A Comparison of Discretization Schemes for Non Negative Diffusion Processes for European Call Options

Alex Vukadinovic

1 Introduction

Non negative diffusion processes are used to model a wide array of financial assets and derivatives. The price of an underlying asset at a time t, given by $S(t)$, say the price of a stock, should remain positive with probability 1 for $t \geq 0$. Consider the simple Geometric Brownian Motion model which can be used to model $S(t)$:

$$dS(t) = rS(t)dt + \sigma S(t)dB(t), \qquad S(0) = S_o \geq 0$$

where the constants r and σ are the interest rate and volatility respectively, and $B(t)$ is standard Brownian motion. The solution to this equation is $S(t) = S_o \exp(rt + \sigma B(t))$ which is clearly non negative for $t \geq 0$. If the analytical solution was unknown, one might instead discretize the equation using the Euler scheme to approximate $S(t)$ by $\tilde{S}_k$:

$$\tilde{S}_{k+1} = \tilde{S}_k + r\tilde{S}_k \Delta t + \sqrt{\Delta t}\sigma \tilde{S}_k Z_{k+1}$$

where Z $\sim N(0,1)$. Since Z is a standard normal random variable, it can take negative values, hence there is a non zero probability that $\tilde{S}_k$ takes a negative value. Here it is clear that even if the analytical solution is a non negative process, the discrete approximation can produce negative values due to discretization error.

In addition to producing unrealistic negative values, for some models the chosen scheme may even break down completely, for example consider the variance part of the Heston model (1) and the same discretization scheme (2):

$$dv(t) = k[\theta - v(t)]dt + \sigma\sqrt{v(t)}dB(t) \tag{1}$$

$$\tilde{v}_{k+1} = \tilde{v}_k - \kappa[\theta - \tilde{v}_k]\Delta t + \sigma\sqrt{\tilde{v}_k \Delta t} Z_{k+1} \tag{2}$$

If $\tilde{v}_k$ takes a negative value, then the scheme can no longer proceed to $\tilde{v}_{k+1}$ since $\sqrt{\tilde{v}_k}$ is not a real number. So the approximate solution only exists up to a random time $t = k$. The probability of the process being negative for the next time step is:

$$\mathbf{P}(\tilde{v}_{k+1} < 0) = \Phi\left(\frac{-\tilde{v}_k + \kappa\Delta t(\theta - \tilde{v}_k)}{\sigma\sqrt{\tilde{v}_k \Delta t}}\right)$$

where Φ is the standard normal cdf. Hence for any $\Delta t > 0$ and $\sigma > 0$ there is always a positive probability that $\tilde{v}_{k+1} < 0$.

Many methods have been proposed to ensure that the discretized solution remains non negative for non negative diffusion processes. One naive approach would be to simply discard the values of the process which are negative and keep only the positive samples. Of course a major problem with this is the wasted computation time needed to generate unused samples. Alternatively, schemes can be modified to only produce positive samples, though one has to be careful when applying naive solutions to this problem. It should still remain the case that the discretized process $\tilde{X}_k$ converges to the actual process $X(t)$ as the size of the largest interval goes to zero. Modifying scheme (2) can involve applying a function $f(\cdot)$ to $\tilde{v}_k$ for example by reflection $f(\tilde{v}_k) = |\tilde{v}_k|$ or absorption $f(\tilde{v}_k) = \tilde{v}_k^+$ to some or all $\tilde{v}_k$ on the RHS. Hence (2) can be rewritten in the form:

$$\tilde{v}_{k+1} = f_1(\tilde{v}_k) - \kappa[\theta - f_2(\tilde{v}_k)]\Delta t + \sigma\sqrt{f_3(\tilde{v}_k)\Delta t} Z_{k+1}$$

Any of the two choices for $f_i(\cdot)$ stated above or the identity $f_i(x) = x$ produce an unbiased estimate $\tilde{v}_k$ for $v(t)$ for Δt small enough. Unfortunately some of these schemes produce large biases for practical sizes of Δt, while other combinations of $f(\cdot)$ produce very little bias for reasonable parameters. An example of such a scheme is the full truncation scheme:

$$\tilde{v}_{k+1} = \tilde{v}_k - \kappa[\theta - \tilde{v}_k^+]\Delta t + \sigma\sqrt{\tilde{v}_k^+ \Delta t} Z_{k+1}$$

which is also strongly convergent in the L^1 norm:

$$\lim_{\Delta t \to 0} sup|v(k) - \tilde{v}_k| \to 0, \qquad t \in [0, T]$$

The above schemes employ the Euler discretization but the problem of negative values is also present in other schemes. The Milstein scheme applied to the Heston model also produces negative values, but still works well with full truncation applied to the volatility. The

presence of a normal random variable in these schemes is the source of the problem, and hence it is natural to try to discretize the process using a non negative random variable instead. Labe, C. et al. (2010) [1] proposes such a method, which under certain conditions, converges to the true solution. This paper will examine aspects of this scheme as compared to the Euler and Milstein schemes with respect to the Heston model.

2 Scheme with Non Negative R.V.

The solutions discussed above centered around modifying the schemes which employ normal random variables, but the paper [1] aims to attack the problem directly. By replacing the normal variable with one which has zero probability of taking a negative value, no further measures are needed to ensure a non negative solution. To illustrate how the solution is implemented, consider a more general two dimensional stochastic differential equation (SDE) of the form:

$$dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dB(t) \qquad X(0) = X_o \tag{3}$$

The conditions which are assumed to be satisfied by (3) are as follows:

Condition 1 The functions $\mu : \mathbb{R}_+ \times \mathbb{R}^2 \to \mathbb{R}^2$ and $\sigma : \mathbb{R}_+ \times \mathbb{R}^2 \to \mathbb{R}^2 \otimes \mathbb{R}^2$ are continuous and $X_o \in \mathbb{R}^2_+$ is a constant vector.

Condition 2 There is an integer m where $0 \leq m \leq 2$ such that for each $X_o \in \mathbf{D} := \mathbb{R}^m_+ \times \mathbb{R}^{2-m}$ there exists a unique weak solution X_t such that $X_t \in \mathbf{D}\ \forall t \geq 0$ almost surely.

Condition 2 states that the for the SDE given by (3), the unique solution must exist so long as if a component of the process is non negative for all t, the initial condition is also non negative for that component.

Condition 3 Let $\omega = (\omega_1, \omega_2)^T$ be a random vector in on a probability space $(\Omega, F, \tilde{\mathbf{P}})$ with mean vector $\nu = (\nu_1, \nu_2)^T$ and co-variance matrix Σ. There exists a continuous function $\tilde{\sigma}(t, x) : \mathbb{R}_+ \times \mathbf{D} \to \mathbb{R}^2 \otimes \mathbb{R}^2$ and a symmetric semi-definate positive $d \times d$ matrix Σ such that $\sigma(t, x)\sigma^T(t, x) = \tilde{\sigma}(t, x)\Sigma\tilde{\sigma}^T(t, x)$ for all $(t, x) \in \mathbb{R}_+ \times \mathbb{R}^2$. Furthermore, there exists a

positive integer n_o such that for $\Delta t = 1/n$, $n \in \mathbb{R}_+$:

$$\inf_{(t,x)\in\mathbb{R}_+\times\mathbf{D}} \left(x + \frac{1}{n}\mu(t,x) - \frac{1}{\sqrt{n}}\tilde{\sigma}(t,x)\nu \right) \in \mathbf{D} \quad \forall n \geq n_o \tag{4}$$

where the infimum is componentwise.

Condition 4 For every $(t,x) \in \mathbb{R}_+ \times \mathbf{D}$, the following holds:

$$\tilde{\mathbf{P}}\left(\tilde{\sigma}(t,x)\omega \in \mathbf{D}\right) = 1 \tag{5}$$

Under the assumption that these four conditions are satisfied, define a distreization scheme for (3) as:

$$\tilde{X}^n_{k+1} = \tilde{X}^n_k + \frac{1}{n}\mu(k/n, \tilde{X}^n_k) + \frac{\tilde{\sigma}(k/n, \tilde{X}^n_k)}{\sqrt{n}}(\omega_k - \nu) \tag{6}$$

where $\tilde{X^m_k}$ is the descretised process at time k/n and ω_k is the k^{th} i.i.d. sample random vector with distribution ω. In [1], the convergence of (6) to the solution of (3) is proven in Theorem 8, stated here:

Theorem (Convergence of Solution): For the scheme given by (6) and conditions 1-4, for an approximating solution $\tilde{X^n_k}$ where the time steps are $1/n$, define the linear interpolation between the discrete solution as:

$$\bar{X}^n(t) = \tilde{X}^n(\lfloor nt \rfloor) + (nt - \lfloor nt \rfloor)(\tilde{X}^n(\lfloor nt \rfloor + 1) - \tilde{X}^n(\lfloor nt \rfloor))$$

then the sequence of approximating processes $(\tilde{X}^n(t))_{n\geq n_o}$ strongly converges to the solution of (3) as $n \to \infty$.

In particular, if $S(t)$ is the price of a security and the expected payoff of a derivative is given by $\mathbf{E}[g(S(t))]$ where g is a continuous on w.r.t $S(t)$ on $t \in [0,T]$ and bounded function, then $\lim_{n\to\infty} \mathbf{E}[g(\tilde{S}^n(t))] = \mathbf{E}[g(S(t))]$.

3 Numerical Results

This section provides a comparison of four numerical schemes applied to the Heston Model for the pricing of European call options. The focus is largely on the bias and computation time of the schemes since the variances of the estimated option prices are similar for all four methods. The prices are estimated based on $N = 1,000,000$ sample paths. For the purpose of comparison, the Euler and Milstein schemes are implemented along with two schemes based on [1]. The Bernoulli method employed in the paper is replicated here as well as an implementation of the scheme with a Gamma random variable, referred to as the Gamma scheme going forward. The Heston Stochastic Volatility Model considered is given by:

$$dv(t) = k[\theta - v(t)]dt + \sigma\sqrt{v(t)}dB^v(t), \quad v(0) = v_o \geq 0 \tag{7}$$

$$dS(t) = S(t)(rdt + \sigma\sqrt{v(t)}dB^s(t), \qquad S(0) = S_o \geq 0 \tag{8}$$

The two dimensional Borwnain motion (B^v, B^s) have instantaneous correlation ρ, i.e. $dB^s(t)dB^v(t) = \rho dt$. Since [1] provides the true option price for a given set of parameters, in order to asses the bias of the four schemes, this paper uses the same parameters, given in Table 1.

Table 1: Parameter Values

ρ	T	σ	κ	θ	r	v_o	S_o	K
-0.3	5	1	2	0.09	0.05	0.09	100	100

For the strike price K, the payoff function for a European call option is $e^{-rT}(S(t) - K)^+$ who's expected value is estimated using crude Monte Carlo. For the given parameters, the true price of the option as given by [1] is 34.9998.

The implementation was done in MATLAB R2016b using the available pseudo-random number generators for the needed distributions. Uniform, Gamma and Normal variables were generated using *rnd*(), *gamrnd*() and *normrnd*() respectively. For each scheme the code was modified only where needed and consideration was given to efficiency in order to

provide a good comparison of computation times.

3.1 Euler and Milstein Schemes

Both the Euler and Milstein schemes have weak order of convergence of 1, but the purpose of implementing both is to compare biases and computation times when applied to the Heston Model. For both methods the log prices are used since these values are allowed to be negative, and full truncation is applied to both schemes for log price and volatility. A simple application of Itō's Lemma produces the process for $\ln(S(t))$ and the Euler discretization with full truncation results in:

$$\ln\tilde{S}_{k+1} = \ln\tilde{S}_k + \frac{1}{n}(r - \frac{1}{2}\tilde{v_k^+}) + \frac{1}{\sqrt{n}}\sqrt{\tilde{v}_k^+}Z_k^S \tag{9}$$

$$\tilde{v}_{k+1} = \tilde{v}_k + \frac{\kappa}{n}(\theta - \tilde{v}_k^+) + \sigma\sqrt{\frac{v_k^+}{n}}Z_k^v \tag{10}$$

The Milstein scheme for the log price reduces to the exact same form as (9), in other words, no correction term appears. The volatility scheme, on the other hand, does differ from the Euler discretization by a correction term. The Milstein scheme for volatility with full truncation is:

$$\tilde{v}_{k+1} = \tilde{v}_k + \frac{\kappa}{n}(\theta - \tilde{v}_k^+) + \sigma\sqrt{\frac{\tilde{v}_k^+}{n}}Z_k^v + \frac{1}{4n}\sigma^2((Z_k^v)^2 - 1) \tag{11}$$

The differences in the biases and computation times are then due to this last term which appears in (11). The results from the Euler and Milstein schemes in Tables 2 and Table 3 indicated biases which are statistically different from zero (at the 95% level) in bold.

Lower and Upper denotes the 95% confidence bounds for the price, and the bias is estimated based on the true value of the option. As the number of time steps n per year increases, both schemes show decrease in bias but the Milstein scheme seems to perform better for almost all n.

Table 2: Euler

n	6	10	20	80	150
Price	35.2974	35.2079	35.0389	35.0734	35.0425
Std	0.0591	0.0587	0.0581	0.0581	0.0581
Bias	**0.2976**	**0.2081**	0.0391	0.0736	0.0427
Lower	35.1815	35.0928	34.9251	34.9595	34.9287
Upper	35.4133	35.323	35.1528	35.1873	35.1563
Time	55.31	62.02	79.68	170.04	268.25

Table 3: Milstein

n	6	10	20	80	150
Price	34.8607	34.9233	34.9145	34.9741	35.0396
Std	0.0601	0.0591	0.0581	0.0583	0.0583
Bias	**-0.1391**	-0.0765	-0.0853	-0.0257	0.0398
Lower	34.743	34.8075	34.8006	34.8598	34.9254
Upper	34.9785	35.0392	35.0285	35.0884	35.1539
Time (s)	54.97	63.67	81.11	194.34	315.61

3.2 Bernoulli Scheme

The choice of Bernoulli random variables follows that of [1] which are denoted here as $(\omega_1^B, \omega_2^B)'$. Choosing Bernoulli random variables described by:

$$\mathbf{P}(\omega_i^B = 0) = \frac{1}{1+\nu_i^2}, \quad \mathbf{P}\left(\omega_i^B = \nu_i + \frac{1}{\nu_i}\right) = \frac{\nu_i^2}{1+\nu_i^2} \tag{12}$$

results in $\mathbf{E}[\omega_i^B] = \nu_i$ and $\mathbf{Var}[\omega_i^B] = 1$. Following [1], the the scheme implemented in this paper sets $\nu_1 = 0.657$ and $\nu_2 = 1$. Also, in order to avoid a second optimization problem needed to find the parameter values for condition 3 to be satisfied, log prices are used for the Bernoulli and Gamma methods. Similarly to the Euler and Milstein scheme, producing Bernoulli pseudo-random variables ω_1^B and ω_3^B with correlation of ρ entails setting $\omega_3^B = \rho\omega_1^B + \sqrt{1-\rho^2}\omega_2^B$. Hence, with the pseudo-random variables written explicitly in terms

of Cholesky decomposition the scheme employed is:

$$\tilde{v}_{k+1} = \tilde{v}_k + \frac{\kappa}{n}(\theta - \tilde{v}_k) + \sigma\sqrt{\frac{\tilde{v}_k}{n}}(\omega_1^B - \nu_1) \tag{13}$$

$$\ln\tilde{S}_{k+1} = \ln\tilde{S}_k + \frac{1}{n}(r - \frac{1}{2}\tilde{v_k}) + \frac{1}{\sqrt{n}}\sqrt{\tilde{v}_k}[\rho(\omega_1^B - \nu_2) + \sqrt{1-\rho^2}(\omega_2^B - \nu_1)] \tag{14}$$

In terms of (3) the coefficients of system (7) and (8) with log prices are

$$\mu(v(t)) = \begin{bmatrix} \kappa(\theta - v(t)) \\ r \end{bmatrix}, \quad \sigma(v(t)) = \begin{bmatrix} \sigma\sqrt{v(t)} & 0 \\ \rho\sqrt{v(t)} & \sqrt{1-\rho^2}v(t) \end{bmatrix}$$

Then for the co-variance matrix Σ of (ω_1^B, ω_3^B) define $\tilde{\sigma}(v(t), S(t)) = diag[\sigma\sqrt{v(t)}, \sqrt{v(t)}]$ as given in condition 3. Since the scheme is formulated in terms of log price, it is only needed to check that the first component of (4) remains positive. By optimizing with respect to $v(t)$, the infimum given in can be shown to have a value of

$$\frac{\kappa\theta}{n} - \frac{\nu_1^2\sigma^2}{4(n-\kappa)} \tag{15}$$

which is positive when $n_o \geq \kappa$ and $0 < \nu_1 \leq 2/\sigma\sqrt{\kappa\theta(1-\kappa/n_o)}$ and is satisfied for the choice of parameters in Table 1. The Bernoulli method seems to provide some major advantages over the Euler and Milstein schemes in terms of computation times and bias.

Table 4: Bernoulli

n	6	10	20	80	150
Price	35.0639	35.0263	35.085	34.9871	35.0051
Std	0.0563	0.0566	0.0576	0.058	0.0578
Bias	0.0641	0.0265	0.0852	-0.0127	0.0053
Lower	34.9536	34.9153	34.9722	34.8735	34.8918
Upper	35.1742	35.1373	35.1978	35.1007	35.1183
Time	21.03	24.61	34.88	95.49	173.48

From Table 4, the computation times are significantly reduced as compared to either scheme. At least in the case of non path dependent options pricing in the Heston model, the Bernoulli scheme offers some significant advantages. The performance increase can be attributed to the fact that Bernoulli pseudo-random variables rely on generating uniform pseudo-random variables which require much less computational resources, as compared to normal pseudo-random variables. The biases are also consistently lower for each n, suggesting the scheme has lower bias overall.

3.3 Gamma Scheme

The Gamma scheme is identical to the Bernoulli scheme except the pseudo-random variables are now under a Gamma distribution. The Gamma pseudo-random variables denoted as (ω_1^G, ω_1^G) are chosen with shape parameters $\alpha_1 = 1/5$, $\beta_1 = \sqrt{5}$ and $\alpha_2 = \beta_2 = 1$. Note that $\mathbf{E}[\omega_1^G] = \sqrt{5}/10$, $\mathbf{E}[\omega_2^G] = 1$ and $\mathbf{Var}[\omega_1^G] = \mathbf{Var}[\omega_2^G] = 1$. Hence similarly to the Bernoulli scheme, condition 3 is satisfied with the parameters chosen in Table 1. The Gamma scheme is not competitive for this set of parameters in the Heston model in both computation time and bias.

Table 5: Gamma

n	6	10	20	80	150	1000
Price	33.4624	33.7945	34.0702	34.5087	34.6479	34.903
Std	0.0583	0.059	0.0583	0.0583	0.0581	0.184
Bias	**-1.5374**	**-1.2053**	**-0.9296**	**-0.4911**	**-0.3519**	-0.0968
Lower	33.3482	33.6788	33.9558	34.3944	34.534	34.5423
Upper	33.5766	33.9101	34.1845	34.6229	34.7617	35.2636
Time	58.63	75.36	109.21	315.81	567.47	3564.36

By replacing Bernoulli pseudo-random variables with the more computationally inten-

sive Gamma pseudo-random variables, the scheme loses the computation time advantage over the Euler and Milstein schemes. Secondly the bias is very large under this scheme for all time steps $6 - 150$. As shown in Table 5, all practical time steps show a bias statistically different from zero at the 95% significance level. For reference, the table also includes the case of 1,000 time steps per year, showing that the scheme does seem to go to zero with large n, although very slowly. Clearly the computation time of about 1 hour is not practical.

4 Conclusions and Final Comparisons

Overall the efficiency of the scheme presented in [1] is very dependent on the distribution of pseudo-random variable chosen, even if all the conditions are satisfied. The use of the Gamma pseudo-random variable performed much worse than the other three schemes. The Bernoulli scheme, on the other hand, shows an advantage in terms of bias and computation for the Heston model employed here. Figures 1 and 2 plot the three best performing schemes with the magnitude of the bias as a function of times steps per year and computation time respectively.

Figure 1: Bias vs Times Steps per Year

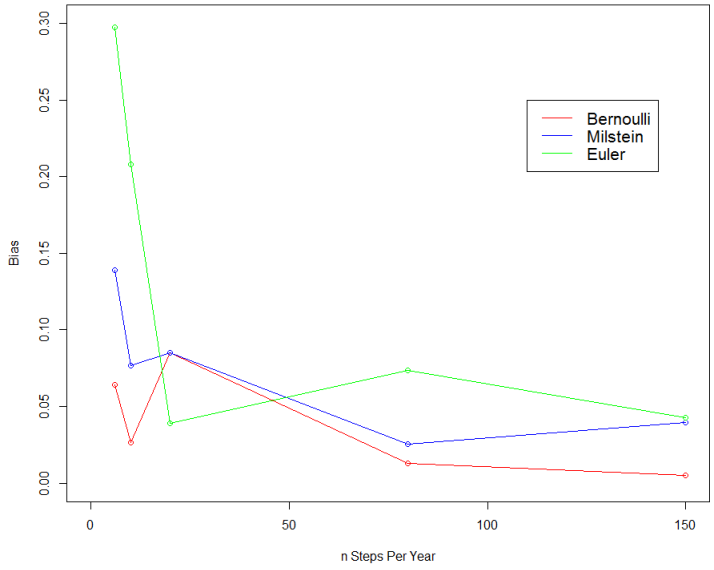

Figure 2: Bias vs Computation Time (s)

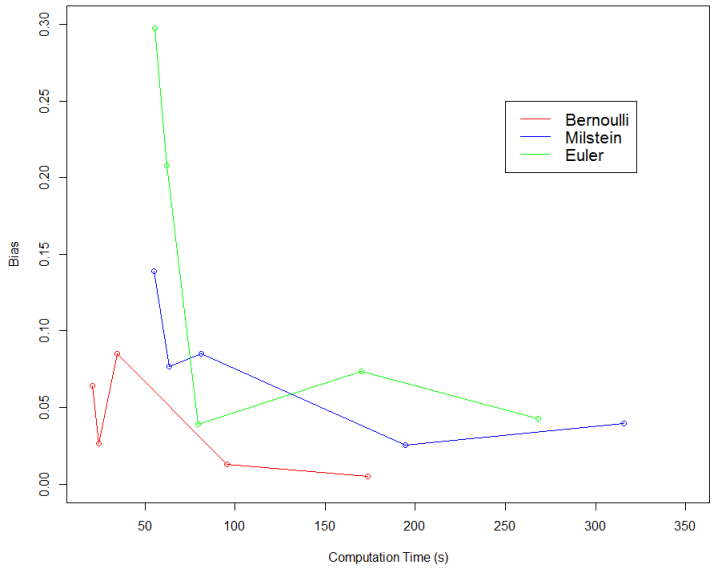

Figure 2 suggest that for the crude Monte Carlo approach, the Bernoulli scheme works best as it provides the lowest computation time for a given level of tolerable bias one might have. The method does provide some drawbacks since the restrictions on the parameters given by condition 3 may be too severe for some applications. In looking for non negative random variables which can be made to satisfy the conditions on the mean and variance, the choices were limited. Lastly, for models is higher dimensions, finding a closed form infimum such as in (15) leading to a restriction on the parameters may not be feasible. Hence the best method would be to choose as set of parameters and use numerical methods to find the infimum in (4).

The Bernoulli scheme provides relatively low bias and computation time for the crude Monte Carlo method. Since the variables only rely on generating uniform random variables, improving computation time with other non negative random variables is not likely. The Bernoulli random variables are likely close to the best choice for this model for a crude Monte Carlo approach.

References

[1] Labbé, C., Remillard, B., & Renaud, J. (2010, November 14). A Simple Discretization Scheme for Nonnegative Diffusion Processes, with Applications to Option Pricing. *SSRN Electronic Journal*. doi:10.2139/ssrn.1619989

[2]Haastrecht, A. V., & Pelsser, A. (2008, September 12). Efficient, Almost Exact Simulation Of The Heston Stochastic Volatility Model. *International Journal of Theoretical and Applied Finance,13*(01), 1-43. doi:10.1142/s0219024910005668