Adaptive stratified sampling

awalee

SUMMARY

Introduction

Modeling risk factors is a fundamental part of risk management and it introduces changes on the day-to-day pricing and hedging on transactions within financial markets. This requires an effective and particularly fast implementation of the numerical methods used in pricing.

To overcome this issue, banks have mainly developed methodologies based on the classic Monte Carlo method, which is very costly in terms of computational time and resources. This cost becomes larger when the number of assets that are analyzed increases, each risk factor can require several thousands of simulated paths, increasing the complexity of the computational process. As an alternative to the usual computing methods, many banks have decided to tackle the issue of Monte Carlo computation cost by using GPUs instead of the usual architecture using only CPUs. This allows the development of a parallel computing framework which is ideal to estimate the parameters used in the derivatives pricing model.

Variance reduction methods are also widely used because they provide a better accuracy. Unfortunately, in practice, such methods can be an overhead due to the fact that either the gain is not significant or the implementation is complicated. But we believe that such methods, if enhanced and adapted, could be of great utility.

In this paper, we will focus on adaptive Stratified Sampling (SS). It is a method of sampling from a population which can be partitioned into subpopulations. It is used in MC to reduce variance. We propose a new method whose aim is to reach the minimal possible variance when using SS. The method takes advantage from sampling from collectively exhaustive and mutually exclusive subpopulations called stratums. This leads to more variance reduction. The challenge is to find the optimal proportion of simulations in each stratum so that the variance reduction can be as effective as possible.

The first two sections will give a formal presentation of MC and SS in a general framework, that is to say, the variable of which we seek to compute the expectation could be any financial asset or metric such as an option payoff or a value at risk…Then we will move to the main subject of the article which is adaptive SS. Finally, we will illustrate the new method by pricing a vanilla option.

1 Monte Carlo

The basic principle of Monte Carlo is to use the Strong Law of Large Numbers: if $(X_k)_{k\geq 1}$ denotes a sequence of independent realizations of an integrable random variable X (i.e. identically distributed random variables) then:

$$\overline{X}_n := \frac{\sum_{k=1}^n X_k}{n} \rightarrow m_X := \mathbb{E}[X] \text{ as n} \rightarrow \infty \text{ almost surely.}$$

The error can be controlled thanks to the Central Limit Theorem which says that: If X is square integrable then:

$$\sqrt{n}(\overline{X}_n - \mathbb{E}[X]) \rightarrow N(0, Var(X)) \text{ as n} \rightarrow \infty \text{ in distribution}$$

where Var(X) is the variance of X.

Therefore, the confidence interval at level $1-\alpha$ of the Monte Carlo simulation is given by:

$$CI_n := \left[\overline{X}_n - q_{1-\frac{\alpha}{2}}\sqrt{\frac{Var(X)}{n}}, \overline{X}_n + q_{1-\frac{\alpha}{2}}\sqrt{\frac{Var(X)}{n}}\right]$$

where $q_{1-\frac{\alpha}{2}}$ is defined as $\mathbb{P}(N(0,1) \leq q_{1-\frac{\alpha}{2}}) := 1 - \frac{\alpha}{2}$.

The rate of convergence for MC is then $O(\frac{1}{\sqrt{n}})$.

We refer to [1] for more details about MC.

2 Stratified Sampling

Stratified sampling is a method of sampling from a population which can be partitioned into subpopulations. It is most used in finance for variance reduction purposes when pricing derivatives with MC.

Next, we will present the method formally.

Principle of stratification sampling

In the following section F will denote the function of interest of which we want to compute the expectation.

The stratification method is based on the following relation: For any partition $(\mathbb{A}_i, i \in \{1,...I\})$ of the space $\mathbb{X}$ of possible sates taken by the random variable X we have:

$$\mathbb{E}[F(X)] = \sum_{i=1}^{I} \mathbb{P}(X \in \mathbb{A}_i)\mathbb{E}[F(X)|X \in \mathbb{A}_i].$$

A usual case is when the probabilities $\mathbb{P}(X \in \mathbb{A}_i), i \in \{1,...I\}$ are known, but the expectations $\mathbb{E}[F(X)|X \in \mathbb{A}_i]$ are not. In such case if we know how to generate a random variable that follows the conditional law $P(X|X \in \mathbb{A}_i)$, an estimator of $\mathbb{E}[F(X)]$ is given by:

$$\sum_{i=1}^{I} \mathbb{P}(X \in \mathbb{A}_i)\frac{1}{n_i}\sum_{k=1}^{n_i} F(X_k^i)$$

where the random variables $((X_k^i), k \leq n_i)$ are independent, identically distributed (i.i.d.) and follow the conditional distribution $P(X|X \in \mathbb{A}_i)$.

The implementation of this method requires the following steps:
- Select the number I of strata to get a partition $(\mathbb{A}_i, i \in \{1,,I\})$ of $\mathbb{X}$.
- Select a random variable X such that $\mathbb{P}(X \in \mathbb{A}_i)$ is explicitly computable and such that we know how to simulate i.i.d. copies of it following the law $P(X|X \in \mathbb{A}_i)$.
- Select the allocation i.e. the number of simulations that we will generate under the law $P(X|X \in \mathbb{A}_i)$, given the constraint that the total number of simulations is n.

In the following, we fix number I of the stratums $(\mathbb{A}_i, i \in \{1,,I\})$ and we assume that the variable of stratification X satisfies all the conditions in the previously mentioned steps.

For more details about sampling techniques, we refer to [2].

We will now consider the method in practice by analysing the allocation policy and the associated estimator.

Allocation policy

Defining an allocation policy means choosing $n_1, \ldots, n_I$ such that:

$$n_1 + \ldots + n_I = n$$

where $n_i, i \in \{1,..., I\}$ is the number of simulations that we sample from the law $\mathbb{P}(X|X \in \mathbb{A}_i)$.

It is equivalent to define $q_1, \ldots, q_I$ such that: $q_i \geq 0, i \in \{1,..., I\}$ and $\sum_{i=1}^{I} q_i = 1$, with every q_i representing the proportion of samples that are allocated to stratum i. We then have $n_i = nq_i, i \in \{1,..., I\}$.

We assume hereafter that all the variables are defined on the same probability space and:

H1:$F(X)$ has a moment of order 2 and hence has finite variance.

H2: $\mathbb{P}(X \in \mathbb{A}_i)$ is explicitly computable and is strictly positive for all elements $\mathbb{A}_i$ of the partition of $\mathbb{X}$.

The followings subsections present the SS estimator and analyze its characteristics.

Stratified sampling estimator

Let $(X_k^i, k \geq 0, i \in \{1,..., I\})$ be some independent random variables such that for all $i \in \{1,..., I\}$, the $(X_k^i, k \geq 0)$ have the same conditional law: $\mathbb{P}(X|X \in \mathbb{A}_i)$.

Let $(q_i, i \in \{1,..., I\})$ be the chosen allocation . The SS estimator is:

$$\hat{\mu}_n(q_{1:I}) = \sum_{i=1}^{I} \frac{p_i}{n_i} \sum_{k=1}^{n_i} F(X_k^i)$$

where $p_i := \mathbb{P}(X \in \mathbb{A}_i)$.

Bias of the estimator

We denote by $\mu_i := \mathbb{E}[F(X)|X \in \mathbb{A}_i]$ and $\sigma_i^2 := Var[F(X)|X \in \mathbb{A}_i]$ and let $\hat{Y}_n$ denote the usual Monte Carlo estimator.

Bias:

Straight calculation shows that:

$$\mathbb{E}[\hat{\mu}_n(q_{1:I})] = \sum_{i=1}^{I} p_i \mu_i.$$

The estimator is therefore unbiased.

Consistency

For all $i \in \{1,..., I\}$, the law of large numbers ensures that when $n \to \infty$:

$$\frac{1}{n_i} \sum_{k=1}^{n_i} F(X_k^i) \to \mathbb{E}[F(X)|X \in \mathbb{A}_i], \text{ almost surely}$$

and:

$$\hat{\mu}_n(q_{1:I}) \to \sum_{i=1}^{I} p_i \mathbb{E}[F(X)|X \in \mathbb{A}_i] = \mathbb{E}[F(X)], \text{ almost surely.}$$

Therefore, the estimator is consistent.

Variance

Let $n_i = nq_i$ be the number of samples in $\mathbb{A}_i$. As samples are independent, we have:

$$\begin{aligned} Var(\hat{\mu}_n(q_{1:I})) &= \textstyle\sum_{i=1}^{I} \frac{p_i^2}{n_i} Var[F(X)|X \in \mathbb{A}_i] \\ &= \frac{1}{n} \textstyle\sum_{i=1}^{I} \frac{p_i^2}{q_i} Var[F(X)|X \in \mathbb{A}_i] \\ &= \frac{1}{n} \textstyle\sum_{i=1}^{I} \frac{p_i^2}{q_i} \sigma_i^2. \end{aligned}$$

In the following section, we present the minimization problem whose aim is to find the optimal allocation $q_{1:I}$ that corresponds to the minimal variance.

Minimization problem

The variance of $\hat{\mu}_n(q_{1:I})$ depends on the allocation policy $q_{1:I}$. We can therefore seek the optimal allocation that makes the variance minimal.

Optimizing the simulation allocation amounts to solve the following minimization problem:

$$\min_{q_{1:I} \in S_I} \sum_{i=1}^{I} \frac{p_i^2}{q_i} \sigma_i^2$$

where

$$S_I := \left\{ q_{1:I} \in \mathbb{R}_+^I | \sum_{i=1}^{I} q_i = 1 \right\}$$

Next, we consider proportional allocation which is a practical and effective allocation policy in terms of reducing variance of the associated estimator.

Proportional allocation

When $q_i = p_i$ i.e. the allocation of each stratum i equals its weight , it is said to be proportional.

In this case, the variance becomes:

$$Var[\hat{\mu}_n(q_{1:I})] = \frac{1}{n} \sum_{i=1}^{I} p_i \sigma_i^2.$$

We will demonstrate that this choice is sub-optimal.

On the one hand:

$$\begin{aligned} Var[F(X)] &= \mathbb{E}[F(X)^2] - \mathbb{E}[F(X)]^2 \\ &= \textstyle\sum_{i=1}^{I} p_i \mathbb{E}[F(X)^2|X \in \mathbb{A}_i] - (\sum_{i=1}^{I} p_i \mu_i)^2 \\ &= \textstyle\sum_{i=1}^{I} p_i \{\sigma_i^2 + \mu_i^2\} - (\sum_{i=1}^{I} p_i \mu_i)^2. \end{aligned}$$

In the last line we used the fact that:

$$\begin{aligned} \mathbb{E}[F(X)^2|X \in \mathbb{A}_i] &= Var[F(X)|X \in \mathbb{A}_i] + \mathbb{E}[F(X)|X \in \mathbb{A}_i]^2 \\ &= \sigma_i^2 + \mu_i^2. \end{aligned}$$

On the other hand, $nVar[\hat{\mu}_n(q_{1:I})] = \sum_{i=1}^{I} p_i \sigma_i^2$.

So by denoting μ_n^{MC} the usual Monte estimator constructed from n i.i.d. copies with the same law as X, and by α_n the difference $nVar[\hat{\mu}_n(q_{1:I})] - nVar[\mu_n^{MC}]$ we can write:

$$\begin{aligned}
\alpha_n &= nVar[\hat{\mu}_n(q_{1:I})] - nVar[\mu_n^{MC}] \\
&= nVar[\hat{\mu}_n(q_{1:I})] - Var[F(X)] \\
&= (\sum_{i=1}^I p_i\mu_i)^2 - \sum_{i=1}^I p_i\mu_i^2 \\
&= 2(\sum_{i=1}^I p_i\mu_i)^2 - 1 \times (\sum_{i=1}^I p_i\mu_i)^2 - \sum_{i=1}^I p_i\mu_i^2 \\
&= 2\sum_{i=1}^I p_i\mu_i \sum_{j=1}^I p_j\mu_j - \sum_{i=1}^I p_i \times (\sum_{i=1}^I p_i\mu_i)^2 - \sum_{i=1}^I p_i\mu_i^2 \\
&= \sum_{i=1}^I p_i(2\mu_i \sum_{j=1}^I p_j\mu_j - (\sum_{i=1}^I p_i\mu_i)^2 - \mu_i^2) \\
&= -\sum_{i=1}^I p_i(\mu_i - \sum_{j=1}^I p_j\mu_j)^2 \leq 0.
\end{aligned}$$

Thus, the SS estimator with proportional allocation has a lower variance than the usual Monte Carlo estimator.

There is another simpler way to see that proportional allocation is a a sub-optimal choice. Let's denote $\sigma(\{X \in \mathbb{A}_i\}, i \in \{1, ..., I\})$ the σ-field spanned by the partition $\{X \in \mathbb{A}_i\}, i \in \{1, ..., I\}$.
We have:

$$\begin{aligned}
\sum_{i=1}^I p_i\sigma_i^2 &= \sum_{i=1}^I \mathbb{E}[1_{\{X \in \mathbb{A}_i\}}(F(X) - \mathbb{E}[F(X)|X \in \mathbb{A}_i])^2] \\
&= \mathbb{E}[(F(X) - \mathbb{E}[F(X)])^2 | \sigma(\{X \in \mathbb{A}_i\}, i \in I)] \\
&\leq \mathbb{E}[(F(X) - \mathbb{E}[F(X)])^2] \\
&= Var[F(X)].
\end{aligned}$$

We just need to multiply by $\frac{1}{n}$ the resulted inequality to get:

$$Var[\hat{\mu}_n(q_{1:I})] \leq Var[\mu_n^{MC}]$$

which proves our claim.

The equality holds if and only if:

$$\mathbb{E}[F(X)|\sigma(\{X \in \mathbb{A}_i\}, i \in I)] = \mathbb{E}[F(X)]$$

or, equivalently if and only if:

$$\mathbb{E}[F(X)|X \in \mathbb{A}_i] = \mathbb{E}[F(X)], i \in \{1, ..., I\}.$$

Therefore, this choice although simple to implement, it always reduces the variance of the estimator and can be very useful since the optimal one is not easy to reach.

This method of stratification is known as the quota method.

The following section is about an adaptive method whose aim is to reach the optimal allocation.

3 Adaptive Stratified sampling

The optimal choice is the solution to the previously mentioned constrained minimization problem. A straightforward application of Cauchy-Schwarz inequality shows that:

$$\begin{aligned}
\sum_{i=1}^I \frac{p_i^2}{q_i}\sigma_i^2 &= \sum_{i=1}^I \frac{p_i^2}{q_i}\sigma_i^2 \times 1 \\
&= \sum_{i=1}^I \frac{p_i^2}{q_i}\sigma_i^2 \times \sum_{i=1}^I \sqrt{q_i}^2 \\
&\geq (\sum_{i=1}^I \frac{p_i\sigma_i}{\sqrt{q_i}}\sqrt{q_i})^2 \\
&= (\sum_{i=1}^I p_i\sigma_i)^2
\end{aligned}$$

Consequently, the solution to the constrained minimization problem of interest is given by:

$$q_i^* = \frac{p_i\sigma_i}{\sum_{j=1}^I p_j\sigma_j}, i \in \{1, ..., I\}$$

and the lower bound in the previous inequality is the associated variance:

$$(\sum_{i=1}^I p_i\sigma_i)^2.$$

Although finding the solution of the optimization problem is relatively straightforward, its implementation is not because the $(\sigma_i, i \in \{1, ..., I\})$ are not known in general. Some attempts have been made to circumvent this problem. In this note, we propose an adaptive method.

More precisely, the conditional variances $\sigma_i^2, i \in \{1, ..., I\}$ are first estimated using some of the available simulations so that we can compute the $q_i^*, i \in \{1, ..., I\}$, then the rest is used to compute the stratified sampling estimator.
This is especially useful in general for some financial payoff $F(X)$ of which the conditional variances has no closed form or requires heavy numerical calculations.

We suggest that for a total number of simulations equal to n, we use n^β for the estimation of $\sigma_i^2, i \in \{1, ..., I\}$ and the rest, $n - n^\beta$, will be used for the computation of the expectation of interest using the optimal allocation estimator. β is a real number in $[0, 1]$.

This β parameterization allows us to seek the best trade-off between the number of simulations used for conditional variances estimation and those used for the computation of the expectation of interest.

In our numerical application this will be applied to the payoff of a vanilla option.

The following section details our methodology for selecting the best β.

4 Numerical Application

We will apply the previous results to the pricing of a vanilla call option. We will consider a geometric Brownian motion (Black-Scholes world) for the underlying stochastic equation.
Therefore the function F will be the payoff of a call option and X will be a standard normal variable.

We will set the volatility to $\sigma = 0.1$, the risk-free rate to $r = 0.05$, the maturity to $T = 1$ and the current underlying price to 50.

By assuming that the risk free rate is constant, the value of the Vanilla Call is:

$$e^{-rT}\mathbb{E}[max(S(T) - K, 0)]$$

where

$$S(T) = S(0)exp\left\{(r - \frac{\sigma^2}{2})T + \sigma\sqrt{T}X\right\}$$

and $\mathbb{E}$ denotes the expectation with respect to the risk neutral probability measure.
X is a standard normal variable.

The number of strata I is set to 10.
The strata $(\mathbb{A}_i, i \in \{1, ..., I\})$ are built as follows:

$$\mathbb{A}_i = \left\{x \in \mathbb{R}, z_{\frac{i-1}{I}} \leq x \leq z_{\frac{i}{I}}\right\}$$

where z_γ is the the γ-quantile of the standard normal distribution.

Thus straightforward calculation shows that:

$$p_i = \frac{1}{I}, i \in \{1, ..., I\}.$$

For a total number of simulations equal to $n = 1000000$, we use n^β for the estimation of the conditional variances with respect to each stratum $\mathbb{A}_i$. During this first process of estimation we will use proportional allocation. This is our best choice as we already showed that it reduces variance. Each stratum will contain

$$\lfloor q_i n^\beta \rfloor = \lfloor p_i n^\beta \rfloor$$

with $p_i = \frac{1}{I} = 0.1$.

We deemed that we must have at least three simulations in each stratum. Since there are ten, we should have $n^\beta \geq 30$. We also deemed that the number of simulations used to compute our payoff should not be less than half the total number of simulation, which corresponds to the condition $n^\beta \leq \frac{n}{2}$.

The previous two conditions roughly imply that the range of β values becomes:

$$0.25 \leq \beta \leq 0.95$$

In order to sample from the conditional law $\mathcal{L}(X|X \in \mathbb{A}_i)$, with $\mathcal{L}$ denoting the probability law $\mathcal{N}(0, 1)$, we begin by sampling u from the standard uniform distribution: $u \sim \mathcal{U}[0, 1]$. Then the random variable:

$$\Phi^{-1}\left(\Phi(z_{\frac{i-1}{I}}) + u(\Phi(z_{\frac{i}{I}}) - \Phi(z_{\frac{i-1}{I}}))\right)$$

with Φ the cumulative distribution function of the standard normal distribution $\mathcal{N}(0, 1)$, has the target conditional law.

Since we have a closed formula for the price of the vanilla call in Black-Scholes model, the comparison between Monte Carlo and adaptive SS will be easy in terms of accuracy. We will compare the two methods in terms of accuracy, variance reduction and computation time. We will run $n = 1000000$ MC simulations. In practice, we cannot afford to price in such costly conditions.

We price a call with the following parameters: The spot is 50, the strike is 50, volatility equals 10% and the risk free rate is 5%.

The closed formula gives a price of 3.40248. This will be our benchmark in terms of accuracy.

The next subsection details the results of our comparison.

Comparison results

The MC estimator is:

$$\frac{\sum_{j=1}^{n} max\left(S(T)_j - K, 0\right)}{n}$$

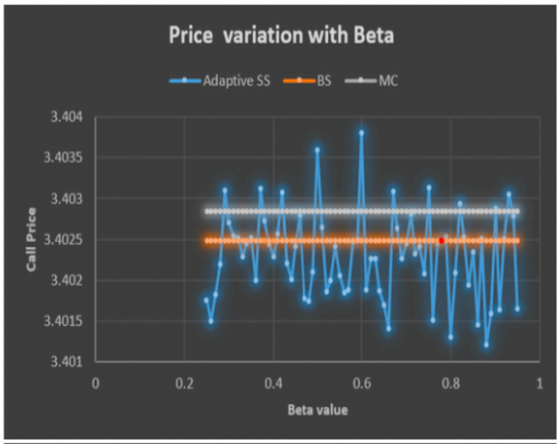
The results may differ depending on the power of the machine used, the compiler and the method of generation of the random uniform variables.

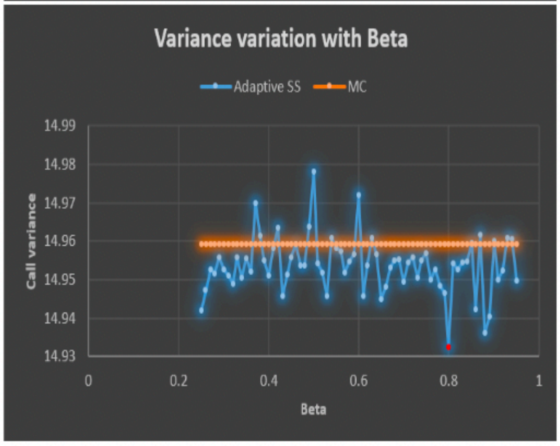
In our case, we used a machine with the following characteristics:
-Processor:Intel(R) Core(TM) i7-8565U CPU @ 1.80GHz 1.99GHz.
-Installed memory (RAM): 8.00 GB (7.82 GB usable).
-System type: 64-bit Operating System, x64-based processor.

The implementation was done in "C++" with Microsoft Visual Studio compiler. We generate the random normal variables using the inverse cumulative distribution function and the uniform random variables are generated by the function "rand()" in the <random> package.

The price we got is 3.40284 with a variance of 14.9593 and this took 0.676 second.

For adaptive SS, the graphs below show the results in terms of pricing and variance reduction.





Next, we will analyse the previous results and see which method is better in our case.

Analysis

We notice that, in terms of accuracy, there are some β values where the new method outperforms MC and others where MC gives better results. The idea is to compare the best performing β with MC.

One striking thing is that the variance is reduced for almost all β values for this method relatively to classic MC.

When it comes to computation time, MC outperforms adaptive SS as it took only 0.676 second whereas adaptive SS took on average 1.302 second. This was expected because conditional expectation computation is an overhead.

We move now to decide which value of β gives the best result and try to come by some general setting that is likely to give the best result when using adaptive SS.

As shown in "Price variation with Beta", accuracy is best for $\beta = 0.78$. So starting by looking around this value for any vanilla payoff seems to be a good idea.

For the variance, as previously mentioned, it is reduced for almost all points. The average of the variance is 14.9537 against 14.9593 for classic MC. Thus adaptive stratified sampling is effective.

The three points with the lowest variance with respect to the other values are in the range $\beta \geq 0.8$, which is expected as for such values the number of simulations by stratum that are used for the estimation of the optimal proportion is relatively high. Thus, the estimation is more efficient. The point $\beta = 0.8$ is the best performing one in terms of variance reduction as its variance is 14.9323.

The best β values are in red in the two graphs.

To conclude the choice of $\beta = 0.78$, which exactly matches the benchmark price (3.40248) and has a variance of 14.9485, is the optimal one as it outperforms MC in accuracy and has a lower variance as well. This point represents the best trade-off between accuracy and variance reduction. We believe that it could be optimal for other vanilla payoffs, or at least it will be a good starting point.

We can state that as a conclusion, some of the methods can be enhanced further and lead to better performances. Depending on the function of which we want to compute the expectation, the performances of the various methods vary . We need to study the payoff function and make a trade-off between the criteria (accuracy, variance reduction and computation time) to choose one of them.

Conclusion

In this paper, we have presented a new method of variance reduction based on stratified sampling and thus have showed that it is possible to propose new derivatives pricing techniques to enhance existing ones.

We first presented classic Monte Carlo and the stratified sampling technique. Then the adaptive stratified sampling method has been illustrated in the case of a vanilla option using a Geometric Brownian Motion for the underlying which is one of the most fundamental processes in quantitative finance.

The proposed method provides a significant variance reduction without compromising accuracy.

Nevertheless, it remains to be seen how we can effectively reach the optimal allocation when the number of possible simulations is relatively low.

References

[1] Gilles Pages. *Introduction to Numerical Probability for Finance.* 2017.

[2] Ravindra SinghNaurang Singh Mangat. *Elements of Survey Sampling.*

A propos d'Awalee

Cabinet de conseil indépendant spécialiste du secteur de la Finance.

Nous sommes nés en 2009 en pleine crise financière. Cette période complexe nous a conduits à une conclusion simple : face aux exigences accrues et à la nécessité de faire preuve de souplesse, nous nous devions d'aider nos clients à se concentrer sur l'essentiel, à savoir leur performance.

Pour accomplir cette mission, nous nous appuyons sur trois ingrédients : habileté technique, savoir-faire fonctionnel et innovation.

Ceci au service d'une ambition : dompter la complexité pour simplifier la vie de nos clients.

«*Run the bank*» avec Awalee !

Contactez-nous

Ronald LOMAS
Partner
rlomas@awaleeconsulting.com
06 62 49 05 97