Project Report I.

Vanilla Option Pricing under Heston-Dynamics

Yunhao Yan (yy2882)

Teammates: Han Lin, Weihao Yan

Summary:

We exploited different pricing methods to price vanilla options with the assumption that the stock price is subject to the dynamics described by the Heston SDE.

In the first part, we implemented the integral transform techniques. Initially we utilize numerical integration to do the Fourier Transform of the asset' characteristic function. Since we use the Black-Scholes Model as the benchmark for Heston Model, which is easy to verify by the analytic formulas, we are very confident about the correctness and accuracy of our results under Heston dynamics. Then Fast Fourier Transform (FFT) is implemented to expediate the integration process. Finally, we implement the Fourier-Cosine Transform to compute the vanilla option's price very efficiently. The Fourier-Cosine Method reduce the algorithm complexity to $O(n)$; in practice, the cosine series converge very quickly, typically within 10^2 summation.

The second part illustrates our efforts to solve the PDE of both Black-Scholes Model and Heston Model. For simplicity, we solve both of them with explicit finite difference method. The BS-PDE is easy to solve and the pricing results fit with transform methods very well. However, the stability of Heston PDE proves to be a problem for us; though we can solve the Heston PDE, its accuracy cannot be refined because of the dilemma of stability and efficiency. Hence the pricing results under Heston PDE is not as accurate as the transform techniques. The problem can be solved by using implicit scheme but we failed to implement it.

The original project materials, including the questions of discretization the Heston Model and some conclusion's proofs, are contained in the third part. We successfully discretize the Heston SDE with predictor-corrector scheme and Andersen' QE scheme. The pricing result proves to be almost exactly what we have computed by FFT. We also explore different schemes such as JK scheme and TG scheme that are mentioned in Andersen (2006) to compare with the QE scheme. Finally, we discuss the choice of Ψ's value's influence towards the pricing result.

1. Integral Transform Technique:

1.1 Characteristic Function (CRF) of Heston Model

The SDE of Heston Model under risk-neutral measure is,

$$dS_t = rS_t dt + \sqrt{v_t} S_t dW_t$$

$$dv_t = b(\theta - v_t)dt + \sigma\sqrt{v_t} dB_t$$

$$corr(B_t, W_t) = \rho$$

Define $\Phi(\mathrm{u}) = \mathrm{E}[\mathrm{e}^{iu \log S_T}]$, which is the characteristic function of the logarithm of final asset price S_T. We have the conclusion that,

$$\Phi(\mathrm{u}) = \frac{\exp\left(iu \log S_0 + iurt + \frac{b\theta t(b - i\rho\sigma u)}{\sigma^2} + \left(\frac{-(u^2 + iu)v_0}{\gamma \coth\frac{\gamma t}{2} + b - i\rho\sigma u}\right)\right)}{\left(\cosh\frac{\gamma t}{2} + \frac{b - i\rho\sigma u}{\gamma}\sinh\frac{\gamma t}{2}\right)^{\frac{2b\theta}{\sigma^2}}}$$

Where $\gamma = \sqrt{\sigma^2(u + iu) + (b - i\rho\sigma u)^2}$

As our benchmark, we state the same conclusion for Black-Scholes Model here,
SDE: $dS_t = rS_t dt + \sigma S_t dW_t$
CRF of the logarithm of final price:

$$\Phi(\mathrm{u}) = \mathrm{e}^{i\left(\log S_0 + \left(r - \frac{\sigma^2}{2}\right)T\right) - \frac{\sigma^2 v^2}{2}T}$$

1.2 Fourier Transform Method

To price a path-independent European option with payoff $V(S_T)$, if the PDF of S_T or $s_T = \log S_T$ under risk-neutral measure is known, its price is straight forward.

$$V_t(S_t) = e^{-r(T-t)} \int_0^{+\infty} V(S_T) f_{S_T}(S_T) dS_T$$

Or for logarithm price,

$$V_t(S_t) = e^{-r(T-t)} \int_{-\infty}^{+\infty} V(e^{s_T}) f_{s_T}(S_T) ds_T$$

In most models, characteristic function of the logarithm price is the simplest to compute. To recover the PDF and hence price the European option, we need the Fourier transform.

That is the PDF $f(x)$,

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-iux}\phi(u)du$$

Where $\phi(u) = \mathrm{E}\left[\mathrm{e}^{iux}\right] = \int_{-\infty}^{+\infty} e^{iux} f(x)dx$. When u → ∞, e^{iux} oscillates frequently so the $\phi(u)$ vanishes when $|u|$ is large. Hence approximately,

$$\begin{aligned} f(x) &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-iux}\phi(u)du \\ &= \frac{1}{2\pi}\int_{-B}^{+B} e^{-iux}\phi(u)du \\ &= \frac{1}{2\pi}\sum_{k=0}^{N-1} e^{-iu_k x}\phi(u_k)\Delta u \end{aligned}$$

Where $u_k = -B + \frac{2B}{N} * k$ and $\Delta \mathrm{u} = \frac{2\mathrm{B}}{\mathrm{N}}$.

Therefore, we can compute the PDF of $\log S_T$ based on Heston Model or BS Model by the numerical integration. The algorithm complexity of the integration is $\mathrm{O}(\mathrm{N}^2)$ since we calculate O(N) points of $f(x)$ and each point is the summation of N items.

1.3 Fast Fourier Transform (FFT)

The process of numerical integration can be accelerated by FFT. Every Fourier transform question can be organized in the following form,

$$\omega_{\mathrm{k}} = \sum_{j=0}^{N-1} e^{-i\frac{2\pi}{N}j\cdot k}\, x_j$$

Where $\mathrm{N} = 2^{\mathrm{m}}$ for some integer m.

Set $\mathrm{s} = e^{-i\frac{2\pi}{N}}$, then we have,

$$\begin{pmatrix} \omega_0 \\ \omega_1 \\ \omega_2 \\ \dots \\ \dots \\ \omega_{N-1} \end{pmatrix} = \begin{pmatrix} s^0 & s^0 & s^0 & \dots & s^0 \\ s^0 & s^1 & s^2 & \dots & s^{(N-1)} \\ s^0 & s^2 & s^4 & \dots & s^{2(N-1)} \\ \dots & \dots & \dots & \dots & \dots \\ s^0 & s^{(N-1)} & s^{2(N-1)} & \dots & s^{(N-1)^2} \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ \dots \\ \dots \\ x_{N-1} \end{pmatrix}$$

We have, ∀k

$$\omega_{\mathrm{k}} = \sum_{j=0}^{N-1} s^{j\cdot k}\, x_j = \sum_{j=0}^{\frac{N}{2}-1} s^{j\cdot k}\, x_{2j} + \mathrm{s}^{\mathrm{k}} \sum_{j=0}^{\frac{N}{2}-1} \mathrm{s}^{j\cdot k} x_{2j+1}$$

Hence the problem is decomposed into compute $\sum_{j=0}^{\frac{N}{2}-1} s^{j\cdot k}\, x_{2j}$ and $\sum_{j=0}^{\frac{N}{2}-1} s^{j\cdot k} x_{2j+1}$, which are two discrete Fourier transform with magnitude $\frac{\mathrm{N}}{2}$. Also, additional plus calculation of step 2N-1 is required.

The total complexity of the former is $\mathrm{O(N)}$ and the total complexity of the later is $2(\mathrm{N}-1)\cdot \log_2 N$. Hence the entire algorithm complexity is reduced to $O(n\,log\,n)$.

1.4 Fourier-Cosine Method

The Fourier-Cosine Method developed by Fang and Oosterlee (2004), utilized the Fourier cosine transform to express the option's price into a series of coefficients. This method reduces the algorithm complexity into $O(n)$.

The value of the option at time t with the logarithm of asset price $x = \log S_t$, is $v(t,x)$. Then,

$$v(t,x) = \int_{-\infty}^{+\infty} v(T,s) f(s|x)\, ds$$

$f(s|x)$ is the PDF of final logarithm distribution. For some appropriate [a, b],

$$v(t,x) \approx \int_a^b v(T,s) f(s|x)\, ds$$
$$= \int_a^b v(T,s) \left(\sum_{k=0}^{\infty} A_k \cos\left(k \frac{y-a}{b-a}\pi\right)\right) ds$$
$$= \sum_{k=0}^{\infty} A_k \int_a^b v(T,s) \cos\left(k \frac{y-a}{b-a}\pi\right) ds$$

$\sum_{k=0}^{\infty} A_k \cos\left(k \frac{y-a}{b-a}\pi\right)$ is the Fourier-Cosine expansion of $f(s|x)$. Hence,

$$A_k = Re\left(\phi\left(\frac{k\pi}{b-a}; x\right) e^{-\frac{ik\pi a}{b-a}}\right)$$

where $\phi(u;x)$ is the characteristic function of $\log(\frac{S_T}{S_t})$ with initial $x = \log S_t$. If we know the characteristic function of $\log(S_T)$, $\phi(u;x)$ will be easy to derive.

Define $V_k = \frac{2}{b-a}\int_a^b v(T,s) \cos\left(k \frac{y-a}{b-a}\pi\right) ds$, then

$$v(t,x) \approx \sum_{k=0}^{N-1} A_k \int_a^b v(T,s) \cos\left(k \frac{y-a}{b-a}\pi\right) ds$$

$$= \frac{b-a}{2}\sum_{k=0}^{N-1} A_k V_k$$

For a European Call option, the V_k term is

$$V_k = \frac{2}{b-a}\int_a^b v(T,s)\cos\left(k\pi\frac{s-a}{b-a}\right)ds$$
$$= \frac{2}{b-a}\int_a^b [\alpha K(e^s-1)]^+\cos\left(k\pi\frac{s-a}{b-a}\right)ds$$

As we will see later, typically $a < 0$, hence,

$$V_k = \frac{2}{b-a}\int_0^b \alpha K(e^s-1)\cdot\cos\left(k\pi\frac{s-a}{b-a}\right)ds$$

So far, we have formulas for V_k and A_k. As long as we choose a truncated interval and truncated series number N, we can calculate the approximating price of the option.

Fang and Oosterlee (2004) proposed an empirical way to select the interval [a, b],

$$\mathrm{a} = \mathrm{c}_1 - L\sqrt{c_2+\sqrt{c_4}}$$
$$\mathrm{b} = \mathrm{c}_1 + L\sqrt{c_2+\sqrt{c_4}}$$

$\mathrm{c_n}$ is the nth cumulants of $\mathrm{x} = \log(\mathrm{S_T})$; $\mathrm{L} = 10$.

For an appropriate truncated series number N, we set the judging condition that $|\frac{b-a}{2}A_kV_k| < 10^{-8}$ to let the summation program stop by itself.

For BS Model:

$$\mathrm{c}_1 = rT; c_2 = \sigma^2 T; \mathrm{c}_4 = 0$$

For Heston Model:

$$\mathrm{c}_1 = rT + \left(1-e^{-bT}\right)\frac{\theta - v_0}{2b} - \frac{1}{2}\theta T$$

$$\mathrm{c}_2 = \frac{1}{8b^3}\begin{pmatrix} \sigma T b e^{-bT}(v_0-\theta)(8b\rho - 4\sigma) + b\rho\sigma\left(1-e^{-bT}\right)(16\theta - 8v_0) \\ +2\theta b T(\sigma^2 + 4b^2 - 4b\rho\sigma) \\ +\sigma^2\left(2v_0 + (\theta - 2v_0)e^{-2bT} + \theta\left(6e^{-bT}-7\right)\right) \\ +8b^2(v_0-\theta)(1-e^{-bT}) \end{pmatrix}$$

c_4 is complicated so we set it to zero.

1.5 Results

Fourier Transform Methods

Parameters Set:

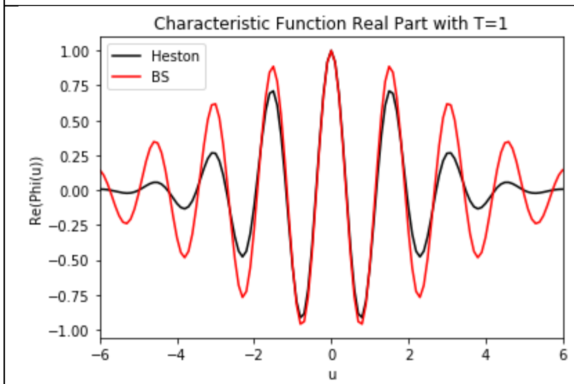
$S_0 = 61.9; r = 0.03;$

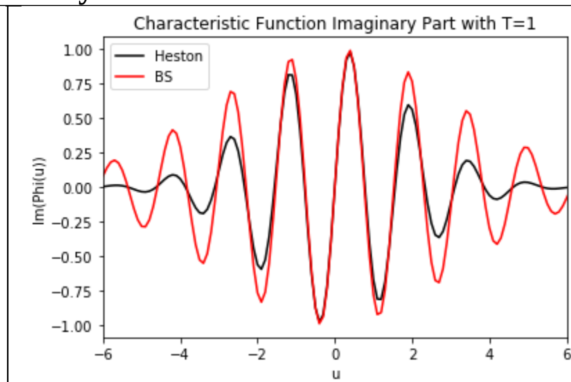
Heston: $b = 0.9;\ \theta = 0.1;\ \rho = -0.5; V_0 = 0.36; \sigma_{V_t} = 0.1;$

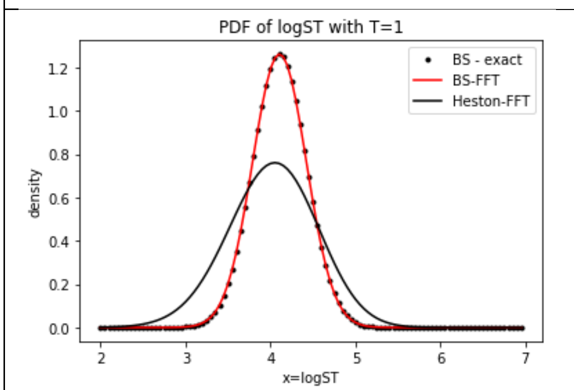
BS: $\sigma_{BS} = \theta = 0.1.$

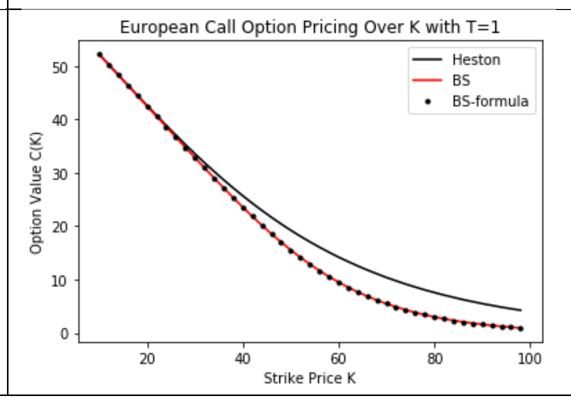
Note: We set the BS Model's volatility equaling the long-term Heston volatility θ, Which means the Heston has larger volatility than BS at first and the effect diminishes as time passes.

Time to Maturity T=1

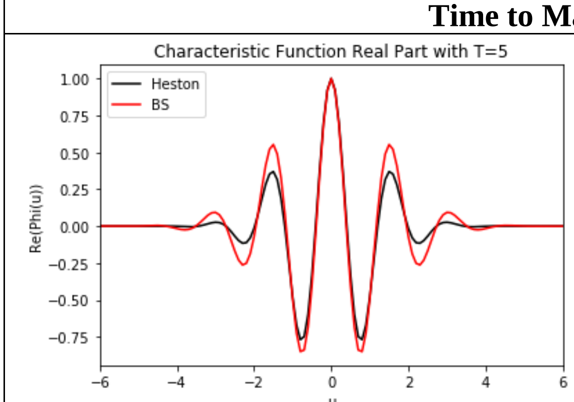


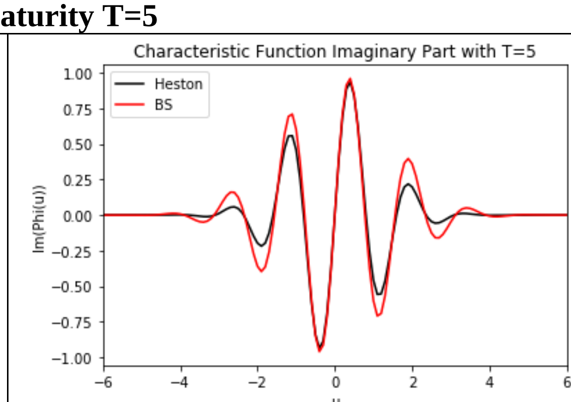


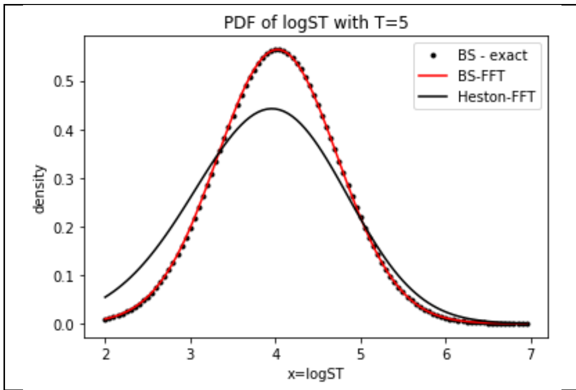


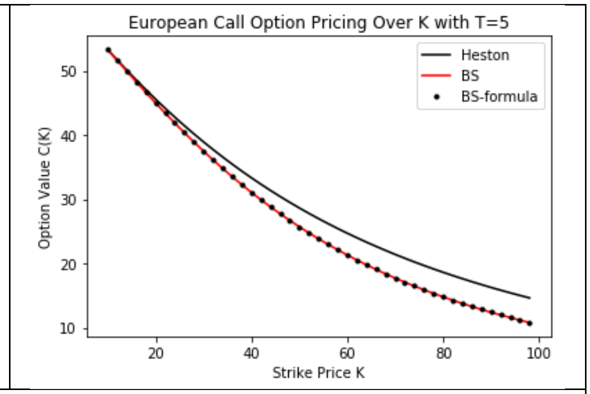


Time to Maturity T=5

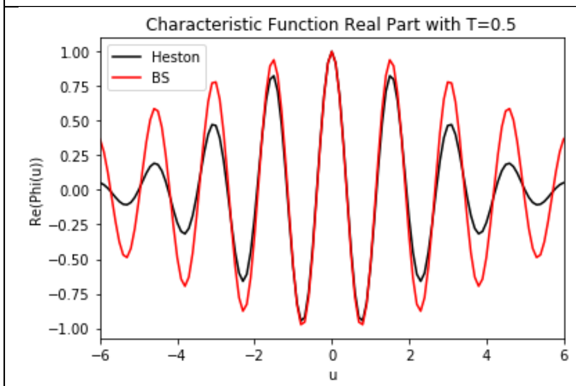


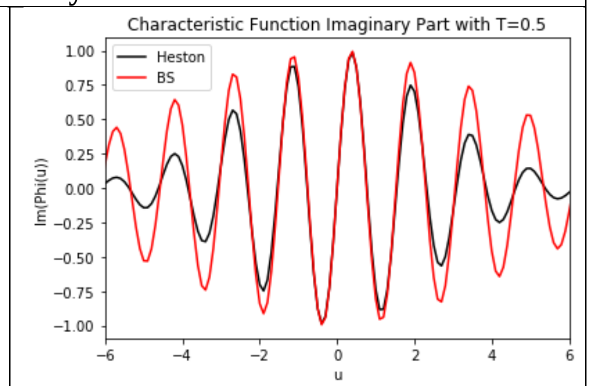


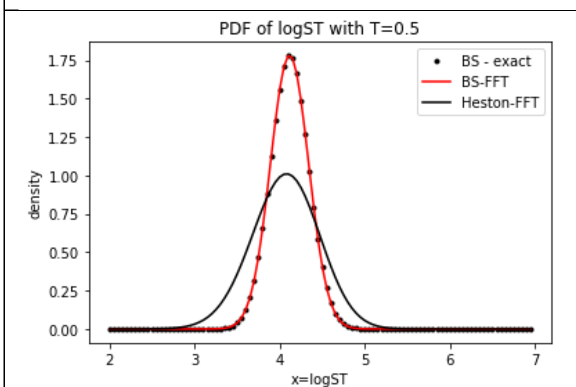


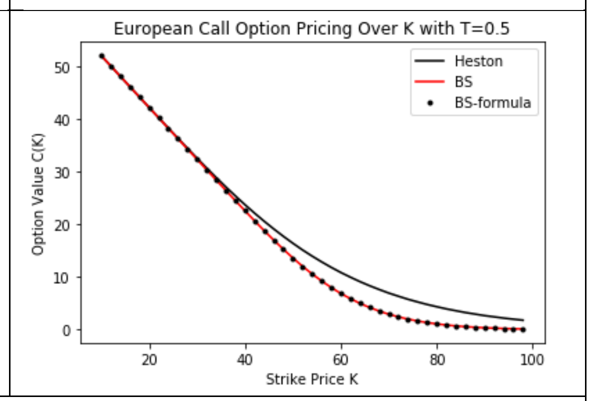


Time to Maturity T=0.5









Comments:

(a) Dot points are analytic results. The benchmark of BS models fits very well with its analytic formulas, so I am also very confident about the correctness and accuracy of pricing options under Heston-dynamics by FFT.

(b) As I anticipated, the option price under BS-dynamics is cheaper than that under Heston-dynamics because the initial volatility in Heston Model is larger than its long-term average. Since I set the long-term average as the volatility in BS Model, I anticipate the volatility in Heston to be larger averagely than in the BS model. Larger volatility means larger value.

(c) The effect of larger initial volatility will diminish in the long-term. As we can see in the PDF at T=0.5, the $\log S_T$ in Heston has wider distribution; however, at T=5, the wider distribution is no longer obvious.

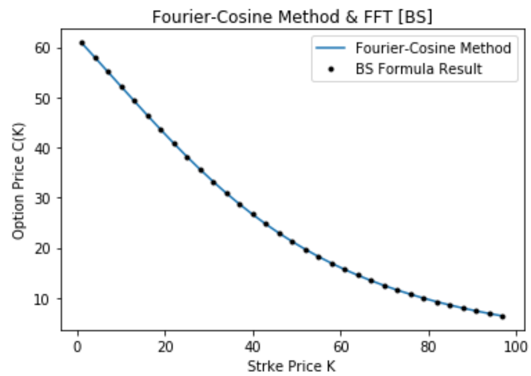
Fourier-Cosine Method

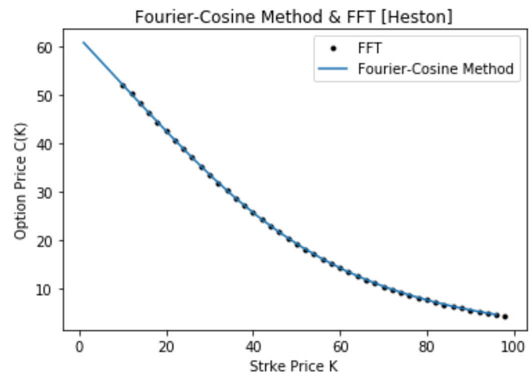
Parameters Set:

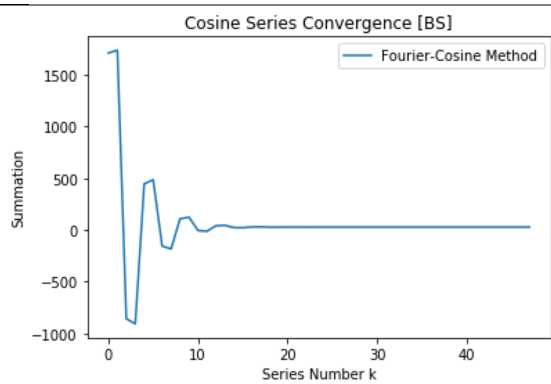
$S_0 = 61.9; r = 0.03;$

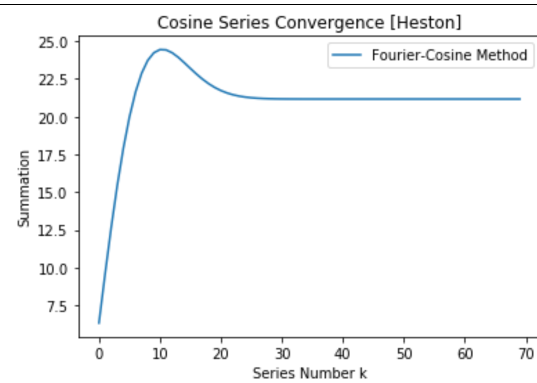
Heston: $b = 0.9;\ \theta = 0.1;\ \rho = -0.5; V_0 = 0.36; \sigma_{V_t} = 0.1;$

BS: $\sigma_{BS} = 0.6.$









Comments:

(a) For BS model, we use the theoretical price to compare with the Fourier Cosine Method results; For Heston model, we use the Fourier Transform results to compare with the Fourier Cosine Method results. The results fit very well so that accuracy of Fourier Cosine Method is guaranteed.

(b) The convergence pictures show that the cosine series converge very quickly. Within 10^2 terms, the series declines below 10^{-8} and the summation iteration break.

2. Numerical Partial Differential Equation

2.1 The PDE method

The drawback of the integral transform methods is obvious—all of they rely on a specific probability density function or characteristic function. But not all models have the closed-formed CRF or PDF as Heston or BS do.

In this case, if we can deduce the option value's PDE from the risk-neutral SDE, we can solve the PDE numerically so that we can give all the options' prices of fixed strike price K with different time to maturity T and current stock price S_t by one effort.

The BS PDE:

$$\frac{\partial v(\tau,x)}{\partial \tau}=\frac{\sigma^2x^2}{2}\frac{\partial^2 v(\tau,x)}{\partial x^2}+rx\frac{\partial v(\tau,x)}{\partial x}-rv(\tau,x)$$

Initial condition: $v(0,x)=(x-K)^+$
Boundary condition: $v(\tau,0)=0$; $\lim_{x\to\infty} v(\tau,x)=x-K$;

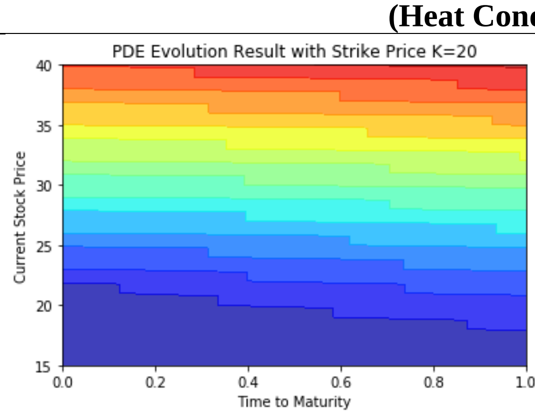
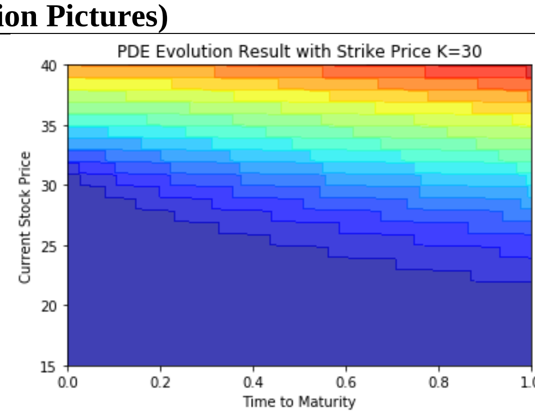
The Heston PDE:

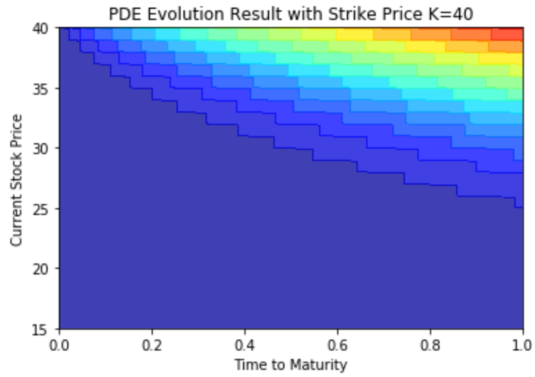
$$\frac{\partial U(\tau,x,v)}{\partial \tau}=\frac{\sigma^2x^2}{2}\frac{\partial^2 U}{\partial x^2}+rx\frac{\partial U}{\partial x}-rU+\rho\sigma v\frac{\partial^2 U}{\partial x\partial v}+\frac{1}{2}\sigma^2 v\frac{\partial^2 U}{\partial v^2}+b(\theta-v)\frac{\partial U}{\partial v}$$

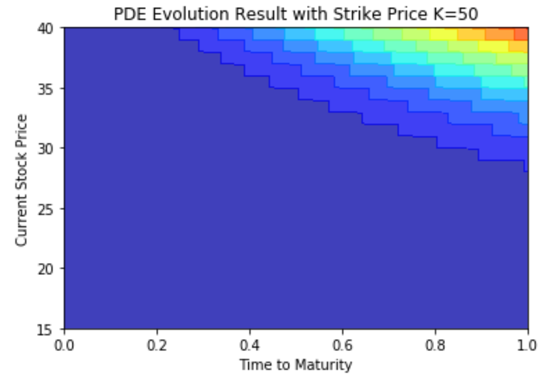
Initial condition: $U(0,x,v)=(x-K)^+$
Boundary condition:
$U(\tau,0,v)=0$; $\lim_{x\to\infty} U(\tau,x,v)=x-K$; $\lim_{v\to\infty} U(\tau,x,v)=x$;

$$\mathrm{r}\frac{\partial U(\tau,x,0)}{\partial x}+\mathrm{b}\theta\frac{\partial U(\tau,x,0)}{\partial v}-\mathrm{r}U(\tau,x,0)=\frac{\partial U(\tau,x,0)}{\partial \tau}$$

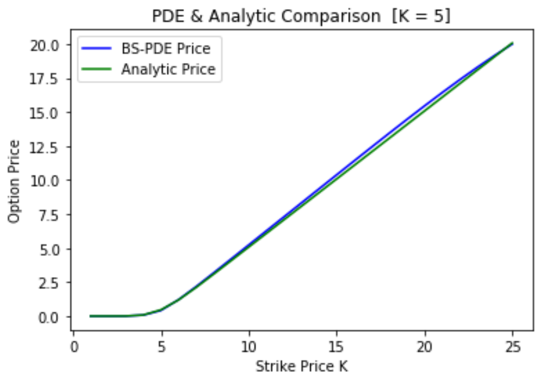
2.2 Results

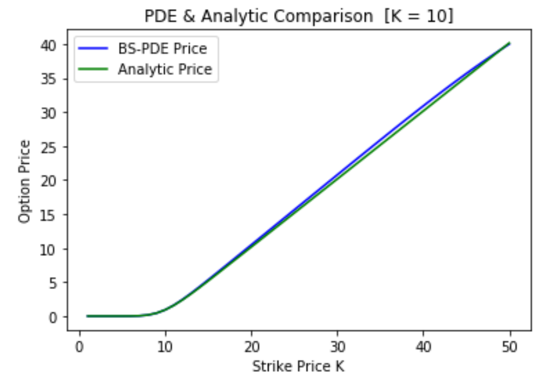
European Call Option Black-Scholes PDE	
Parameter Set: $\mathrm{K}=20;\tau_{max}=1;\ \mathrm{r}=0.03;\sigma=0.3;\mathrm{d}\tau=0.0001;\ \mathrm{dS}=0.5;\mathrm{S}_{max}=5\cdot K$	
The Evolution Results for Different Strike K **(Heat Conduction Pictures)**	
	

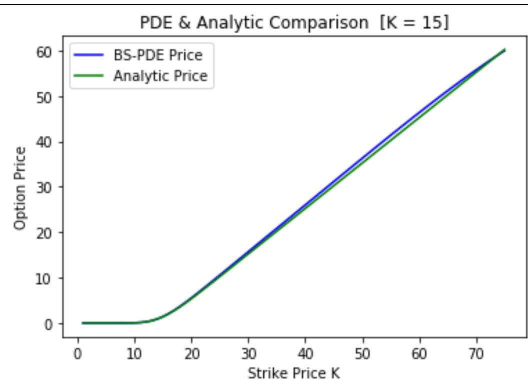


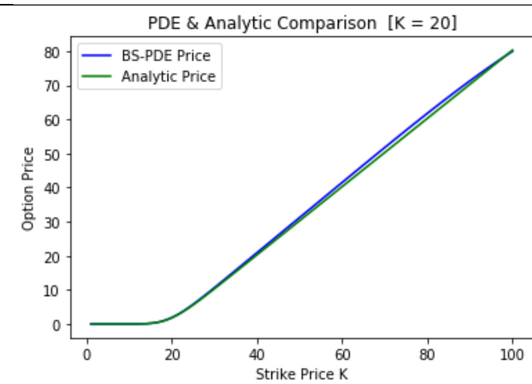


Comparison of Analytic Results with PDE Results









Comments:

(a) The numerical solution for Black-Scholes PDE fit very well the the analytic results. Hence the accuracy of the solution has been justified. Since the evolution of the PDE can give all the option prices with different initial S_t and time to maturity at one single effort, it is more convenient than the integral transform methods.

(b) The heat conduction picture of the option value is very reasonable—the longer the time to maturity as well as the larger initial stock price indicate a higher option price.

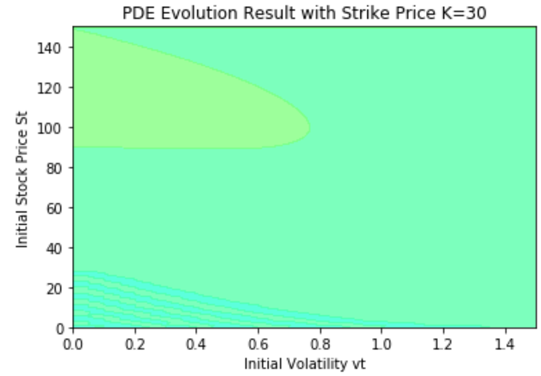
European Call Option Heston PDE

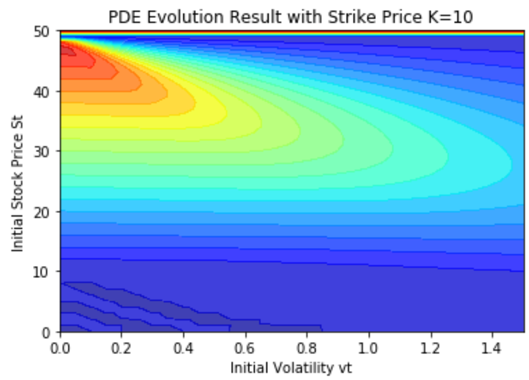
Parameters Set:

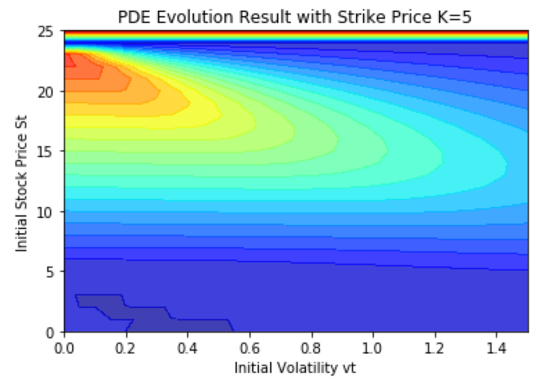
$S_0 = 61.9; r = 0.03;$ T=0.1

Heston: $b = 0.9;\ \theta = 0.1;\ \rho = -0.5; V_0 = 0.36; \sigma_{V_t} = 0.1;$

The Evolution Results for Different Strike K at Fixed T=0.1
(Initial S_t & Initial v_t)







Comments:

Although we can give a evolution picture here, the option pricing accuracy is dubious here. By checking some of the initial points (S_t, v_t) with fixed T, we found that the option's value typically have 1%-3% error compared with the transform techniques.

This is not a small error since sometimes even we adopted a BS model will give a more accurate result than solving Heston PDE. The cause is that we adopted explicit finite difference hence the stability of the PDE is questionable. We are forced to set very small dt and relatively large dv to compute the result. The large dv results in the inaccuracy here.

The problem can be solved by first using transform of the coordinate and adopting an implict scheme.

3. Monte Carlo Simulation

3.1 Predictor-Corrector Scheme for Price Process

(1) To prove (1.3), apply ito lemma to $e^{-rt}S_t$.

$$
\begin{aligned}
&d\log(e^{-rt}S_t)\\
&= d(-rt + \log(S_t)) = -rdt + \frac{1}{S_t}dS_t - \frac{1}{2S_t^2}dS_t dS_t\\
&= \sqrt{V_t}\left(\rho dB_t + \sqrt{1-\rho^2}dW_t\right) - \frac{1}{2}V_t
\end{aligned}
$$

(2) To prove (1.4), integrate (1.2).

$$
\begin{aligned}
&\sigma\int_t^{t+\Delta t}\sqrt{V_s}\,dB_s = V_{t+\Delta t} - V_t - b\theta t + \int_t^{t+\Delta t} bV_s\,ds\\
&\int_t^{t+\Delta t}\sqrt{V_s}\,dB_s = \frac{1}{\sigma}\left(V_{t+\Delta t} - V_t - b\theta t + \int_t^{t+\Delta t} bV_s\,ds\right)
\end{aligned}
$$

(3) To prove (1.5), integrate (1.3) and use (1.4).

$$
\begin{aligned}
&\log\left(e^{-r(t+\Delta t)}S_{t+\Delta t}\right) - \log(e^{-rt}S_t)\\
&= -\frac{1}{2}(V_{t+\Delta t} - V_t) + \rho\int_t^{t+\Delta t}\sqrt{V_s}\,ds + \sqrt{1-\rho^2}\int_t^{t+\Delta t}\sqrt{V_s}dW_s\\
&= -\frac{1}{2}(V_{t+\Delta t} - V_t) + \sqrt{1-\rho^2}\int_t^{t+\Delta t}\sqrt{V_s}dW_s + \frac{\rho}{\sigma}\left(V_{t+\Delta t} - V_t - b\theta\Delta t + b\int_t^{t+\Delta t}V_s ds\right)\\
&= \left(\frac{\rho b}{\sigma} - \frac{1}{2}\right)\int_t^{t+\Delta t}V_s ds + \frac{\rho}{\sigma}(V_{t+\Delta t} - V_t - b\theta\Delta t) + \sqrt{1-\rho^2}\int_t^{t+\Delta t}\sqrt{V_s}dW_s
\end{aligned}
$$

Hence,

$$
\log\left(\frac{e^{-r(t+\Delta t)}S_{t+\Delta t}}{e^{-rt}S_t}\right) = \left(\frac{\rho b}{\sigma} - \frac{1}{2}\right)\int_t^{t+\Delta t}V_s ds + \frac{\rho}{\sigma}(V_{t+\Delta t} - V_t - b\theta\Delta t) + \sqrt{1-\rho^2}\int_t^{t+\Delta t}\sqrt{V_s}dW_s
$$

(4) The stochastic integral in the (1.5) is a central Gaussian distribution.

Recall that if the integral part is deterministic, it will be a Paley-Wiener integral so that it is a zero-mean gaussian variable. Here though V_t is not a deterministic function, V_t is purely driven by the Brownian motion B_t which is independent of the Brownian motion W_t.

Conditioned one specific trajectory of V_t, for example $V^0(t)$, because V_t and W_t are independent, knowing the trajectory of $V^0(t)$ will not influence the W_t. Therefore,

$$(X|V^0(t)) = \left(\int_t^{t+\Delta t} \sqrt{V_s} dW_s \,\middle|\, V^0(t)\right)$$

Is a Paley-Wiener integral hence a Gaussian variable with zero-mean and $\int_t^{t+\Delta t} V_s ds$ variance.

(5) Predictor-Corrector Scheme.

$$\int_t^{t+\Delta t} V_s ds \approx \frac{\Delta t}{2}(V_{t+\Delta t} + V_t)$$

Using the approximation in (1.5),

$$\begin{aligned}
&\log\left(e^{-r(t+\Delta t)} S_{t+\Delta t}\right) \\
&= \log(e^{-rt}S_t) + \left(\frac{\rho b}{\sigma} - \frac{1}{2}\right)\int_t^{t+\Delta t} V_s ds + \frac{\rho}{\sigma}(V_{t+\Delta t} - V_t - b\theta\Delta t) + \sqrt{1-\rho^2}\int_t^{t+\Delta t}\sqrt{V_s}dW_s \\
&= \log(e^{-rt}S_t) + K_0 + K_1 V_t + K_2 V_{t+\Delta t} + \sqrt{K_3(V_{t+\Delta t} + V_t)}N
\end{aligned}$$

The last term is because $\int_t^{t+\Delta t}\sqrt{V_s}dW_s$ is a Gaussian random variable with variance,

$$\int_t^{t+\Delta t} V_s ds \approx \frac{\Delta t}{2}(V_{t+\Delta t} + V_t)$$

3.2 Modified Euler Scheme for Volatility Process

It is impossible to implement the Euler scheme because V_{t+1} has a positive probability to negative. However, it will be used into the next iteration under the square root term.

Instead, using

$$V_{t+\Delta t} = V_t + b(\theta - V_t^+)\Delta t + \sigma\sqrt{\Delta t V_t^+}N$$

Will solve the issue because it forces the V_t^+ to be non-negative. Correspondingly, I believe (1.6) should also be modified as,

$$\log\left(e^{-r(t+\Delta t)} S_{t+\Delta t}\right) = \log(e^{-rt}S_t) + K_0 + K_1 V_t + K_2 V_{t+\Delta t} + \sqrt{K_3(V_{t+\Delta t} + V_t)^+}N$$

With an additional plus notation in the square root term.

3.3 Anderson's Scheme for Volatility Process

(1) Differentiate $e^{bt}V_t$

$$d\left(e^{bt}V_t\right) = be^{bt}V_t dt + e^{bt}dV_t = e^{bt}b\theta dt + \sigma e^{bt}\sqrt{V_t}dB_t$$

Integrate the RHS and LHS over $[t, t+\Delta t]$,

$$e^{b(t+\Delta t)}V_{t+\Delta t} - e^{bt}V_t = b\theta \int_t^{t+\Delta t} e^{bs} ds + \sigma \int_t^{t+\Delta t} e^{bs}\sqrt{V_s} dB_s$$
$$V_{t+\Delta t} - e^{-b\Delta t}V_t = \theta(1 - e^{-b\Delta t}) + \sigma e^{-b\Delta t} \int_t^{t+\Delta t} e^{bs}\sqrt{V_s} dB_s$$
$$V_{t+\Delta t} = e^{-b\Delta t}V_t + \theta(1 - e^{-b\Delta t}) + \sigma e^{-b\Delta t} \int_t^{t+\Delta t} e^{bs}\sqrt{V_s} dB_s$$

As we have argued in 1.1.4, $\int_t^{t+\Delta t} \sqrt{V_s} dB_s$, though not a Paley-Wiener integral, is a zero-mean Gaussian variable. Take expectation both sides,

$$E[V_{t+\Delta t}|F_t] = e^{-b\Delta t}V_t + \theta(1 - e^{-b\Delta t})$$
$$Var[V_{t+\Delta t}|F_t] = \mathrm{E}\left[\sigma e^{-b\Delta t} \int_t^{t+\Delta t} e^{bs}\sqrt{V_s} dB_s\right]^2 = \sigma^2 e^{-2b\Delta t} E\left[\int_t^{t+\Delta t} e^{2bs} V_s ds \,\middle|\, F_t\right]$$

Since,

$$E\left[\int_t^{t+\Delta t} V_s ds \,\middle|\, F_t\right]$$
$$= \int_t^{t+\Delta t} e^{2bs}\left(e^{-bs}V_t + \theta(1 - e^{-bs})\right) ds$$
$$= \frac{\mathrm{V_t}}{b}\left(e^{b(t+\Delta t)} - e^{bt}\right) + \frac{\theta}{b}\left(\frac{1}{2}\left(e^{2b(t+\Delta t)} - e^{2bt}\right) - \left(e^{b(t+\Delta t)} - e^{bt}\right)\right)$$

We have,

$$Var[V_{t+\Delta t}|F_t]$$
$$= \sigma^2 e^{-2b\Delta t} E\left[\int_t^{t+\Delta t} e^{2bs} V_s ds \,\middle|\, F_t\right]$$
$$= \sigma^2 e^{-2b\Delta t}\left(\frac{\mathrm{V_t}}{b}(e^{b\Delta t} - 1) + \frac{\theta}{b}\left(\frac{1}{2}(e^{2b\Delta t} - 1) - (e^{b\Delta t} - 1)\right)\right)$$
$$= \sigma^2 e^{-b\Delta t}\frac{\mathrm{V_t}}{b}(1 - e^{-b\Delta t}) + \sigma^2\frac{\theta}{b}\left(\frac{1}{2}(1 - e^{-2b\Delta t}) - e^{-b\Delta t}(1 - e^{-b\Delta t})\right)$$
$$= \frac{V_t\sigma^2 e^{-b\Delta t}}{b}(1 - e^{-b\Delta t}) + \frac{\theta\sigma^2}{2b}(1 - e^{-b\Delta t})^2$$

(2) To simulate $V_{t+\Delta t}$ using a standard normal variable Z, we need to match its mean and variance with conditioned $V_{t+\Delta t}$. That is,

$$V_{t+\Delta t} \approx a(c+Z)^2$$

$$E[a(c+Z)^2] = ac^2 + a$$

$$E[(a(c+Z)^2)^2] = a^2(c^4 + 6c^2 + 3)$$

$$Var[a(c+Z)^2] = 2a^2 + 4a^2c^2$$

Set,

$$E[a(c+Z)^2] = E[V_{t+\Delta t}|V_t]$$

$$Var[a(c+Z)^2] = Var[V_{t+\Delta t}|V_t]$$

We will have,

$$E[V_{t+\Delta t}|V_t] = a(c^2+1) \rightarrow a = \frac{E[V_{t+\Delta t}|V_t]}{c^2+1}$$

$$Var[V_{t+\Delta t}|V_t] = 2a^2(1+2c^2) = \frac{2E[V_{t+\Delta t}|V_t]^2}{(1+c^2)^2}(1+2c^2)$$

$$\Psi(1+c^2)^2 = 2(1+2c^2)$$

Where $\Psi = \frac{Var[V_{t+\Delta t}|V_t]}{E[V_{t+\Delta t}|V_t]^2}$. If $\Psi < 2$,

$$c^2 = (2\Psi^{-1}-1) + \sqrt{2\Psi^{-1}(2\Psi^{-1}-1)}$$

Since $c^2 \geq 0$.

(3) When the variance is close to zero, the central parameter of the non-centered χ^2 distribution can not be approximated well by a normal distribution variable.

Define,

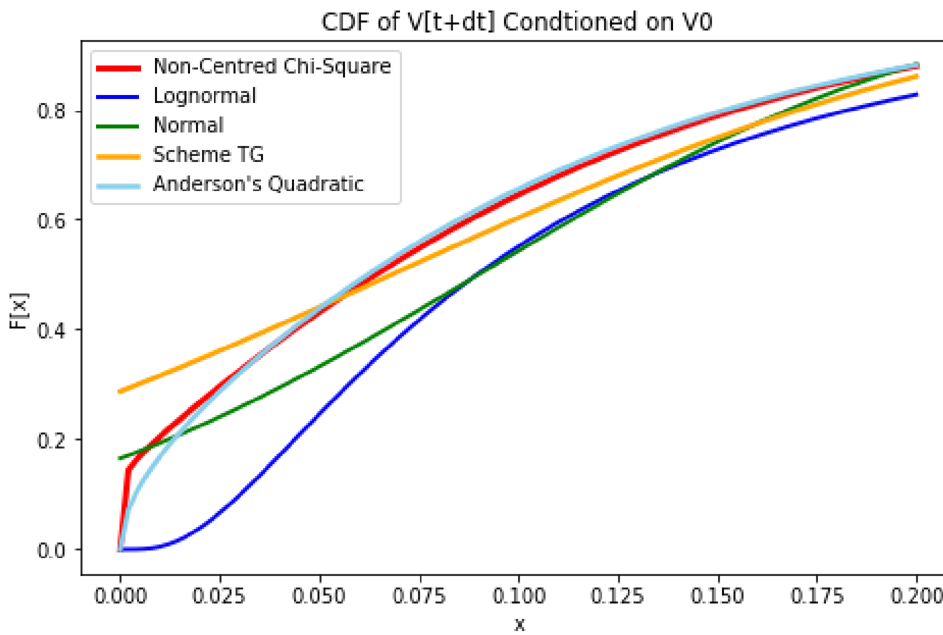
$$d = \frac{4b\theta}{\sigma^2}$$

$$n(t, t+\Delta t) = \frac{4be^{-b\Delta t}}{\sigma^2(1-e^{-b\Delta t})}$$

Then $[V_{t+\Delta t}|V_t] \times \frac{n(t,t+\Delta t)}{e^{-\kappa(T-t)}}$ is distributed as a non-central χ^2distribution with d degree of freedom and $V(t) \cdot n(t, t+\Delta t)$ non-centrality parameter. For small V_t, the non-centrality parameter approaches zero, and the distribution of $V_{t+\Delta t}$ became proportional to that of an ordinary χ^2distribution with d degrees of freedom, whose density is,

$$f_{\chi^2}(x; d) = \frac{e^{-\frac{x}{2}} x^{\frac{d}{2}-1}}{2^{\frac{d}{2}}\,\Gamma\left(\frac{v}{2}\right)}$$

The $x^{\frac{d}{2}-1}$ term indicates the density will be large around 0, who decays slower than a regular Gaussian approximation. We can see it intuitively from the picture below,



The picture shows the CDF of different distributions of approximating the exact non-centered χ^2distribution (the red line). All of the distributions have been matched for their first and second moments. We can see that though the Andersen's quadratic approach simulate the exact distribution very well, its CDF decay to zero when x → 0. Hence, we need a modified scheme to deal with case.

With the combination of Dirac $\delta(x)$ and PDF of an exponential distribution, we have,

$$E[V_{t+\Delta t}|V_t] = \frac{1-p}{\beta} \rightarrow \beta = \frac{1-p}{E[V_{t+\Delta t}|V_t]}$$
$$E\left[V_{t+\Delta t}^2|V_t\right] = \frac{2(1-p)}{\beta^2}$$

$$Var[V_{t+\Delta t}|V_t] = \frac{2(1-p)}{\beta^2} - \left(\frac{1-p}{\beta}\right)^2$$
$$\therefore \frac{Var[V_{t+\Delta t}|V_t]}{E[V_{t+\Delta t}|V_t]^2}(1-p)^2 = 2(1-p) - (1-p)^2$$
$$\therefore \Psi = \frac{1+p}{1-p}$$
$$\therefore p = \frac{\Psi - 1}{\Psi + 1} \ (if\ \Psi > 1)$$

The follow inverse function method gives the correct simulation of the variable $[V_{t+\Delta t}|V_t]$,

$$F^{-1}(u) = 1_{\{p<u\leq 1\}}\beta^{-1}\log(\frac{1-p}{1-u})$$

The meaning of the expression $p\delta_0(x) + (1-p)\beta e^{-\beta x}1_{\{x>0\}}$ is with probability p, $[V_{t+\Delta t}|V_t]$ equals 0, with probability $(1-p)$, $[V_{t+\Delta t}|V_t]$ follow an exponential distribution with mean $\frac{1}{\beta}$. Observing the inverse formula $F^{-1}(u)$, conditioning on $\{u > p\}$ or $\{u < p\}$,

If $\{u < p\}$: $F^{-1}(u) = 0$ with probability $\Pr(u < p) = p$.
If $\{u > p\}$: $F^{-1}(u) = \beta^{-1}\log(\frac{1-p}{1-u})$, since u is a uniform distribution and $u > p$,the distribution of $\frac{1-p}{1-u}$ is identical to the distribution of $\frac{1}{1-v}$, where $v \sim uniform(0,1)$. The expression of $\beta^{-1}\log(\frac{1}{1-v})$ is exactly the inverse method of an exponential random variable with mean $\frac{1}{\beta}$.

As a conclusion, $F^{-1}(u) = 1_{\{p<u\le 1\}}\beta^{-1}\log(\frac{1-p}{1-u})$ is the inversed method of generating a random variable with the combined density $p\delta_0(x) + (1-p)\beta e^{-\beta x}1_{\{x>0\}}$.

(4) The simulations and stability analysis will be presented in the following results section

3.4 Other Discretization Scheme Explored

Beyond the discretization schemes in this project, we also explored other discretization scheme mentioned in Anderson (2006).

First, the KJ Scheme in Kahl, C. and P. Jackel (2005), using the implicit Milstein scheme and coupled with their IJK discretization of stock price process, which dictates:

$$\log\left(\frac{S_{t+\Delta t}}{S_t}\right) = -\frac{\Delta t}{4}(V_{t+\Delta t} + V_t) + \rho\sqrt{V_t \Delta t}Z_V + \frac{1}{2}\left(\sqrt{V_{t+\Delta t}} + \sqrt{V_t}\right)\left(\sqrt{\Delta t}Z_S - \rho\sqrt{\Delta t}Z_V\right) + \frac{1}{4}\sigma\rho\Delta t(Z_V^2 - 1)$$

$$V_{t+\Delta t} = \frac{V_t + b\theta\Delta t + \sigma\sqrt{V_t}\sqrt{\Delta t}Z_V + \frac{1}{4}\sigma^2\Delta t(Z_V^2 - 1)}{1 + b\Delta t}$$

Where Z_v and Z_x are two independent standard normal variables. However, truncating a positive V_t^+ is also required in this scheme, similar to the modified Euler Scheme in the project.

Second, Scheme TG in Anderson (2006). Similar to approximating the non-centered χ^2 distribution with a random variable $\sigma(\mu + Z)^2$ and matching their first and second cumulants, the random variable in Scheme TG is,

$$(\mu + \sigma \cdot Z)^+$$

Which is a truncated Gaussian variable. The solution for μ and σ is,

$$\mu = \frac{E[V_{t+\Delta t}|V_t]}{\frac{N'(r(\Psi))}{r(\Psi)} + N(r(\Psi))}$$

$$\sigma = \frac{E[V_{t+\Delta t}|V_t]}{N'\big(r(\Psi)\big) + r(\Psi)N\big(r(\Psi)\big)}$$

Where N(x) is the CDF of standard normal distribution, and r(x) fits the following equation,

$$\mathrm{r(x)N'(r(x)) + N(r(x))(1 + r(x)^2)} = (1 + x)\Big(N'\big(r(x)\big) + r(x)N\big(r(x)\big)\Big)^2$$

3.5 Results

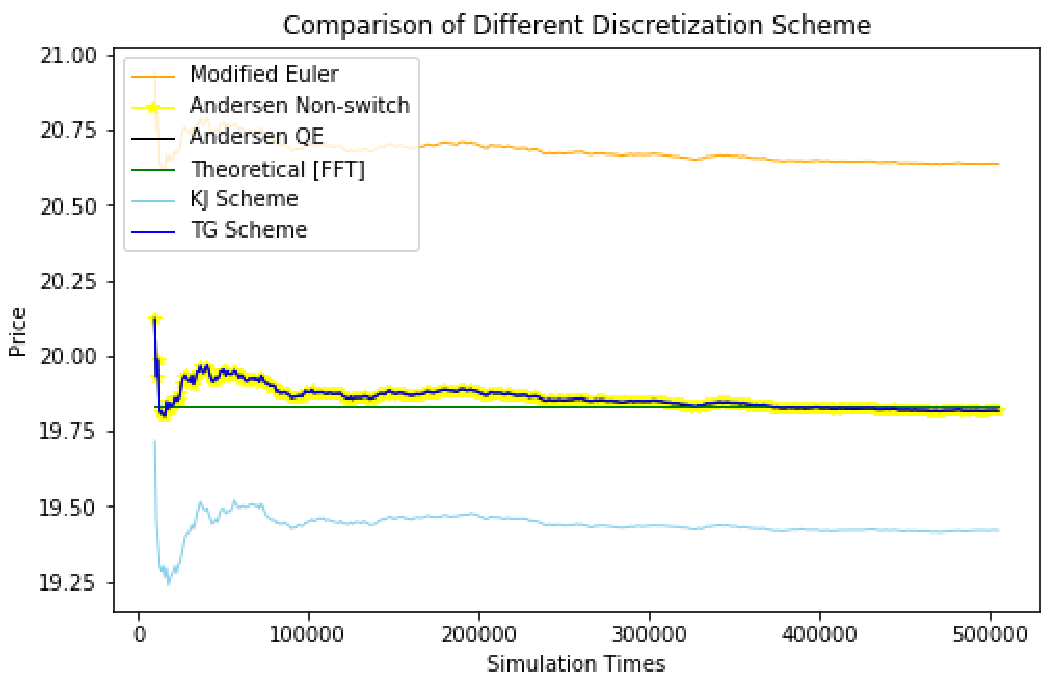
Comparison of Different Discretization Scheme

Original Parameter Set:

$S_0 = 61.9; r = 0.03; K=50;$

$\theta = 0.1; b = 0.9; \sigma = 0.1; T = 1; dt = 1/24.$

Convergence Picture:



Numerical Results:

Method	Option Price
Exact (FFT)	**19.833**
Modified Euler	**20.648**
Andersen's Quadratic	**19.826**
Andersen's QE (Not Active)	**19.826**
KJ Scheme	**19.416**
TG Scheme	**19.829**

Comment:

based on the original parameter set, the QE scheme never switch. The condition $\Psi < 1.5$ always holds so that the variable is always approximated by a $\sigma(\mu + Z)^2$ variable; the combination of Dirac density and exponential distribution is never activated.

The result is not surprising because we have the condition $2b\theta > \sigma^2$. According to Anderson (2006), the condition indicates that $V_{t+\Delta t}$ can never reach zero. If the Dirac and exponential scheme should be activated, the V_t process has a positive probability to reach zero. Hence, the no-switching phenomenon here is reasonable. We will explore the switching scheme by changing parameters in the following section.

All the results are compared with a theoretical option price calculated by Fourier Transformation. We believe the Fourier Transformation result is accurate enough to be a benchmark for comparing simulations. From the picture, we have the conclusion that, the QE scheme and TG scheme are obviously better than the JK or Modified Euler.

Ψ's Stability Analysis

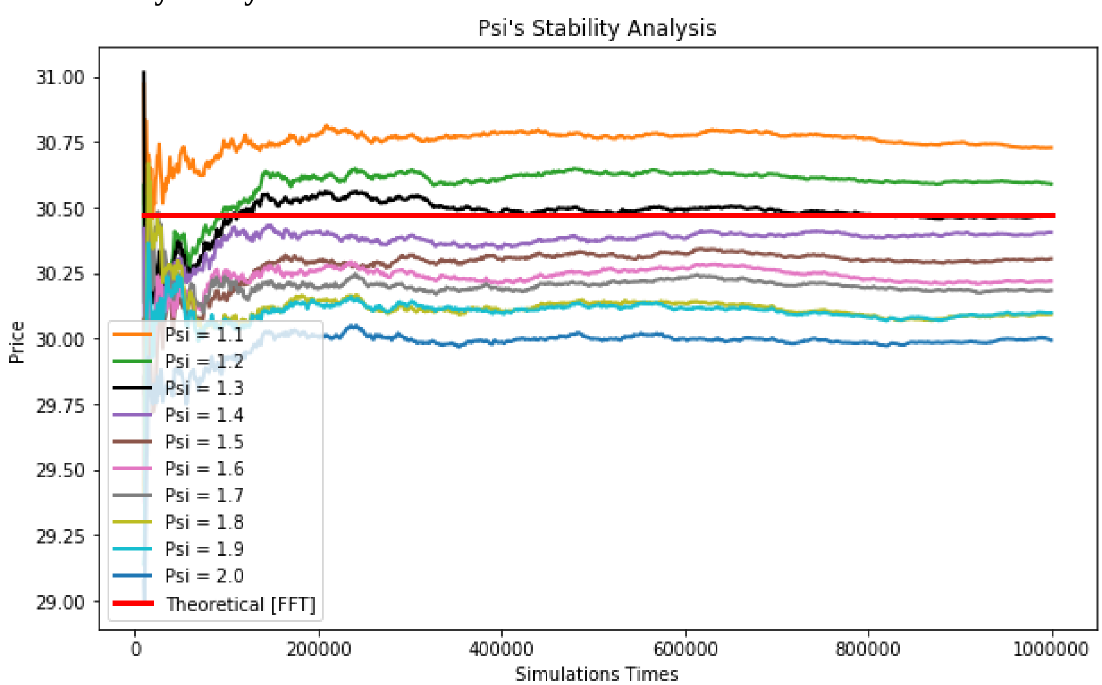
Changed Parameter Set:

$S_0 = 61.9$; $r = 0.03$; K=50;

$V0 = 0.36$, $r = 0.03$, $\sigma = 0.4$, $b = 0.2$, $\rho = -0.5$, $\theta = 0.1$, T $= 4$

Note that: This set of parameters satisfies $2b\theta < \sigma^2$; therefore, we anticipate switch here.

Ψ's Stability Analysis Picture:



Comment:

From the picture, the choice of Ψ indeed influence our final simulation value slightly. Since we know the exact option value that is calculated by FFT, we can tell $\Psi = 1.3$ is the best choice for this set of parameters.

The simulation error related to the choice of Ψ is roughly confined within $\pm$ 0.25 so it is acceptable.

References:

[1] Andersen, Leif B.G., Efficient Simulation of the Heston Stochastic Volatility Model (January 23, 2007). Available at SSRN: https://ssrn.com/abstract=946405 or http://dx.doi.org/10.2139/ssrn.946405

[2] Fang Fang and Cornelis W.Oosterlee. Pricing early exercise and discrete barrier options based on Fourier-Cosine series expansions. SIAM Journal on Scientific Computing, 8(2):1-18, Winter 2004.

[3] Kahl, C. and P. Jackel (2005), “Fast strong approximation Monte Carlo scheme for stochastic volatility models,” Working Paper, ABN AMRO and University of Wuppertal.

[4] Ali Hirsa, Computational Methods in Finance

[5] Gilles Pages, Introduction to Numerical Probability for Finance

(Next Page is Report II)

Project Report II.

Importance Sampling and Variance Reduction

Yunhao Yan (yy2882)

Teammates: Han Lin, Weihao Yan

Summary:

For simulation efficiency, variance reduction techniques are necessary. In this section, we formulated the idea of importance sampling and implement it into the Monte Carlo simulation of deep out-of-the-money option.

Although the best importance sampling distribution can be proved, it is not practical because of the high cost of inversing a distribution function. Therefore, a Gaussian variable is typically a good choice for importance sampling. We utilized the Stochastic Gradient Descent Method to determine the optimal mean parameter θ^*. Here we compare the simulation efficiency of non-importance sampling, $\theta_1 = -\frac{\left(\log\left(\frac{S_0}{K}\right)+rT\right)}{\sigma\sqrt{T}}$, optimal θ^* and the exact price given by BS formula.

1. Importance Sampling

1.1 Basic Formulation and proofs

If $\theta = E_f[h(X)]$, where X is a random variable following the density f(x), and $h(X)$ is the statistical quantity we are interested in, then we have,

$$\theta = E_f[h(X)] = \mathrm{E_g}\left[\frac{h(X)f(X)}{g(X)}\right]$$

In the RHS context, the X follows the density g(x). Here is the proof,

$$\mathrm{E_g}\left[\frac{h(X)f(X)}{g(X)}\right] = \int \frac{h(x)f(x)}{g(x)} dG(x) = \int h(x)\, dF(x) = E_f[h(X)]$$

The variance reduced, if not amplified, is

$$\begin{aligned} &\mathrm{Var_f}[h(X)] - Var_g\left[\frac{h(X)f(X)}{g(X)}\right] \\ &= E_f[h(X)^2] + \theta^2 - \left(\mathrm{E_g}\left[\frac{h^2(X)f^2(X)}{g(X)^2}\right] + \theta^2\right) \\ &= E_f[h(X)^2] - \mathrm{E_g}\left[\frac{h^2(X)f^2(X)}{g(X)^2}\right] \\ &= E_f[h(X)^2] - E_f\left[\frac{h^2(X)f(X)}{g(X)}\right] \end{aligned}$$

$$= \int h^2(x)(1 - \frac{f(x)}{g(x)})f(x)dx$$

From the expression $E_f[h(X)^2] - E_f\left[\frac{h^2(X)f(X)}{g(X)}\right]$, we know if the variance is indeed reduced, then we must have

$$E_f[h(X)^2] > E_f\left[\frac{h^2(X)f(X)}{g(X)}\right]$$

1.2 In the Black-Scholes Context

$$\mathrm{S_T} = S_0 e^{\left(r-\frac{\sigma^2}{2}\right)T+\sigma\sqrt{T}Z}$$

To replace Z~Normal(0,1) with Y~Normal(θ, 1), we have,

$$E_f[h(Z)] = E_g\left[\frac{h(Z)f(Z)}{g(Z)}\right]$$

$$E_g\left[\frac{h(Z)f(Z)}{g(Z)}\right] = E_f\left[h(Z+\theta)\frac{\frac{1}{\sqrt{2\pi}}\ e^{-\frac{(Z+\theta)^2}{2}}}{\frac{1}{\sqrt{2\pi}}\ e^{-\frac{(Z)^2}{2}}}\right] = E_f\left[h(Z+\theta)e^{-\frac{1}{2}\theta(2Z+\theta)}\right]$$

$$= E_f\left[h(Z+\theta)e^{-Z\theta-\frac{1}{2}\theta^2}\right] = \mathrm{e}^{\frac{1}{2}\theta^2}E_f\left[h(Z+\theta)e^{-\theta(Z+\theta)}\right]$$

To choose a good θ, one choice is to set

$$E_f\left[S_0 e^{\left(r-\frac{\sigma^2}{2}\right)T+\sigma\sqrt{T}(Z+\theta)}\right] = K$$

The $\left(r-\frac{\sigma^2}{2}\right)T+\sigma\sqrt{T}(Z+\theta)$ is a normal distribution with mean $\left(r-\frac{\sigma^2}{2}\right)T+\sigma\sqrt{T}\theta$ and variance $\sigma^2 T$, Therefore,

$$E_f\left[S_0 e^{\left(r-\frac{\sigma^2}{2}\right)T+\sigma\sqrt{T}(Z+\theta)}\right]$$
$$= S_0 e^{\left(r-\frac{\sigma^2}{2}\right)T+\sigma\sqrt{T}\theta+\frac{1}{2}\sigma^2 T}$$
$$= S_0 e^{rT+\sigma\sqrt{T}\theta}$$

Set $S_0 e^{rT+\sigma\sqrt{T}\theta} = K$, we have,

$$\theta = -\frac{\log\left(\frac{S_0}{K}\right)+rT}{\sigma\sqrt{T}}$$

1.3 The Best Sampling Distribution

If the original X is non-negative, let's variated the reduction expression,

$$\delta\ \mathrm{J[g(x)]} = \delta\int_0^{+\infty} h^2(x)\frac{f^2(x)}{g(x)}dx = -\int_0^{+\infty} h^2(x)\frac{f^2(x)}{g^2(x)}(\delta g)dx$$

since g(x) is a density, we must have,

$$\int_0^{+\infty}\delta g(x)dx = 0$$

Observing that if $g(x) = \frac{h(x)\cdot f(x)}{\int_0^{+\infty} h(s)\cdot f(s)ds}$,

$$\delta\, J[g(x)] = -\int_0^{+\infty} h^2(x) \frac{f^2(x)}{\left(\frac{h(x)\cdot f(x)}{\int_0^{+\infty} h(s)\cdot f(s)ds}\right)^2} (\delta g) dx$$

$$= -\left(\int_0^{+\infty} h(s)\cdot f(s) ds\right)^2 \int_0^{+\infty} \left(\frac{h(x)\cdot f(x)}{h(x)\cdot f(x)}\right)^2 \delta g(x) dx = 0$$

Hence $g(x) = \frac{h(x)\cdot f(x)}{\int_0^{+\infty} h(s)\cdot f(s)ds}$ is the optimal distribution to reduce the variance for a non-negative random variable. But the formula is not practical because of the high cost of inverse a unknow function.

2. Stochastic Gradient and Optimal θ^*

Since we want to maximize the variance reduction, we want to minimize the variance of this quantity

$$E_g\left[\frac{h(Z)f(Z)}{g(Z)}\right]$$

Hence minimize the quantity,

$$\min\left\{E_g\left[\frac{h(Z)f(Z)}{g(Z)}\right]^2\right\}$$

$$= \min\{e^{-\theta^2} E\left[h^2(Z+\theta)e^{-2\theta Z}\right]\}$$

$$= \min\left\{E\left[\left(S_0 * e^{(r-\sigma^2)T+\sigma\sqrt{T}(Z+\theta)} - K\right)^2 e^{-2\theta Z-\theta^2}\right]\right\}$$

Stochastic Gradient Method uses the gradient of a specific variable and iterates for many times instead of computing the arduous gradient of the expectation.

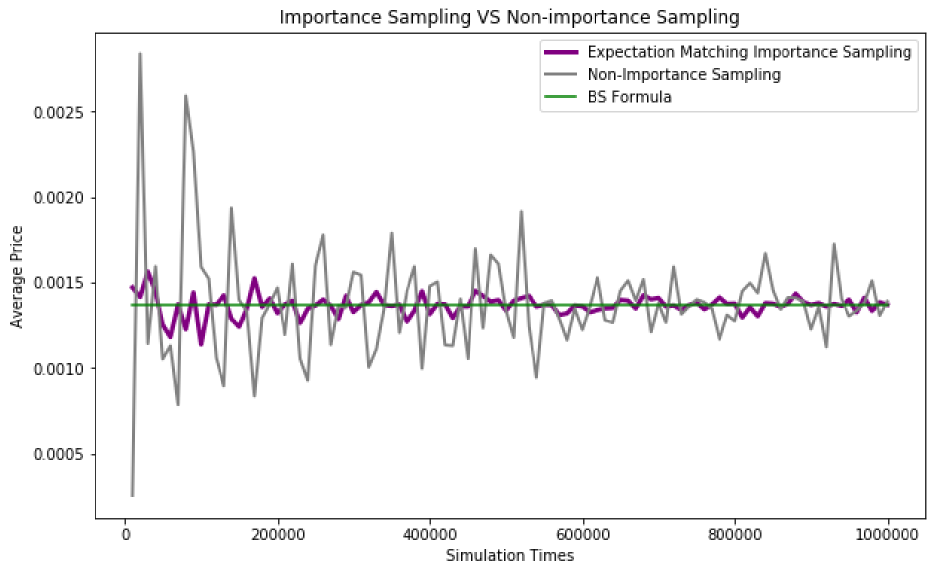
In this case, $K(\theta) = h^2(Z+\theta)e^{-2\theta Z-\theta^2}$,

$$\frac{\partial K(\theta)}{\partial \theta} = 2h * h' e^{-2\theta Z-\theta^2} - 2(Z+\theta)h^2 e^{-2\theta Z-\theta^2}$$

Hence, we generate standard normal variable Z and update θ using the above equation.

3. Results

Importance Sampling with Expectation Matching θ

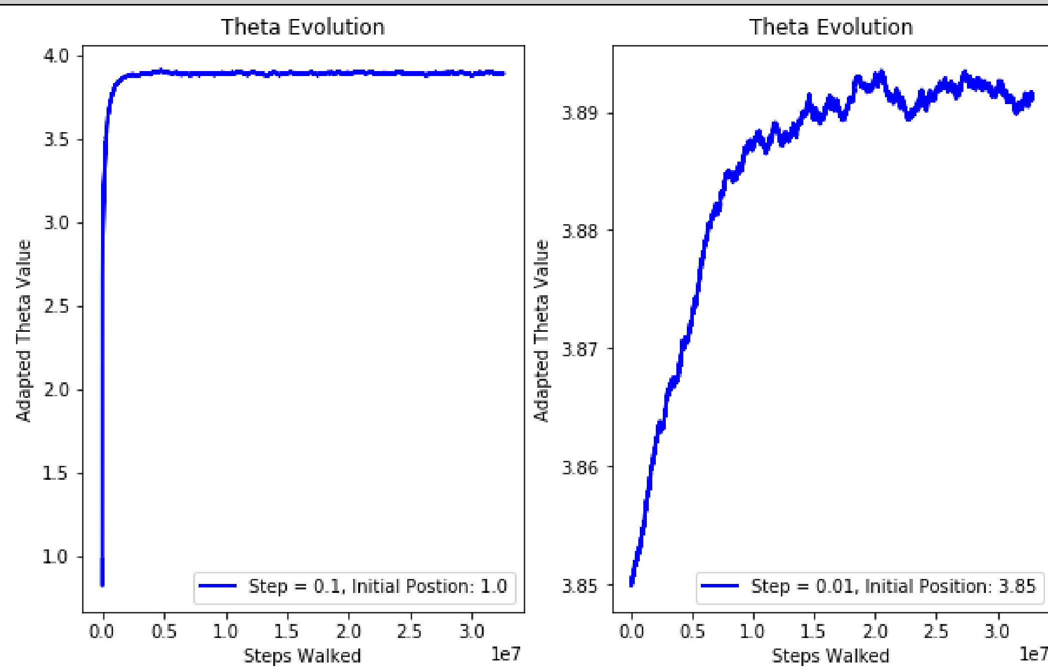


Comment:

From the picture, we can see that using the θ s. t. the expectation of risk-neutral S_T matching the strike price K, will reduce the variance considerably.

$$\theta_1 = -\frac{\log\left(\frac{S_0}{K}\right)+rT}{\sigma\sqrt{T}} = 0.823371152441798$$

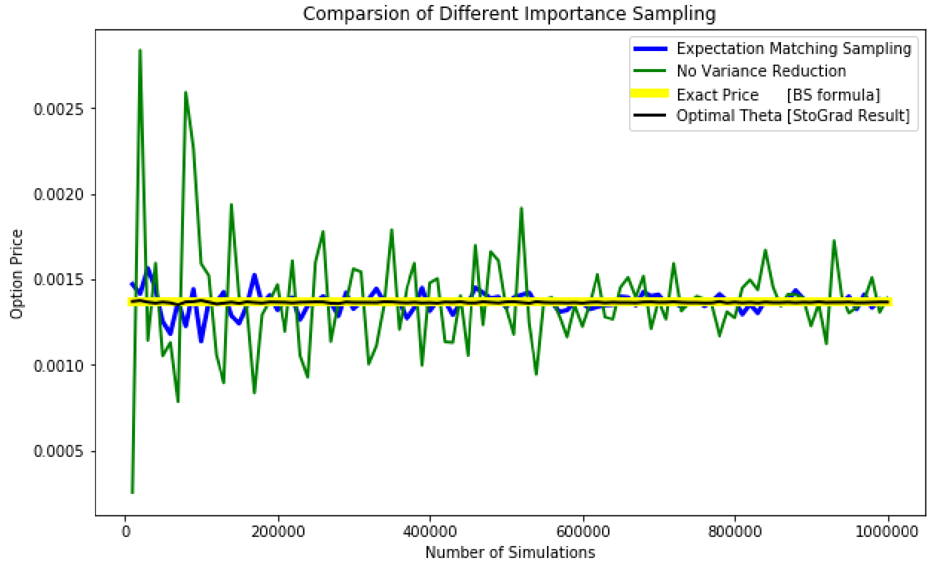
Stochastic Gradient Descent



Comment:

The optimal parameter we found is l $\theta^* = 3.8950840816731374$

A Comprehensive Comparison of Simulation Efficiency



Comment:

The optimal θ^* reduce the variance greatly than the arbitrary expectation match θ. The option's pricing line is almost on the analytic formula line. Find the optimal θ is very meaning for large scale Monte Carlo Simulation.

References:

[1] Gilles Pages, Introduction to Numerical Probability for Finance

[2] Ali Hirsa, Computational Methods in Finance

[3] Kunmiao Liang, Theoretical Mechanics, Chapter 8 Calculus of Variations.