EFFICIENT RISK MANAGEMENT IN MONTE CARLO

Luca Capriotti

The 7th Summer School in Mathematical Finance,
African Institute for Mathematical Sciences,
Muizenberg, Cape Town, South Africa, February 20-22, 2014

Outline

Module 1: Efficient Monte Carlo Sampling

- General Concepts: Why do we need Monte Carlo Methods?
  - 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
  - Classical Techniques
    - Antithetic Variates
    - Control Variates
    - Stratified Sampling
    - Importance Sampling
  - LSIS: Least Square Importance Sampling
    - Case Study I: LSIS and the Libor Market Model

Outline (cont'd)

Module 2: Risk Management and MC: Classical Approaches

- Hedges and Price Sensitivities (Greeks)
  - Shortcoming of the Finite Differences ('no time to think') approach
- Likelihood Ratio Method
  - Basic Idea: Differentiating the Probability Distribution
  - Pros and Cons
  - Case Study II: Reducing the Variance of Likelihood Ratio Greeks
- Pathwise Derivative Method
  - Basic Idea: Differentiating the Estimator "Path-by-Path"
  - Problems with Discontinuous Payouts
  - Handling Discontinuous Payouts
  - Examples
  - Pros and Cons

Outline (cont'd)

Module 3:
Risk Management by Adjoint Algorithmic Differentiation I

- Pathwise Derivative Method
- Algebraic Adjoint Approaches
- Adjoint Algorithmic Differentiation (AAD)
- AAD as a Design Paradigm
- AAD enabled Monte Carlo Engines
- First Applications
- Case Study: Adjoint Greeks for the Libor Market Model
- Conclusions

Outline (cont'd)

Module 4:
Risk Management by Adjoint Algorithmic Differentiation II

- Correlation Greeks and Binning Techniques
- Case Study: Correlation Greeks for Basket Default Contracts

Module 1:

Efficient Monte Carlo Sampling

Monte Carlo Methods

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

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

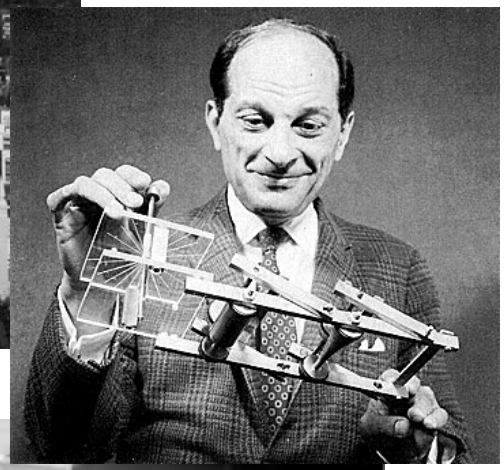
Number 247 SEPTEMBER 1949 *Volume 44*

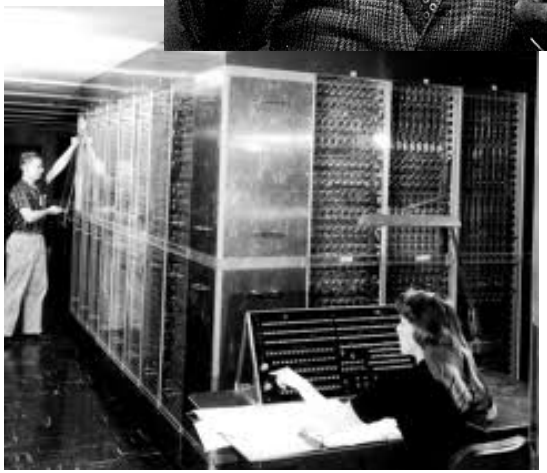
THE MONTE CARLO METHOD

NICHOLAS METROPOLIS AND S. ULAM
Los Alamos Laboratory

We shall present here the motivation and a general description of a method dealing with a class of problems in mathematical physics. The method is, essentially, a statistical approach to the study of differential equations, or more generally, of integro-differential equations that occur in various branches of the natural sciences.

ALREADY in the nineteenth century a sharp distinction began to appear between two different mathematical methods of treating physical phenomena. Problems involving only a few particles were studied in classical mechanics, through the study of systems of ordinary differential equations. For the description of systems with very many particles, an entirely different technique was used, namely, the method of statistical mechanics. In this latter approach, one does not concentrate on the individual particles but studies the properties of *sets of particles*. In pure mathematics an intensive study of the properties of sets of points was the subject of a new field. This is the so-called theory of sets, the basic theory of integration, and the twentieth century development of the theory of probabilities prepared the formal apparatus for the use of such models in theoretical physics, i.e., description of properties of aggregates of points rather than of individual points and their coordinates.





Using random numbers to solve multi dimensional problems with probabilistic methods

Multidimensional Integrals

- Many multidimensional problems in applied mathematics/natural sciences can be formulated as cubature problems:

$$I = \int f(\mathbf{x})\, d\mathbf{x}$$

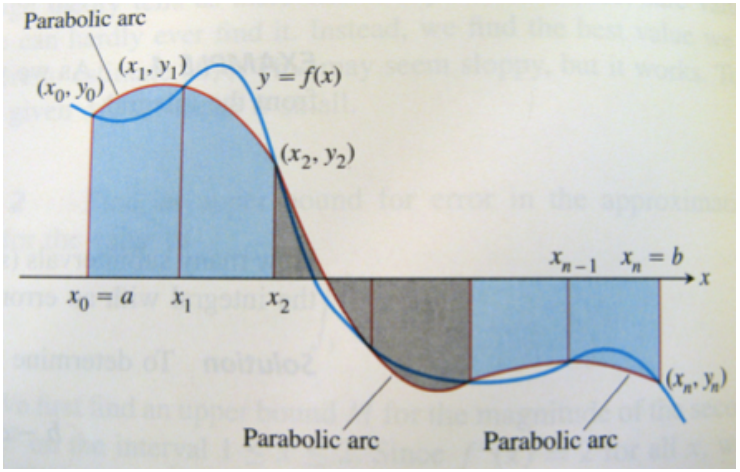
- Expectation values are notable examples:

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

$$I = \int d\mathbf{x}\, \underbrace{G(\mathbf{x})P(\mathbf{x})}_{f(\mathbf{x})}$$

Multidimensional Integrals (cont'd)

- Numerical quadrature formulas are very efficient in 1 dimension:



$$I = \int f(x)dx$$

$$\hat{I} = \sum_{i=1}^{n} w_i f(x_i)$$

Error: $\left|I - \hat{I}\right| = O(n^{-r})$

$r = 2$ Trapezoid $\left(f^{(2)}(x) < \infty\right)$

$r = 4$ Simpson's $\left(f^{(4)}(x) < \infty\right)$

Multidimensional Integrals (cont'd)

- Quadrature in higher dimension:

$$\hat{I} = \sum_{i_1=1}^{n} \dots \sum_{i_m=1}^{n} w_1^{(i_1)} \dots w_m^{(i_m)} f(x_1^{(i_1)}, \dots, x_m^{(i_m)})$$

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

Multidimensional Integrals (cont'd)

- Error (Bahvalov's theorem):

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

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

$m =$	1	2	3	...
$N =$	10	10^2	10^3	...

CURSE OF DIMENSIONALITY

Stochastic (Monte Carlo) Sampling

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

$$I = \int f(\mathbf{x})\, d\mathbf{x} = \int d\mathbf{x}\, G(\mathbf{x}) P(\mathbf{x})$$

$$\Downarrow$$

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

- Law of large numbers:

$$I \simeq \bar{I} = \frac{1}{N_{\mathrm{MC}}} \sum_{i=1}^{N_{\mathrm{MC}}} G(\mathbf{x}_i) \quad \mathbf{x}_i \sim P(\mathbf{x})$$

Central Limit Theorem and Monte Carlo Uncertainties

$$I \simeq \frac{1}{N_{\mathrm{MC}}} \sum_{i=1}^{N_{\mathrm{MC}}} G(\mathbf{x}_i) \pm \frac{\Sigma}{\sqrt{N_{\mathrm{MC}}}}$$

- Variance of the Estimator:

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

- Monte Carlo Estimator of the Variance:

$$\Sigma^2 \simeq \frac{1}{N_{\mathrm{MC}}} \sum_{i=1}^{N_{\mathrm{MC}}} \left(G(\mathbf{x}_i) - \bar{I} \right)^2$$

Breaking the spell: Monte Carlo vs. Quadrature

- Quadrature in high dimension m :

$$\text{Quadrature Error} = O(N^{-1/m})$$

(for a quadrature that does not require function smoothness)

- Monte Carlo:

$$\text{MC Error} = O(N_{\text{MC}}^{-1})$$

(finite 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...

$$I \simeq \frac{1}{N_{\mathrm{MC}}} \sum_{i=1}^{N_{\mathrm{MC}}} G(\mathbf{x}_i) \pm \frac{\Sigma}{\sqrt{N_{\mathrm{MC}}}}$$

- We need to cope with the Variance:

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

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: $$G(x) = \frac{1}{\sqrt{x}}$$

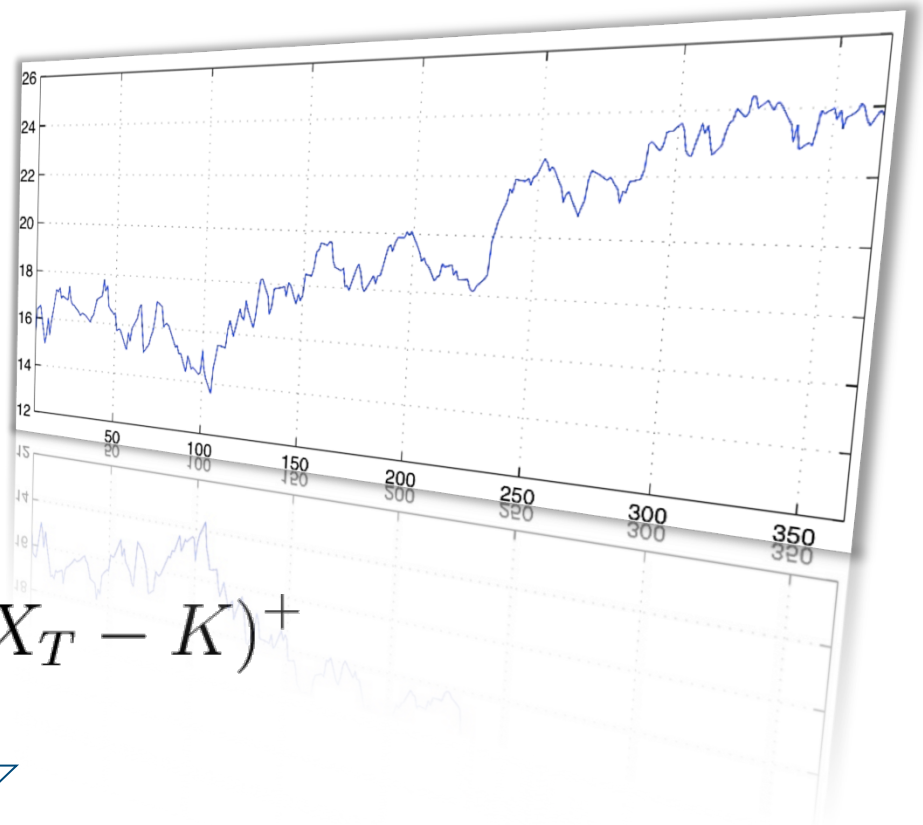
$$\mathrm{Var}\left[G(x)\right] = \int_0^1 dx \frac{1}{x} - \left(\int_0^1 dx \frac{1}{\sqrt{x}}\right)^2 = \infty$$

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:

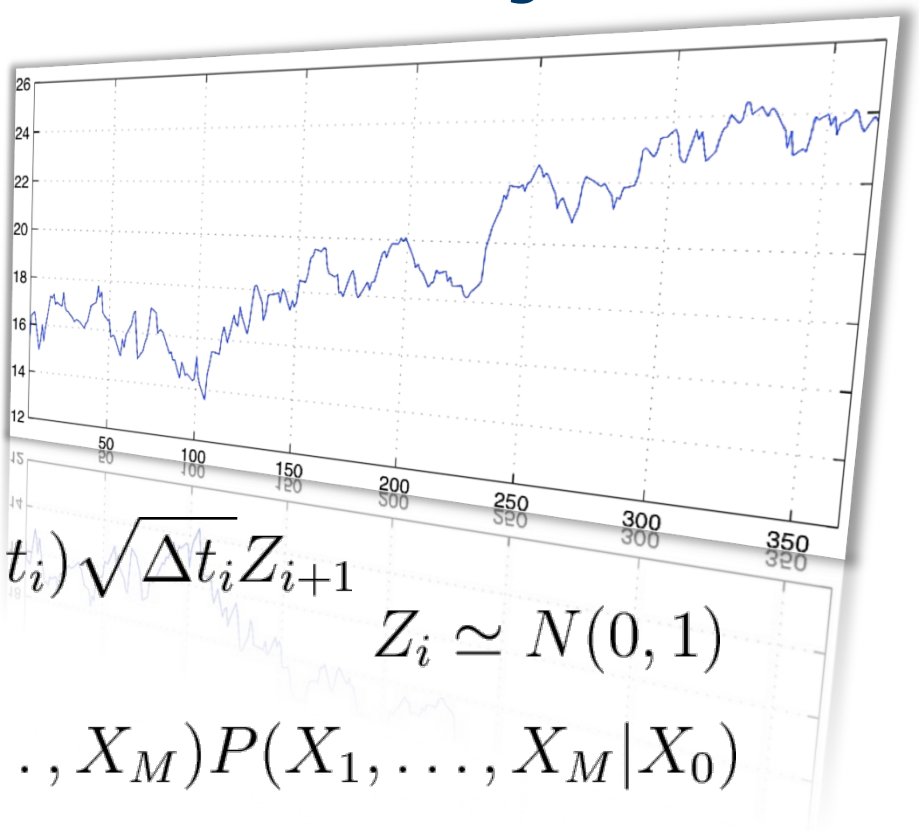
$$X_{i+1} = X_i + \mu(X_i, t_i)\Delta_i + \sigma(X_i, t_i)\sqrt{\Delta t_i}Z_{i+1}$$

$$Z_i \simeq N(0,1)$$

$$V = \int dX_1 \dots dX_M G(X_1, \dots, X_M) P(X_1, \dots, X_M | X_0)$$

$$= \int dZ_1 \dots dZ_M G(Z_1, \dots, Z_M) P(Z)$$

$$P(Z) = (2\pi)^{-M/2} \exp\left(-Z \cdot Z/2\right)$$



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.

- **Stratified Sampling**

  Constraining the fraction of samples drawn from different subset (strata) of the sample space to be equal to the theoretically expected value.

- **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

J. M. Hammersley, and K. W. Morton, A new Monte Carlo technique: Antithetic variates. Proc. Cambridge Philosophical Society **52**, 449 (1956).

- The intuition:

  - Consider two i.d. variates X, Y with expectation μ, variance σ^2 and covariance $\sigma^2_{XY} = \mathbb{E}[(X-\mu)(Y-\mu)]$

  - Compare 3 estimators of the mean of the common distribution:

$$X_i \qquad Y_i \qquad \frac{X_i + Y_i}{2}$$

$$\mathrm{Var}[X_i] = \mathrm{Var}[Y_i] = \sigma^2$$

$$\mathrm{Var}[\frac{X_i + Y_i}{2}] = \frac{\sigma^2_X + \sigma^2_Y + 2\sigma^2_{XY}}{4} = \frac{\sigma^2}{2}(1 + \rho_{XY})$$

In presence of negative correlation grouping helps to remove noise

Antithetic Variates (cont'd)

- In a typical context: $V = \mathbb{E}[G(.)]$

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

$\tilde{G} = G(1 - U_1, \ldots, 1 - U_d) \qquad \tilde{G} = G(-Z_1, \ldots, -Z_d)$

The transformed random variates have the same distribution

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

$$\sigma^2_{\text{AV}} = \text{Var}[\frac{G + \tilde{G}}{2}] = \frac{1}{2}\big(\text{Var}[G] + \text{Cov}[\text{G}, \tilde{\text{G}}]\big)$$

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:

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

$$\Downarrow$$

$$\mathrm{Cov}[G, \tilde{G}] < 0$$

- 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.

$$I = \int_0^1 e^{-x} dx \qquad I = \int_0^1 e^{-(x-1/2)^2} dx$$

- 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

H. Kahn and A. W. Marshall, Methods of reducing sample size in Monte Carlo computations, Oper. Res. **1**, 263 (1953).

- 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:

$$Y^i_{\mathrm{CV}}(\alpha) = Y_i - \alpha(X_i - \mu_X)$$

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

Variance:

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

Optimal for: $\alpha^\star = \dfrac{\mathrm{Cov}[\mathrm{X},\mathrm{Y}]}{\mathrm{Var}[\mathrm{X}]}$

$$\mathrm{Var}[Y_{\mathrm{CV}}(\alpha^\star)] = \mathrm{Var}[Y](1-\rho^2_{XY})$$

Efficacy of Control Variates

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

$$\alpha^{\star} = \frac{\mathrm{Cov}[\mathrm{X}, \mathrm{Y}]}{\mathrm{Var}[\mathrm{X}]}$$

$$\mathrm{Var}[Y_{\mathrm{CV}}(\alpha^{\star})] = \mathrm{Var}[Y](1 - \rho_{XY}^{2})$$

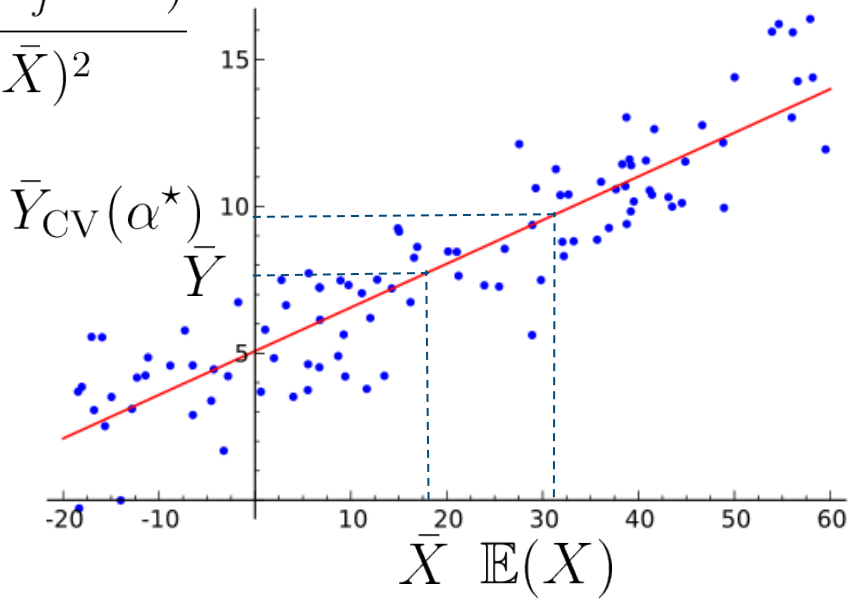
- Computational cost of CV estimator is roughly the same of the standard one (control are usually cheap to compute).
- Therefore, as long as $\rho_{XY} \neq 0$, CV are always beneficial (CV never hurt). Sign of correlation is irrelevant.
- However, the variance reduction increases sharply with $|\rho_{XY}|$.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

ρ	1	0.99	0.95	0.9	0.8	0.7
VR	∞	50	10	5	2.7	1.96

Geometric interpretation

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

Linear Regression Slope



$$\bar{Y} \rightarrow \bar{Y}_{\mathrm{CV}} = \bar{Y} - \alpha^\star(\bar{X} - \mathbb{E}(X))$$

$\bar{Y}_{\mathrm{CV}}(\alpha^\star)$ 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:

$$Y^i_{\mathrm{CV}}(\alpha) = Y_i - \alpha^T (X_i - \mu_X)$$

$$\alpha, X_i \in \mathbb{R}^d$$

Linear Regression coefficient

$$\alpha^\star = \Sigma_X^{-1} \Sigma_{XY}$$

$$(\Sigma_X)_{j,k} = \mathrm{Cov}[X^j, X^k] \qquad (\Sigma_{X,Y})_j = \mathrm{Cov}[X^j, Y]$$

Optimal Variance:

$$\mathrm{Var}[Y_{\mathrm{CV}}(\alpha^\star)] = \mathrm{Var}[Y](1 - R^2)$$

$$R^2 = \Sigma_{XY}^T \Sigma_X^{-1} \Sigma_{XY} / \sigma_Y^2$$

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.

Stratified Sampling

W.G. Cochran, Sampling Techniques, John Wiley and Sons, 1977.

Suppose we want to compute

$$\mu_Y = \mathbb{E}[Y]$$

and that the sample space of Y can be decomposed in subsets (strata):

$$\mathbb{P}(\bigcup_{i=1}^{K} A_i) = 1 \qquad A_i \cap A_j = \emptyset$$

Then:

$$\mu_Y = \sum_{i=1}^{K} p(Y \in A_i)\mathbb{E}[Y_i | Y \in A_i]$$

Strata can be defined in terms of another random variable X

$$\mu_Y = \sum_{i=1}^{K} p(X \in A_i)\mathbb{E}[Y_i | X \in A_i]$$

Stratified Sampling (Cont'd)

$$\mu_Y = \sum_{i=1}^{K} p(X \in A_i)\mathbb{E}[Y_i | X \in A_i]$$

Stratified Estimator:

$$\hat{Y}_{\mathrm{ST}}(n) = \sum_{i=1}^{K} p_i \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{i,j} = \frac{1}{n} \sum_{i=1}^{K} \sum_{j=1}^{n_i} Y_{i,j}$$

$$n_i = np_i$$

Clearly removes the inter-strata variability.

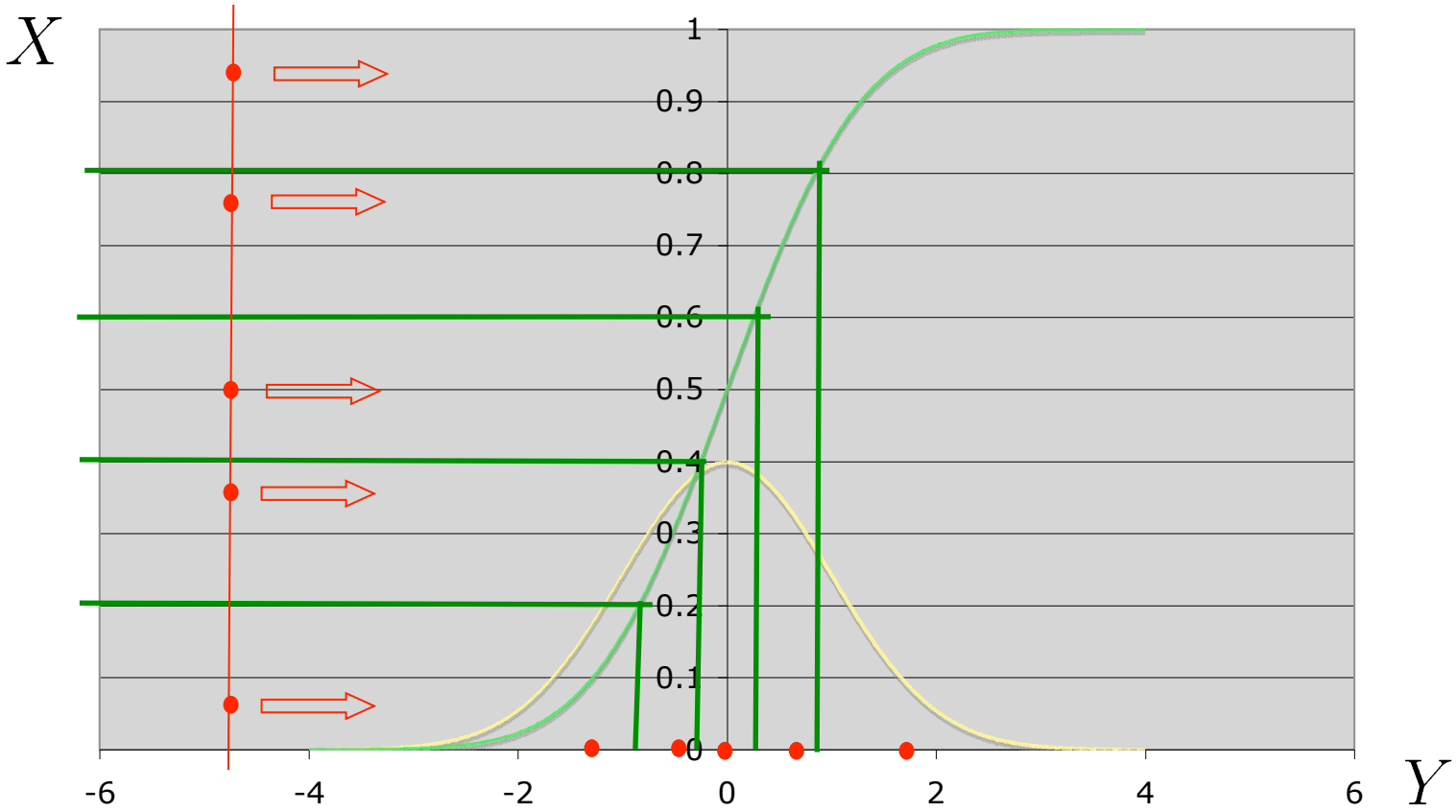
Unbiased Estimator:

$$\mathbb{E}[\hat{Y}_{\mathrm{ST}}] = \frac{1}{n} \sum_{i=1}^{K} \sum_{j=1}^{n_i} p_i \mathbb{E}[Y_{i,j}] = \frac{1}{n} \sum_{i=1}^{K} p_i \mu_i = \mu_Y$$

$$\mu_i = \mathbb{E}[Y | X_i \in A_i]$$

Stratified Sampling (cont'd)

Stratifying a Normal Random Variable



Reducing the Variance by Sampling in a more regular pattern.

Stratified Sampling (Cont'd)

The number of samples in each stratum does not need to be proportional to each stratum's probability:

$$\hat{Y}_{\mathrm{ST}} = \frac{1}{n}\sum_{i=1}^{K}\frac{p_i}{q_i}\sum_{j=1}^{n_i} Y_{i,j} \qquad n_i = q_i n$$

- By optimizing the variance over q_i (e.g. in a set of short pre-simulations) it is possible to find an allocation that is at least as effective as proportional allocation.
- Pre-simulations need to be accurate enough to be reliable as a suboptimal choice of allocation may lead to an increase of variance.
- This is guaranteed not to happen for proportional allocation.

Stratified Sampling (Cont'd)

Variance analysis:

$$\mathrm{Var}[\hat{Y}_{\mathrm{ST}}] = \mathrm{Var}[\sum_{i=1}^{K} \frac{p_i}{n_i} \sum_{j=1}^{n_i} Y_{i,j}] = \frac{1}{n} \sum_{i=1}^{K} p_i \sigma_i^2$$

Compare with the Crude estimator:

$$\sigma_i^2 = \mathrm{Var}[Y|X \in A_i]$$

$$\mathrm{Var}[\hat{Y}_{\mathrm{CR}}] = \mathrm{Var}[\hat{Y}_{\mathrm{ST}}] + \mathrm{Var}[\mathbb{E}[Y|X \in A_i]]$$

$$\mathbb{E}[\mathrm{Var}[Y|X \in A_i]]$$

Stratified Sampling removes the Inter-strata Variance

What's left is the Intra-startum variance:

$$\mathrm{Var}[\hat{Y}_{\mathrm{ST}}] = \frac{1}{n}\mathrm{Var}[Y_{\mathrm{res}}] \qquad Y_{\mathrm{res}} = Y - \mathbb{E}[Y|X \in A_i]$$

Importance Sampling

H. Kahn, Modifications of the Monte Carlo method, Seminar on Scientific Computations, IBM Applied Science Dept., (1949).

$$V = E_P\left[G(Z)\right] = \int_D dZ\; G(Z)\, P(Z)$$

$$Z = (Z_1, \ldots, Z_d)$$

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:

$$\tilde{V} = \frac{1}{N_{\mathrm{MC}}} \sum_{i=1}^{N_{\mathrm{MC}}} W(Z_i) G(Z_i) \qquad Z_i \sim \tilde{P}(Z)$$

$$W(Z) = P(Z)/\tilde{P}(Z)$$

Likelihood Ratio Weight

Importance Sampling (cont'd)

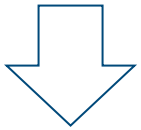
$$\tilde{V} = \frac{1}{N_{\mathrm{MC}}} \sum_{i=1}^{N_{\mathrm{MC}}} W(Z_i) G(Z_i) \qquad Z_i \sim \tilde{P}(Z)$$

$$W(Z) = P(Z)/\tilde{P}(Z)$$

The variance of the new estimator

$$\tilde{\Sigma}^2 = \int_D dZ \; \left(W(Z)\, G(Z) - V\right)^2 \; \tilde{P}(Z)$$

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:

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

$$\tilde{\Sigma}^2 = \int_D dZ \ \left(W(Z)\, G(Z) - V\right)^2 \ \tilde{P}(Z)$$

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:

$$I = \int_0^1 \frac{dx}{\sqrt{x}} \qquad P(x) = 1(0 \le x \le 1) \quad G(x) = \frac{1}{\sqrt{x}}$$

$$\mathrm{Var}\big[G(x)\big] = \int_0^1 dx \frac{1}{x} - \Big(\int_0^1 dx \frac{1}{\sqrt{x}}\Big)^2 = \infty$$

Optimal Sampling density: $\tilde{P}(x) = \dfrac{G(x)P(x)}{I} = \dfrac{1}{2\sqrt{x}}$

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

IS Estimator: $\tilde{G}(x) = W(x)G(x) = 2$

$$\mathrm{Var}[\tilde{G}(x)] = 0 \quad \text{Optimal!}$$

LSIS: Least Squares Importance Sampling

L.C., Least Squares Importance Sampling for Monte Carlo Securities Pricing, Quantitative Finance (2008) **[available on ssrn]**.

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

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

LSIS (Cont'd)

Minimize the Variance

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

… or equivalently minimize

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

Guess for V

MC estimator:

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

A LEAST SQUARES PROBLEM!

LSIS (Cont'd)

Algorithm:

1) Choose a trial probability density and an initial value of the parameters θ.

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

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

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

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

LSIS (Cont'd)

Correlated Sampling makes the approach practical:

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

A limited number of paths is necessary to determine the optimal θ.

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

$$S_2(\theta) \qquad\qquad S_2(\theta')$$

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

European Call

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

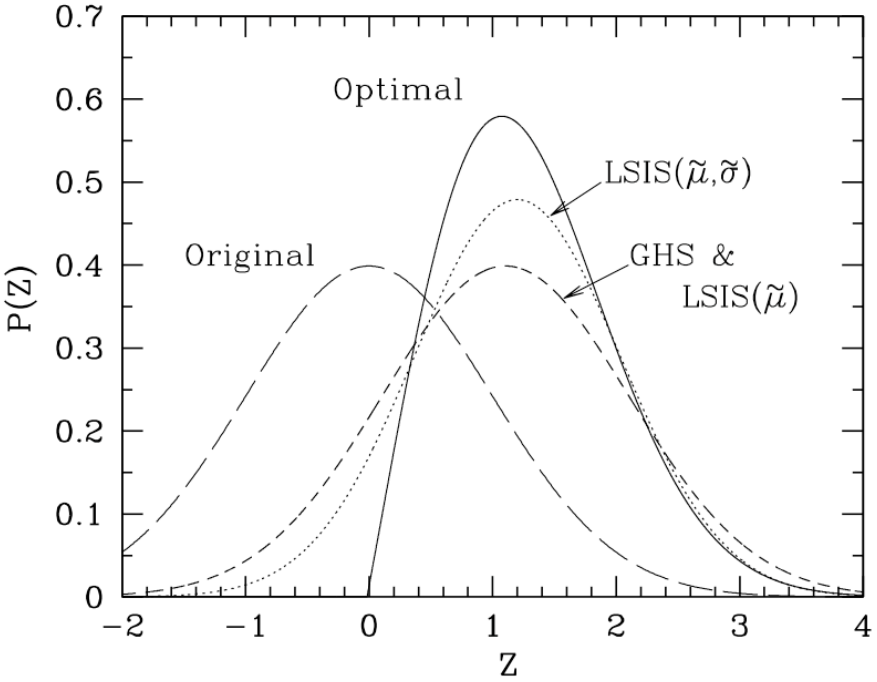
$$P(Z) = (2\pi)^{-1/2} \exp\left(-Z^2/2\right)$$

Trial Density

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

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

$$N_{\mathrm{MC}} \simeq 50$$



σ	K	LSIS($\tilde{\mu}$)	LSIS($\tilde{\mu},\tilde{\sigma}$)	RM	GHS
0.1	30	104(1)	1700(100)	112(4)	100(1)
	50	7.8(1)	15(1)	7.8(4)	7.8(1)
	60	33.5(5)	84(5)	31(2)	33.5(5)
0.3	30	16.4(1)	51(1)	16.8(4)	14.8(2)
	50	9.9(5)	27(1)	11(2)	9.9(1)
	60	15.6(1)	35(1)	15.2(4)	14.2(1)

Variance Reduction

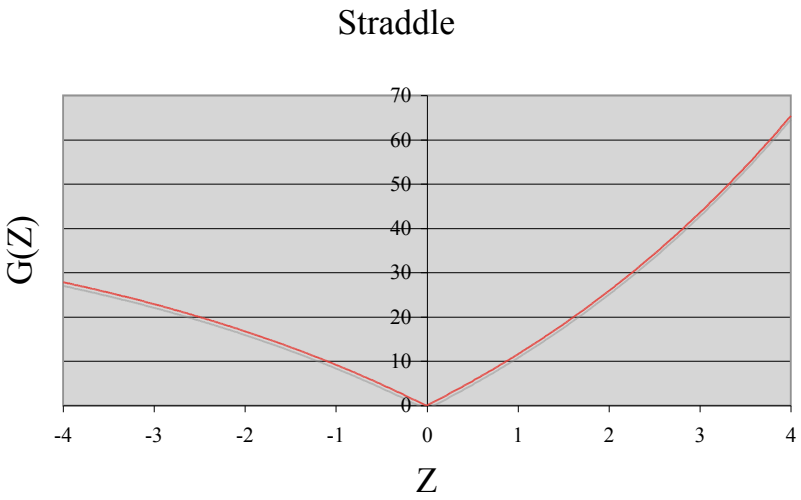
$$\mathrm{VR} = \left(\frac{\sigma(\text{Crude MC})}{\sigma(\text{IS})}\right)^2$$

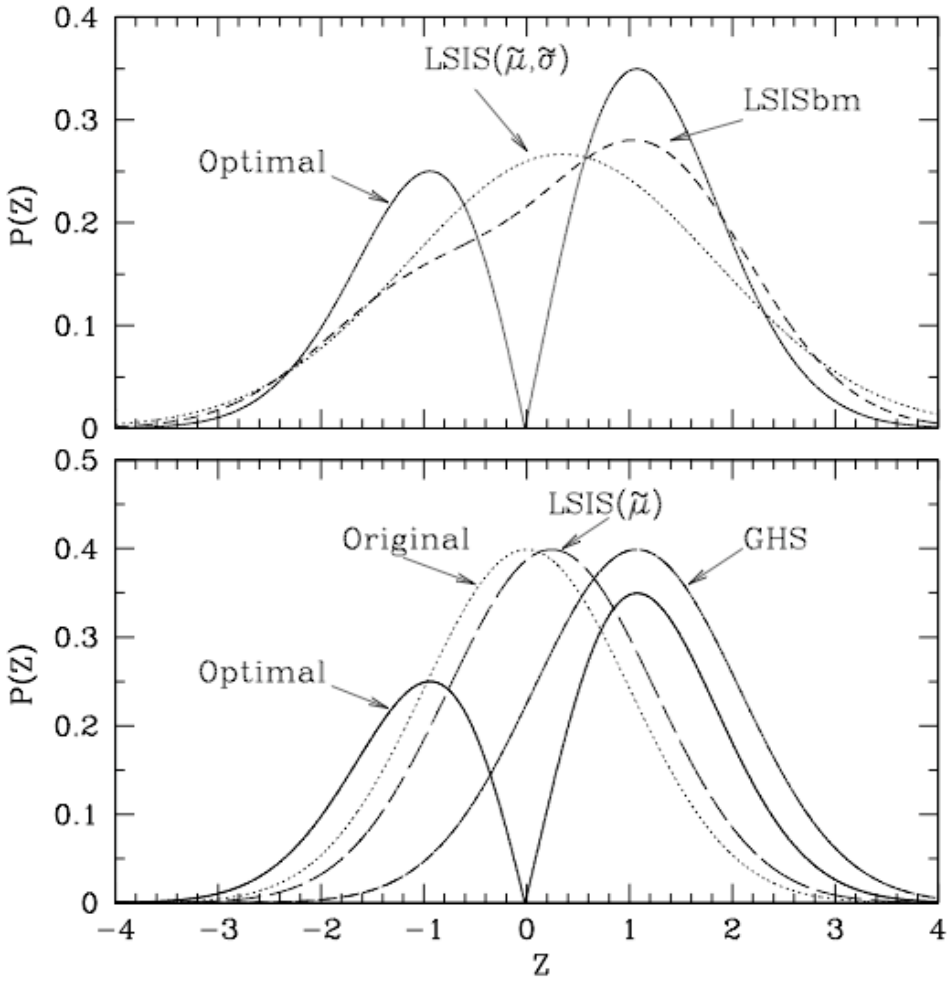
B. Arouna, J. Comp. Finance (2003).

P. Glasserman et al., Math. Finance (1999).

European Straddle

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

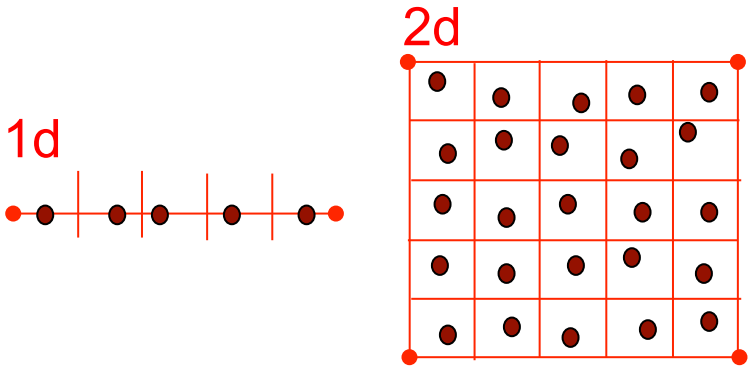




Bimodal Ansatz:

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

LSIS + Stratified Sampling (LSIS+)



Too many sample points to fill the space in high dimensions

I can stratify one-dimensional projections!

P. Glasserman et al., Math. Finance (1999).

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

$X^{(i)}$: 1-d Stratified Normal

$Y^{(i)}$: $N(0, I_d)$

$$\xi \propto \mu(\text{LSIS})$$

LSIS+

Asian Option with Stratified Sampling:

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

M	σ	K	VR (LSIS)	VR (LSIS+)
16	0.3	45	8.8(1)	950(20)
		50	9.9(1)	1225(15)
		55	13.6(7)	1900(100)
64	0.3	45	9.0(2)	1060(30)
		50	10.3(5)	1290(30)
		55	12.5(5)	1320(100)

Computational Speed-Up of 3 orders of magnitude!

Case Study I:

LSIS for Libor Market Model

L.C., Wilmot Magazine 2007 **[available on ssrn]**.

Euler Discretization:

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

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

Risk-Neutral Drift

This fits in the general framework:

$$V = E_P\left[G(Z)\right] = \int_D dZ\; G(Z)\, P(Z)$$

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

Trial Density $\quad \tilde{P}_{\tilde{\mu}}(Z) = (2\pi)^{-d/2}\; e^{-(Z-\tilde{\mu})^2/2}$

Linear parameterization of the drift (knot points)

Caplet

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

T_m (years)	K	N_k	LSIS	LSIS+
1.0	0.04	1	11.4(1)	1349(1)
1.0	0.055	1	13.3(2)	2300(2)
1.0	0.07	1	20.2(1)	4126(4)
2.5	0.04	1	14.0(1)	1189(1)
2.5	0.055	1	15.5(1)	897(1)
2.5	0.07	1	18.1(1)	1831(1)
5.0	0.040	1	12.7(1)	235.2(5)
5.0	0.060	1	12.5(1)	237.0(5)
5.0	0.080	1	14.5(1)	193.3(4)
7.0	0.04	1	7.9(3)	40.0(1)
7.0	0.055	1	8.5(4)	43.7(1)
7.0	0.07	1	8.5(4)	40(1)

Speed-Ups:
2-3 order of magnitude

$$N_{\mathrm{MC}} \simeq 100$$

Swaptions

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

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

Speed-Ups:

1-2 order of magnitude

Straddle

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

T_m (years)	K	N_k	LSIS	LSIS (MM)
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)

MM guess provides sizable improvements

Speed-Ups:
1 order of magnitude

Bi-Modal ansatz

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

Module 1: Summary

- Efficient Monte Carlo Sampling
  - Multidimensional Integrals, the curse of dimensionality and MC
  - Computational efficiency: Variance and Statistical Uncertainties
- Variance Reduction Techniques (Classic Approaches)
  - **Antithetic Variates**
    Pros: Easy to implement
    Cons: Limited Benefits especially in high dimensions
  - **Control Variates**
    Pros: They (virtually) never hurt. Maybe very powerful
    Cons: Difficult in general to find suitable controls
  - **Stratified Sampling**
    Pros: They never hurt (when using proportional allocation)
    Cons: Strata are hard to find for multidimensional problems
  - **Importance Sampling**
    Pros: Powerful technique, especially for rare events
    Cons: Requires a good trial density

Module 1: Summary (cont'd)

- 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
  - **Case Study I**: LSIS for Libor Market Models

References:

- P. Glasserman, Monte Carlo Methods in Financial Engineering, Springer.
- L.C., Least Squares Importance Sampling for Monte Carlo Securities Pricing, Quantitative Finance (2008) **[available on ssrn]**.
- L.C., Least Squares Importance Sampling for Libor Market Models, Wilmott Magazine (2009) **[available on ssrn]**.

Module 2:

Efficient Risk Management in Monte Carlo: Classical Approaches

Outline

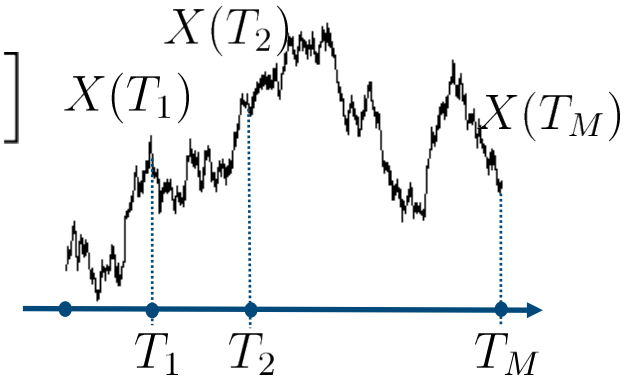
- Hedges and Price Sensitivities (Greeks)
  - Shortcoming of the Finite Differences ('no time to think') approach
- Likelihood Ratio Method
  - Basic Idea: Differentiating the Probability Distribution
  - Pros and Cons
  - Case Study II: Reducing the Variance of Likelihood Ratio Greeks
- Pathwise Derivative Method
  - Basic Idea: Differentiating the Estimator "Path-by-Path"
  - Problems with Discontinuous Payouts
  - Handling Discontinuous Payouts
  - Examples
  - Pros and Cons

Greeks in Monte Carlo

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

$$dX_t = rX_t dt + \sigma X_t dW_t$$

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



Model Parameters: $\theta = (X(T_0), \sigma, r, \ldots)$

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

$$\frac{\partial V(\theta)}{\partial \theta_1} = \frac{\partial V(\theta)}{\partial \sigma} \Rightarrow \text{Vega}$$

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

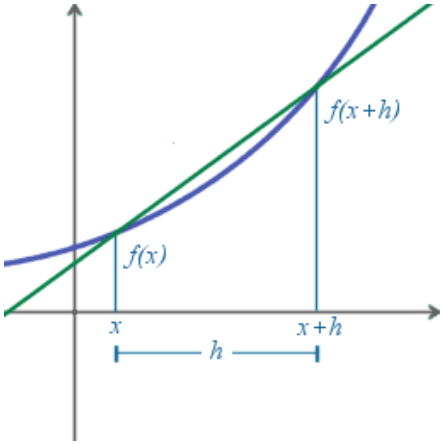
Finite Differences Approaches (`Bumping')

$$\frac{\partial V(\theta)}{\partial \theta} \simeq \frac{V(\theta + h) - V(\theta)}{h}$$

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

Sources of Error:

Bias



Variance

$$\mathrm{Var}\left[\frac{V(\theta + h) - V(\theta)}{h}\right]$$

Variance of Finite Difference Estimators

$$\frac{\partial V(\theta)}{\partial \theta} \simeq \frac{V(\theta+h)-V(\theta)}{h}$$

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

MC estimators

$$\mathrm{Var}\left[\tilde{V}(\theta+h)-\tilde{V}(\theta)\right] \underset{h\to 0}{\begin{cases} O(1) \\ O(h) \\ O(h^2) \end{cases}}$$

Infinite Variance of the Sensitivity Estimator

Finite Variance of the Sensitivity Estimator

Variance of Finite Difference Estimators (cont'd)

$$\mathrm{Var}[\frac{\tilde{V}(\theta+h)-\tilde{V}(\theta)}{h}] = \frac{1}{h^2}\mathrm{Var}[\tilde{V}(\theta+h)-\tilde{V}(\theta)]$$

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

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

$$\mathrm{Var}[\tilde{V}(\theta+h)]+\mathrm{Var}[\tilde{V}(\theta)] \to 2\mathrm{Var}[\tilde{V}(\theta)]$$

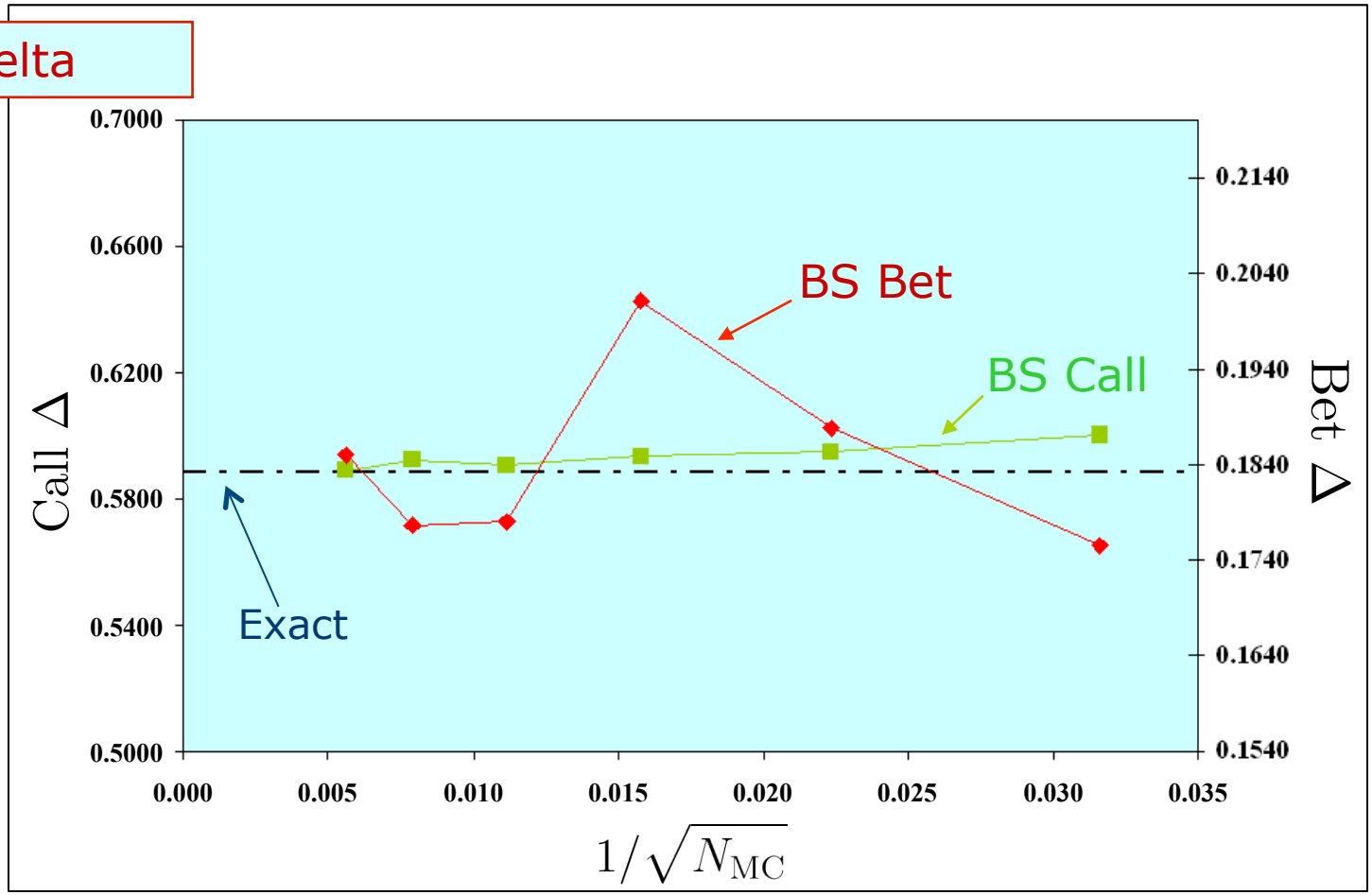
($\mathrm{Var}[\tilde{V}(\theta)]$continuous)

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

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

$$\exists \kappa \ s.t. \ |G(x)-G(y)| \leq \kappa \, ||x-y| \ \forall \ x,y$$

Delta



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

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

- 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 $\int dx\, G(x)\, \partial_\theta P_\theta(x)$?

$$\int dx\, G(x) \partial_\theta P_\theta(x) = \int dx\, G(x) \frac{\partial_\theta P_\theta(x)}{P_\theta(x)} P_\theta(x)$$

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

LRM weight

The 'sign' problem of the LRM estimators

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

- 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

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

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)

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

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

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

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 49. 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 II:

Reducing the Variance of Likelihood Ratio Greeks in MC

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

- Gaussian Copula Models

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

Market Implied Marginals

Probability Density Function

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

LRM weight:

$$m_i(x_i) = dM_i(x_i)/dx_i$$

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

$$Z_i = \Phi^{-1}(M_i(x_i)) \qquad \partial_\theta Z_i = \frac{\partial_\theta M_i(x_i)}{\phi(\Phi^{-1}(M_i(x_i)))}$$

Reducing the Variance of Likelihood Ratio Greeks in MC

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

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

$$X_i = \sigma_i \sqrt{T} Z_i \qquad \mu_i = \log S_i^0 + (r - \sigma_i^2/2)T$$

$$Z_i = \frac{\log S_i/S_i^0 - (r - \sigma_i^2/2)T}{\sigma_i \sqrt{T}} \qquad \hat{\Sigma}_{ij} = \sigma_i \sigma_j \Sigma_{ij}$$

Delta Weight

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

Vega Weight

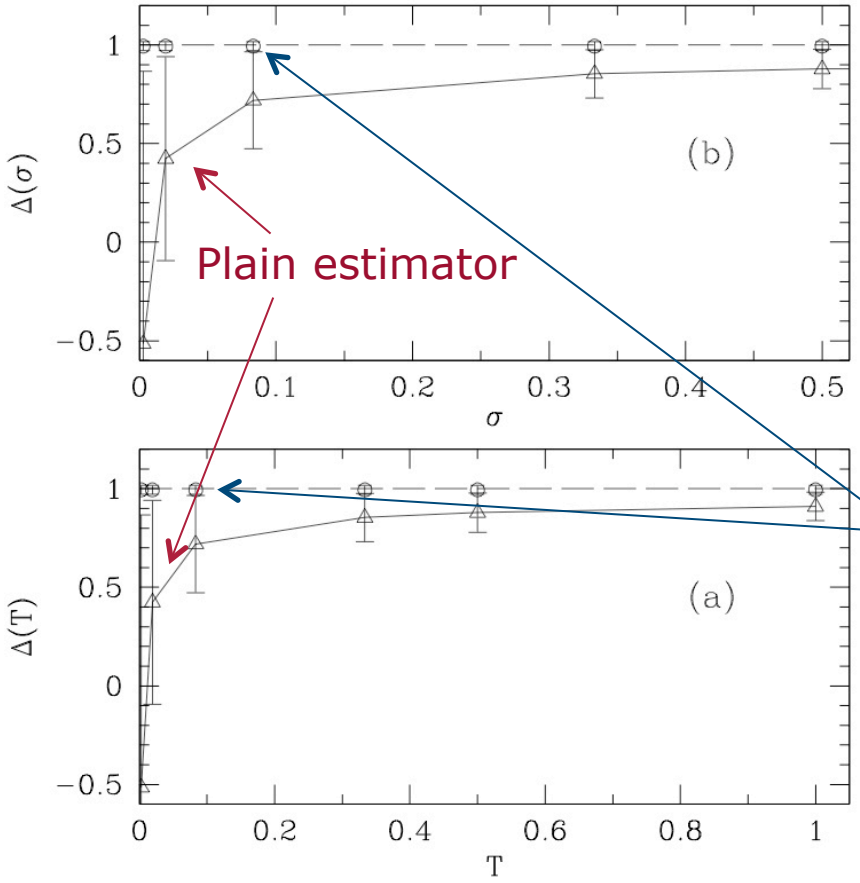
$$\Omega_{\sigma_i}(Z) = \left(\frac{Z_i}{\sigma_i} - \sqrt{T}\right)[\Sigma^{-1} Z]_i - \frac{1}{\sigma_i}$$

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



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

$$\mathrm{Var}[\Omega_{S_i^0}] \underset{\sigma_i\sqrt{T}\to 0}{\longrightarrow} \infty$$

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

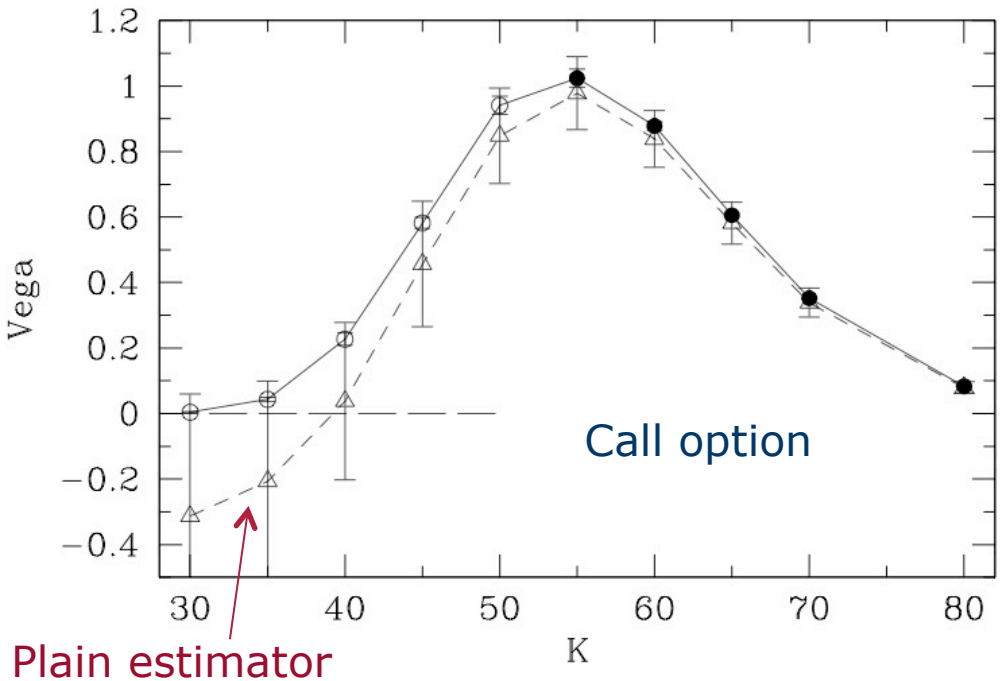
Deep in the money call option

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

K	AV	AV+CV	AV+LSIS
30	1.3(1)	$4(1)10^4$	11(1)
40	1.7(1)	150(10)	7.0(6)
50	2.1(1)	22(2)	8.0(8)
60	2.1(2)	5.5(6)	80(10)
70	1.9(2)	5.2(5)	340(40)
80	1.7(3)	1.7(3)	1100(100)

Pathwise Derivative Method

- Monte Carlo Expectation Values

$$V(\theta) = \mathbb{E}_{\mathbb{Q}}\Big[P\big(X(T_1),\ldots,X(T_M)\big)\Big]$$

… and sensitivities

$$\frac{\partial V(\theta)}{\partial \theta_k} = \mathbb{E}_{\mathbb{Q}}\Big[\frac{\partial P\big(X(\theta)\big)}{\partial \theta_k}\Big]$$

Lipschitz

- Pathwise Derivative Estimator

$$\bar{\theta}_k \equiv \frac{\partial P\big(X(\theta)\big)}{\partial \theta_k} = \sum_{j=1}^{N\times M} \frac{\partial P\big(X\big)}{\partial X_j} \times \frac{\partial X_j(\theta)}{\partial \theta_k}$$

Chain Rule

Payout Derivatives

Tangent Process

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:

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

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

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

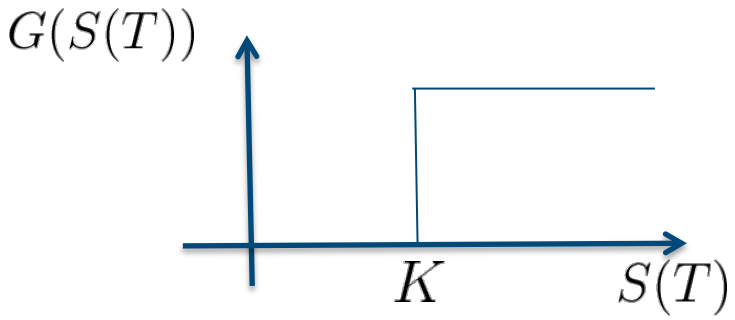
Tangent Process

$$\frac{\partial S(T)}{\partial \sigma} = S(T)\left(\sqrt{T}Z - \sigma T\right)$$

Pathwise Derivative Method: Simple Examples

What about digitals?

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



NOT LIPSCHITZ!

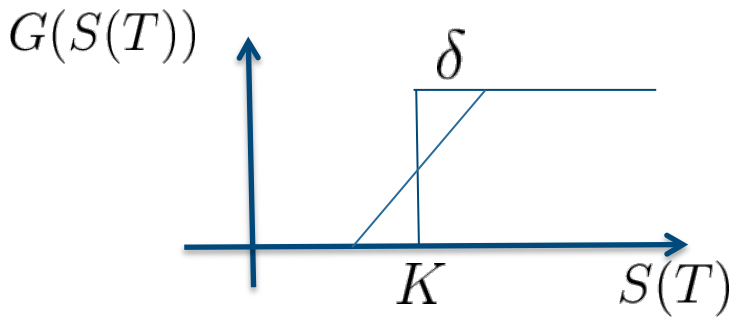
$$\frac{\partial G(S(T))}{\partial S(T)} = 0 \quad (S(T) \neq K) \quad \Rightarrow \quad \frac{\partial V}{\partial S_0} = 0$$

Actually…

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

Which cannot be sampled with Monte Carlo

Pathwise Derivative Method: Simple Examples



Homework assignment:

Smooth out the payout above with 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 δ .

Pathwise Derivative Method: Simple Examples

Asian option:

$$G(X_1, \ldots, X_M) = e^{-rt_M}(\bar{X} - K)^+$$

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

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:

$$\bar{\theta}_{X_0} = \sum_{i=1}^{M} \frac{\partial G(X)}{\partial X_i}\frac{\partial X_i}{\partial X_0} = \frac{\partial G(X)}{\partial \bar{X}} \sum_{i=1}^{M} \frac{\partial \bar{X}}{\partial X_i}\frac{\partial X_i}{\partial X_0} = e^{-rT}\theta(\bar{S} - K)\frac{\bar{X}}{X_0}$$

Homework assignment:

Work out the estimator for Vega.

Pathwise Derivative Method: Libor Market Model

Log Euler scheme:

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

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

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:

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

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

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

The variance of the Finite Difference Estimators is asymptotically equal to the variance of the Pathwise Derivative Estimator

Pathwise Derivative Method: Challenges

$$\bar{\theta}_k \equiv \frac{\partial P\big(X(\theta)\big)}{\partial \theta_k} = \sum_{j=1}^{N\times M} \frac{\partial P\big(X\big)}{\partial X_j} \times \frac{\partial X_j(\theta)}{\partial \theta_k}$$

Payout Derivatives

Tangent Process

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

Module 2: Summary

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

References

- J. M. Hammersley, and K. W. Morton, *A new Monte Carlo Technique: Antithetic Variates*, Proc. Cambridge Philosophical Society **52**, 449 (1956).
- H. Kahn and A. W. Marshall, *Methods of reducing sample size in Monte Carlo computations*, Oper. Res. **1**, 263 (1953).
- W.G. Cochran, *Sampling Techniques*, John Wiley and Sons, (1977).
- H. Kahn, *Modifications of the Monte Carlo method,* Seminar on Scientific Computations, IBM Applied Science Dept., (1949).
- Luca Capriotti, *Reducing the Variance of Likelihood Ratio Greeks in Monte Carlo*, Proceedings of Winter Simulation Conference (2008).

Also available at **www.luca-capriotti.net** or **ssrn**