\hat{h} = \frac{1}{M} \sum_{k=1}^M \hat{h}_k \quad \text{et} \quad \hat{\sigma}_h^2 = \frac{1}{M-1} \sum_{k=1}^M (\hat{h}_k - \hat{h})^2 \quad (7)$$

- ▷ Par conséquent, lorsque M est assez grand, pour obtenir la caractéristique gaussienne de \hat{h}_k , la variable

$$T_M = \frac{\hat{h} - h}{\sqrt{\frac{\hat{\sigma}_{\hat{h}}^2}{M}}}$$

suit une t-distribution de Student avec $M - 1$ degrés de liberté.

- ▷ La précision sur h est maintenant définie par

$$P\left(\left|\hat{h} - h\right| < a\right) = P\left(\left|\frac{\hat{h} - h}{\sqrt{\frac{\hat{\sigma}_{\hat{h}}^2}{M}}}\right| < \frac{a}{\sqrt{\frac{\hat{\sigma}_{\hat{h}}^2}{M}}}\right) \quad (8)$$

- ▷ L'intervalle de confiance de h , de niveau $100(1 - \alpha)\%$, est $[\hat{h} - a, \hat{h} + a]$

$$\text{où } a = t_{1-\alpha, M-1} \sqrt{\frac{\hat{\sigma}_{\hat{h}}^2}{M}}$$

Amélioration de la Qualité des Simulations MC



- ▷ Nous avons vu que, pour obtenir une précision adéquate, il faut augmenter de façon importante le nombre de simulations Monte Carlo.
- ▷ Avec les puissances de calcul moderne disponibles aujourd'hui, le nombre de simulations n'est pas vraiment un facteur dominant.
- ▷ Toutefois pour des situations complexes, où il faut simuler dynamiquement tout un processus (plutôt qu'une seule variable), le temps de calcul devient un enjeu très important.
- ▷ Il faut alors chercher à réduire le nombre de simulations nécessaires.

- ▷ Nous allons examiner 4 techniques très utilisées dont la première pour améliorer la précision et les trois autres pour réduire le temps de calcul.

1. Rééchantillonnage quadratique

2. Variables antithétiques

3. Variables de contrôle

4. Échantillonnage préférentiel

- ▷ Quand on génère des variables aléatoires, on obtient des statistiques empiriques qui ne coïncident en général pas avec les statistiques du modèle.
- ▷ Il faut donc transformer les données de façon à ce que les paramètres empiriques égalent les paramètres théoriques.



- ▷ La technique du **rééchantillonnage quadratique** introduite par Barraquand en 1995, nous permet d'y remédier.

- ▷ Soit $\underline{X} = (X_1, \dots, X_n)^T$ un vecteur aléatoire à n dimensions de moyenne
$$\underline{m}_X = E[\underline{X}] = (E[X_1], E[X_2], \dots, E[X_n])^T \quad (9)$$

- ▷ Et la matrice covariance de la variable \underline{X} est donnée par

$$\begin{aligned} \Lambda_X &= E \left[\left(\underline{X} - \underline{m}_X \right) \left(\underline{X} - \underline{m}_X \right)^T \right] \\ &= E \left[\underline{X} \underline{X}^T \right] - \underline{m}_X \underline{m}_X^T \end{aligned} \quad (10)$$

- ▷ Pour estimer \underline{m}_X et Λ_X , nous pouvons exécuter M simulations et les statistiques d'échantillon obtenues sont

$$\begin{aligned} \hat{\underline{m}}_X &= \frac{1}{M} \sum_{k=1}^M \underline{X}^k \\ \hat{\Lambda}_X &= \frac{1}{M} \sum_{k=1}^M \left(\underline{X}^k - \hat{\underline{m}}_X \right) \left(\underline{X}^k - \hat{\underline{m}}_X \right)^T \\ &= \frac{1}{M} \sum_{k=1}^M \underline{X}^k \left(\underline{X}^k \right)^T - \hat{\underline{m}}_X \hat{\underline{m}}_X^T \end{aligned}$$

où \underline{X}^k est le vecteur obtenu à partir de la k^{ieme} simulation.

- ▷ A partir de la loi des grands nombres lorsque M est très grand, $\hat{\underline{m}}_{\underline{X}}$ et $\hat{\underline{\Lambda}}_{\underline{X}}$ deviennent proches de $\underline{m}_{\underline{X}}$ et $\underline{\Lambda}_{\underline{X}}$ avec une plus grande précision.
- ▷ Lorsque M est petit, cette précision peut être très mauvaise.
- ▷ Cependant, nous pouvons modifier les données X de telle sorte que la moyenne de l'échantillon et la matrice de covariance coïncident avec la moyenne statistique et la matrice de covariance.

▷ Pour cela on définit la matrice $H = \sqrt{\underline{\Lambda}_{\underline{X}}} \left(\sqrt{\hat{\underline{\Lambda}}_{\underline{X}}} \right)^{-1}$.

▷ Et on modifie les données X en construisant le vecteur Y comme suit

$$Y = H \left(\underline{X} - \hat{\underline{m}}_{\underline{X}} \right) + \underline{m}_{\underline{X}} \quad (11)$$

▷ Pour les simulations $k = 1, \dots, M$, le nouveau vecteur échantillon devient alors

$$Y^k = H \left(\underline{X}^k - \hat{\underline{m}}_{\underline{X}} \right) + \underline{m}_{\underline{X}}$$

▷ et la moyenne d'échantillon de \underline{Y}^k devient

$$\hat{\underline{m}}_{\underline{Y}} = \frac{1}{M} \sum_{k=1}^M \underline{Y}^k$$

- ▷ En développant cette somme, nous obtenons l'égalité entre la moyenne d'échantillon de \underline{Y} et la moyenne statistique de \underline{X}

$$\widehat{m}_{\underline{Y}} = m_{\underline{X}}$$

- ▷ De même, nous pouvons calculer la matrice de covariance de l'échantillon de \underline{Y} et on obtient

$$\widehat{\Lambda}_{\underline{Y}} = \Lambda_{\underline{X}}$$

Note: La transformation en (11) permet d'obtenir la moyenne empirique de \underline{Y}^k identique à la moyenne statistique de \underline{X} et la matrice de covariance empirique de \underline{Y}^k égale à la matrice de covariance statistique de \underline{X} .

Remarque(s) : Pour effectuer la transformation en (11), nous devons connaître la matrice de covariance statistique de \underline{X} . Dans de telles situations, cette transformation améliore la précision des calculs obtenus par les simulations de Monte Carlo.



- ▷ Si on veut obtenir la moyenne d'une variable aléatoire à grande variance, il faut un nombre élevé de simulations pour atteindre la précision désirée.
- ▷ Par contre, quand la variance est faible, le nombre de simulations nécessaires est faible.
- ▷ Il est donc utile d'effectuer des transformations permettant de réduire la variance de la variable obtenue.
- ▷ La méthode de réduction de variance la plus simple est l'utilisation de la méthode de **Variables Antithétiques**.
- ▷ Il s'agit de générer N variables aléatoires X_i et en construire N autres variables ayant la même densité de probabilité mais corrélées négativement avec les variables générées qu'on notera X_i^a

$$\begin{aligned} E [(X_k^a - m_X) (X_j - m_X)] &= -\delta_{k,j} \sigma_X^2 \\ &= \begin{cases} -\sigma_X^2 & \text{si } k = j \\ 0 & \text{si } k \neq j \end{cases} \end{aligned}$$

- ▷ La variance de la moyenne empirique obtenue de ces $2N$ variables est en effet plus faible que celle obtenue de $2N$ variables indépendantes.
- ▷ Tenant en compte que les X_k sont indépendantes ainsi que les X_k^a (et donc tous les $2N$ variables X_k et X_j^a sont indépendantes exceptées les paires X_k et X_k^a), la variance de la moyenne empirique \hat{m}_a est égale à

$$\begin{aligned} \text{Var} [\hat{m}_a] &= E \left[(\hat{m}_a - m_X)^2 \right] \\ &= \frac{1}{4N} \sigma_X^2 \end{aligned}$$

où la moyenne est obtenue comme suit

$$\hat{m}_a = \frac{1}{2N} \sum_{k=1}^N (X_k + X_k^a)$$

Note: Par contre, la variance empirique de $2N$ variables aléatoires indépendantes est égale à $\frac{1}{2N} \sigma_X^2$. On voit donc qu'on améliore de 2 fois la variance de la moyenne empirique tout en générant seulement une moitié de variables aléatoires.

Remarque(s) : Dans certaines situations, on ne peut pas obtenir une variable antithétique parfaite c'est-à-dire 100% négativement corrélée avec celles générées, on se contentera alors des variables ayant la même densité mais partiellement corrélées négativement avec celles générées.

Exemple:

- ▷ Soit X une variable aléatoire uniforme sur un segment $[a, b]$, c'est-à-dire $U(a, b)$, alors $X^a = b + a - X$ est une variable parfaitement antithétique de X . Effectivement on voit que X^a est aussi uniformément répartie entre $[a, b]$ et négativement corrélée à X .
- ▷ Soit $X \sim N(m, \sigma^2)$ une variable aléatoire gaussienne de moyenne m et de variance σ^2 . Alors $X^a = 2m - X$ est une variable antithétique de même densité et négativement corrélée à X .



- ▷ L'objectif des simulations est de réduire la variance de l'estimateur utilisé.
- ▷ Nous avons vu qu'en utilisant les variables antithétiques, cette variance est réduite de façon importante et au meilleur des cas on a une réduction de 50%.
- ▷ La technique des variables antithétiques exploite les propriétés de corrélations négatives des variables.
- ▷ Si on peut introduire d'autres variables pas nécessairement antithétiques mais négativement corrélées avec les variables générées, on peut toujours réduire la variance de l'estimateur.
- ▷ Soit X une variable aléatoire et Y une autre variable **corrélée** avec X appelée **variable de contrôle de X** . On construit la nouvelle variable

$$Z = X + \alpha(Y - E[Y]) \quad (12)$$

- ▷ La moyenne statistique de Z est identique à celle de X :

$$m_Z = m_X$$

- ▷ Au lieu d'estimer la moyenne de X , il revient de façon équivalente à calculer la moyenne de Z .

Note: L'objectif est de choisir le α qui va réduire le plus la variance de la variable aléatoire Z définie en (12).

- ▷ Ce choix optimum, qu'on notera α^* , doit se faire comme suit

$$\frac{d\text{Var}(Z)}{d\alpha} = 0 \quad (13)$$

- ▷ Par suite, le choix optimum est donné par la valeur suivante

$$\alpha^* = -\frac{\text{Cov}(X, Y)}{\text{Var}(Y)} \quad (14)$$

- ▷ Ainsi, en calculant la variance de la variable Z à partir de (12) on obtient

$$\text{Var}(Z) = \text{Var}(X) - \frac{[\text{Cov}(X, Y)]^2}{\text{Var}(Y)} \quad (15)$$

Remarque(s): La formule (15) montre que $\text{Var}(Z)$ est inférieure à $\text{Var}(X)$. Ce qui satisfait l'objectif qu'on s'était fixé à la note précédente.

- ▷ Dans le cas de plusieurs variable de contrôle Y_1, Y_2, \dots, Y_k de X , on construit la nouvelle variable Z comme suit

$$Z = X + \alpha_1(Y_1 - E[Y_1]) + \dots + \alpha_k(Y_k - E[Y_k]) \quad (16)$$

- ▷ Écrite sous forme vectorielle, la variable aléatoire Z s'écrit

$$Z = X + \underline{\alpha}^T (\underline{Y} - \underline{m}_Y)$$

- ▷ Utilisant le même argument que celui dans (13) (mais cette fois ci sous forme vectorielle), on obtient le vecteur optimal suivant

$$\underline{\alpha}^* = -\underline{\Lambda}_Y^{-1} \underline{P}, \text{ où } \underline{P} = E \left[(\underline{Y} - \underline{m}_Y) (X - \underline{m}_X) \right]$$

- ▷ La variance de la variable aléatoire Z devient

$$\text{Var}(Z) = \text{Var}(X) - \underline{P}^T \underline{\Lambda}_Y^{-1} \underline{P}$$

Remarque(s) : La matrice $\underline{\Lambda}_Y$ est définie positive, son inverse l'est également, donc la forme quadratique $\underline{P}^T \underline{\Lambda}_Y^{-1} \underline{P}$ est une quantité positive. Ce qui signifie que $\text{Var}(Z)$ est plus petite que $\text{Var}(X)$.



- ▷ La technique d'**échantillonnage préférentiel** ("**importance sampling**") consiste à effectuer des changements de variables afin de réduire la variance des variables à simuler.
- ▷ Soit X une variable aléatoire de fonction de densité de probabilité $f_X(x)$ et Y obtenue par une transformation sur X , c'est à dire $Y = h(X)$. On a

$$E[Y] = \int h(t)f_X(t)dt = \int h(y)\frac{f_X(y)}{g_Y(y)}g_Y(y)dy \quad (17)$$

- ▷ L'intégrale (17) peut être interprétée comme la valeur moyenne d'une variable aléatoire Z obtenue par une transformation de la variable aléatoire Y ayant comme fonction de densité de probabilité $g_Y(y)$, où

$$Z(Y) = h(Y)\frac{f_X(Y)}{g_Y(Y)}, \text{ et } E[Z] = E[Y] \quad (18)$$

- ▷ La moyenne empirique de Z est par conséquent

$$\hat{E}[Z] = \hat{m}_Z = \frac{1}{N^*} \sum_{k=1}^{N^*} Z(Y_k)$$

où les N^* variables aléatoires Y_k sont indépendantes et de même fonction de densité de probabilité $g_Y(y)$.



- ▷ La **précision** de l'estimateur $\hat{E}[Z]$ est donnée par

$$N^* \geq \left[Q^{-1} \left(\frac{\eta}{2} \right) \right]^2 \frac{1}{k^2} \frac{\text{Var}(Z)}{(E[Z])^2}, \text{ avec } Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{u^2}{2}} du \quad (19)$$

- ▷ Pour réduire N^* , il faut chercher à réduire $\text{Var}(Z)$. Le rapport de réduction d'une telle précision est par définition

$$\lambda = \frac{\text{Var}(Z)}{\text{Var}(X)} \quad (20)$$

- ▷ Le problème à résoudre est de déterminer $g_Y(y)$ permettant de réduire sensiblement $\text{Var}(Z)$.
- ▷ Pour lectures complémentaires, voir [1, 2, 3, 4, 5, 6, 7, 8].



- [1] Dupire.
Monte Carlo simulation : Methodologies and applications for pricing and risk management.
Risk, 1998.
- [2] P. Glasserman.
Monte Carlo methods in financial engineering.
Springer, 2003.
- [3] P. Jackel.
Monte Carlo Methods in finance.
Wiley, 2002.
- [4] E. P. Kloeden, P. E. and H. Schurz.
Numerical solution of SDE through computer experiments.
Springer, 2nd edition, 1997.



- [5] P. E. Kloeden and E. Platen.
Numerical solution of stochastic differential equations.
Springer, 1992.
- [6] E. P. R. S. Lapeyre, B.
Méthodes de Monte Carlo pour les équations de transport et de diffusion.
Springer, Mathématiques et Applications 29, 1998.
- [7] S. T. W. V. Press, W. and B. Flannery.
Numerical Recipes in C: The Art of Scientific Computing.
Cambridge University Press, 1992.
- [8] R. Y. Rubinstein.
Simulation and the Monte Carlo Method.
Wiley, 1981.