Two topics in:

EFFICIENT RISK MANAGEMENT IN MONTE CARLO

Luca Capriotti

Quantitative Strategies - Credit Suisse

Baruch College
New York April 2, 9 2012
Outline

Module 1: Efficient Monte Carlo Sampling

- What are Monte Carlo methods good for:
  - Multidimensional Integrals and the curse of dimensionality
  - How Stochastic approaches break the spell in many cases
  - The main Limitation of Monte Carlo: Variance and Statistical Uncertainties

- Variance Reduction Techniques
  - Antithetic Variates
  - Control Variates
  - Importance Sampling
  - Case Study I: Least Squares Importance Sampling (LSIS)
Outline (cont’d)

Module 2: Efficient Risk Management in Monte Carlo

- Hedges and Price Sensitivities (Greeks)
  - Shortcoming of the Finite Differences (‘no time to think’) approach
- Classical Approaches
  - Likelihood Ratio Method
  - Case Study II: Reducing the Variance of Likelihood Ratio Greeks
  - Pathwise Derivative Method
- Modern Approaches
  - Adjoint Methods
  - The new deal: Adjoint Algorithmic Differentiation (AAD)
  - Case Study III: Real Time Counterparty Credit Risk Management in Monte Carlo
Module 1:
Efficient Monte Carlo Sampling
Monte Carlo Methods

(what has a the Principality of Monaco anything to do to scientific computations?)

Using random numbers to solve multi dimensional problems with probabilistic methods

Luca Capriotti – Efficient Risk Management in Monte Carlo
Multidimensional Integrals

- Many multidimensional problems in applied mathematics/natural sciences can be formulated as multidimensional integrals:
  \[ I = \int f(x) \, dx \]

- Numerical quadrature formulas are very efficient in 1 dimension:
  \[
  \hat{I} = \sum_{i=1}^{n} w_i f(x_i)
  \]
  Error: \[ |I - \hat{I}| = O(n^{-r}) \]

  - \( r = 2 \) Trapezoid (\( f^{(2)}(x) < \infty \))
  - \( r = 4 \) Simpson’s (\( f^{(4)}(x) < \infty \))
Multidimensional Integrals (cont’d)

- Quadrature in higher dimension:

\[
\hat{I} = \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \ldots \sum_{i_m=1}^{n} w_1^{(i_1)} w_2^{(i_2)} \ldots w_m^{(i_m)} f(x_1^{(i_1)}, \ldots, w_m^{(i_m)})
\]

With \( N \) function evaluations we can sample \( n = N^{1/m} \) points along each dimension.

Error (Bahvalov’ s thm):

\[
|I - \hat{I}| = O(n^{-r}) = O(N^{-r/m})
\]

\( (\partial^r f / \partial x_1^r < \infty, \ldots, \partial^r f / \partial x_m^r < \infty) \)

\[
\begin{array}{cccc}
m = & 1 & 2 & 3 & \ldots \\
N = & 10 & 10^2 & 10^3 & \ldots \\
\end{array}
\]

CURSE OF DIMENSIONALITY
Stochastic (Monte Carlo) Sampling

- It is always possible to choose a probability density function $P$ so that:

$$I = \int f(x) \, dx = \int dx \, G(x) \, P(x)$$

$$I = \mathbb{E}_P[G(x)]$$

- Law of large numbers:

$$I \sim \bar{I} = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} G(x_i) \quad x_i \sim P(x)$$
Central Limit Theorem and Monte Carlo Uncertainties

\[ I \approx \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} G(x_i) \pm \frac{\sum}{\sqrt{N_{MC}}} \]

- Variance of the Estimator:

\[ \Sigma^2 = \mathbb{E}_P[G(x)^2] - \mathbb{E}_P[G(x)]^2 \]

- Monte Carlo Estimator of the Variance:

\[ \Sigma^2 \approx \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} \left( G(x_i) - \bar{I} \right)^2 \]
Breaking the spell: Monte Carlo vs. Quadrature

- Quadrature in high dimension \( m \):

\[
\text{Quadrature Error} = O\left( N^{-1/m} \right)
\]

(for a quadrature that does not require function smoothness)

- Monte Carlo:

\[
\text{MC Error} = O\left( N_{MC}^{-1} \right)
\]

(finete estimator variance)

No dependency on the dimensionality of the problem, only on the number of MC samples
Variance and Statistical Uncertainty

- All done then? Well not quite...

\[
I \sim \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} G(x_i) \pm \frac{\sum}{\sqrt{N_{MC}}}
\]

- We need to cope with the Variance:

\[
\sum^2 = \mathbb{E}_P[G(x)^2] - \mathbb{E}_P[G(x)]^2
\]

If the variance is large one may require a large number of MC paths to reach an acceptable level of accuracy. If the variance is infinite the MC estimator is not very useful.
Infinite Variance Example:

- Consider this example:
  \[ I = \int_0^1 \frac{dx}{\sqrt{x}} \]

- The simplest MC calculation involves sampling \( x \) uniformly:
  \[ P(x) = 1(0 \leq x \leq 1) \]

- And use the MC estimator:
  \[ G(x) = \frac{1}{\sqrt{x}} \]
  \[ \text{Var}[G(x)] = \int_0^1 \frac{1}{x} = \infty \]

**Homework assignment:** simulate and plot the estimator and its error as a function of the number of MC replications.
Wait a second! What Integrals are we talking about?

- Simulation of an SDE:

\[
dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t
\]

Call Option:

\[
B(0, T)(X_T - K)^+
\]

\[
V = \mathbb{E}[B(0, T)(X_T - K)^+] = \int dX_T P(X_T|X_0)B(0, T)(X_T - K)^+
\]
Wait a second! What Integrals are we talking about?

- ‘Path’ simulation:

\[
dX_t = \mu(X_t, t)\,dt + \sigma(X_t, t)\,dW_t
\]

Euler Discretization:

\[
X_{i+1} = X_i + \mu(X_i, t_i)\Delta_i + \sigma(X_i, t_i)\sqrt{\Delta t_i}Z_{i+1} \quad Z_i \sim N(0, 1)
\]

\[
V = \int dX_1 \ldots dX_M G(X_1, \ldots, X_M) P(X_1, \ldots, X_M | X_0)
\]

\[
= \int dZ_1 \ldots dZ_M G(Z_1, \ldots, Z_M) P(Z)
\]

\[
P(Z) = (2\pi)^{-M/2} \exp\left(-Z \cdot Z / 2\right)
\]
Variance Reduction Techniques

Classical Methods:

- **Antithetic Variates:**
  Exploiting the reduction in variance that results when negatively correlated samples are produced and grouped together.

- **Control Variates:**
  Using the information on the error in estimates of known quantities to reduce the error in an estimate of a correlated unknown quantity.

- **Importance Sampling**
  Modifying the sampling distribution to increase the likelihood of sampling estimators where they are larger or more rapidly varying (i.e., where it matters most).
Antithetic Variates


- The intuition:
  - Consider two i.d. variates $X, Y$ with expectation $\mu$, variance $\sigma^2$ and covariance $\sigma_{XY}^2 = \mathbb{E}[(X - \mu)(Y - \mu)]$
  - Compare 3 estimators of the mean of the common distribution:
    \[
    X_i, \quad Y_i \quad \frac{X_i + Y_i}{2} \quad \frac{\sigma^2}{2} \quad \frac{\sigma^2 + \sigma_Y^2 + 2\sigma_{XY}^2}{4} \quad \frac{\sigma^2}{2} \left(1 + \rho_{XY}\right)
    \]
    
    In presence of negative correlation grouping helps to remove noise
Antithetic Variates (cont’d)

- In a typical context:

\[ V = \mathbb{E}[G(.)] \]

\[
G(U_1, \ldots, U_d) \quad U_i \sim \text{Unif}[0, 1] \quad G(Z_1, \ldots, Z_d) \quad Z_i \sim N(0, 1)
\]

\[
\tilde{G} = G(1 - U_1, \ldots, 1 - U_d) \quad \tilde{G} = G(-Z_1, \ldots, -Z_d)
\]

The transformed random variates have the same distribution

\[
\bar{G}_{AV} = \frac{1}{N_{MC}} \sum_{j=1}^{N_{MC}} \frac{G_j + \tilde{G}_j}{2} \quad [\bar{G}_{AV} - V] \Rightarrow N(0, \sigma_{AV}^2 / N_{MC})
\]

\[
\sigma_{AV}^2 = \text{Var}\left[\frac{G + \tilde{G}}{2}\right] = \frac{1}{2} \left(\text{Var}[G] + \text{Cov}[G, \tilde{G}]\right)
\]
When are Antithetic Variates beneficial?

- Computational cost of the AV estimator is roughly twice than the cost of the plain estimator.
- Therefore AV are beneficial if:

\[
\text{Var}[\bar{G}_{\text{AV}}] < \text{Var}\left[\frac{1}{2N_{\text{MC}}} \sum_{j=1}^{2N_{\text{MC}}} G_j\right]
\]

\[
\text{Cov}[G, \tilde{G}] < 0
\]

- This happens if \(G(Z_1, \ldots, Z_d)\) or \(G(U_1, \ldots, U_d)\) are monotonic functions. In high dimension this is a big constraint.
- If their Taylor expansion contains only odd powers the AV estimator has zero variance.
Antithetic Variates (Examples)

Homework assignment:

- Compute analytically the variance of the simple and AV estimators for the integrals below and verify your results implementing the MC sampling.

\[ I = \int_{0}^{1} e^{-x} \, dx \quad I = \int_{0}^{1} e^{-(x-1/2)^2} \, dx \]

- Investigate the efficacy of AV for a 6 months straddle on a lognormal equity asset for different levels of the moneyness and volatility (assume zero interest rates and dividends).
Control Variates


- We want to calculate $\mu_Y = \mathbb{E}[Y]$ and we happen to know $\mu_X = \mathbb{E}[X]$ for a random variable that we believe to be correlated.

- Introduce the CV estimator:

$$ Y_{CV}^i(\alpha) = Y_i - \alpha(X_i - \mu_X) $$

$$ \mathbb{E}[Y_{CV}(\alpha)] = \mathbb{E}[Y] \quad \text{(Unbiased)} $$

Variance:

$$ \text{Var}[Y_{CV}(\alpha)] = \text{Var}[Y] - 2\alpha \text{Cov}[X, Y] + \alpha^2 \text{Var}[X] $$

Optimal for:

$$ \alpha^* = \frac{\text{Cov}[X, Y]}{\text{Var}[X]} $$

$$ \text{Var}[Y_{CV}(\alpha^*)] = \text{Var}[Y](1 - \rho_{XY}^2) $$
Efficacy of Control Variates

\[ Y_{CV}^i = Y_i - \alpha^*(X_i - \mu_X) \]

\[ \alpha^* = \frac{\text{Cov}[X, Y]}{\text{Var}[X]} \]

\[ \text{Var}[Y_{CV}(\alpha^*)] = \text{Var}[Y](1 - \rho_{XY}^2) \]

- Computational cost of CV estimator is roughly the same of the standard one (control are usually cheap to compute).
- Therefore, as long as \( \rho_{XY} \neq 0 \), CV are always beneficial (CV never hurt). Sign of correlation is irrelevant.
- However, the variance reduction increases sharply with \( |\rho_{XY}| \). Hence the benefit of CV is significant only for relatively high level of correlations.

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>1</th>
<th>0.99</th>
<th>0.95</th>
<th>0.9</th>
<th>0.8</th>
<th>0.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>VR</td>
<td>( \infty )</td>
<td>50</td>
<td>10</td>
<td>5</td>
<td>2.7</td>
<td>1.96</td>
</tr>
</tbody>
</table>
Geometric interpretation

\[ \alpha^* = \frac{\text{Cov}[X, Y]}{\text{Var}[X]} = \frac{\sum_{j=1}^{N_{MC}} (X_j - \bar{X})(Y_j - \bar{Y})}{\sum_{j=1}^{N_{MC}} (X_j - \bar{X})^2} \]

Linear Regression Slope

\[ \bar{Y} \rightarrow \bar{Y}_{CV} = \bar{Y} - \alpha^* (\bar{X} - \mathbb{E}(X)) \]

\(\bar{Y}_{CV}(\alpha^*)\) is the value fitted by the line at \(\mathbb{E}(X)\)

In this example \(\bar{X}\) underestimates \(\mathbb{E}(X)\) and \(\bar{Y}\) is adjusted upwards accordingly.
Multiple Controls:

\[ Y_{CV}^i(\alpha) = Y_i - \alpha^T(X_i - \mu_X) \]

\[ \alpha, X_i \in \mathbb{R}^d \]

Linear Regression coefficient

\[ \alpha^* = \Sigma_X^{-1} \Sigma_{XY} \]

\[ (\Sigma_X)_{j,k} = \text{Cov}[X^j, X^k] \quad (\Sigma_{X,Y})_j = \text{Cov}[X^j, Y] \]

Optimal Variance:

\[ \text{Var}[Y_{CV}(\alpha^*)] = \text{Var}[Y](1 - R^2) \]

\[ R^2 = \Sigma_{XY}^T \Sigma_X^{-1} \Sigma_{XY} / \sigma_Y^2 \]
Control Variates (Examples)

**Homework assignment:**

- Apply CV to the integral below using as control the linear term in its Taylor expansion. Verify numerically the variance of the CV estimators.

\[
I = \int_0^1 e^{-x} \, dx
\]

- Consider the calculation of arithmetic Asian call and put options on a lognormal asset sampled every week for 6 months. Study the efficacy of the following CVs: underlying asset sampled at expiry; European option with the same expiry; geometric Asian call. Produce scatter plots of the arithmetic Asian payoff versus each of the controls considered and plot the best linear regression line passing through the sample average of the arithmetic Asian payoff and the control.
Importance Sampling


\[ V = E_P [G(Z)] = \int_D dZ \ G(Z) \ P(Z) \]
\[ Z = (Z_1, \ldots, Z_d) \]

A simple identity

\[ \int_D dZ \ G(Z) \ P(Z) = \int_D dZ \ \frac{G(Z)P(Z)}{\tilde{P}(Z)} \tilde{P}(Z) \]

A new weighted estimator:

\[ \tilde{V} = \frac{1}{N_{\text{MC}}} \sum_{i=1}^{N_{\text{MC}}} W(Z_i)G(Z_i) \quad Z_i \sim \tilde{P}(Z) \]

\[ W(Z) = \frac{P(Z)}{\tilde{P}(Z)} \]

Likelihood Ratio Weight
Importance Sampling (cont’d)

\[ \tilde{V} = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} W(Z_i)G(Z_i) \quad Z_i \sim \tilde{P}(Z) \]

\[ W(Z) = \frac{P(Z)}{\tilde{P}(Z)} \]

The variance of the new estimator

\[ \tilde{\Sigma}^2 = \int_D dZ \ (W(Z)G(Z) - \tilde{V})^2 \tilde{P}(Z) \]

critically depends on the choice of the sampling probability distribution.

Choose \( \tilde{P}(Z) \) to make the variance \( \tilde{\Sigma} \) as small as possible.
Optimal Sampling Distribution:

\[ P_{\text{opt}}(Z) = \frac{1}{V} G(Z)P(Z) \]

\[ \tilde{\Sigma}^2 = \int_D dZ \ (W(Z)G(Z) - V)^2 \tilde{P}(Z) \quad \text{Zero Variance!} \]

\[ \tilde{V} = \frac{1}{N_{\text{MC}}} \sum_{i=1}^{N_{\text{MC}}} W(Z_i)G(Z_i) = \frac{1}{N_{\text{MC}}} \sum_{i=1}^{N_{\text{MC}}} V \]

Too bad we don’t know \( V \)!

Nonetheless we can still try to find a sampling distribution that is as close as possible to the optimal one.
Importance Sampling of Singular Integrals

Let’s consider again an infinite variance integral:

\[ I = \int_0^1 \frac{dx}{\sqrt{x}} \quad P(x) = 1(0 \leq x \leq 1) \quad G(x) = \frac{1}{\sqrt{x}} \]

\[ \text{Var}[G(x)] = \int_0^1 \frac{1}{x} = \infty \]

Sampling density: \[ \tilde{P}(x) = \frac{1}{2\sqrt{x}} \]

Likelihood Ratio weight: \[ W(x) = P(x)/\tilde{P}(x) = 2\sqrt{x} \]

Estimator: \[ \tilde{G}(x) = 2 \]

\[ \text{Var}[\tilde{G}(x)] = 0 \]

Optimal density \[ G(x)P(x)/I \]
Case Study:

Least Squares Importance Sampling (LSIS)

Importance Sampling is generally formulated as an optimization problem

Trial sampling density: 

\[ \tilde{P}_\theta(Z) \]

Set of optimization parameters

\[ \tilde{\Sigma}^2 = \int_{D} dZ \ (W(Z)G(Z) - V)^2 \tilde{P}(Z) \]

ORIGINAL MEASURE

\[ \tilde{\Sigma}_\theta^2 = E_P [W_\theta(Z)G^2(Z)] - E_P [G(Z)]^2 \]
Least Squares Importance Sampling (LSIS)

Minimize the Variance

\[ \widetilde{\Sigma}_{\theta}^2 = E_P \left[ W_\theta(Z)G^2(Z) \right] - E_P \left[ G(Z) \right]^2 \]

... or equivalently minimize

\[ S_2(\theta) = E_P \left[ \left( W_\theta(Z)^{1/2}G(Z) - V_T \right)^2 \right] \]

MC estimator:

\[ \sim \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} \left( W_\theta(Z_i)^{1/2}G(Z_i) - V_T \right)^2 \]

\[ \sum_{i=1}^{N_{MC}} \left( y_i - f_\theta(x_i) \right)^2 \]

A LEAST SQUARES PROBLEM!
Least Squares Importance Sampling (LSIS)

Algorithm:

1) Choose a trial probability density and an initial value of the parameters $\theta$.

2) Generate a suitable number $N_{MC}$ of replications of the random variates $Z_i$.

3) Set: $x_i \rightarrow Z_i, y_i \rightarrow V_T, f_\theta(x_i) \rightarrow W_\theta(Z_i)^{1/2}G(Z_i)$

4) Feed the pairs $(x_i, y_i)$ into a non linear Least Square Fitter (e.g., Levenberg-Marquardt) to determine the optimal $\theta$. 

$$\sum_{i=1}^{N_{MC}} \left( y_i - f_\theta(x_i) \right)^2$$
Least Squares Importance Sampling (LSIS)

Correlated Sampling makes the approach practical:

\[
S_2(\theta) \simeq \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} \left( W_\theta(Z_i)^{1/2} G(Z_i) - V_T \right)^2
\]

A limited number of paths is necessary to determine the optimal \( \theta \).

In fact, the configurations \( Z_i \) are FIXED. So, the difference between

\[
S_2(\theta) \quad S_2(\theta')
\]

is much more accurate than the MC estimate of each of them.
European Call

\[ G(Z) = e^{-rT} \left( X_0 \exp \left[ \left( r - \frac{\sigma^2}{2} \right) T + \sigma \sqrt{T}Z \right] - K \right)^+ \]

\[ P(Z) = (2\pi)^{-1/2} \exp \left( -Z^2 / 2 \right) \]

Trial Density

\[ \tilde{P}_{\mu}(Z) = (2\pi)^{-d/2} e^{-(Z-\mu)^2 / 2} \]

\[ \tilde{P}_{\mu,\tilde{\sigma}}(Z) = (2\pi\tilde{\sigma}^2)^{-1/2} e^{-(Z-\mu)^2 / 2\tilde{\sigma}^2} \]

\[ N_{MC} \approx 50 \]

<table>
<thead>
<tr>
<th>( \sigma )</th>
<th>( K )</th>
<th>LSIS(( \mu ))</th>
<th>LSIS(( \mu,\tilde{\sigma} ))</th>
<th>RM</th>
<th>GHS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>30</td>
<td>104(1)</td>
<td>1700(100)</td>
<td>112(4)</td>
<td>100(1)</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>7.8(1)</td>
<td>15(1)</td>
<td>7.8(4)</td>
<td>7.8(1)</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>33.5(5)</td>
<td>84(5)</td>
<td>31(2)</td>
<td>33.5(5)</td>
</tr>
<tr>
<td>0.3</td>
<td>30</td>
<td>16.4(1)</td>
<td>51(1)</td>
<td>16.8(4)</td>
<td>14.8(2)</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>9.9(5)</td>
<td>27(1)</td>
<td>11(2)</td>
<td>9.9(1)</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>15.6(1)</td>
<td>35(1)</td>
<td>15.2(4)</td>
<td>14.2(1)</td>
</tr>
</tbody>
</table>

Variance Reduction

\[ VR = \left( \frac{\sigma(\text{Crude MC})}{\sigma(\text{IS})} \right)^2 \]


European Straddle

\[ G(Z) = e^{-rT} X_0 \exp \left[ \left( r - \frac{\sigma^2}{2} \right) T + \sigma \sqrt{T} Z \right] - K \]

Bimodal Ansatz:

\[ \tilde{P}(Z) = (2\pi)^{-d/2} \left[ w_a e^{-\left( Z - \mu_a \right)^2 / 2} + w_b e^{-\left( Z - \mu_b \right)^2 / 2} \right] \]
Stratified Sampling

Stratifying a Normal Random Variable

Reducing the Variance by Sampling in a more regular pattern.
LSIS + Stratified Sampling (LSIS+)

1d

2d

Too many sample points to fill the space in high dimensions

I can stratify one-dimensional projections!


\[ Z^{(i)} = \xi X^{(i)} + (I_d - \xi \xi^t)Y^{(i)} \]

1-d Stratified Normal

\[ N(0, I_d) \]

\[ \xi \propto \mu(\text{LSIS}) \]

LSIS+
Asian Option with Stratified Sampling:

\[ G(Z) = e^{-rT} \left( \frac{1}{M} \sum_{i=1}^{M} X_i - K \right)^+ \]

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>0.3</td>
<td>45</td>
<td>8.8(1)</td>
<td>950(20)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>9.9(1)</td>
<td>1225(15)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>55</td>
<td>13.6(7)</td>
<td>1900(100)</td>
</tr>
<tr>
<td>64</td>
<td>0.3</td>
<td>45</td>
<td>9.0(2)</td>
<td>1060(30)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>10.3(5)</td>
<td>1290(30)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>55</td>
<td>12.5(5)</td>
<td>1320(100)</td>
</tr>
</tbody>
</table>

Computational Speed-Up of 3 orders of magnitude!
Libor Market Model Setting

Euler Discretization:

\[
\frac{L_i(n+1)}{L_i(n)} = \exp \left[ (\mu_i(L(n)) - ||\sigma_i(n)||^2/2) h_e + \sigma_i^T(n) Z(n+1) \sqrt{h_e} \right]
\]

\[
\mu_i(L(t)) = \sum_{j=\eta(t)}^i \frac{\sigma_i^T \sigma_j h L_j(t)}{1 + h L_j(t)}
\]

Risk-Neutral Drift

This fits in the general framework:

\[
V = E_P [G(Z)] = \int_D dZ \ G(Z) \ P(Z)
\]

\[
P(Z) = N(0, I_d) \equiv (2\pi)^{-d/2} \ e^{-Z^2/2}
\]

Trial Density

\[
\tilde{P}_\tilde{\mu}(Z) = (2\pi)^{-d/2} \ e^{-(Z-\tilde{\mu})^2/2}
\]

Linear parameterization of the drift (knot points)
Caplet

\[ C_h(T_m) = \left( \prod_{i=0}^{m} \frac{1}{1 + hL_i(T_i)} \right) h(L_m(T_m) - K)^+ \]

<table>
<thead>
<tr>
<th>( T_m ) (years)</th>
<th>( K )</th>
<th>( N_k )</th>
<th>LSIS</th>
<th>LSIS+</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>0.04</td>
<td>1</td>
<td>11.4(1)</td>
<td>1349(1)</td>
</tr>
<tr>
<td>1.0</td>
<td>0.055</td>
<td>1</td>
<td>13.3(2)</td>
<td>2300(2)</td>
</tr>
<tr>
<td>1.0</td>
<td>0.07</td>
<td>1</td>
<td>20.2(1)</td>
<td>4126(4)</td>
</tr>
<tr>
<td>2.5</td>
<td>0.04</td>
<td>1</td>
<td>14.0(1)</td>
<td>1189(1)</td>
</tr>
<tr>
<td>2.5</td>
<td>0.055</td>
<td>1</td>
<td>15.5(1)</td>
<td>897(1)</td>
</tr>
<tr>
<td>2.5</td>
<td>0.07</td>
<td>1</td>
<td>18.1(1)</td>
<td>1831(1)</td>
</tr>
<tr>
<td>5.0</td>
<td>0.040</td>
<td>1</td>
<td>12.7(1)</td>
<td>235.2(5)</td>
</tr>
<tr>
<td>5.0</td>
<td>0.060</td>
<td>1</td>
<td>12.5(1)</td>
<td>237.0(5)</td>
</tr>
<tr>
<td>5.0</td>
<td>0.080</td>
<td>1</td>
<td>14.5(1)</td>
<td>193.3(4)</td>
</tr>
<tr>
<td>7.0</td>
<td>0.04</td>
<td>1</td>
<td>7.9(3)</td>
<td>40.0(1)</td>
</tr>
<tr>
<td>7.0</td>
<td>0.055</td>
<td>1</td>
<td>8.5(4)</td>
<td>43.7(1)</td>
</tr>
<tr>
<td>7.0</td>
<td>0.07</td>
<td>1</td>
<td>8.5(4)</td>
<td>40.0(1)</td>
</tr>
</tbody>
</table>

Speed-Ups:
2-3 order of magnitude

\[ N_{MC} \approx 100 \]
Swaptions

\[ V(T_n) = \sum_{i=n+1}^{M+1} B(T_n, T_i) h(S_n(T_n) - K)^+ \]

<table>
<thead>
<tr>
<th>( T_n ) (years)</th>
<th>( T_{M+1} )</th>
<th>( K )</th>
<th>( N_k )</th>
<th>LSIS</th>
<th>LSIS+</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>1.5</td>
<td>0.04</td>
<td>3</td>
<td>6.8(3)</td>
<td>35.2(8)</td>
</tr>
<tr>
<td>0.5</td>
<td>1.5</td>
<td>0.055</td>
<td>3</td>
<td>10.5(4)</td>
<td>143(2)</td>
</tr>
<tr>
<td>0.5</td>
<td>1.5</td>
<td>0.07</td>
<td>3</td>
<td>21.2(6)</td>
<td>209(2)</td>
</tr>
<tr>
<td>0.5</td>
<td>2.5</td>
<td>0.04</td>
<td>3</td>
<td>7.0(3)</td>
<td>41.9(9)</td>
</tr>
<tr>
<td>0.5</td>
<td>2.5</td>
<td>0.055</td>
<td>3</td>
<td>9.8(3)</td>
<td>149(2)</td>
</tr>
<tr>
<td>0.5</td>
<td>2.5</td>
<td>0.07</td>
<td>3</td>
<td>18.6(5)</td>
<td>427(2)</td>
</tr>
<tr>
<td>0.5</td>
<td>5.5</td>
<td>0.04</td>
<td>3</td>
<td>6.8(3)</td>
<td>50(1)</td>
</tr>
<tr>
<td>0.5</td>
<td>5.5</td>
<td>0.055</td>
<td>3</td>
<td>8.5(3)</td>
<td>106(1)</td>
</tr>
<tr>
<td>0.5</td>
<td>5.5</td>
<td>0.07</td>
<td>3</td>
<td>12.0(4)</td>
<td>148(1)</td>
</tr>
<tr>
<td>1.0</td>
<td>6.0</td>
<td>0.04</td>
<td>3</td>
<td>8.0(4)</td>
<td>144(2)</td>
</tr>
<tr>
<td>1.0</td>
<td>6.0</td>
<td>0.055</td>
<td>3</td>
<td>8.6(3)</td>
<td>165(2)</td>
</tr>
<tr>
<td>1.0</td>
<td>6.0</td>
<td>0.07</td>
<td>3</td>
<td>12.7(4)</td>
<td>654(3)</td>
</tr>
<tr>
<td>2.0</td>
<td>7.0</td>
<td>0.04</td>
<td>3</td>
<td>9.2(3)</td>
<td>70(1)</td>
</tr>
<tr>
<td>2.0</td>
<td>7.0</td>
<td>0.055</td>
<td>3</td>
<td>9.7(3)</td>
<td>139(1)</td>
</tr>
<tr>
<td>2.0</td>
<td>7.0</td>
<td>0.09</td>
<td>3</td>
<td>13.9(4)</td>
<td>140(1)</td>
</tr>
<tr>
<td>5.0</td>
<td>10.0</td>
<td>0.04</td>
<td>5</td>
<td>7.3(4)</td>
<td>76(1)</td>
</tr>
<tr>
<td>5.0</td>
<td>10.0</td>
<td>0.055</td>
<td>5</td>
<td>7.4(3)</td>
<td>72(2)</td>
</tr>
<tr>
<td>5.0</td>
<td>10.0</td>
<td>0.09</td>
<td>5</td>
<td>7.5(4)</td>
<td>197(2)</td>
</tr>
</tbody>
</table>

Speed-Ups: 1-2 order of magnitude
Straddle

\[ S_{th}(T_m) = \left( \prod_{i=0}^{m} \frac{1}{1 + hL_i(T_i)} \right) h |L_m(T_m) - K| \]

\begin{tabular}{cccccc}
\hline
\textbf{\( T_m \) (years)} & \textbf{\( K \)} & \textbf{\( N_k \)} & \textbf{LSIS} & \textbf{LSIS (MM)} \\
\hline
1.0 & 0.04 & 1 & 2.0(2) & 11.6(5) \\
1.0 & 0.05 & 1 & 1.4(1) & 6.4(3) \\
1.0 & 0.055 & 1 & 1.1(1) & 6.1(3) \\
1.0 & 0.06 & 1 & 1.0(1) & 4.0(2) \\
1.0 & 0.07 & 1 & 1.1(1) & 2.6(1) \\
5.0 & 0.04 & 1 & 6.8(4) & 24.6(8) \\
5.0 & 0.05 & 1 & 5.1(3) & 21.2(7) \\
5.0 & 0.055 & 1 & 4.0(3) & 14.8(5) \\
5.0 & 0.06 & 1 & 3.5(3) & 15.5(6) \\
5.0 & 0.07 & 1 & 2.9(2) & 15.8(6) \\
5.0 & 0.09 & 1 & 2.0(2) & 10.0(4) \\
\hline
\end{tabular}

MM guess provides sizable improvements

Speed-Ups:

1 order of magnitude

Bi-Modal ansatz

\[ \tilde{P}(Z) = \left( 2\pi \right)^{-d/2} \left[ w_a \ e^{-\left( Z - \mu_a \right)^2 / 2} + w_b \ e^{-\left( Z - \mu_b \right)^2 / 2} \right] \]
Module 1: Summary

- Efficient Monte Carlo Sampling
  - Multidimensional Integrals, the curse of dimensionality and MC
  - Computational efficiency: Variance and Statistical Uncertainties

- Variance Reduction Techniques
  - **Antithetic Variates**
    Pros: Easy to implement
    Cons: Limited Benefits especially in high dimensions
  - **Control Variates**
    Pros: They never hurt. Maybe very powerful.
    Cons: Difficult in general to find suitable controls.
  - **Importance Sampling**
    Pros: Powerful technique, especially for rare events
    Cons: Requires a good trial density
Module 1: Summary (cont’d)

- **Case Study: Least Squares Importance Sampling (LSIS)**
  - Simple Importance Sampling strategy based on a quick LS Optimization.
  - Can be combined with Stratification for further efficiency gains (LSIS+).
  - LSIS can be used with non-Gaussian/multi-modal trial densities.
  - LSIS and LSIS+ can result in computational savings of orders of magnitude.

**References:**
Module 2:
Efficient Risk Management in Monte Carlo
Outline

- Hedges and Price Sensitivities (Greeks)
  - Shortcoming of Finite Differences (‘no time to think’) approach

- Classical Approaches
  - Likelihood Ratio Method
  - Case Study II: Reducing the Variance of Likelihood Ratio Greeks
  - Pathwise Derivative Method

- Modern Approaches
  - Adjoint Approaches
  - The new deal: Adjoint Algorithmic Differentiation (AAD)
  - Case Study III: Real Time Counterparty Credit Risk Management in Monte Carlo
Greeks in Monte Carlo

- Hedging, i.e. making your portfolio neutral to moves of market risk factors, requires the calculation of price sensitivities.

\[ V(\theta) = \mathbb{E}_Q \left[ G_\theta(X(T_1, \theta), \ldots, X(T_M, \theta)) \right] \]

Model Parameters:

\[ \theta = (X(T_0), \sigma, r, \ldots) \]

\[ \frac{\partial V(\theta)}{\partial \theta_0} = \frac{\partial V(\theta)}{\partial X(T_0)} \quad \Rightarrow \text{Delta} \]

\[ \frac{\partial V(\theta)}{\partial \theta_1} = \frac{\partial V(\theta)}{\partial \sigma} \quad \Rightarrow \text{Vega} \]

\[ \frac{\partial V(\theta)}{\partial \theta_2} = \frac{\partial V(\theta)}{\partial r} \quad \Rightarrow \text{IR Risk} \]

\[ dX_t = rX_t dt + \sigma X_t dW_t \]
Finite Differences Approaches (‘Bumping’)

\[
\frac{\partial V(\theta)}{\partial \theta} \approx \frac{V(\theta + h) - V(\theta)}{h}
\]

- **Pros:**
  - easy to implement
- **Cons:**
  - time consuming (at least 2x evaluation time)
  - accuracy of risk

**Sources of Error:**
- Bias
- Variance

\[
\text{Var} \left[ \frac{V(\theta + h) - V(\theta)}{h} \right]
\]
Variance of Finite Difference Estimators

\[
\frac{\partial V(\theta)}{\partial \theta} \approx \frac{V(\theta + h) - V(\theta)}{h}
\]

\[
\text{Var} \left[ \frac{\tilde{V}(\theta + h) - \tilde{V}(\theta)}{h} \right] = \frac{1}{h^2} \text{Var} [\tilde{V}(\theta + h) - \tilde{V}(\theta)]
\]

- Infinite Variance of the Sensitivity Estimator
- Finite Variance of the Sensitivity Estimator

MC estimators

Luca Capriotti – Efficient Risk Management in Monte Carlo
Variance of Finite Difference Estimators (cont’d)

\[
\text{Var}\left[ \frac{\tilde{V}(\theta + h) - \tilde{V}(\theta)}{h} \right] = \frac{1}{h^2} \text{Var}[\tilde{V}(\theta + h) - \tilde{V}(\theta)]
\]

\[
\text{Var}[\tilde{V}(\theta + h) - \tilde{V}(\theta)]
\]

- \( O(1) \): independent sampling of the base and bumped estimator.

\[
\text{Var}[\tilde{V}(\theta + h)] - \text{Var}[\tilde{V}(\theta)] \to 2\text{Var}[\tilde{V}(\theta)]
\]

- \( O(h) \): same random seed.

- \( O(h^2) \): estimator Lipschitz continuous (and other minor technical conditions, see e.g. Glasserman’s book).

\[ \exists \kappa \ s.t. \ |G(x) - G(y)| \leq \kappa \|x - y\| \ \forall \ x, y \]
Homework assignment:
Compute the variance of the estimators for the examples above as a function of the finite increment. Perform the calculation using the same and a different random seed in the base and perturbed evaluation.
Greeks without Bumping (Classical Approaches)

- **Likelihood Ratio Method**
  - Differentiation of the pdf associated with the stochastic process.
  - Greeks obtained by multiplying the payoff by a suitable weight.

- **Pathwise Derivative Method**
  - Differentiate both the process and the payout through the chain rule.
  - Equivalent to standard correlated bumping as the bump goes to zero.

- **Malliavin weights method**
  - Stochastic Calculus of Variations allows to derive a generalized Integration by Parts Formula.
  - As in the LRM, the Greeks are obtained by re-weighing the payoff. The weight is generally a complicated stochastic integral.
Greeks without Bumping (Classical Approaches)

- **Likelihood Ratio Method**
  - Differentiation of the pdf associated with the stochastic process.
  - Greeks obtained by multiplying the payoff by a suitable weight.

- **Pathwise Derivative Method**
  - Differentiate both the process and the payout through the chain rule.
  - Equivalent to standard correlated bumping as the bump goes to zero.

- **Malliavin weights method**
  - Stochastic Calculus of Variations allows to derive a generalized Integration by Parts Formula.
  - As in the LRM, the Greeks are obtained by re-weighing the payoff. The weight is generally a complicated stochastic integral.
Likelihood Ratio Method

- Isolate the dependence on the parameters in the density function

\[ V(\theta) = \mathbb{E}_{P_{\theta}}[G(X_1, \ldots, X_M)] = \int dx P_{\theta}(x)G(x) \]

- If the probability density function is regular enough

\[ \partial_{\theta} \int dx \ G(x)P_{\theta}(x) = \int dx \ G(x)\partial_{\theta}P_{\theta}(x) \]

- How to sample \( \int dx \ G(x)\partial_{\theta}P_{\theta}(x) \)?

\[ \int dx \ G(x)\partial_{\theta}P_{\theta}(x) = \int dx \ G(x)\frac{\partial_{\theta}P_{\theta}(x)}{P_{\theta}(x)}P_{\theta}(x) \]

\[ \frac{\partial V(\theta)}{\partial \theta} = \mathbb{E}_{P_{\theta}}[G(X)\Omega(X)] \quad \Omega_{\theta}(x) = \partial_{\theta} \log P_{\theta}(x) \]

LRM weight
The ‘sign’ problem of the LRM estimators

$$\frac{\partial V(\theta)}{\partial \theta} = \mathbb{E}_{P_\theta} \left[ G(X) \Omega(X) \right] \quad \Omega_\theta(x) = \partial_\theta \log P_\theta(x)$$

- LRM weights have zero expectation value. Hence they have no definite sign.

$$\mathbb{E}_P[\Omega_\theta(X)] = \partial_\theta \int dx P_\theta(x) = \partial_\theta 1 = 0$$

- This can give rise to poor variance properties whenever the configurations with opposite sign have similar weight in the MC average

$$\bar{\theta}_k = \frac{1}{N_{\text{MC}}} \sum_{i=1}^{N_{\text{MC}}} \Omega_{\theta_k}(X[i])G(X[i])$$

so that the final outcome is the result of the cancellation of two comparable and not necessarily highly correlated quantities.
Likelihood Ratio Method (cont’d)

\[
\frac{\partial V(\theta)}{\partial \theta} = \mathbb{E}_{P_\theta} \left[ G(X) \Omega(X) \right] \quad \Omega_\theta(x) = \partial_\theta \log P_\theta(x)
\]

Pros:
- Simultaneous valuation of Value and Sensitivities:
  - Generally more efficient (for a given number of replications) than repeating the simulation from scratch.
- Does not require regularity conditions on the payoff.
- No finite difference bias.

Cons:
- Requires explicit knowledge of the density function.
- The variance of the estimator is difficult to predict a priori. A large variance can destroy the computational benefit of resampling with LRM weights instead of repeating the simulation.
Likelihood Ratio Method: Examples

- BS Setup
  
  \[ S = S_0 \exp \left[ (r - \sigma^2/2)T + \sigma \sqrt{T} Z \right] \]

  \[ P_\theta(S) = \frac{1}{S \sigma \sqrt{T}} \phi(Z(S)) \quad Z(S) = \frac{\log S/S_0 - (r - \sigma^2/2)T}{\sigma \sqrt{T}} \]

  Delta Weight
  
  \[ \Omega_{S_0}(S) = \partial_{S_0} \log P_\theta(S) = \frac{Z(S)}{\sigma \sqrt{T} S_0} \]

  Vega Weight
  
  \[ \Omega_{\sigma}(S) = \partial_{\sigma} \log P_\theta(S) = \frac{Z(S)^2 - 1}{\sigma} - Z(S) \sqrt{T} \]

- The variance of the Delta (Vega) weights diverge as \( \sigma \sqrt{T} \to 0 \) \( (\sigma \to 0) \).

**Homework assignment:**

Verify the formulas above and use them to implement the calculation of Delta and Vega for the call and bet analyzed in slide 48. Compare the variance of the LRM estimators with the variance of the finite differences ones.
Likelihood Ratio Method: Examples (cont’d)

- BS Path Dependent Setup (Asian Options)

\[ G(S_1, \ldots, S_M) = e^{-rT}(\bar{S} - K) \quad \bar{S} = \frac{1}{M} \sum_{i=1}^{M} S_i \]

Markov property

\[ P_\theta(S_1, S_2, \ldots, S_M) = P_\theta^{(1)}(S_1|S_0)P_\theta^{(2)}(S_2|S_1) \ldots P_\theta^{(M)}(S_M|S_{M-1}) \]

\[ P_\theta^i(S_i|S_{i-1}) = \frac{1}{S_i\sigma\sqrt{t_i - t_{i-1}}} \phi(Z_i(S_i|S_{i-1})) \]

\[ Z_i(S_i|S_{i-1}) = \frac{\log S_i/S_{i-1} - (r - \sigma^2/2)(t_i - t_{i-1})}{\sigma\sqrt{t_i - t_{i-1}}} \]
Likelihood Ratio Method: Examples (cont’d)

Delta Weight
\[ \Omega_{S_0} = \frac{Z_1}{S_0 \sigma \sqrt{t_1}} \]

Vega Weight
\[ \Omega_{\sigma} = \sum_{i=1}^{M} \left[ \frac{Z_i^2 - 1}{\sigma} - Z_i \sqrt{t_i - t_{i-1}} \right] \]

- The variance of both estimators diverge for small times step and volatility.

Homework assignment:
Verify the formulas above and implement them for an at the money option. Study the variance of the estimators as a function of the sampling interval and asset volatility.
Case Study: Reducing the Variance of Likelihood Ratio Greeks in MC

L.C., Proceedings Winter Simulation Conference 2008 [available on ssrn].

- Gaussian Copula Models

\[ F(x) = \prod_{i=1}^{N} \int_{-\infty}^{x_i} dy_i P(y_1, \ldots, y_N) = \Phi_N(\Phi^{-1}(M_1(x_1)), \ldots, \Phi^{-1}(M_N(x_N)); \Sigma) \]

Market Implied Marginals

Probability Density Function

\[ P(x) = \phi_N(\Phi^{-1}(M_1(x_1)), \ldots, \Phi^{-1}(M_N(x_N))) \prod_{i=1}^{N} \frac{m_i(x_i)}{\phi(\Phi^{-1}(M_i(x_i)))} \]

LRM weight:

\[ \Omega_\theta(x) = \sum_{i=1}^{N} \partial_\theta \log m_i(x_i) - Z(x)^T(\Sigma^{-1} - I)\partial_\theta Z(x) \]

\[ Z_i = \Phi^{-1}(M_i(x_i)) \]

\[ \partial_\theta Z_i = \frac{\partial_\theta M_i(x_i)}{\phi(\Phi^{-1}(M_i(x_i)))} \]
Reducing the Variance of Likelihood Ratio Greeks in MC

- Lognormal Marginals

\[ S_i = S_i^0 \exp[(r - \sigma_i^2/2)T + \sigma_i \sqrt{T} Z_i] \]

\[ \Omega_\theta(Z) = -\frac{1}{2} \text{Tr}[\hat{\Sigma}^{-1} \partial_\theta \hat{\Sigma}] + \frac{1}{2} X \hat{\Sigma}^{-1} (\partial_\theta \hat{\Sigma}) \hat{\Sigma}^{-1} X + X \hat{\Sigma}^{-1} \partial_\theta \mu_i \]

\[ X_i = \sigma_i \sqrt{T} Z_i \]

\[ \mu_i = \log S_i^0 + (r - \sigma_i^2/2)T \]

\[ Z_i = \frac{\log S_i/S_i^0 - (r - \sigma_i^2/2)T}{\sigma_i \sqrt{T}} \]

\[ \hat{\Sigma}_{ij} = \sigma_i \sigma_j \Sigma_{ij} \]

**Delta Weight**

**Vega Weight**

\[ \Omega_{S_i^0} = \frac{[\Sigma^{-1} Z]_i}{\sigma_i \sqrt{T} S_i^0} \]

\[ \Omega_{\sigma_i}(Z) = \left( \frac{Z_i}{\sigma_i} - \sqrt{T} \right) [\Sigma^{-1} Z]_i - \frac{1}{\sigma_i} \]

**Homework assignment:**

Work out the formulas above and verify their consistency with the one dimensional case.
Reducing the Variance of Likelihood Ratio Greeks in MC

- Antithetic Variates Solve the problem divergence of the variance of Delta estimators

\[ \Omega_{S_i^0} = \frac{\left[ \sum^{-1} Z \right]_i}{\sigma_i \sqrt{T} S_i^0} \]

\[ \text{Var}[\Omega_{S_i^0}] \xrightarrow{\sigma_i \sqrt{T} \to 0} \infty \]

\[ \text{Var}[\Omega_{S_i^0}^{ant}] = \text{Var}\left[ \frac{\left[ \sum^{-1} (Z - Z) \right]_i}{2\sigma_i \sqrt{T} S_i^0} \right] = 0 \]

Deep in the money call option
Reducing the Variance of Likelihood Ratio Greeks in MC

- Unfortunately do not help much for Vega because the estimator is not odd

\[ \Omega_{\sigma_i}(Z) = \left( \frac{Z_i}{\sigma_i} - \sqrt{T} \right) [\Sigma^{-1} Z]_i - \frac{1}{\sigma_i} \]

- Control Variates and LSIS provide significant improvements

Natural controls:

\[ \partial_{\sigma_i} \mathbb{E}[S_i] = 0 \quad \mathbb{E}[\Omega_{\sigma_i}] = 0 \]

Variance Reductions

<table>
<thead>
<tr>
<th>K</th>
<th>AV</th>
<th>AV+CV</th>
<th>AV+LSIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>1.3(1)</td>
<td>4(1)10^4</td>
<td>11(1)</td>
</tr>
<tr>
<td>40</td>
<td>1.7(1)</td>
<td>150(10)</td>
<td>7.0(6)</td>
</tr>
<tr>
<td>50</td>
<td>2.1(1)</td>
<td>22(2)</td>
<td>8.0(8)</td>
</tr>
<tr>
<td>60</td>
<td>2.1(2)</td>
<td>5.5(6)</td>
<td>80(10)</td>
</tr>
<tr>
<td>70</td>
<td>1.9(2)</td>
<td>5.2(5)</td>
<td>340(40)</td>
</tr>
<tr>
<td>80</td>
<td>1.7(3)</td>
<td>1.7(3)</td>
<td>1100(100)</td>
</tr>
</tbody>
</table>
Pathwise Derivative Method

- Monte Carlo Expectation Values
  \[ V(\theta) = \mathbb{E}_Q\left[g(X(T_1), \ldots, X(T_M))\right] \]
  ... and sensitivities
  \[ \frac{\partial V(\theta)}{\partial \theta_k} = \mathbb{E}_Q\left[\frac{\partial}{\partial \theta_k} g(X(\theta))\right] \]

- Pathwise Derivative Estimator
  \[ \bar{\theta}_k \equiv \frac{\partial g(X(\theta))}{\partial \theta_k} = \sum_{j=1}^{N \times M} \frac{\partial g(X)}{\partial X_j} \times \frac{\partial X_j}{\partial \theta_k} \]
Pathwise Derivative Method: Simple Examples

BS Delta

\[ S(T) = S_0 \exp \left[ (r - \sigma^2/2)T + \sigma \sqrt{T} Z \right] \]

Call Option:

\[ G(S(T)) = e^{-rT} \left( S(T) - K \right)^+ \]

Pathwise estimator:

\[ \bar{\theta}_{S_0} = \frac{\partial G(S(T))}{\partial S_0} = \frac{\partial G(S(T))}{\partial S(T)} \frac{\partial S(T)}{\partial S_0} \]

Payout Derivative

\[ \frac{\partial G(S(T))}{\partial S(T)} = e^{-rT} \theta(S(T) - K) \]

Tangent Process

\[ \frac{\partial S(T)}{\partial S_0} = \frac{S(T)}{S_0} \]
Pathwise Derivative Method: Simple Examples

BS Vega

\[ S(T) = S_0 \exp \left[ (r - \sigma^2/2)T + \sigma \sqrt{T} Z \right] \]

Call Option:

\[ G(S(T)) = e^{-rT} (S(T) - K)^+ \]

Pathwise estimator:

\[ \bar{\theta}_{\sigma} = \frac{\partial G(S(T))}{\partial \sigma} = \frac{\partial G(S(T))}{\partial S(T)} \frac{\partial S(T)}{\partial \sigma} \]

Payout Derivative

\[ \frac{\partial G(S(T))}{\partial S(T)} = e^{-rT} \theta(S(T) - K) \]

Tangent Process

\[ \frac{\partial S(T)}{\partial \sigma} = S(T) (\sqrt{T} Z - \sigma T) \]
Pathwise Derivative Method: Simple Examples

What about digitals?

\[ G(S(T)) = e^{-rT} \theta(S(T) - K) \]

Actually...

\[ \frac{\partial G(S(T))}{\partial S(T)} = 0 \quad (S(T) \neq K) \quad \Rightarrow \quad \frac{\partial V}{\partial S_0} = 0 \]

\[ \frac{\partial G(S(T))}{\partial S(T)} = e^{-rT} \delta(S(T) - K) \]

Which cannot be sampled with Monte Carlo
Homework assignment:
Smooth out the payout above with a standard call spread and implement the calculation of delta by means of the pathwise derivative method. Compare the value with the analytical result. Calculate the bias and the variance of the estimator as a function of the call spread width $\delta$. 
Pathwise Derivative Method: Simple Examples

Asian option:

\[ G(X_1, \ldots, X_M) = e^{-r t_M} (\bar{X} - K)^+ \]

\[ \bar{X} = \frac{1}{M} \sum_{i=1}^{M} X(t_i) \]

LN model:

\[ X(t_i) = X(t_{i-1}) \exp([r - \sigma^2/2](t_i - t_{i-1}) + \sigma \sqrt{t_i - t_{i-1}} Z_i] \]

Tangent process:

\[ \frac{\partial X(t_i)}{\partial X_0} = \frac{\partial X(t_{i-1})}{\partial X_0} \exp([r - \sigma^2/2](t_i - t_{i-1}) + \sigma \sqrt{t_i - t_{i-1}} Z_i] = \frac{X(t_i)}{X_0} \]

Pathwise Estimator for Delta:

\[ \bar{\theta}_X = \sum_{i=1}^{M} \frac{\partial G(X)}{\partial X_i} \frac{\partial X_i}{\partial X_0} = \frac{\partial G(X)}{\partial \bar{X}} \sum_{i=1}^{M} \frac{\partial \bar{X}}{\partial X_i} \frac{\partial X_i}{\partial X_0} = e^{-r T} \theta(S - K) \frac{\bar{X}}{X_0} \]

Homework assignment:

Work out the estimator for Vega.
Pathwise Derivative Method: Libor Market Model

Log Euler scheme:

\[
\frac{L_i(n + 1)}{L_i(n)} = \exp \left[ \left( \mu_i(L(n)) - \|\sigma_i(n)\|^2/2 \right) h_e + \sigma_i^T(n)Z(n + 1)\sqrt{h_e} \right]
\]

Delta tangent process:

\[
\Delta_{i,k}(t) = \frac{\partial L_i(t)}{\partial L_k(0)}
\]

\[
\Delta_{i,k}(n + 1) = \Delta_{i,k}(n) \frac{L_i(n + 1)}{L_i(n)} + L_i(n + 1) \sum_{j=1}^{N} \frac{\partial \mu_i(n)}{\partial L_j(n)} \Delta_{j,k}(n) h
\]

Matrix Recursion:

\[
\Delta(n + 1) = D(n) \Delta(n) \quad \Delta(0) = I
\]

Homework assignment:

Work out the recursion for Vega.
Pathwise Derivative Method: Variance of the Estimators

Finite Difference Estimator:

\[
\frac{P(X(\theta^{(k)})[i_{MC}]) - P(X(\theta)[i_{MC}])}{\Delta \theta} \rightarrow \frac{\partial P(X(\theta)[i_{MC}])}{\partial \theta_k}
\]

\[
\theta^{(k)} = (\theta_1, \ldots, \theta_k + \Delta \theta, \ldots)
\]

\[
\text{Var} \left[ \frac{\partial P(X(\theta))}{\partial \theta_k} \right] = \lim_{\Delta \theta \to 0} \text{Var} \left[ \frac{P(X(\theta^{(k)})) - P(X(\theta))}{\Delta \theta} \right]
\]

The variance of the Finite Difference Estimators is asymptotically equal to the variance of the Pathwise Derivative Estimator.
Pathwise Derivative Method: Challenges

\[ \bar{\theta}_k \equiv \frac{\partial g(X(\theta))}{\partial \theta_k} = \sum_{j=1}^{N \times M} \frac{\partial g(X)}{\partial X_j} \times \frac{\partial X_j}{\partial \theta_k} \]

Since the variance of the estimator is comparable to the one of finite differences, all this is worth the hassle if the resulting computational time is significantly lower than the one of Bumping.

We need an efficient way to calculate:

1. Simulation of the Tangent Process
2. Derivatives of the Payout
“Algebraic” Adjoint Methods

Giles and Glasserman’s `Smoking Adjoint’, Risk Magazine 2006
Leclerc et al., Risk Magazine 2009
Joshi et al., several preprints

Libor Market Model & Swaptions
Concentrate on the efficient Simulation of the Tangent Process

In a nutshell:

1. Formulate the propagation of the Tangent process in terms of Linear Algebra Ops
2. Optimize the computation time by rearranging the order of the computations
3. Implement the rearranged sequence of operations
Algebraic Adjoint Methods

\[ \bar{\theta} = \frac{\partial P(X(N))}{\partial X(N)}^T \Delta(N) \]

Tangent Process

Matrix Recursion

\[ \Delta(n + 1) = D(n) \Delta(n) \quad \Delta(0) = I \]

Matrix Matrix Forward Recursion

\[ O(N^3) \]

Matrix Vector Backward Recursion

\[ \bar{\theta} = \frac{\partial P(X(N))}{\partial X(N)}^T D(N-1) \cdots D(1) \Delta(0) \]

\[ O(N^2) \]
Algebraic Adjoint Methods

Arbitrary number of sensitivities at a fixed small cost

Giles and Glasserman, Risk Magazine 2006
Limitations of Algebraic Adjoint Methods

- LMM is bit of an ad-hoc application ...
  
  - Difficult to generalize to Path Dependent Options
  
  - The required Algebraic Analysis is in general cumbersome
  
  - Not general enough for all the applications in Finance
  
  - The derivatives required are often not available in closed form
  
  - What about the derivatives of the Payout?
Algorithmic Adjoint Approaches: AAD

- Adjoint implementations can be seen as instances of a programming technique known as Adjoint Algorithmic Differentiation (AAD)
- In general AAD allows the calculation of the gradient of an algorithm at a cost that is a small constant ($\sim 4$) times the cost of evaluating the function itself, independent of the number of input variables.

The Payoff estimator is a mapping of the form

$$\theta \rightarrow g(X(\theta))$$

AAD gives all the Risk estimators for a small fixed cost

$$\bar{\theta}_k \equiv \frac{\partial g(X(\theta))}{\partial \theta_k}$$
How does AAD work anyway?

\[ Y = \text{FUNCTION}(X) \]

\[ X \rightarrow \ldots \rightarrow U \rightarrow V \rightarrow \ldots \rightarrow Y \]

**Adjoint Rule**

\[ \bar{V}_k = \sum_{j=1}^{m} \bar{Y}_j \frac{\partial Y_j}{\partial V_k} \]

**Propagation Rule**

\[ \bar{U}_i = \sum_{k} \bar{V}_k \frac{\partial V_k}{\partial U_i} \]

\[ \bar{X} \leftarrow \ldots \leftarrow \bar{U} \leftarrow \bar{V} \leftarrow \ldots \leftarrow \bar{Y} \]

**Main Result**

\[ \bar{X} = \text{FUNCTION}_B(X, \bar{Y}) \]

**Lin. Comb. Jacobian Rows at a small fixed cost**

\[ \bar{X}_i = \sum_{j=1}^{m} \bar{Y}_j \frac{\partial Y_j}{\partial X_i} \]
AAD as a Design Paradigm

- AAD can be used as a **design paradigm** even for large inhomogeneous algorithms
- Addresses both aspects of the implementation of the Pathwise Derivative Method

\[
\frac{\partial g(X)}{\partial X_j} \quad \frac{\partial X_j(\theta)}{\partial \theta_k} \quad \text{L.C., Journ. Comp. Fin. (2011)}
\]

- Linear combination of the rows of the Jacobian
- All the Greeks at a cost that is a small (~4) multiple of the PV estimator
Diffusive Setting

\[
\begin{align*}
&\text{PROP}_0 \\
&\text{PROP}_1 \\
&\text{PROP}_N \\
\end{align*}
\]

\[
\begin{align*}
&P_0(X) \\
&X(T_1) \cdots X(T_M) \\
&X(t_{N_S}) \\
&\{X(t_m)\}_{m \leq N_S-1} \\
&\vdots \\
&P(T_1) \cdots P(T_M) \\
\end{align*}
\]

\[
\begin{align*}
&P(X(T_1) \cdots X(T_M)) \\
&\tilde{X}(T_1) \cdots \tilde{X}(T_M) \\
&\tilde{X}(t_{N_S}) \\
&\{\tilde{X}(t_m)\}_{m \leq N_S-1} \\
&\vdots \\
&P(X(t_1) \cdots X(t_M)) \\
\end{align*}
\]
Lognormal Example

\[ X(t_{n+1}) = \text{PROP}_n(X(t_n), \theta) \]

- **Step 1**
  \[ \mu = r X(t_n) \]

- **Step 2**
  \[ \Sigma = \sigma X(t_n) \]

- **Step 3**
  \[ X(t_{n+1}) = \mu \Delta t + \Sigma \sqrt{\Delta t} Z \]
  \[ (\bar{X}(t_n), \bar{\theta}) = \text{PROP}_b(., \bar{X}(t_{n+1})) \]

- **Step 1**
  \[ \bar{X}(t_n) + = \bar{\mu} r \bar{\theta}_r + = \bar{\mu} X(t_n) \]

- **Step 2**
  \[ \bar{X}(t_n) + = \bar{\Sigma} \sigma \bar{\theta}_\sigma + = \bar{\Sigma} X(t_n) \]

- **Step 3**
  \[ \bar{\mu} = \bar{X}(t_{n+1}) \Delta t \]
  \[ \bar{\Sigma} = \bar{X}(t_{n+1}) \sqrt{\Delta t} Z \]
Best of Asian Option

- Full Delta and Vega calculation for just twice the cost to calculate the PV

L.C. & Mike Giles, preprint (2011)
Back to the LMM test ground

- Full Delta and Vega calculation for just twice the cost to calculate the PV
- Similar results for both Euler and predictor corrector discretizations
- 100x savings in typical applications

L.C. & Mike Giles, preprint (2011)
Case Study:
Real Time Counterparty Risk Management in Monte Carlo

- Risk manage CVA/DVA is challenging because all the trades facing the same counterparty must be valued at the same time, typically with Monte Carlo

- AAD is naturally suited for this task

\[
V_{CVA} \approx \sum_{i=1}^{N_O} \mathbb{E}\left[ \mathbb{I}(T_{i-1} < \tau_c \leq T_i) D(0, T_i) \times \text{LGD}(T_i) \left( \text{NPV}(T_i) - C \left( R(T_i^-) \right) \right)^+ \right]
\]

Luca Capriotti – Efficient Risk Management in Monte Carlo
Real Time Counterparty Risk Management in Monte Carlo

- A new challenge: Rating Dependent Payoffs

\[ P(T_i, R(T_i), X(T_i)) = \sum_{r=0}^{N_R} \tilde{P}_i(X(T_i); r) \delta_{r, R(T_i)} \]

Rating Transition Markov Chain model (Jarrow, Lando and Turnbull '97)

\[ R(T_i) = \sum_{r=1}^{N_R} \mathbb{I}\left(\tilde{Z}_i^R > Q(T_i, r)\right) \]

- The Rating state space is discrete (hence the Payoff is non Lipschitz)
- The Pathwise Derivative method gives only part of the Risk
Real Time Counterparty Risk Management in Monte Carlo

Singular Contribution:

$$\partial_{\theta_k} P(T_i, \tilde{Z}_i, X(T_i)) = - \sum_{r=1}^{N_R} \left( \tilde{P}_i(X(T_i); r) - \tilde{P}_i(X(T_i); r - 1) \right) \delta \left( \tilde{Z}_i^R = Q(T_i, r; \theta) \right) \partial_{\theta_k} Q(T_i, r; \theta)$$

Can be integrated out analytically

$$\bar{\theta}_k = - \sum_{r=1}^{N_R} \frac{\phi(Z^*, Z^X_i; \rho_i)}{\sqrt{i} \phi(Z^X_i, \rho^X_i)} \partial_{\theta_k} Q(T_i, r; \theta) \times \left( \tilde{P}_i(X(T_i); r) - \tilde{P}_i(X(T_i); r - 1) \right)$$

Variance Reduction vs. Bumping:

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$\text{VR}[Q(1,1)]$</th>
<th>$\text{VR}[Q(1,2)]$</th>
<th>$\text{VR}[Q(1,3)]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>24</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>0.01</td>
<td>245</td>
<td>165</td>
<td>125</td>
</tr>
<tr>
<td>0.001</td>
<td>2490</td>
<td>1640</td>
<td>1350</td>
</tr>
</tbody>
</table>
Real Time Counterparty Risk Management in Monte Carlo

- Test Application: Calculation of risk for the CVA of a portfolio of 5 commodity swaps over a 5 years horizon (over 600 risks)

- Bumping: ~ 1h 40 min
- AAD: ~ 10 sec

L.C., J. Lee and M. Peacock, Risk (2011)
Module 2: Summary

- **Hedges and Price Sensitivities (Greeks)**
  - Shortcoming of Finite Differences (‘no time to think’) approach

- **Classical Approaches:**
  - **Likelihood Ratio Method:**
    - **Pros:**
      - Generally more efficient (for a given number of replications) than repeating the simulation from scratch.
      - Does not require regularity conditions on the payoff.
      - No finite difference bias.
    - **Cons:**
      - Requires explicit knowledge of the density function.
      - The variance of the estimator is difficult to predict a priori. A large variance can destroy the computational benefit of resampling with LRM weights instead of repeating the simulation.
Module 2: Summary (cont’d)

- **Case Study II: Reducing the Variance of Likelihood Ratio Greeks**
  - Antithetic Variates solve the problem of the divergence of Variance for Delta.
  - Control Variates and LSIS are very effective for basket pricing.

- **Pathwise Derivative Method**
  
  **Pros:**
  - Generally more efficient (for a given number of replications) than repeating the simulation from scratch.
  - No finite difference bias.
  - No surprises from the Variance

  **Cons:**
  - Requires Lipschitz continuity of the payoff.
  - The efficiency gains might not be enough to justify the cost of implementing technique.
Module 2: Summary (cont’d)

- **Modern Approaches: Adjoint Algorithmic Differentiation (AAD):**
  - Algebraic Adjoint approaches can be seen as specific instances of a more general paradigm: Adjoint Algorithmic Differentiation (AAD).
  - AAD can be employed to evaluate efficiently option sensitivities for virtually any model and financial security encountered in practice.
  - AAD allows the calculation of the Greeks in at most 4 times the time necessary for the calculation of the P&L of the portfolio.
  - Risk is calculated orders of magnitude faster than standard bumping, thus producing a significant reduction in infrastructure costs, and allowing “real time” monitoring of Risk and more effective hedging strategies.
  - **Case Study III: Real Time Counterparty Credit Risk Management in Monte Carlo**
References


Also available at [www.luca-capriotti.net](http://www.luca-capriotti.net) or [ssrn](http://ssrn.com)