Rare event simulation related to financial risks

Gang Liu
gang.liu@polytechnique.edu

CMAP, Ecole Polytechnique, France
joint work with
Ankush Agarwal, Stefano De Macro, Emmanuel Gobet

April 18, 2016
Colloque Jeunes Probabilistes et Statisticiens, Les Houches

Table of contents

Introduction

Questions:

- estimate $p := \mathbb{P}(X \in A)$ and $\mathbb{E}(\varphi(X)|X \in A)$ when $p < 10^{-5}$
- sample from $X|X \in A$
- compute sensitivity like $\frac{\partial_\theta \mathbb{E}(\varphi(X^\theta)\mathbf{1}_{X^\theta \in A})}{\mathbb{E}(\varphi(X^\theta)\mathbf{1}_{X^\theta \in A})}$

Introduction

Questions:

- estimate $p := \mathbb{P}(X \in A)$ and $\mathbb{E}(\varphi(X)|X \in A)$ when $p < 10^{-5}$
- sample from $X|X \in A$
- compute sensitivity like $\frac{\partial_\theta \mathbb{E}(\varphi(X^\theta)\mathbf{1}_{X^\theta \in A})}{\mathbb{E}(\varphi(X^\theta)\mathbf{1}_{X^\theta \in A})}$

Simple Monte Carlo: $(X_n)_{n \geq 1}$ i.i.d. copies of X, by CLT

$$\sqrt{N}(S_N - \mathbb{P}(X \in A)) \to \mathcal{N}(0, p(1-p))$$

where $S_N = \frac{1}{N}\sum_{k=1}^N 1_{\{X_k \in A\}}$
95% confidence interval: $(S_N - 1.96\sqrt{\frac{p(1-p)}{N}}, S_N + 1.96\sqrt{\frac{p(1-p)}{N}})$

Introduction

Questions:

- estimate $p := \mathbb{P}(X \in A)$ and $\mathbb{E}(\varphi(X)|X \in A)$ when $p < 10^{-5}$
- sample from $X|X \in A$
- compute sensitivity like $\frac{\partial_\theta \mathbb{E}(\varphi(X^\theta)\mathbf{1}_{X^\theta \in A})}{\mathbb{E}(\varphi(X^\theta)\mathbf{1}_{X^\theta \in A})}$

Simple Monte Carlo: $(X_n)_{n\geq 1}$ i.i.d. copies of X, by CLT

$$\sqrt{N}(S_N - \mathbb{P}(X \in A)) \to \mathcal{N}(0, p(1-p))$$

where $S_N = \frac{1}{N}\sum_{k=1}^N 1_{\{X_k \in A\}}$

95% confidence interval: $(S_N - 1.96\sqrt{\frac{p(1-p)}{N}}, S_N + 1.96\sqrt{\frac{p(1-p)}{N}})$

But $\frac{\sqrt{p(1-p)}}{\sqrt{N}p} \approx \frac{1}{\sqrt{Np}}$ is large for small p, which means large relative error.

Importance sampling

Classic technique: importance sampling

Importance sampling

Classic technique: importance sampling

Example: X follows $\mathcal{N}(0,1)$, to estimate $\mathbb{P}(X>5)$, we define another probability $\mathbb{Q}$ by

$$\frac{d\mathbb{Q}}{d\mathbb{P}}(X) = \exp\{aX - \frac{1}{2}a^2\}$$

under $\mathbb{Q}$ X follows $\mathcal{N}(a,1)$, so with $a=5$ and (X_n) i.i.d copies of $\mathcal{N}(5,1)$

$$\mathbb{P}(X>5) = \mathbb{E}^{\mathbb{Q}}(1_{X>5}\frac{d\mathbb{P}}{d\mathbb{Q}}) \approx \frac{1}{N}\sum_{n=1}^{N} 1_{X_n>5}\frac{d\mathbb{P}}{d\mathbb{Q}}(X_n)$$

Importance sampling

Classic technique: importance sampling

Example: X follows $\mathcal{N}(0,1)$, to estimate $\mathbb{P}(X > 5)$, we define another probability $\mathbb{Q}$ by

$$\frac{d\mathbb{Q}}{d\mathbb{P}}(X) = \exp\{aX - \frac{1}{2}a^2\}$$

under $\mathbb{Q}$ X follows $\mathcal{N}(a,1)$, so with $a = 5$ and (X_n) i.i.d copies of $\mathcal{N}(5,1)$

$$\mathbb{P}(X > 5) = \mathbb{E}^{\mathbb{Q}}(1_{X>5}\frac{d\mathbb{P}}{d\mathbb{Q}}) \approx \frac{1}{N}\sum_{n=1}^{N} 1_{X_n>5}\frac{d\mathbb{P}}{d\mathbb{Q}}(X_n)$$

Unfortunately, in general case, it's not easy to design such a new probability. When X is a complicated random system(stochastic process, random matrix, random graph, etc), new techniques need to be found.

Reformulation using conditional probabilities

Classic technique: importance sampling. However, in general it is difficult to implement this method.

We define a series of nested subsets of the entire probability space $\mathbb{S}$

$$\mathbb{S} := A_0 \supset \cdots \supset A_k \supset \cdots \supset A_n := A$$

$$\mathbb{P}(X \in A) = \prod_{k=1}^{n} \mathbb{P}(X \in A_k | X \in A_{k-1})$$

Reformulation using conditional probabilities

Classic technique: importance sampling. However, in general it is difficult to implement this method.

We define a series of nested subsets of the entire probability space $\mathbb{S}$

$$\mathbb{S} := A_0 \supset \cdots \supset A_k \supset \cdots \supset A_n := A$$

$$\mathbb{P}(X \in A) = \prod_{k=1}^{n} \mathbb{P}(X \in A_k | X \in A_{k-1})$$

Question: how to estimate $\mathbb{P}(X \in A_k | X \in A_{k-1})$?

Reformulation using conditional probabilities

Classic technique: importance sampling. However, in general it is difficult to implement this method.

We define a series of nested subsets of the entire probability space $\mathbb{S}$

$$\mathbb{S} := A_0 \supset \cdots \supset A_k \supset \cdots \supset A_n := A$$

$$\mathbb{P}(X \in A) = \prod_{k=1}^{n} \mathbb{P}(X \in A_k | X \in A_{k-1})$$

Question: how to estimate $\mathbb{P}(X \in A_k | X \in A_{k-1})$?

Existing methods: splitting/restart, interacting particles system(IPS).
We propose an new method using the ergodicity of Markov chain

Definition of shaking transformation

Definition: Given a random object X(variable, process, $\cdots$), $\mathcal{K}(\cdot)$ is a reversible shaking transformation for X if:

$$(X, \mathcal{K}(X)) \stackrel{\mathrm{d}}{=} (\mathcal{K}(X), X). \quad (1)$$

We also write $\mathcal{K}(X) = K(X, Y)$, where K is deterministic and Y is independent of X

Definition of shaking transformation

Definition: Given a random object X(variable, process, $\cdots$), $\mathcal{K}(\cdot)$ is a reversible shaking transformation for X if:

$$(X, \mathcal{K}(X)) \stackrel{\mathrm{d}}{=} (\mathcal{K}(X), X). \tag{1}$$

We also write $\mathcal{K}(X) = K(X, Y)$, where K is deterministic and Y is independent of X

Examples:

- If X is a standard normal variable

$$K(X, \mathbf{N}(0,1)) = \rho X + \sqrt{1-\rho^2}\mathbf{N}(0,1), -1 \leq \rho \leq 1$$

- If X is a standard Brownian motion

$$K(X, G') = (\int_0^t \rho_s dX_s + \int_0^t \sqrt{1-\rho_s^2} dG'_s)_{0 \leq t \leq T}$$

Shaking with rejection and conditional invariance

Let $k \in \{0, 1, \cdots, n-1\}$, define the shaking with rejection $\mathcal{M}_k^{\mathcal{K}}$ by

$$\mathcal{M}_k^{\mathcal{K}}(X) = \begin{cases} \mathcal{K}(X) & \text{if } \mathcal{K}(X) \in A_k \\ X & \text{if } \mathcal{K}(X) \notin A_k. \end{cases} \tag{2}$$

Shaking with rejection and conditional invariance

Let $k \in \{0, 1, \cdots, n-1\}$, define the shaking with rejection $\mathcal{M}_k^{\mathcal{K}}$ by

$$\mathcal{M}_k^{\mathcal{K}}(X) = \begin{cases} \mathcal{K}(X) & \text{if } \mathcal{K}(X) \in A_k \\ X & \text{if } \mathcal{K}(X) \notin A_k. \end{cases} \tag{2}$$

Proposition (conditional invariance)

Let $k \in \{0, 1, \cdots, n-1\}$. The distribution of X conditionally on $\{X \in A_k\}$ is invariant w.r.t. the random transformation $\mathcal{M}_k^{\mathcal{K}}$: i.e. for any bounded (random) measurable $\varphi : \mathbb{S} \to \mathbb{R}$, we have

$$\mathbb{E}\big(\varphi(\mathcal{M}_k^{\mathcal{K}}(X))|X \in A_k\big) = \mathbb{E}\big(\varphi(X)|X \in A_k\big). \tag{3}$$

POP(Parallel One-Path) method

Birkhoff's theorem for ergodic Markov chain $(Z_i)_{i \geq 0}$ with a unique invariant distribution π:

$$\frac{1}{N} \sum_{i=0}^{N-1} f(Z_i) \underset{N \to +\infty}{\longrightarrow} \int f \mathrm{d}\pi \qquad a.s.$$

POP(Parallel One-Path) method

Birkhoff's theorem for ergodic Markov chain $(Z_i)_{i\geq 0}$ with a unique invariant distribution π:

$$\frac{1}{N}\sum_{i=0}^{N-1} f(Z_i) \underset{N\to+\infty}{\longrightarrow} \int f \mathrm{d}\pi \qquad a.s.$$

Observation: the conditional invariance of $\mathcal{M}_k^{\mathcal{K}}$ with respect to $X|X \in A_k$ enables to use the ergodic property of Markov chain

POP(Parallel One-Path) method

Birkhoff's theorem for ergodic Markov chain $(Z_i)_{i\geq 0}$ with a unique invariant distribution π:

$$\frac{1}{N}\sum_{i=0}^{N-1} f(Z_i) \underset{N\to+\infty}{\longrightarrow} \int f \mathrm{d}\pi \qquad a.s.$$

Observation: the conditional invariance of $\mathcal{M}_k^{\mathcal{K}}$ with respect to $X|X\in A_k$ enables to use the ergodic property of Markov chain

Given an initial position $X_{k,0}\in A_k$, we define $X_{k,i} := \mathcal{M}_k^{\mathcal{K}}(X_{k,i-1})$

$$\mathbb{E}(\varphi(X)|X\in A_k) \approx \frac{1}{N}\sum_{i=0}^{N-1}\varphi(X_{k,i})$$

With $\varphi \equiv \mathbf{1}_{A_{k+1}}$, $\mathbb{P}(X\in A_{k+1}|X\in A_k) \approx \frac{1}{N}\sum_{i=0}^{N-1}\mathbf{1}_{A_{k+1}}(X_{k,i})$

POP(Parallel One-Path) method

Birkhoff's theorem for ergodic Markov chain $(Z_i)_{i\geq 0}$ with a unique invariant distribution π:

$$\frac{1}{N}\sum_{i=0}^{N-1} f(Z_i) \underset{N\to+\infty}{\longrightarrow} \int f \mathrm{d}\pi \qquad a.s.$$

Observation: the conditional invariance of $\mathcal{M}_k^{\mathcal{K}}$ with respect to $X|X \in A_k$ enables to use the ergodic property of Markov chain

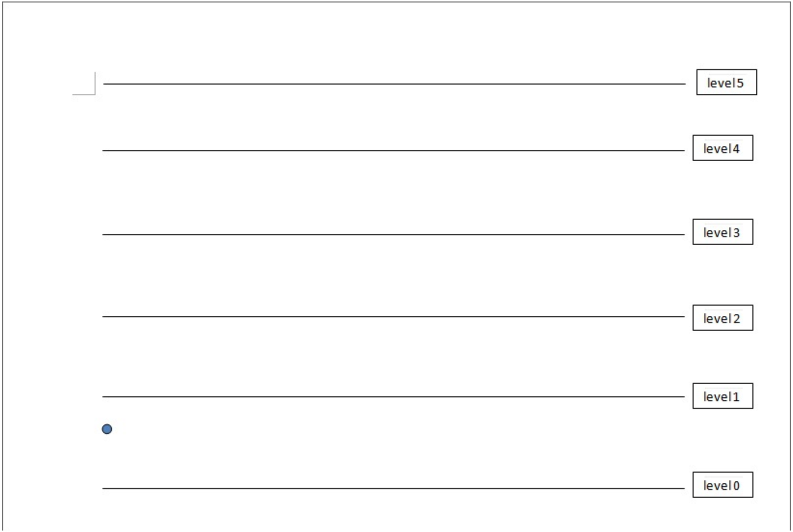
Given an initial position $X_{k,0} \in A_k$, we define $X_{k,i} := \mathcal{M}_k^{\mathcal{K}}(X_{k,i-1})$

$$\mathbb{E}(\varphi(X)|X \in A_k) \approx \frac{1}{N}\sum_{i=0}^{N-1} \varphi(X_{k,i})$$

With $\varphi \equiv \mathbf{1}_{A_{k+1}}$, $\mathbb{P}(X \in A_{k+1}|X \in A_k) \approx \frac{1}{N}\sum_{i=0}^{N-1} \mathbf{1}_{A_{k+1}}(X_{k,i})$

Our estimators for each $P(X \in A_{k+1}|X \in A_k)$ can be made independent!

POP playing



convergence POP

For all finite dimension cases, we can prove POP method converges almost surely using a short proof for Markov chain's ergodicity from Asmussen and Glynn (2011)

For convergence rate(Łatuszyński et al. (2013))

$$\theta = \pi(f), \hat{\theta} = \frac{1}{N} \sum_{i=1}^{N} f(X_i)$$

under some stronger assumptions, there exists constant C such that

$$\mathbb{E}(\hat{\theta} - \theta)^2 \leq \frac{C}{N}$$

Explicit shaking construction

- If X is a standard normal variable

$$K_\rho(X, \mathbf{N}(0,1)) = \rho X + \sqrt{1-\rho^2}\mathbf{N}(0,1), -1 \le \rho \le 1$$

- If X is a standard Brownian motion

$$K_\rho(X, G') = (\int_0^t \rho_s dX_s + \int_0^t \sqrt{1-\rho_s^2} dG'_s)_{0\le t\le T}$$

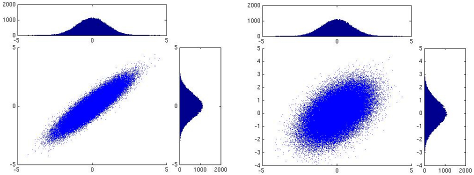


Figure: Shaking $N(0,1)$ with $\rho = 0.9$ and $\rho = 0.5$

Explicit shaking construction

- For a Gamma distribution $Ga \sim Gamma(\alpha, \beta)$, i.e

$$\mathbb{P}(Ga \in dx) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx, x > 0$$

The transformation is (see Dufresne (1998))

$$K(Ga) = Ga * Beta(\alpha(1-p), \alpha p) + Gamma(\alpha p, \beta)$$

In particular, it applies for exponential variable with $\alpha = 1$

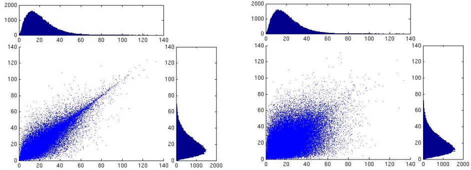


Figure: Shaking $Gamma(2.5, 0.12)$ with $p = 0.1$ and $p = 0.5$

Shaking list

- Poisson variable $P \sim \mathcal{P}(\lambda)$: $\mathcal{K}(P) = Binomial(P, p) + \mathcal{P}((1-p)\lambda)$
- Bernoulli variable $B \sim Bernoulli(q)$: $qP(1,0) = (1-q)P(0,1)$

$$Y \stackrel{\mathrm{d}}{=} f(X) \implies \mathcal{K}_Y(\cdot) = f(\mathcal{K}_X(f^{-1}(\cdot)))$$

- Uniform U: $-\ln U \stackrel{\mathrm{d}}{=} Exp(1)$
- Cauchy C: $\frac{1}{\pi}\arctan(C) + \frac{1}{2}$ is uniform
- $\chi^2(k)$ R_k: $R_k \stackrel{\mathrm{d}}{=} 2Gamma(\frac{k}{2}, 1)$

Other shakings

- if $Y = f(X_1, X_2, \cdots, X_n)$, shake Y through shaking all the X_i's
- Metropolis-Hasting Gibbs type shaking
- Given a large number of r.v.'s, we can also only shake a randomly sampled part of them

Shaking transformation for stochastic process

- Compound Poisson process: Let $X_t = \sum_{k=1}^{N_t} Y_k$ be a $CPP(\lambda, \mu)$

$$\text{CPP decomposition: } X_t = X_t^a + X_t^b$$

where $X^a \stackrel{\mathrm{d}}{=} CPP((1-p)\lambda, \mu)$ and $X^b \stackrel{\mathrm{d}}{=} CPP(p\lambda, \mu)$

$$K(X, Z) = (X_t^a + Z_t)_{0 \le t \le T}, Z_t \stackrel{\mathrm{d}}{=} CPP(p\lambda, \mu)$$

- Let Y be a pure jump process with inter-arrival $(A_n)_{n\geq 1}$ and $(B_n)_{n\geq 1}$, shake all the A_n's and B_n's $\implies$ shake Y.
- Conditional shaking, keep inter-arrival $(A_n)_{n\geq 1}$, only shake $(B_n)_{n\geq 1}$.

Others possibilities: To shake a Levy process for exampke, we can apply shaking transformations for the underlying Brownian motion and compound Poisoon process.

Adaptive POP method

Question: How to choose the values of intermediary levels?

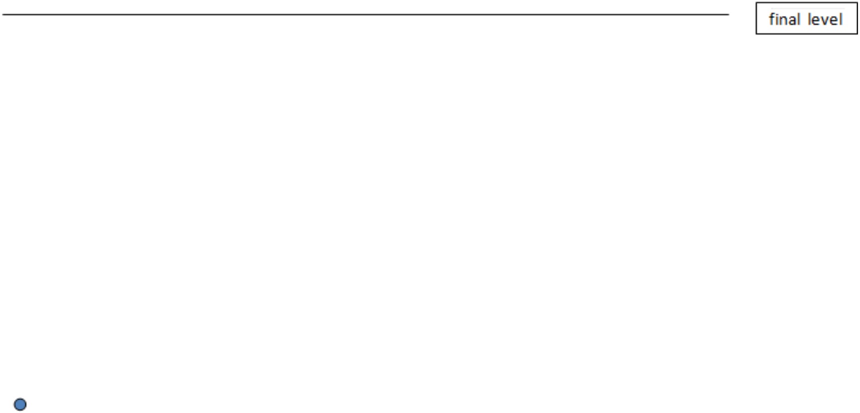
Adaptive POP method

Question: How to choose the values of intermediary levels?

In case that no additional information is available about the model, we can choose our nested subset on the run, i.e. in an adaptive way.

We propose an adaptive version of POP method and prove it converges almost surely.

Adaptive POP playing with 50% quantile



Extreme scenario generation and sensitivity

Extreme scenario generation: recall the Markov chain defined by

$$X_{k,0} \in A_k, X_{k,i} := \mathcal{M}_k^{\mathcal{K}}(X_{k,i-1})$$

we have $\|\mathcal{L}(X_{k,i}) - X|X \in A_k\|_{TV} \to 0$

Extreme scenario generation and sensitivity

Extreme scenario generation: recall the Markov chain defined by

$$X_{k,0} \in A_k, X_{k,i} := \mathcal{M}_k^{\mathcal{K}}(X_{k,i-1})$$

we have $\|\mathcal{L}(X_{k,i}) - X|X \in A_k\|_{TV} \to 0$

Sensitivity: by likelihood method or Malliavin calculus, there exists some ϕ such that

$$\frac{\partial_\theta \mathbb{E}(\varphi(X^\theta)\mathbf{1}_{X^\theta \in A})}{\mathbb{E}(\varphi(X^\theta)\mathbf{1}_{X^\theta \in A})} = \frac{\mathbb{E}(\phi \mathbf{1}_{X^\theta \in A})}{\mathbb{E}(\varphi(X^\theta)\mathbf{1}_{X^\theta \in A})} = \frac{\mathbb{E}(\phi | X^\theta \in A)}{\mathbb{E}(\varphi(X^\theta) | X^\theta \in A)}$$

which can be evaluated using only one Markov chain

Oscillation of Orstein-Ulhenbeck process

$$dY_t = \lambda(\mu - Y_t)dt + \sigma dW_t, Y_0 = 0, \lambda = 1, \mu = 0, \sigma = 1, T = 1$$

$$\mathbb{P}(\max_{0 \le l \le 100} \tilde{Y}_{t_l} > 1.6 \text{ and } \min_{0 \le l \le 100} \tilde{Y}_{t_l} < -1.6)$$

7×10^9 MC simulation gives $[3.9709, 4.3691] \times 10^{-7}$
Set $L_i = 1.6 * (\frac{i}{5})^{1/2}$ and $A_i = (\max_{0 \le l \le 100} \tilde{Y}_{t_l} > L_i \text{ and } \min_{0 \le l \le 100} \tilde{Y}_{t_l} < -L_i)$

100 runs for each parameter:

IPS: $M = 10^5$

	mean	std	std/mean
$\rho = 0.9$	4.01e-07	1.23e-07	0.31
$\rho = 0.75$	4.10e-07	1.67e-07	0.41
$\rho = 0.5$	2.44e-07	4.76e-07	1.95

POP: $N = 10^5$

	mean	std	std/mean
$\rho = 0.9$	4.14e-07	2.68e-08	0.06
$\rho = 0.75$	4.18e-07	4.60e-08	0.11
$\rho = 0.5$	4.29e-07	1.26e-07	0.29

Oscillation of Orstein-Ulhenbeck process

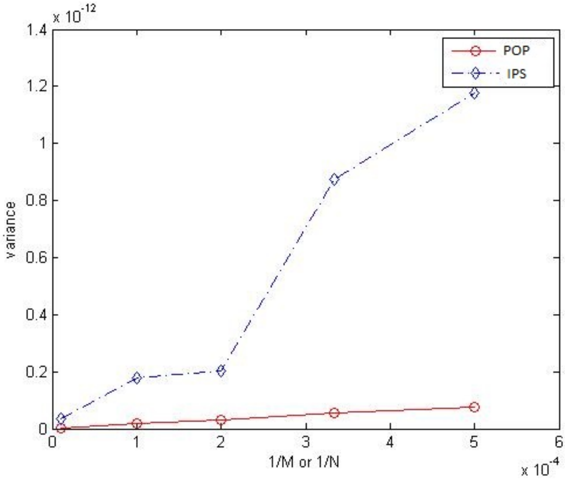


Figure: Variance for two methods

Model misspecification and robustness

Real world: σ_+ when spot is lower than past M-4, σ_- in the other case.

Trader thinks it's a constant volatility σ_-, hedging the payoff $(S_T - K)_+$ with BS. With $T = 1$, $S_0 = 10$, $\sigma_- = 0.2$, $\sigma_+ = 0.27$ $K = 10$ and $L = -2.4$, what is the probability that the trader's P&L is less than L?

The crude Monte Carlo method with 5×10^8 simulations provides a 99% confidence interval $[2.93, 3.34] \times 10^{-6}$.

Model misspecification and robustness

$M = N = 10^5$:

with prefixed intermediary levels $L_k = \frac{k}{5} * L, k = 1, 2, 3, 4, 5$

	IPS			POP		
	mean ($\times 10^{-6}$)	std. ($\times 10^{-7}$)	std./mean	mean ($\times 10^{-6}$)	std. ($\times 10^{-7}$)	std./mean
$\rho = 0.9$	3.10	5.29	0.17	3.13	2.07	0.07
$\rho = 0.7$	3.23	13.3	0.41	3.11	3.98	0.13
$\rho = 0.5$	2.79	25.9	0.93	3.18	8.44	0.27

adaptive methods:

	IPS			POP		
	mean ($\times 10^{-6}$)	std. ($\times 10^{-7}$)	std./mean	mean ($\times 10^{-6}$)	std. ($\times 10^{-7}$)	std./mean
$\rho = 0.9$	3.06	4.95	0.16	3.18	2.42	0.08
$\rho = 0.7$	2.98	11.1	0.37	3.10	3.71	0.12
$\rho = 0.5$	2.45	23.6	0.96	3.06	7.27	0.24

Typical scenario leading to large hedging loss

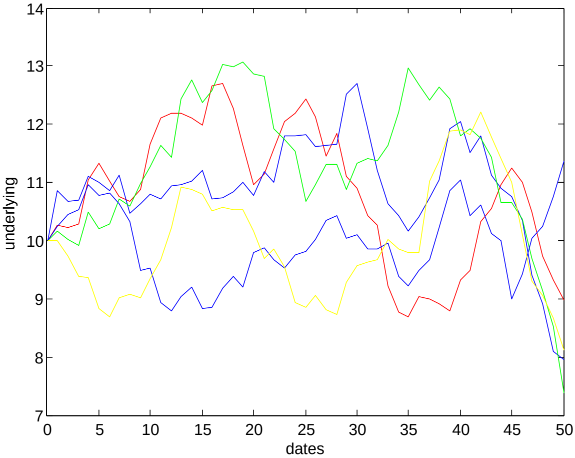


Figure: Typical paths of the underlying stock price which lead to large hedging loss

Thank You