Monte Carlo Simulation

Kaiza Amouh

March 9, 2020

Outline

Outline

1. Monte Carlo Simulation
   - Monte Carlo Principle
   - Applications of Monte Carlo
2. Variance Reduction
3. Quasi Monte Carlo : low discrepancy sequences
4. American Monte Carlo

Principle of Monte Carlo Simulation

The Monte Carlo method is based on the Strong Law of Large Numbers (SLLN)

Strong Law of Large Numbers (SLLN)

If $(X_k)_{k\geq 1}$ denotes a sequence of independent copies of an integrable random variable X, defined on a probability space$(\Omega, \mathcal{A}, \mathbb{P})$

$$\mathbb{P}(d\omega) - a.s. \quad \overline{X}_N(\omega) := \frac{X_1(\omega) + ... + X_N(\omega)}{N} \xrightarrow{N \to +\infty} m_x := \mathbb{E}(X)$$

Important issues related to Monte Carlo

- What is the rate of convergence of the method ?
- How can the resulting error be controlled ?

Rate of convergence

- The (weak) rate of convergence in the *SLLN* is ruled by the *Central Limit Theorem (CLT)*

$$\sqrt{N}\left(\overline{X}_N - m_x\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \sigma_x^2\right) \quad as \quad N \to +\infty$$

where $\sigma_x^2 = Var(X) := \mathbb{E}\left[\left(X - \mathbb{E}(X)\right)^2\right]$

- The quadratic rate of convergence (*i.e.* the rate in $L^2(\mathbb{P})$) is exactly

$$\left\|\overline{X}_N - m_x\right\|_2 = \frac{\sigma_x}{\sqrt{N}}$$

a.s. rate of convergence : *Law of the Iterated Logarithm (LIL)*

$$\limsup_{N \to +\infty} \sqrt{\frac{N}{2log(logN)}}\left(\overline{X}_N - m_x\right) = \pm\sigma_x$$

Data-driven control of the error : confidence interval

- Using the *CLT*, for every real numbers $a < b$, we have

$$\lim_N \mathbb{P}\left[\sqrt{N}\left(\frac{\overline{X}_N - m_x}{\sigma_x}\right) \in [a,b]\right] \xrightarrow{N\to+\infty} \mathbb{P}\left(\mathcal{N}(0,1) \in [a,b]\right) = \Phi(b) - \Phi(a)$$

  where Φ denotes the distribution function of the standard normal distribution

- Let $\alpha \in (0,1)$ denote a *confidence level* (close to 1), and let a_α be the two-sided α-quantile defined as the unique solution to the equation

$$\mathbb{P}\left(|\mathcal{N}(0,1)| \leq a_\alpha\right) = \alpha \quad i.e. \quad 2\Phi(a_\alpha) - 1 = \alpha$$

- Then one defines the random interval

$$J_N^\alpha := \left[\overline{X}_N - a_\alpha \frac{\sigma_x}{\sqrt{N}} \ , \ \overline{X}_N + a_\alpha \frac{\sigma_x}{\sqrt{N}}\right]$$

Data-driven control of the error : confidence interval

Probability of laying in the confidence interval

$$\mathbb{P}\left(m_x \in J_N^\alpha\right) = \mathbb{P}\left(\frac{\sqrt{N}}{\sigma_x}\left|\overline{X}_N - m_x\right| \le a_\alpha\right) \xrightarrow{N \to +\infty} \mathbb{P}\left(\left|\mathcal{N}(0,1)\right| \le a_\alpha\right) = \alpha$$

- At this stage, σ_x is unknown and must be estimated
- Applying the *SLLN* to the sequence of integrable random variables $\left(X_k^2\right)_{k \ge 1}$ yields

$$\overline{V}_N := \frac{1}{N-1}\sum_{k=1}^{N}\left(X_k - \overline{X}_N\right)^2 \xrightarrow{N \to +\infty} Var(X) = \sigma_x^2$$

- We hence perform a *companion Monte Carlo simulation* to estimate the variance σ^2

Vanilla Option Pricing in a Black-Scholes model

- A European *vanilla* option with maturity $T > 0$ is an option related to a European payoff $h_T := h(X_T)$ which only depends on X at time T
- In a complete market, the option premium at any time $t \in [0, T]$ is

$$V_t = e^{-r(T-t)}\mathbb{E}\left(h(X_T)/\mathcal{F}_t\right)$$

Examples of Vanilla options with strike K

$$\text{Vanila Call} : \ h(X_T) = (X_T - K)^+$$

$$\text{Best-of Call} : \ h(X_T, Y_T) = (max(X_T, Y_T) - K)^+$$

$$\text{Exchange Call Spread} : \ h(X_T, Y_T) = ((X_T - Y_T) - K)^+$$

Greeks computation in a Black-Scholes model

- The aim is to derive some representation of sensitivities as expectations
- These expectations can be evaluated using a Monte Carlo simulation in parallel with the premium computation
- Let $X_0 = x > 0$, and $X_t = x \; exp \left[(r - \frac{\sigma^2}{2})t + \sigma W_t \right]$
- Consider for every $x \in (0, \infty)$ the option price $f(x) := \mathbb{E}\left(h(X_T)\right)$

Proposition

- If $h : (0, +\infty) \to \mathbb{R}$ is differentiable and h' has a polynomial growth, then f is differentiable and

$$\forall x > 0, \quad f'(x) = \mathbb{E}\left(h'(X_T) \frac{X_T}{x} \right)$$

- If $h : (0, +\infty) \to \mathbb{R}$ is simply a Borel function with polynomial growth, then f is still differentiable and

$$\forall x > 0, \quad f'(x) = \mathbb{E}\left(h(X_T) \frac{W_T}{x \sigma T} \right)$$

Proof of the Proposition

- The first part can be obtained using the permutation of differentiation and integrals, together with the fact that

$$\frac{\partial}{\partial x} h(X_T) = h'(X_T)\frac{\partial X_T}{\partial x} = h'(X_T)\frac{X_T}{x}$$

- Under this assumption, let $\mu := r - \frac{\sigma}{2}$. We have:

$$\begin{aligned} f'(x) &= \int_{\mathbb{R}} h'\left(xe^{(\mu T+\sigma\sqrt{T}u)}\right) e^{(\mu T+\sigma\sqrt{T}u)} e^{-\frac{u^2}{2}} \frac{du}{\sqrt{2\pi}} \\ &= \frac{1}{x\sigma\sqrt{T}} \int_{\mathbb{R}} \frac{\partial h\left(xe^{(\mu T+\sigma\sqrt{T}u)}\right)}{\partial u} e^{-\frac{u^2}{2}} \frac{du}{\sqrt{2\pi}} \\ &= -\frac{1}{x\sigma\sqrt{T}} \int_{\mathbb{R}} h\left(xe^{(\mu T+\sigma\sqrt{T}u)}\right) \frac{\partial e^{-\frac{u^2}{2}}}{\partial u} \frac{du}{\sqrt{2\pi}} \end{aligned}$$

where we used an integration by part, taking advantage of the fact that : due to the polynomial growth of h, we have:

$$\lim_{|u|\to+\infty} h\left(xe^{(\mu T+\sigma\sqrt{T}u)}\right) e^{-\frac{u^2}{2}} = 0$$

Proof of the Proposition

We can then write:

$$f'(x) = \frac{1}{x\sigma\sqrt{T}} \int_{\mathbb{R}} h\left(xe^{(\mu T + \sigma\sqrt{T}u)}\right) ue^{-\frac{u^2}{2}} \frac{du}{\sqrt{2\pi}}$$

$$= \frac{1}{x\sigma T} \int_{\mathbb{R}} h\left(xe^{(\mu T + \sigma\sqrt{T}u)}\right) \sqrt{T}ue^{-\frac{u^2}{2}} \frac{du}{\sqrt{2\pi}}$$

$$= \frac{1}{x\sigma T}\mathbb{E}\left(h(X_T)W_T\right)$$

Computing the Value-at-Risk by Monte Carlo simulation

- Let X be a real-valued random variable, representative of a *loss*.
- For a given confidence level $\alpha \in (0,1)$, the *Value-at-Risk* at level α is any real number $V@R_{\alpha,X}$ satisfying

$$\mathbb{P}(X \leq V@R_{\alpha,X}) = \alpha \in (0,1)$$

- For convenience, one often assumes that the lowest solution of the above equation is the $V@R_{\alpha,X}$
- One naive way to compute $V@R_{\alpha,X}$ is to estimate the empirical distribution function of a Monte Carlo simulation at some points ξ lying in a grid $\Gamma := \{\xi_i, i \in I\}$

$$\widehat{F(\xi)_N} := \frac{1}{N}\sum_{k=1}^{N} \mathbf{1}_{\{X_k \leq \xi\}}, \quad \xi \in \Gamma$$

- Then one solves the equation $F(\hat{\xi})_N = \alpha$ (using an interpolation step of course)

Outline

Variance Reduction - Motivations

- For a large enough N, we have:

$$\mathbb{P}\left(m_x \in \left[\overline{X}_N - a_\alpha \frac{\sigma_x}{\sqrt{N}} \quad , \quad \overline{X}_N + a_\alpha \frac{\sigma_x}{\sqrt{N}}\right]\right) \approx 2\Phi(a_\alpha) - 1 = \alpha$$

- In numerical probability, we adopt a reverse point of view based on the needed accuracy $\epsilon > 0$
- To make $\overline{X}_N$ enter a confidence interval $[m_x - \epsilon, m_x + \epsilon]$ with a confidence level α, one needs to process a Monte Carlo simulation of size

$$N \geq N^x(\epsilon, \alpha) = \frac{a_\alpha^2 Var(X)}{\epsilon^2}$$

- For a given accuracy, $N^x(\epsilon, \alpha)$ grows linearly with the variance of X
- For a given variance, $N^x(\epsilon, \alpha)$ grows like the inverse of the square of the accuracy

Variance Reduction - (not so!) Naive approach

- Assume we know 2 random variables $X, X' \in L^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ satisfying $\mathbb{E}(X) = \mathbb{E}(X') = m \in \mathbb{R}$
- If both X and X' can be simulated with an equivalent cost (complexity), then we choose the one with the lowest variance !

Practical implementation

Assume there exists a random variable $Y \in L^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ such that

- $\mathbb{E}(Y)$ can be computed at a very low cost by a deterministic method (closed form, numerical analysis method)
- The random variable $(X - Y)$ can be simulated with the same complexity as X
- The variance $Var(X - Y) < Var(X)$

Then the random variable $X' = X - Y + \mathbb{E}(Y)$ can be simulated instead of X, and we have

$$\mathbb{E}(X') = \mathbb{E}(X - Y) + \mathbb{E}(Y) = \mathbb{E}(X) \quad and \quad Var(X') = Var(X - Y) < Var(X)$$

The random variable Y is called a *static control variate* for X

Variance Reduction - Pseudo Control Variate

- In option pricing, payoffs are usually non-negative
- For any random variable Y satisfying the two first conditions, together with $0 \leq Y \leq X$, then Y is a *good candidate* to reduce variance.
- The closer Y is to X, the smaller $(X - Y)$ is, and the smaller the variance is.

The *Jensen* inequality is an efficient tool to design pseudo-control variate, when dealing with path-dependent or multi-asset options.

Jensen inequality

Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ be a random variable and let $g : \mathbb{R} \to \mathbb{R}$ be a convex function. Suppose X and $g(X)$ are integrable. Then, for any sub-α-field $\mathcal{B}$ of $\mathcal{A}$,

$$g\left(\mathbb{E}(X/\mathcal{B})\right) \leq \mathbb{E}\left(g(X)/\mathcal{B}\right) \qquad \mathbb{P} - a.s.$$

In particular considering $\mathcal{B} = \{\emptyset, \Omega\}$ yields the above inequality for regular expectation i.e.

$$g\left(\mathbb{E}(X)\right) \leq \mathbb{E}\left(g(X)\right)$$

Pseudo Control Variate - Example of Basket options

- We consider a call option on a basket of d risky assets, with strike K

$$h_T = \left(\sum_{i=1}^{d} \alpha_i X_T^i - K \right)^+$$

where $\left(X^1, ..., X^d\right)$ models the price of d traded risky assets on a market and α_i are some positive weights satisfying $\sum_{1 \leq i \leq d} \alpha_i = 1$

- Then the convexity of the exponential implies that

$$0 \leq e^{\sum_{1 \leq i \leq d} \alpha_i log\left(X_T^i\right)} \leq \sum_{i=1}^{d} \alpha_i X_T^i$$

so that

$$h_T \geq h'_T := \left(e^{\sum_{1 \leq i \leq d} \alpha_i log\left(X_T^i\right)} - K \right)^+ \geq 0$$

Pseudo Control Variate - Example of Basket options

- The motivation for this example is that a (possibly correlated) d-dimensional Black-Scholes model, $\sum_{1\leq i\leq d}\alpha_i log\left(X_T^i\right)$ still has a normal distribution.
- Therefore, the Call-like European option h_T' has a closed form.
- The correlated d-dimensional Black-Scholes model can be defined by the following system of SDE's:

$$dX_t^i = X_t^i\left(rdt + \sum_{j=1}^{i}\sigma_{ij}dW_t^j\right), \quad t\in[0,T], \quad X_0^i := x_i > 0, \quad i = 1,...,d$$

where $W = \left(W^1,...,W^d\right)$ is a standard d-dimensional Brownian motion and $\sigma = [\sigma_{ij}]_{1\leq i,j\leq d}$ is a lower triangular matrix.
- Its closed form solution is given by:

$$X_t^i = x_i \exp\left(\left(r - \frac{\sigma_{i.}^2}{2}\right)t + \sum_{j=1}^{i}\sigma_{ij}W_t^j\right), \quad where \quad \sigma_{i.}^2 = \sum_{j=1}^{i}\sigma_{ij}^2$$

Example of Basket options - Implementation

- Step 1 : Compute analytically the expectation $\mathbb{E}(e^{-rT})h'_T$.

$$\sum_{1\leq i\leq d} \alpha_i log\left(X_T^i/x_i\right) \stackrel{d}{=} \mathcal{N}\left(\left(r - \frac{1}{2}\sum_{1\leq i\leq d}\alpha_i\sigma_{i.}^2\right)T \;\; ; \;\; \alpha'\sigma\sigma'\alpha T\right)$$

where α is the column vector with components $\alpha_i, i = 1, ..., d$

$$e^{-rT}\mathbb{E}(h'_T) = Call_{BS}\left(\prod_{i=1}^{d} x_i^{\alpha_i}, K, \left(r - \frac{1}{2}\sum_{1\leq i\leq d}\alpha_i\sigma_{i.}^2 + \frac{1}{2}\alpha'\sigma\sigma'\alpha\right), \sqrt{\alpha'\sigma\sigma'\alpha}, T\right)$$

- Step 2 : Joint simulation of the couple $\left(h_T, h'_T\right)$

$$e^{-rT}\left(h_T - h'_T\right) = e^{-rT}\left(\left(\sum_{i=1}^{d}\alpha_i X_T^i - K\right)^+ - \left(e^{\sum_{1\leq i\leq d}\alpha_i log\left(X_T^i\right)} - K\right)^+\right)$$

Pseudo Control Variate - Example of Asian options

$$Let \quad X_t = X_0 \exp\left(\left(r - \frac{\sigma^2}{2}\right)t + \sigma W_t\right) \quad and \quad h_T = \phi\left(\frac{1}{T}\int_0^T X_t dt\right)$$

h_T is a generic Asian payoff with ϕ is non-negative, and X_t a regular Black-Scholes dynamics

The (standard) Jensen inequality applied to the probability measure $\frac{1}{T}\mathbf{1}_{[0,T]}(t)dt$ implies

$$\begin{aligned}\tfrac{1}{T}\int_0^T X_t dt \quad &\geq X_0 \exp\left(\tfrac{1}{T}\int_0^T \left(\left(r - \tfrac{\sigma^2}{2}\right)t + \sigma W_t\right)dt\right)\\ &= X_0 \exp\left(\left(r - \tfrac{\sigma^2}{2}\right)\tfrac{T}{2} + \tfrac{\sigma}{T}\int_0^T W_t dt\right)\end{aligned}$$

Now

$$\int_0^T W_t dt = TW_T - \int_0^T s dW_s = \int_0^T (T-s)dW_s$$

Pseudo Control Variate - Example of Asian options

Then we have

$$\frac{1}{T}\int_0^T W_t dt \overset{d}{=} \mathcal{N}\left(0\ ,\ \frac{1}{T^2}\int_0^T s^2 ds\right) \overset{d}{=} \mathcal{N}\left(0\ ,\ \frac{T}{3}\right)$$

This suggests to rewrite the right hand side of the above inequality in a "Black-Scholes asset" style, namely:

$$\frac{1}{T}\int_0^T X_t dt \geq X_0 e^{-\left(\frac{r}{2}+\frac{\sigma^2}{12}\right)T} \exp\left(\left(r-\frac{\sigma^2}{6}\right)T + \sigma\frac{1}{T}\int_0^T W_t dt\right)$$

This naturally leads to introduce the so-called *Kemna-Vorst* pseudo-control variate

$$h_T^{KV} := \phi\left(X_0 e^{-\left(\frac{r}{2}+\frac{\sigma^2}{12}\right)T} \exp\left(\left(r-\frac{\sigma^2}{6}\right)T + \sigma\frac{1}{T}\int_0^T W_t dt\right)\right)$$

which is clearly of Black-Scholes type and satisfies $h_T \geq h_T^{KV}$

Example of Asian options - Implementation

- Step 1 : The random variable h_T^{KV} is an admissible control variate as soon as the vanilla option related to the payoff $\phi(X_T)$ has a closed form

$$e^{-rT}\mathbb{E}\left(\phi\left(X_T\right)\right) = premium_{BS}^{\phi}\left(X_0, T, r, \sigma\right)$$

then we have

$$e^{-rT}\mathbb{E}\left(h_T^{KV}\right) = premium_{BS}^{\phi}\left(X_0 e^{-\left(\frac{r}{2}+\frac{\sigma^2}{12}\right)T}, T, r, \frac{\sigma}{\sqrt{3}}\right)$$

- Step 2 : One has to simulate independent copies of the couple $\left(h_T, h_T^{KV}\right)$, i.e. the couple

$$\left(\left(W_t\right)_{t\in[0,T]} \quad , \quad \frac{1}{T}\int_0^T W_t dt\right)$$

Example of Asian options - Computation of time integrals

- In order to perform this simulation, one needs to compute time integrals in both payoffs
- Numerical integration methods are applied, such as the mid-point quadrature formula

$$\frac{1}{T}\int_0^T f(t)dt \approx \frac{1}{n}\sum_{k=1}^{n} f\left(\frac{2k-1}{2n}T\right)$$

- Keep in mind that the functions f of interest are here given by

$$f(t) = \phi\left(X_0 \exp\left(\left(r - \frac{\sigma^2}{2}\right)t + \sigma W_t(\omega)\right)\right) \quad or \quad f(t) = W_t(\omega)$$

for the first and the second payoff function respectively.

- Their regularity is α-Holder, $\alpha < \frac{1}{2}$ (locally for the payoff h_T)

Pseudo Control Variate - Example of Best-of-call options

- The payoff of a *Best-of-call* option is given by

$$h_T = \left(max\left(X_T^1, X_T^2\right) - K\right)^+$$

- Using the convexity inequality (an application of Jensen inequality)

$$\sqrt{ab} \leq max(a, b) \implies h_T \geq h'_T := \left(\sqrt{X_T^1 X_T^2} - K\right)^+$$

- In a 2-dimensional Black-Scholes model, the option with payoff h'_T has a closed form
- The procedure can be improve by noting that more generally, $a^\theta b^{1-\theta} \leq max(a, b)$ when $\theta \in (0, 1)$

Using parity equations to produce control variates

- In derivatives pricing, one can always build a control variate as soon as the payoff of interest satisfies a so-called *parity equation*
- These *parity equations* are model free, so they can be applied for various specifications of the dynamics of the underlying asset.
- We denote by $(S_t)_{t\in[0,T]}$ the risky asset
- We work under the risk-neutral probability:

$$\left(e^{-rt}S_t\right)_{t\in[0,T]} \quad \textit{is a martingale on the scenarii space } (\Omega, \mathcal{A}, \mathbb{P})$$

Vanilla Call-Put parity ($d = 1$)

- Let the premia of *Call* and *Put* option be denoted as follows

$$Call_0 = e^{-rT}\mathbb{E}\left((S_T - K)^+\right) \quad and \quad Put_0 = e^{-rT}\mathbb{E}\left((K - S_T)^+\right)$$

- Since $(S_T - K)^+ - (K - S_T)^+ = S_T - K$ and $(e^{-rt}S_t)_{t\in[0,T]}$ is a martingale,

$$Call_0 - Put_0 = S_0 - e^{-rT}K$$

- So that $Call_0 = \mathbb{E}(X) = \mathbb{E}(X')$ with

$$X := e^{-rT}(S_T - K)^+ \quad and \quad X' := e^{-rT}(K - S_T)^+ + S_0 - e^{-rT}K$$

- As a result, one sets $Y = X - X' = e^{-rT}S_T - S_0$
- Y turns out to be the terminal value of a martingale null at time 0 !

Asian Call-Put parity ($d = 1$)

- We assume that the averaging phase runs from T_0 to T

$$Call_0^{As} = e^{-rT}\mathbb{E}\left(\left(\frac{1}{T - T_0}\int_{T_0}^{T} Stdt - K\right)^{+}\right)$$

$$Put_0^{As} = e^{-rT}\mathbb{E}\left(\left(K - \frac{1}{T - T_0}\int_{T_0}^{T} Stdt\right)^{+}\right)$$

- Using *Fubini's Theorem* and the fact that $(e^{-rt}S_t)_{t\in[0,T]}$ is a martingale, we have:

$$Call_0^{As} - Put_0^{As} = S_0\frac{1 - e^{-r(T-T_0)}}{r(T - T_0)} - e^{-rT}K$$

Asian Call-Put parity ($d = 1$)

- Once again we can write $Call_0^{As} = \mathbb{E}(X) = \mathbb{E}(X')$ with

$$X := e^{-rT}\left(\frac{1}{T-T_0}\int_{T_0}^{T} Stdt - K\right)^+$$

$$X' := S_0\frac{1-e^{-r(T-T_0)}}{r(T-T_0)} - e^{-rT}K + e^{-rT}\left(K - \frac{1}{T-T_0}\int_{T_0}^{T} Stdt\right)^+$$

- Which leads to

$$Y = e^{-rT}\frac{1}{T-T_0}\int_{T_0}^{T} Stdt - S_0\frac{1-e^{-r(T-T_0)}}{r(T-T_0)}$$

Antithetic method : negatively correlated variables

- We now assume that X and Y have not only the same expectation m_X but also the same variance
- If X and Y are negatively correlated, we set

$$\chi := \frac{X+Y}{2}$$

- We have:

$$Var(\chi) = \frac{1}{4}\left(Var(X) + Var(Y) + 2Cov\,(X,Y)\right) = \frac{Var(X) + Cov\,(X,Y)}{2}$$

- The size $N^X(\epsilon,\alpha)$ and $N^\chi(\epsilon,\alpha)$ of the simulation using X and χ respectively, to enter a given interval $[m-\epsilon, m+\epsilon]$ is given by

$$N^X = \left(\frac{a_\alpha}{\epsilon}\right)^2 Var(X) \quad and \quad N^\chi = \left(\frac{a_\alpha}{\epsilon}\right)^2 Var(\chi)$$

Antithetic method : complexity analysis

- Let κ be the complexity for simulating X. Simulating χ hence requires 2κ complexity
- One would prefer simulating χ only if $2\kappa N^{\chi} < \kappa N^{X} \Leftrightarrow Cov(X, Y) < 0$

Proposition : Antithetic generation

- Let $\phi, \psi : (\mathbb{R}, \mathcal{B}or(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}or(\mathbb{R}))$ be two monotone functions with the same monotony
  Let $T : (\mathbb{R}, \mathcal{B}or(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}or(\mathbb{R}))$ be a nonincreasing transform and let $Z : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ be a random variable.
  Assume that $\phi(Z), \psi(Z), \psi(T(Z)) \in L^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$. Then
  $$Cov\left(\phi(Z), \psi(Z)\right) \geq 0 \quad and \quad Cov\left(\phi(Z), \psi\left(T(Z)\right)\right) \leq 0$$
- If furthermore $\psi = \phi$ and $Z \overset{d}{=} T(Z)$, then $X = \phi(Z)$ and $Y = \phi\left(T(Z)\right)$ are identically distributed and satisfy $Cov(X, Y) \leq 0$
  In that case, the random variables X and Y are called ***antithetic***

Antithetic method : Application

The antithetic random variable method applies in two situations

- The symmetric random variable Z : $\longrightarrow Z \stackrel{d}{=} -Z$ i.e. $T(z) = -z$
- The $[0, L]$-valued random variable Z such that $Z \stackrel{d}{=} L - Z$ i.e. $T(z) = L - z$

European option pricing in a B&S model

For $Z \rightsquigarrow \mathcal{N}(0,1)$, the payoff is $h_T = \phi(X_T) = \phi\left(X_0 e^{\left(r - \frac{\sigma^2}{2}\right)T} + \sigma\sqrt{T}Z\right)$

The function $z \mapsto \phi\left(X_0 e^{\left(r - \frac{\sigma^2}{2}\right)T} + \sigma\sqrt{T}z\right)$ is monotone, and $W_T \stackrel{d}{=} -W_T$

Uniform distribution on the unit interval : $U \rightsquigarrow \mathcal{U}([0,1])$

If ϕ is monotone on $[0,1]$, then $Var\left(\frac{\phi(U)+\phi(1-U)}{2}\right) \leq \frac{1}{2} Var\left(\phi(U)\right)$

Basic principle of importance sampling

- Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E})$ be an E-valued random variable?
- Let μ be a σ-finite *reference* measure on $(E, \mathcal{E})$ so that

$$\exists \textit{ a density } f : (E, \mathcal{E}) \to (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+)) \textit{ such that } \mathbb{P}_X = f.\mu$$

- In practice, we will assume $E = \mathbb{R}$ and μ=Lebesgue measure
- Let $h \in L^1(\mathbb{P}_X)$. Then,

$$\mathbb{E}\left(h(X)\right) = \int_{\mathbb{R}} h(x)\mathbb{P}_X(dx) = \int_{\mathbb{R}} h(x)f(x)\mu(dx)$$

- For any μ-a.s. positive probability density g defined on $(\mathbb{E}, \mu)$, on has

$$\mathbb{E}\left(h(X)\right) = \int_{\mathbb{R}} h(x)f(x)\mu(dx) = \int_{\mathbb{R}} \frac{h(x)f(x)}{g(x)} g(x)\mu(dx) = \mathbb{E}\left(\frac{h(Y)f(Y)}{g(Y)}\right)$$

Importance sampling's variance reduction

$$
\begin{aligned}
\mathbb{E}\left[\left(\frac{h(Y)f(Y)}{g(Y)}\right)^2\right] &= \mathbb{E}\left[\left(\frac{hf}{g}(Y)\right)^2\right] \\
&= \int_{\mathbb{R}} \left(\frac{h(x)f(x)}{g(x)}\right)^2 g(x)\mu(dx) \\
&= \int_{\mathbb{R}} h(x)^2 \frac{f(x)}{g(x)} f(x)\mu(dx) \\
&= \mathbb{E}\left(h(X)^2 \frac{f}{g}(X)\right)
\end{aligned}
$$

- Simulating $\frac{hf}{g}(Y)$ rather than $h(X)$ will reduce the variance if and only if

$$
\mathbb{E}\left(h(X)^2 \frac{f}{g}(X)\right) < \mathbb{E}\left(h(X)^2\right)
$$

How to design and implement importance sampling ?

- The goal is to replace X by Y so that $\frac{h}{fg}(Y)$ is "*closer*" than $h(X)$ to their common mean
- Consider a *Call* with strike K and $X_0 \ll K$ (deep out-of-the-money)
- Most scenarios ω would yield $(X_T(\omega) - K)^+ = 0$
- Small number of events with positive payoffs $\Rightarrow$ rough estimation of $\mathbb{E}\left((X_T - K)^+\right)$
- If we can switch from $(X_t)_{t\in[0,T]}$ to $(Y_t)_{t\in[0,T]}$ so that Y_T takes most of its values in $[K, +\infty)$, then

$$\mathbb{E}\left((X_T - K)^+\right) = \mathbb{E}\left((Y_T - K)^+ \frac{f}{g}(Y_T)\right)$$

Parametric importance sampling

- We introduce a family of random variables $(Y_\theta)_{\theta \in \Theta}$ such that $g_\theta > 0$
- Assume $\exists\ \theta_0 \in \Theta$ such that $Y_{\theta_0} = X$
- The problem becomes a parametric optimization problem

$$\min_{\theta \in \Theta} \left\{ \mathbb{E}\left[\left(h(Y_\theta) \frac{f}{g_\theta}(Y_\theta) \right)^2 \right] = \mathbb{E}\left(h(X)^2 \frac{f}{g_\theta}(X) \right) \right\}$$

- Of course, there is no reason why the solution to the above problem should be θ_0 (unless the parametric model is inappropriate)

Example with Cameron-Martin formula

- In the 1-dimensional Black-Scholes model, the premium of an option with payoff ϕ is

$$e^{-rT}\mathbb{E}\left(\phi(X_T)\right) = \mathbb{E}\left(h(Z)\right) = \int_{\mathbb{R}} h(z)e^{-\frac{z^2}{2}}\frac{dz}{\sqrt{2\pi}}$$

where $Z \overset{d}{=} \mathcal{N}(0,1)$ and $h(z) = e^{-rT}\phi\left(X_0 e^{\left(r-\frac{\sigma^2}{2}\right)T+\sigma\sqrt{T}z}\right)$

- From now on, we will focus on

$$\mathbb{E}\left(h(Z)\right) = \int_{\mathbb{R}} h(z)f(z)dz \quad \textit{where} \quad f(z) = \frac{e^{\frac{z^2}{2}}}{\sqrt{2\pi}}$$

- The idea is to introduce the parametric family

$$Y_\theta = Z + \theta, \quad \theta \in \Theta := \mathbb{R}$$

Example with Cameron-Martin formula

$$\textit{With } g_\theta(y) = \frac{e^{-\frac{(y-\theta)^2}{2}}}{\sqrt{2\pi}}\ ,\quad \textit{we have } \frac{f}{g_\theta}(y) = e^{-\theta y + \frac{\theta^2}{2}}$$

Cameron-Martin formula

$$\mathbb{E}\left(h(Z)\right) = e^{\frac{\theta^2}{2}}\,\mathbb{E}\left(h(Y_\theta)e^{-\theta Y_\theta}\right) = e^{\frac{\theta^2}{2}}\,\mathbb{E}\left(h(Z+\theta)e^{-\theta(Z+\theta)}\right)$$
$$= e^{-\frac{\theta^2}{2}}\,\mathbb{E}\left(h(Z+\theta)e^{-\theta Z}\right)$$

The next step is to choose a "good" θ which significantly reduces the variance

$$\min_{\theta\in\mathbb{R}}\left[e^{\frac{\theta^2}{2}}\,\mathbb{E}\left(h^2(Z)e^{-\theta Z}\right) = e^{-\theta^2}\,\mathbb{E}\left(h^2(Z+\theta)e^{-2\theta Z}\right)\right]$$

Heuristic Sub-optimal approaches

We could "re-center" the simulation of X around K by replacing Z by $Z+\theta$

$$\mathbb{E}\left(X_0 \exp\left(\left(r-\frac{\sigma^2}{2}\right)T+\sigma\sqrt{T}\left(Z+\theta\right)\right)\right)=K \Longrightarrow \theta := -\frac{log(X_0/K)+rT}{\sigma\sqrt{T}}$$

Or solve the similar, although slightly different equation

$$\mathbb{E}\left(X_0 \exp\left(\left(r-\frac{\sigma^2}{2}\right)T+\sigma\sqrt{T}\left(Z+\theta\right)\right)\right)=e^{-rT}K \Longrightarrow \theta := -\frac{log(X_0/K)}{\sigma\sqrt{T}}$$

Or search for θ such that $\mathbb{P}\left(X_0 \exp\left(\left(r-\frac{\sigma^2}{2}\right)T+\sigma\sqrt{T}\left(Z+\theta\right)\right)<K\right)=\frac{1}{2}$

$$\Longrightarrow \theta := -\frac{log(X_0/K)+\left(r-\frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}$$

Outline

Quasi-Random Numbers

- Quasi-Random numbers are generated through purely deterministic sequences, and these sequences don't even attempt to emulate the behavior of *independent uniform random variables.*
- *They rather cover the space in d dimensions with fewer gaps than independent random variables would normally.*
- *Pseudo-Random Numbers may result in large variance when used in a Monte Carlo simulation*
- *The so-called Quasi-Random Numbers present excellent properties in terms of variance reduction.*
- *The essential property of such a sequence is the so-called Low Discrepancy*

Discrepancy of a Sequence

Let $(\xi_n)_{x \geq 1}$ be a $[0,1]^d$-valued sequence. One defines the discrepancy of (ξ_n) as follows:

Discrepancy at the origin or "Star Discrepancy"

$$D_n^*(\xi) := \sup_{x \in [0,1]^d} \left| \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{[\![0,x]\!]}(\xi_k) - \prod_{i=1}^{d} x^i \right|$$

Extreme Discrepancy

$$D_n^\infty(\xi) := \sup_{x,y \in [0,1]^d} \left| \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{[\![x,y]\!]}(\xi_k) - \prod_{i=1}^{d} \left(y^i - x^i\right) \right|$$

Portemanteau Theorem

The following assertions are equivalent

1. $(\xi_{x\geq 1})$ is uniformly distributed on $[0,1]^d$
2. For every $x \in [\![0,1]\!] := [0,1]^d$,
$$\frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{[\![0,x]\!]}(\xi_k) \to \lambda_d([\![0,x]\!]) := \prod_{i=1}^{d} x^i \quad as\ n \to +\infty$$
3. $D_n^{\infty}(\xi) \to 0 \quad as\ n \to +\infty$
4. $D_n^{*}(\xi) \to 0 \quad as\ n \to +\infty$

Portemanteau Theorem

The following assertions are equivalent

1. $(\xi_{x\geq 1})$ is uniformly distributed on $[0,1]^d$
2. (Weyl's criterion) For every integer $p \in \mathbb{N}^d \ \{0\}$
$$\frac{1}{n}\sum_{k=1}^{n} e^{2i\pi(p|\xi_k)} \to 0 \quad \textit{as } n \to +\infty \quad (\textit{where } i^2 = -1).$$
3. (Bounded Riemann integrable function) For every bounded λ_d-a.s. continuous Lebesgue-measurable function $f : [0,1]^d \to \mathbb{R}$
$$\frac{1}{n}\sum_{k=1}^{n} f(\xi_k) \to \int_{[0,1]^d} f(x)\lambda_d(dx) \quad \textit{as } n \to +\infty$$

Van der Corput and Halton sequence

Let $p_1, ..., p_d$ be the first d prime numbers (or simply, d pairwise prime numbers).
the $d-$dimensional Halton sequence is defined, for every $n \geq 1$, by:

$$\xi_n = (\Phi_{p_1}(n), ...; \Phi_{p_d}(n))$$

where the so-called "*radical inverse functions*" Φ_p are defined by

$$\Phi_p(n) = \sum_{k=0}^{r} \frac{a_k}{p^{k+1}} \;\; with \;\; n = a_0 + a_1 p + ... + a_r p^r$$

and $a_i \in \{0, ..., (p-1)\}$, $a_r \neq 0$, denotes the p-adic expansion of n

Van der Corput and Halton sequence

The discrepancy of Halton sequence can be bounded using the *Chinese Remainder Theorem* (a.k.a *Theoreme chinois* in French)

$$D_n^*(\xi) \leq \frac{1}{n}\left(d + \prod_{i=1}^{d}\left((p_i - 1)\frac{log\ n}{2\ log\ p_i} + \frac{p_i+2}{2}\right)\right), \quad n \geq 1$$

Van der Corput sequence ($d = 1$)

When $d = 1$, the sequence $(\Phi_p(n))_{n\geq1}$ is called the *the p-adic Van der Corput sequence* (and the integer p needs not to be prime).
For $d = 2$, one easily checks that the first terms of the $VdC(2)$ sequence are as follows

$$\xi_1 = \frac{1}{2},\ \xi_2 = \frac{1}{4},\ \xi_3 = \frac{3}{4},\ \xi_4 = \frac{1}{8},\ \xi_5 = \frac{5}{8},\ \xi_1 = \frac{3}{8},\ \xi_7 = \frac{7}{8}...$$

The Kakutani sequences

Based on Kakutani adding machine : p-adic addition on $[0,1]$

Principle of p-adic addition on $[0,1]$

A p-adic addition is a binary operation defined on the set of p-adic expansions.
Let $\oplus_p$ denote this addition
If $x, y \in [0,1]$ have their respective regular p-adic expansions as

$$x = \overline{0, x_1 x_2 \dots x_k \dots}^p \quad \textit{and} \quad y = \overline{0, y_1 y_2 \dots y_k \dots}^p$$

Then

$$(x \oplus_p y)_k = (x_k + y_k)\,\mathbf{1}_{\{x_{k-1}+y_{k-1} \le p-1\}} + (1 + x_k + y_k)\,\mathbf{1}_{\{x_{k-1}+y_{k-1} \ge p\}}$$

With the convention $x_1 = y_1 = 0$

Convention: $1 = \overline{0.(p-1)(p-1)(p-1)\dots}^p$

Example: $0.12333\ldots \oplus_{10} 0.412777\ldots = 0.535011\ldots$

The Kakutani sequences

For every $y \in [0,1]$, one defines the associated p-adic rotation with angle y by
$$T_{p,y}(x) := x \oplus_p y$$

Construction of the sequence

Let $p_1, ..., p_d$ denote the first d prime numbers, $y_1, ..., y_d \in (0,1)$, where y_i is a p_i-adic rational number satisfying $y_i \geq \frac{1}{p_i}$, $i = 1, ..., d$ and $x_1, ..., x_d \in [0,1]$. Then the Kakutani sequence $(\xi)_{n\geq 1}$ is defined by

$$\xi_n := \left(T^{n-1}_{p_i,y_i}(x_i)\right)_{1\leq i\leq d}, \quad n \geq 1$$

$$D^*_n(\xi) \leq \frac{1}{n}\left(1 + \prod_{i=1}^{d}\left((p_i - 1)\left\lfloor \frac{log(p_i n)}{log(p_i)}\right\rfloor\right)\right) = O\left(\frac{(log\ \ n)^d}{n}\right) \ \ as \ \ n \to +\infty$$

The Faure sequences

Let p be the smallest prime integer not lower than d (i.e. $p \geq d$).

d-dimensional Faure sequence for every $n \geq 1$

$$\xi_n = \left(\Phi_p(n-1),\ C_p\left(\Phi_p(n-1), ...,\ C_p^{d-1}\left(\Phi_p(n-1)\right)\right)\right)$$

where Φ_p still denotes the p-adic radical inverse functionand, for every p-adic rational number u with (regular) p-adic expansion $u = \sum_{k \geq 0} u_k p^{-(k+1)} \in [0,1]$

$$C_p(u) = \sum_{k\geq 0} \underbrace{\left(\sum_{j \geq k} \binom{j}{k} u_j \ mod.\ p\right)}_{\in \{0,...,(p-1)\}} p^{-(k+1)}$$

The discrepancy at the origin satisfies

$$D_n^*(\xi) \leq \frac{1}{n}\left(\frac{1}{d!}\left(\frac{p-1}{2logp}\right)^d (logn)^d + O\left((logn)^{d-1}\right)\right)$$

The Niederreiter sequences

Let $q \geq d$ be the smallest primary integer not lower than d (a primary integer reads $q = p^r$ with p prime).

The $(0, d)$-Niederreiter sequence is defined for $n \geq 1$ by:

$$\xi_n = (\Psi_{q,1}(n-1), \Psi_{q,2}(n-1), ..., \Psi_{q,d}(n-1))$$

where

$$\Psi_{q,i}(n) := \sum_j \psi^{-1}\left(\sum_k C^{(i)}_{(j,k)}\Psi(a_k)\right) q^{-j}$$

and $\Psi : \{0, ..., (q-1)\} \to \mathbb{F}_q$ is a one-to-one correspondence between $\{0, ..., (q-1)\}$ and the finite field $\mathbb{F}_q$ with cardinal q satisfying $\Psi(0) = 0$ and

$$C^{(i)}_{(j,k)} = \binom{k}{j-1}\Psi(i-1)$$

Particular cases of Niederreiter sequences

Sobol sequence

When $q = 2^r$, with $2^{r-1} < d \leq 2^r$, the Neiderreiter sequence coincides with Sobol's sequence

Faure sequence

When p is the lowest prime number not lower than d, the Neiderreiter sequence coincides with the Faure sequence

- The sequences of this family all have discrepancy satisfying an upper bound with a structure similar to that of the Faure sequence.

Outline

Bermudan options

- Let us consider $0 = t_0 < t_1 < ... < t_n = T$ a discrete subdivision of $[0, T]$
- A Bermudan option gives the right to the buyer to exercise at any date $t_0, ..., t_n$ and pays $f\left(S_{t_k}\right)$ at time t_k
- Let $\left(\tilde{V}_t\right)_{0 \leq t \leq T}$ denote the associated hedging portfolio. Then :
  - at date $T = t_n$ we have $V_{t_n} = f\left(S_{t_n}\right)$
  - at date $T = t_{n-1}$ we have

$$\begin{aligned} V_{t_{n-1}} &= max\left[f\left(S_{t_{n-1}}\right)\ ;\ e^{-r(t_n - t_{n-1})}\mathbb{E}\left(f\left(S_{t_n}\right)/\mathcal{F}_{t_{n-1}}\right)\right] \\ &= max\left[f\left(S_{t_{n-1}}\right)\ ;\ e^{-r(t_n - t_{n-1})}\mathbb{E}\left(V_{t_n}/\mathcal{F}_{t_{n-1}}\right)\right] \end{aligned}$$

  - The same way we get $\forall k \in \{0, ..., n-1\}$

$$\begin{cases} V_T &= f(S_T) \\ V_{t_k} &= max\left[f\left(S_{t_k}\right)\ ;\ e^{-r(t_{k+1} - t_k)}\mathbb{E}\left(V_{t_{k+1}}/\mathcal{F}_{t_k}\right)\right] \end{cases}$$

Remarks

- Note that the process $(S_t)_{0 \leq t \leq T}$ is Markovian then

$$\mathbb{E}\left(V_{t_{k+1}}/\mathcal{F}_{t_k}\right) = \mathbb{E}\left(V_{t_{k+1}}/S_{t_k}\right)$$

- A Bermudan option is more expensive than an European option.
- If we let $n \to +\infty$, then the price of the Bermudan option tends to the price of an American option

Martingale Stopping theorem

Now if we consider $\forall k \in \{0, ..., n-1\}$

$$\begin{cases} V_T & = f(S_T) \\ V_{t_k} & = max\left[f\left(S_{t_k}\right)\ ;\ e^{-r(t_{k+1}-t_k)}\mathbb{E}\left(V_{t_{k+1}}/\mathcal{F}_{t_k}\right)\right] \end{cases}$$

Then

$$V_{t_k} = \underset{\tau \in \{t_k, \ldots, t_n\}}{sup} e^{-r(\tau - t_k)}\mathbb{E}\left[f\left(S_\tau\right)/S_{t_k}\right]$$

Theorem

The stopping time

$$\tau_k^* = inf\left\{t_i \in \{t_k, ..., t_n\} | f\left(S_{t_i}\right) \geq e^{-r(t_{i+1}-t_i)}\mathbb{E}\left[f\left(S_{t_{i+1}}\right)/S_{t_k}\right]\right\}$$

satisfies

$$V_{t_k} = e^{-r(\tau_k^* - t_k)}\mathbb{E}\left[f\left(S_{\tau_k^*}\right)/S_{t_k}\right]$$

Longstaff-Schwarz algorithm

- The sequence $(\tau_k^*)_{0\leq k\leq n}$ satisfies the dynamic programming principle

$$\begin{cases} \tau_n^* & = T \\ \tau_k^* & = t_k \mathbf{1}_{B_k} + \tau_{k+1}^* \mathbf{1}_{B_k^c}, \quad \textit{for } 0 \leq k \leq n-1 \end{cases}$$

where

$$B_k = \left\{ f\left(S_{t_k}\right) \geq \mathbb{E}\left[e^{-r(\tau_{k+1}^* - t_k)} f\left(S_{\tau_{k+1}^*}\right) \Big/ S_{t_k} \right] \right\}$$

- The expectation can be approximated by a regression, based on a basis function $(P_l)_{l\geq 1}$

$$\mathbb{E}\left[e^{-r(\tau_{k+1}^* - t_k)} f\left(S_{\tau_{k+1}^*}\right) \Big/ S_{t_k} \right] = \sum_{l\geq 1} \alpha_{k,l} P_l\left(S_{t_k}\right)$$

Regression

- Using the definition of the conditional expectation, the sequence $(\alpha_{k,l})_{l\geq 1}$ is the sequence that minimizes the distance

$$\mathbb{E}\left[\left(e^{-r(\tau^*_{k+1}-t_k)}f\left(S_{\tau^*_{k+1}}\right)-\sum_{l\geq 1}\alpha_{k,l}P_l\left(S_{t_k}\right)\right)^2\right]$$

- In practice, we need to truncate the sum $\sum_{l\geq 1}\alpha_{k,l}P_l\left(S_{t_k}\right)$ and approximate it by

$$\sum_{l\geq 1}^{L}\alpha_{k,l}P_l\left(S_{t_k}\right), \quad where \;\; L>1$$

Longstaff-Schwartz algorithm

- Simulate $\left(S^j_{t_0}, ..., S^j_{t_n}\right)_{1\leq j\leq M}$, M copies of $(S_{t_0}, ..., S_{t_n})$
- For all $1 \leq j \leq M$ we set $\tau_{j,n} := t_n = T$
- Then compute the sequence $\left(\alpha^j_{k,l}\right)_{1\leq l\leq L}$ that minimizes

$$\frac{1}{M}\sum_{j=1}^{M}\left[\left(e^{-r(\tau^*_{j,k+1}-t_k)}f\left(S^j_{\tau_{j,k+1}}\right)-\sum_{l\geq 1}^{L}\alpha_{k,l}P_l\left(S^j_{t_k}\right)\right)^2\right]$$

- For all $j \in \{1, ..., M\}$ we define

$$\tau_{j,k} = t_k \mathbf{1}_{A_{j,k}} + \tau_{j,k+1}\mathbf{1}_{A^c_{j,k}}, \quad for \;\; 0 \leq k \leq n-1$$

where

$$A_{j,k} = \left\{ f\left(S^j_{t_k}\right) \geq \sum_{l=1}^{L}\alpha^j_{k,l}P_l\left(S^j_{t_k}\right)\right\}$$

Price approximation

- For $k = 0$ the price of the Bermudan option is approximated by

$$\frac{1}{M}\sum_{j=1}^{M} e^{-r\tau_{j,0}} f\left(S^{j}_{\tau_{j,0}}\right)$$

- The Longstaff Schwarz algorithm converges in $\mathcal{L}^2$ as $L \to +\infty$
- For a fixed L, it converges almost surely as $M \to +\infty$

Computing the coordinates $(\alpha_{k,l})_{1\leq l\leq L}$ - Basic approach

- For a fixed time step t_k, this approach consists on simply solving the following system

$$\begin{pmatrix} P_1\left(S^1_{t_k}\right) & \dots & P_L\left(S^1_{t_k}\right) \\ \vdots & \vdots & \vdots \\ P_1\left(S^M_{t_k}\right) & \dots & P_L\left(S^M_{t_k}\right) \end{pmatrix} \begin{pmatrix} \alpha_{k,1} \\ \vdots \\ \alpha_{k,L} \end{pmatrix} = \begin{pmatrix} e^{-r(\tau_{1,k+1}-t_k)f\left(S^1_{\tau_{1,k+1}}\right)} \\ \vdots \\ e^{-r(\tau_{M,k+1}-t_k)f\left(S^M_{\tau_{M,k+1}}\right)} \end{pmatrix}$$

- Advantage : easy to implement
- Drawback : not a high accuracy

Computing the coordinates $(\alpha_{k,l})_{1\leq l\leq L}$ - Optimal approach

- For a fixed time step t_k we aim at computing the sequence $\left(\alpha^j_{k,l}\right)_{1\leq l\leq L}$ that minimizes

$$\frac{1}{M}\sum_{j=1}^{M}\left[\left(e^{-r(\tau^*_{j,k+1}-t_k)}f\left(S^j_{\tau_{j,k+1}}\right)-\sum_{l\geq 1}^{L}\alpha_{k,l}P_l\left(S^j_{t_k}\right)\right)^2\right]$$

- We differentiate the above quantity with respect to α_{k,l_0} and solve

$$\sum_{j=1}^{M}\left(e^{-r(\tau^*_{j,k+1}-t_k)}f\left(S^j_{\tau_{j,k+1}}\right)-\sum_{l\geq 1}^{L}\alpha_{k,l}P_l\left(S^j_{t_k}\right)\right)P_{l_0}\left(S^j_{t_k}\right)=0$$

Computing the coordinates $(\alpha_{k,l})_{1 \leq l \leq L}$ - Optimal approach

- This is equivalent to solving

$$\sum_{l=1}^{L} \left(\sum_{j=1}^{M} P_l\left(S^j_{t_k}\right) P_{l_0}\left(S^j_{t_k}\right) \right) \alpha_{k,l} = \sum_{j=1}^{M} e^{-r(\tau_{j,k+1}-t_k)} f\left(S^j_{\tau_{j,k+1}}\right) P_{l_0}\left(S^j_{t_k}\right)$$

- Let $H_{l_0,l} = \sum_{j=1}^{M} P_l\left(S^j_{t_k}\right) P_{l_0}\left(S^j_{t_k}\right)$. We need to solve

$$\sum_{l=1}^{L} H_{l_0,l}\alpha_{k,l} = \sum_{j=1}^{M} e^{-r(\tau_{j,k+1}-t_k)} f\left(S^j_{\tau_{j,k+1}}\right) P_{l_0}\left(S^j_{t_k}\right)$$

Computing the coordinates $(\alpha_{k,l})_{1 \leq l \leq L}$ - Optimal approach

We can write this in a matrix equation given by

$$H_{\alpha_k} = \sum_{j=1}^{M} e^{-r(\tau_{j,k+1} - t_k)} f\left(S^j_{\tau_{j,k+1}}\right) P\left(S^j_{t_k}\right)$$

where

$$\alpha_k = (\alpha_{k,1}, ..., \alpha_{K,L}) \quad \textit{and} \quad P\left(S^j_{t_k}\right) = \left(P_1\left(S^j_{t_k}\right), ..., P_L\left(S^j_{t_k}\right)\right)$$