Optimal Importance Sampling in Securities Pricing (Yi Su and Michael C. Fu)

6-601-09 Simulation Monte Carlo

Geneviève Gauthier

HEC Montréal

Formulation and Settings I

Assumptions

1. The financial market is arbitrage-free.
   - There is an equivalent probability measure Q.
2. European contingent claim paying X_T at time T.
   1. $\mathrm{E}^Q\left[X_T^2\right] < \infty$.
3. $\{r_t : t \geq 0\}$ is the risk free spot rate process
4. $C_T = \exp\left(-\int_0^T r_s ds\right) X_T$ is the present value of the payoff.
5. The time t price of the claim is $C_0 = \mathrm{E}^Q\left[C_T\right]$.

Formulation and Settings I

Example

1. The underlying asset follows a GBM

$$dS_t = rS_t dt + \sigma S_t dW_t^Q$$

where W^Q is a Q standard Brownian motion.

2. For a call option, $X_T = \max(S_T - K; 0)$
3. $C_T = \exp(-rT) X_T$ is the present value of the payoff.

Formulation and Settings I

Change of measure

1. $C_0 = \mathrm{E}^{Q}\left[C_T\right] = \mathrm{E}^{Q^*}\left[\frac{dQ}{dQ^*} C_T\right]$
   1. where $\frac{dQ}{dQ^*}$ is the Radon-Nikodym derivative.
2. Formulation with last course's notation:

$$
\begin{aligned}
C_0 &= \int C(s) f_S(s)\, ds \\
&= \int C(s) \frac{f_S(s)}{\phi(s)} \phi(s)\, ds \\
&= \mathrm{E}^{Q^*}\left[C(S_T) \frac{f_S(S_T)}{\phi(S_T)}\right]
\end{aligned}
$$

   1. S_T price of the underlying asset at time T.
   2. $\frac{f_S(S_T)}{\phi(S_T)}$ is the Radon-Nikodym derivative.

Formulation and Settings I

Variance reduction

1. The goal is to minimize the variance of the price estimator which involves

$$
\begin{aligned}
& \mathrm{Var}^{Q^*}\left[\frac{dQ}{dQ^*}C_T\right] \\
= & \mathrm{E}^{Q^*}\left[\left(\frac{dQ}{dQ^*}C_T\right)^2\right] - \left(\mathrm{E}^{Q^*}\left[\frac{dQ}{dQ^*}C_T\right]\right)^2 \\
= & \mathrm{E}^{Q^*}\left[\left(\frac{dQ}{dQ^*}C_T\right)^2\right] - C_0^2
\end{aligned}
$$

2. It is not possible to minimize the variance if we do not restrict Q^* to belong to a family of measures.

Formulation and Settings II

Variance reduction

3. $\{Q(\theta) : \theta \in \Theta\}$ is the family of measures we consider
   1. θ is the parameter
   2. Θ is a compact set
   3. $\forall \theta \in \Theta$, $Q(\theta)$ is absolutely continuous with respect to Q.
4. The variance reduction is then reduced to

$$\min_{\theta \in \Theta} \mathrm{E}^{Q(\theta)} \left[\left(\frac{dQ}{dQ(\theta)} C_T \right)^2 \right].$$

Formulation and Settings I

Example

1. The underlying asset follows a GBM

$$dS_t = rS_t dt + \sigma S_t dW_t^Q$$

where W^Q is a Q standard Brownian motion.

2. For a call option, $X_T = \max(S_T - K; 0)$
3. $C_T = \exp(-rT) X_T$ is the present value of the payoff.
4. Let

$$W_t^{Q(\theta)} = W_t^Q - \theta t.$$

and assume that we simulate

$$\begin{aligned} dS_t &= rS_t dt + \sigma S_t dW_t^{Q(\theta)} \\ &= (r - \sigma\theta) S_t dt + \sigma S_t dW_t^Q \\ &= \lambda S_t dt + \sigma S_t dW_t^Q \end{aligned}$$

instead of

$$dS_t = rS_t dt + \sigma S_t dW_t^Q$$

Formulation and Settings II

Example

5. If we use $W^{Q(\theta)}$ instead of W^Q, then the likelihood ratio is

$$\frac{f(w)}{\phi(w)} = \frac{\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{T}}\exp\left(-\frac{w^2}{2T}\right)}{\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{T}}\exp\left(-\frac{(w-\theta T)^2}{2T}\right)} = \exp\left(-\theta w + \frac{1}{2}\theta^2 T\right).$$

Moreover, under $Q(\theta)$, $C_T = \exp(-rT)X_T$ is a function of θ while it is not the case under Q.

Formulation and Settings I

Vazquez-Abad & Dufresne

1. Vazquez-Abad & Dufresne attack the minimization problem by applying

$$\frac{\partial}{\partial\theta}\mathrm{E}^{Q(\theta)}\left[\left(\frac{dQ}{dQ(\theta)}C_T\right)^2\right].$$

2. Under some technical conditions (we discuss this topic further in this presentation),

$$\frac{\partial}{\partial\theta}\mathrm{E}^{Q(\theta)}\left[\left(\frac{dQ}{dQ(\theta)}C_T\right)^2\right] = \mathrm{E}^{Q(\theta)}\left[\frac{\partial}{\partial\theta}\left(\frac{dQ}{dQ(\theta)}C_T\right)^2\right].$$

   1. Requires derivatives for both C_T and $\frac{dQ}{dQ(\theta)}$ since C_T does depend on θ under $Q(\theta)$.

Formulation and Settings I

Su & Fu

1. Since

$$
\begin{aligned}
V(\theta) &= \mathrm{E}^{Q(\theta)}\left[\left(\frac{dQ}{dQ(\theta)} C_T\right)^2\right] \\
&= \mathrm{E}^{Q}\left[\frac{dQ(\theta)}{dQ}\left(\frac{dQ}{dQ(\theta)} C_T\right)^2\right] \\
&= \mathrm{E}^{Q}\left[\frac{dQ}{dQ(\theta)} C_T^2\right],
\end{aligned}
$$

the minimization problem is numerically easier to solve when dealing with $\frac{\partial}{\partial\theta}\mathrm{E}^{Q}\left[\frac{dQ}{dQ(\theta)} C_T^2\right]$ since C_T^2 does not depend on θ under the measure Q.

Stochastic approximation I

General context

1. **Gradient-based stochastic approximation** (SA) (like Vazquez-Abad & Dufresne)

$$\theta^* = \arg\min_{\theta \in \Theta} V(\theta)$$

where $V(\theta) = \mathrm{E}^Q \left[\frac{dQ}{dQ(\theta)} C_T^2 \right]$ via the following iterative scheme

$$\theta_{n+1} = \Pi_\Theta \left(\theta_n - a_n \widehat{g}_n \right)$$

where

1. θ_n is the nth iteration,
2. $\widehat{g}_n$ represents an estimate of the gradient $\nabla V(\theta)$,
3. $\{a_n : n \in \{1, 2, 3, ...\}\}$ is a positive sequence of numbers converging to zero,
4. Π_Θ is a projection on Θ.

Stochastic approximation II

General context

2. The main difference between Su & Fu and Vazquez-Abad & Dufresne if the form of $V(\theta)$ used in the **infinitesimal perturbation analysis (IPA) estimator**:

   1. Vazquez-Abad & Dufresne:

   $$V(\theta) = \mathrm{E}^{Q(\theta)}\left[\left(\frac{dQ}{dQ(\theta)}C_T\right)^2\right].$$

   2. Su & Fu:

   $$V(\theta) = \mathrm{E}^{Q}\left[\frac{dQ}{dQ(\theta)}C_T^2\right].$$

Stochastic approximation I

Technicalities

1. **Assumption 1**. $L(\theta) = \frac{dQ}{dQ(\theta)}$ is piecewise differentiable on Θ.
2. **Intuition**. Differentiation inside $\mathrm{E}^Q\left[L(\theta)\, C_T^2\right]$ leads the IPA estimator $C_T^2 \frac{\partial L}{\partial \theta}(\theta)$.
   1. that is, $\frac{\partial V}{\partial \theta}(\theta) = \frac{\partial}{\partial \theta}\mathrm{E}^Q\left[C_T^2 L(\theta)\right] = \mathrm{E}^Q\left[C_T^2 \frac{\partial L}{\partial \theta}(\theta)\right]$

Definition

The **infinitesimal perturbation analysis (IPA) estimator** is $C_T^2 \frac{\partial L}{\partial \theta}(\theta)$.

The following Theorem shows that under some suitable conditions, the IPA estimator is unbiased (under the measure Q).

Stochastic approximation II

Technicalities

Theorem

Unbiasness of the IPA estimator. *If*
(a) Assumption 1 holds,
(b) $\exists M(\theta)$ *s.t.* $\|L(\theta+\Delta\theta) - L(\theta)\| < M(\theta)\|\Delta\theta\|$ $Q-a.s.$ *uniformly as* $\Delta\theta \longrightarrow 0$,
and either
(i) $\exists\delta > 0$, $E^Q\left[C_T^{2+2\delta}\right] < \infty$, *and* $E^Q\left[M(\theta)^{1+\frac{1}{\delta}}\right] < \infty$ *or*
(ii) $E^Q\left[C_T^2 M(\theta)\right] < \infty$
then $C_T^2 \frac{\partial L}{\partial\theta}(\theta)$ *is an unbiased estimator of* $\frac{\partial V}{\partial\theta}(\theta)$ *under measure* Q.

Stochastic approximation I

Proof

1. The proof is based on the Dominated convergence theorem

Theorem

Dominated convergence theorem (DCT).

1. If $X_1, X_2, ..., X, Y$ are Borel-measurable,
2. $X_n \to X$ $P-$a.s.,
3. $\forall n$, $|X_n| < Y$ and
4. Y is integrable,

2. then X is integrable and $\lim_{n\to\infty} \mathrm{E}^P [X_n] = \mathrm{E}^P [\lim_{n\to\infty} X_n] = \mathrm{E}^P [X]$.

Stochastic approximation II

Proof

3. **Proof of the Theorem.** If (ii) $\mathrm{E}^Q\left[C_T^2 M(\theta)\right] < \infty$, then

$$\begin{aligned}
\frac{\partial V}{\partial \theta}(\theta) &= \frac{\partial}{\partial \theta}\mathrm{E}^Q\left[C_T^2 L(\theta)\right] \\
&= \lim_{\Delta\theta \to 0} \frac{\mathrm{E}^Q\left[C_T^2 L(\theta+\Delta\theta)\right] - \mathrm{E}^Q\left[C_T^2 L(\theta)\right]}{\Delta\theta} \\
&= \lim_{\Delta\theta \to 0} \mathrm{E}^Q\left[C_T^2 \frac{L(\theta+\Delta\theta) - L(\theta)}{\Delta\theta}\right] \\
&= \mathrm{E}^Q\left[C_T^2 \lim_{\Delta\theta \to 0} \frac{L(\theta+\Delta\theta) - L(\theta)}{\Delta\theta}\right] \quad \text{(DCT)} \\
&= \mathrm{E}^Q\left[C_T^2 \frac{\partial L}{\partial \theta}(\theta)\right].
\end{aligned}$$

Stochastic approximation III

Proof

4. If (i) $\exists\delta > 0$, $\mathrm{E}^Q\left[C_T^{2+2\delta}\right] < \infty$ and $\mathrm{E}^Q\left[M(\theta)^{1+\frac{1}{\delta}}\right] < \infty$ then (ii) is satisfied since Hölder's inequality[1] implies that

$$\begin{aligned} & \mathrm{E}^Q\left[C_T^2 M(\theta)\right] \\ = & \left(\mathrm{E}^Q\left[C_T^{2(1+\delta)}\right]\right)^{\frac{1}{1+\delta}} \left(\mathrm{E}^Q\left[(M(\theta))^{\frac{1+\delta}{\delta}}\right]\right)^{\frac{\delta}{1+\delta}}. \ \square \end{aligned}$$

[1] $\mathrm{E}\left[|XY|\right] \le \left(\mathrm{E}\left[|X|^{\frac{1}{p}}\right]\right)^p \left(\mathrm{E}\left[|Y|^{\frac{1}{q}}\right]\right)^q$, $p, q \geqslant 0$, $p + q = 1$.

Stochastic approximation I

Convexity

1. The use of the first derivative to find an optimum is OK if the function is convex.
2. That is the goal of the second theorem.

Theorem

Convexity. *If*
(a) $L(\theta)$ *and* C_T *satisfy the conditions of the previous theorem and, in addition,*
(b) $\frac{\partial^2 L}{(\partial\theta)^2}(\theta) > 0$ $Q-a.s.$ *and*
(c) $\exists G(\theta)$ *such that* $\left\| \frac{\partial L}{\partial\theta}(\theta + \Delta\theta) - \frac{\partial L}{\partial\theta}(\theta) \right\| < G(\theta) \|\Delta\theta\|$ $Q-a.s.$ *uniformly as* $\Delta\theta \to 0$, *and*
(d) $\mathrm{E}^Q\left[C_T^2 G(\theta)\right] < \infty$ *(I think there is a typo in the paper)*
then $V(\theta)$ *is a convex function for* θ.

Stochastic approximation II

Convexity

Proof. From Theorem 1,

$$\frac{\partial}{\partial\theta}\mathrm{E}^{Q}\left[C_T^2 L(\theta)\right] = \mathrm{E}^{Q}\left[C_T^2 \frac{\partial L}{\partial\theta}(\theta)\right].$$

Stochastic approximation III

Convexity

Therefore,

$$
\begin{aligned}
&\frac{\partial^2}{\partial\theta^2}\mathrm{E}^Q\left[C_T^2 L(\theta)\right] \\
&= \lim_{\Delta\theta\to 0}\frac{\frac{\partial}{\partial\theta}\mathrm{E}^Q\left[C_T^2 L(\theta)\right]\Big|_{\theta=\theta+\Delta\theta} - \frac{\partial}{\partial\theta}\mathrm{E}^Q\left[C_T^2 L(\theta)\right]}{\Delta\theta} \\
&= \lim_{\Delta\theta\to 0}\frac{\mathrm{E}^Q\left[C_T^2\frac{\partial L}{\partial\theta}(\theta+\Delta\theta)\right]\Big| - \mathrm{E}^Q\left[C_T^2\frac{\partial L}{\partial\theta}(\theta)\right]}{\Delta\theta} \\
&= \lim_{\Delta\theta\to 0}\mathrm{E}^Q\left[C_T^2\frac{\frac{\partial L}{\partial\theta}(\theta+\Delta\theta) - \frac{\partial L}{\partial\theta}(\theta)}{\Delta\theta}\right] \\
&= \mathrm{E}^Q\left[C_T^2\lim_{\Delta\theta\to 0}\frac{\frac{\partial L}{\partial\theta}(\theta+\Delta\theta) - \frac{\partial L}{\partial\theta}(\theta)}{\Delta\theta}\right] \quad \text{(by DCT)} \\
&= \mathrm{E}^Q\left[C_T^2\frac{\partial^2 L}{(\partial\theta)^2}(\theta)\right] > 0. \ \square
\end{aligned}
$$

Stochastic approximation I

Practical consideration

1. Although derived under the measure Q, implementation of the gradient estimator can also be carried out under an alternative measure such as $Q(\theta)$.
2. In this case, the IPA estimator $C_T^2 \frac{\partial L}{\partial \theta}(\theta)$ becomes $C_T^2 \frac{\partial L}{\partial \theta}(\theta) L(\theta)$ since

$$
\begin{aligned}
\frac{\partial V}{\partial \theta}(\theta) &= \frac{\partial}{\partial \theta} \mathrm{E}^Q \left[C_T^2 L(\theta) \right] \\
&= \mathrm{E}^Q \left[C_T^2 \frac{\partial L}{\partial \theta}(\theta) \right] \quad \text{(Theorem 1)} \\
&= \mathrm{E}^{Q(\theta)} \left[C_T^2 \frac{\partial L}{\partial \theta}(\theta) \frac{dQ}{dQ(\theta)} \right] \\
&= \mathrm{E}^{Q(\theta)} \left[C_T^2 \frac{\partial L}{\partial \theta}(\theta) L(\theta) \right].
\end{aligned}
$$

Stochastic approximation II

Practical consideration

3. This is likely to be advantageous in the same situation in which the change of measure for estimating the price itself is beneficial, since the gradient estimator also contains the term C_T^2.

Definition

IPA-$Q(\theta)$ estimator is $C_T^2 \frac{\partial L}{\partial \theta}(\theta) L(\theta)$.

Since $\mathrm{E}^{Q(\theta)}\left[C_T^2 \frac{\partial L}{\partial \theta}(\theta) L(\theta)\right] = \frac{\partial V}{\partial \theta}(\theta)$, then the IPA-$Q(\theta)$ estimator $C_T^2 \frac{\partial L}{\partial \theta}(\theta) L(\theta)$ is unbiased for $\frac{\partial V}{\partial \theta}(\theta)$ under $Q(\theta)$.

Application to diffusion I

1. The underlying asset price process follows

$$dS_t = \mu\left(S_t, t\right) dt + \sigma\left(S_t, t\right) dW_t^Q$$

   where W^Q is a $Q-$Brownian motion.
2. Set $W_t^{Q(\theta)} = W_t^Q - \theta t$
3. By Girsanov's theorem,
   1. there is a measure $Q\left(\theta\right)$ under which $W^{Q(\theta)}$ is a Brownian motion,
   2.

$$\begin{aligned} L\left(\theta\right) &= \frac{dQ}{dQ\left(\theta\right)} \\ &= \exp\left(-\theta W_T^{Q(\theta)} - \frac{1}{2}\theta^2 T\right) \\ &= \exp\left(-\theta W_T^Q + \frac{1}{2}\theta^2 T\right) \end{aligned}$$

Application to diffusion II

4. **Intuitively**, according to the notation of previous presentation,

$$\begin{aligned} C_0 &= \mathrm{E}^{Q}\left[C\left(W_T\right)\right] \\ &= \int C\left(w\right) f\left(w\right) dw \\ &= \int C\left(w\right) \frac{f\left(w\right)}{\phi\left(w\right)} \phi\left(w\right) dw \\ &= \mathrm{E}^{Q(\theta)}\left[C\left(W_T^{Q(\theta)}\right) \frac{f\left(W_T^{Q(\theta)}\right)}{\phi\left(W_T^{Q(\theta)}\right)}\right]. \end{aligned}$$

Therefore, $L\left(\theta\right) = \frac{f\left(W_T^Q\right)}{\phi\left(W_T^Q\right)}$ is the ratio of two densities, f being the density of W_T^Q under Q and ϕ being its density under $Q\left(\theta\right)$.

Application to diffusion III

Since, $W_T^Q \overset{Q}{\sim} N(0, T)$ and $W_T^Q \overset{Q(\theta)}{\sim} N(\theta T, T)$, then

$$
\begin{aligned}
L(\theta) &= \frac{\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{T}}\exp\left(-\frac{1}{2}\frac{\left(W_T^Q\right)^2}{T}\right)}{\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{T}}\exp\left(-\frac{1}{2}\frac{\left(W_T^Q-\theta T\right)^2}{T}\right)} \\
&= \exp\left(-\frac{1}{2}\frac{\left(W_T^Q\right)^2}{T}+\frac{1}{2}\frac{\left(W_T^Q-\theta T\right)^2}{T}\right) \\
&= \exp\left(-\theta W_T^Q+\frac{1}{2}\theta^2 T\right)
\end{aligned}
$$

Application to diffusion IV

5. Note that

$$\begin{aligned}\frac{\partial L}{\partial \theta}(\theta) &= \frac{\partial}{\partial \theta}\exp\left(-\theta W_T^Q + \frac{1}{2}\theta^2 T\right)\\ &= \left(-W_T^Q + \theta T\right)\exp\left(-\theta W_T^Q + \frac{1}{2}\theta^2 T\right)\\ &= \left(-W_T^Q + \theta T\right) L(\theta)\\ &= -W_T^{Q(\theta)} L(\theta).\end{aligned}$$

Application to diffusion V

6.

$$\begin{aligned}\frac{\partial^2 L}{\partial\theta^2}(\theta) &= \frac{\partial}{\partial\theta}\left(-W_T^Q+\theta T\right)L(\theta)\\ &= T\exp\left(-\theta W_T^Q+\frac{1}{2}\theta^2 T\right)\\ &\quad+\left(-W_T^Q+\theta T\right)^2\exp\left(-\theta W_T^Q+\frac{1}{2}\theta^2 T\right)\\ &= \left(T+\left(-W_T^Q+\theta T\right)^2\right)L(\theta)\\ &> 0\ Q-\text{a.s. (Assumption of Theorem 2)}\end{aligned}$$

Stochastic approximation I

IPA-Q(theta)

1. As discussed before, it is usually preferable to use the IPA-Q(θ) estimator $C_T^2 \frac{\partial L}{\partial \theta}(\theta) L(\theta)$ for the gradient.
2. **Example**. For deep out-of-the-money options, C_T will be zero most of the time under the measure Q, and this could lead to a large variance when estimating the gradient.

Stochastic approximation II

IPA-Q(theta)

3. For the Brownian motion setting, since
   1. $L(\theta) = \exp\left(-\theta W_T^{Q(\theta)} - \frac{1}{2}\theta^2 T\right)$ and
   2. $\frac{\partial L}{\partial \theta}(\theta) = -W_t^{Q(\theta)} L(\theta)$,

then the IPA-Q(θ) estimator is

$$
\begin{aligned}
& C_T^2 \frac{\partial L}{\partial \theta}(\theta) L(\theta) \\
= \quad & C_T^2 \left(-W_T^{Q(\theta)} L(\theta)\right) L(\theta) \\
= \quad & -C_T^2 W_T^{Q(\theta)} \left(\exp\left(-\theta W_T^{Q(\theta)} - \frac{1}{2}\theta^2 T\right)\right)^2 \\
= \quad & -C_T^2 W_T^{Q(\theta)} \exp\left(-2\theta W_T^{Q(\theta)} - \theta^2 T\right)
\end{aligned}
$$

Stochastic approximation I

Convergence

Theorem

Convergence (Fu 1990). *If*
(a) $\theta \in \Theta$,
(b) $\frac{\partial V}{\partial \theta}$ *is continuous in* θ,
(c) V *is convex and therefore as a unique minimum in* $\theta^* \in \Theta$ *where* Θ *is a compact set,*
(d) $\theta_{n+1} = \theta_n - a_n g_n(\theta_n)$
(e) $sup_{\theta \in \Theta} \mathrm{E}\left[g_n^2(\theta)\right] < K < \infty$
(f) $\mathrm{E}\left[g_n(\theta_n) \mid \mathcal{F}_n\right] = \frac{\partial V}{\partial \theta}(\theta_n) + \beta_n$
(g) where $\sum_{i=n}^{\infty} |a_i \beta_i| < \infty$, $\sum_{n=1}^{\infty} a_n = \infty$, $\sum_{n=1}^{\infty} a_n^2 < \infty$
then $\theta_n \to \theta^*$ *a.s.*

Stochastic approximation I

Algorithm

1. **Optimization stage - find θ^***
    1. **Initialization**: set $\theta = \theta_0$ and $\varepsilon > 0$
    2. **Loop**: for $n = 1$ to N_1
        1. (N_1 is the maximal number of iterations to determine θ^*)
    3. **Loop**: for $i = 1$ to N_2
        1. N_2 is the sample size required to estimate the gradient
        2. Generate a sample according to $dS_t = (\mu(S_t, t) + \theta_n \sigma(S_t, t))\, dt + \sigma(S_t, t)\, dW_t^{Q(\theta_n)}$.
        3. End of the inner loop.
    4. The IPA-$Q(\theta_n)$ estimator: $g_n(\theta_n) = \frac{1}{N_2} \sum_{i=1}^{N_2} -C_{T,i}^2 W_{T,i}^{Q(\theta_n)} \exp\left(-2\theta W_{T,i}^{Q(\theta_n)} - \theta_n^2 T\right)$.
    5. $\theta_{n+1} = \theta_n - a_n g_n(\theta_n)$.

Stochastic approximation II

Algorithm

6. **stopping criterion**: if $|a_n g_n(\theta_n)| < \varepsilon$, then exit the loop.

2. Set $\theta^* = \theta_{n+1}$
3. **Pricing stage - under** $Q(\theta^*)$
    1. For $i = 1$ to N_3
        1. N_3 is the sample size for the pricing purpose.
    2. Generate a sample according to $dS_t = (\mu(S_t, t) + \theta^* \sigma(S_t, t))\, dt + \sigma(S_t, t)\, dW_t^{Q(\theta^*)}$.
    3. The point estimator for the price is

$$\begin{aligned}\widehat{C}_0 &= \frac{1}{N_3} \sum_{i=1}^{N_3} C(S_{T,i})\, L(\theta^*, i) \\ &= \frac{1}{N_3} \sum_{i=1}^{N_3} C(S_{T,i}) \exp\left(-\theta^* W_{T,i}^{Q(\theta^*)} - \frac{1}{2}(\theta^*)^2 T\right)\end{aligned}$$

Stochastic approximation III

Algorithm

4. The estimator of the variance of the price estimator is

$$\widehat{\mathrm{Var}}^{Q(\theta^*)}\left[\widehat{C}_0\right] = \left(\frac{1}{N_3}\sum_{i=1}^{N_3} C^2\left(S_{T,i}\right)\exp\left(-2\theta^* W_{T,i}^{Q(\theta^*)} - (\theta^*)^2\, T\right)\right) - \widehat{C}_0^2$$

5. The margin of error is

$$\text{Margin } = z_{\frac{\alpha}{2}}\sqrt{\frac{\widehat{\mathrm{Var}}^{Q(\theta^*)}\left[\widehat{C}_0\right]}{N_3}}$$

Computational experiments I

Asian Option

1. The underlying asset price process: $dS_t = rS_t dt + \sigma S_t dW_t^Q$
2. Asian option: $C_T = \exp(-rT)\max(A_T - K; 0)$ $A_T = \frac{1}{N-N_0}\sum_{i=N_0+1}^{N} S_{i\frac{T}{N}}$
3. Set $\theta = \frac{\lambda-\mu}{\sigma}$
   1. Under $Q(\theta)$, $dS_t = \lambda S_t dt + \sigma S_t dW_t^{Q(\theta)}$

Computational experiments II
Asian Option

4. The two estimator that are compared:
   1. Vazquez-Abad & Dufresne: $V(\theta) = \mathrm{E}^{Q(\theta)}\left[(L(\theta)\, C_T)^2\right]$ which implies that the IPA estimator is $\frac{\partial}{\partial\theta}(L(\theta)\, C_T)^2 = 2L(\theta)\, C_T \left(C_T \frac{\partial L}{\partial\theta}(\theta) + L(\theta)\frac{\partial}{\partial\theta} C_T\right)$.
   2. Fu & Su: $V(\theta) = \mathrm{E}^{Q}\left[L(\theta)\, C_T^2\right]$ which implies that the IPA estimator is $\frac{\partial L}{\partial\theta}(\theta)\, C_T^2$ and the IPA-$Q(\theta)$ estimator is $L(\theta)\frac{\partial L}{\partial\theta}(\theta)\, C_T^2$.

Computational experiments III

Asian Option

1. No optimization on the θ
2. $N_3 = 50\ 000$

Table I, p.40: Asian call option

	IPA-VD		IPA-Q (λ)		
	$\frac{\partial V}{\partial \lambda}$	CI	$\frac{\partial V}{\partial \lambda}$	CI	VR
0.2	-175.5	15.7	-178,8	4.28	13
0.3	-93.4	9.2	-96.2	2.06	20
0.4	-38.7	7.3	-40.6	1.44	26
0.5	3.89	8.3	3.83	1.89	19
0.6	45.44	12.0	48.39	3.46	12
0.7	94.88	22.2	104.97	7.41	9
0.8	168.82	41.6	190.81	16.86	6

$S_0 = 50,\ K = 50,\ \sigma = 0.2,\ r = 0,05,\ T = 1$ daily average

Computational experiments IV

Asian Option

1. **Initial value**: λ_0 satisfies $S_0 = \exp\left(-\lambda_0 T\right) K$ so that the expected terminal stock price would be at the strike price.
2. $N_1 = 20$ is the maximal number of iterations to determine θ^*
3. $N_2 = 50$ is the sample size required to estimate the gradient
4. $\varepsilon = 0.001$ (stopping criteria)
    1. There is at most 1000 paths devoted to the determination of the optimal measure $Q\left(\theta\right)$
5. $a_n = a_0 n^{-0.75}$, $a_0 = \left|\frac{1}{g_0\left(\lambda_0\right)}\right|$
6. Ad hoc restriction: $|\Delta\lambda| \leq 0.2$
7. $N_3 = 10\ 000$
8. $S_0 = 50,\ \sigma = 0.2,\ r = 0, 05,\ T = 1$

Computational experiments V

Asian Option

9. Optimal λ^* is taken from Vazquez-Abad & Dufresne obtained by an extensive brute-force search

Table II, p.41: Asian call option

	IS via SA/IPA-$Q(\lambda)$				IS via optimal λ^*		
K	Price	CI	λ	N_1^*	Price	CI	λ^*
30	20.407	0.134	0.26	15	20.407	0.135	0.25
45	8.320	0.114	0.43	20	8.318	0.115	0.40
50	5.675	0.096	0.53	19	5.672	0.096	0.50
55	3.713	0.076	0.55	20	3.718	0.076	0.60
75	0.575	0.022	0.79	18	0.574	0.022	0.80

References I

- Su, Y and C. Fu (2002). Optimal Importance Sampling in Securities Pricing, Journal of Computational Finance vol. 5, no. 4 27-50.
- Vazquez-Abad, F. and D. Dufresne (1998). Accelerated Simulation for Pricing Asian Options, *Proceedings of the Winter Simulation Conference*, 1493-1500.