MAFS5250 – Computational Methods for Pricing Structured Products

Topic 5 - Monte Carlo simulation

5.1 General formulation of the Monte Carlo procedure
- Expected value and variance of the estimate
- Multistate extension – correlated random samples
- Computational efficiency
- Calculating the Greeks

5.2 Variance reduction techniques
- Antithetic variates method
- Control variate method
- Importance sampling

5.3 Valuation of American options
- Method of parameterization of the early exercise boundary
- Linear regression method via basis functions

5.1 General formulation of the Monte Carlo procedure

A wide class of European style derivative pricing problems lead to the evaluation of the following expectation functional

$$E_t[f(Z(T;t,z))].$$

- The stochastic process Z describes the price evolution of one or more underlying financial variables such as asset prices and interest rates, under the respective risk neutral probability distribution.

- The process Z has the initial value z at time t, and the function f specifies the value of the derivative at the expiration time T.

- The Monte Carlo method is basically a numerical procedure for estimating the expected value of a random variable, and so it leads itself naturally to derivative pricing problem represented as expectations.

Simulation procedure

The simulation procedure involves generating random variables with a given probability density and using the law of large numbers to take the average of these values as an estimate of the expected value of the random variable.

Consider a Brownian motion with drift rate μ and volatility σ, where

$$dx_t = \mu\, dt + \sigma\, dB_t.$$

Here, B_t denotes the standard Brownian motion with zero drift rate and unit variance rate. The Euler discretized scheme is given by

$$x_{t+\Delta t} = x_t + \mu\Delta t + \sigma(B_{t+\Delta t} - B_t).$$

Provided that σ is a constant, the Euler scheme provides $O(\Delta t)$ approximation to the stochastic differential equation for x_t (see Appendix A3). The random increment $B_{t+\Delta t} - B_t$ has mean zero and variance Δt, so it can be simulated by random samples of $\sqrt{\Delta t}\epsilon$, where ϵ is a sample from a standard normal distribution.

In the context of derivative pricing, the Monte Carlo procedure involves the following steps.

(i) Simulate sample paths of the underlying state variables in the derivative model such as asset prices and interest rates over the life of the derivative, according to the risk neutral probability distributions.

(ii) For each simulated sample path, evaluate the discounted cash flows of the derivative.

(iii) Take the sample average of the discounted functional of the cash flows over all sample paths.

Simulated calculation of a European call option

The numerical procedure requires the computation of the expected payoff of the call option at expiry, $E_t[\max(S_T - X, 0)]$, based on the information available at time t (which is S_t in the current context) and discounted to the present value at time t, namely,

$$e^{-r(T-t)} E_t[\max(S_T - X, 0)].$$

Assuming that the asset price S_t follows the Geometric Brownian motion, where the price dynamics under the risk neutral measure is given by

$$\frac{dS_t}{S_t} = (r - q)\, dt + \sigma\, dB_t$$

or

$$d \ln S_t = \left(r - q - \frac{\sigma^2}{2} \right) dt + \sigma\, dB_t.$$

Here, σ is the volatility, r is the riskless interest rate, and q is the dividend yield.

Over time increment Δt, the Euler approximation scheme gives

$$\ln \frac{S_{t+\Delta t}}{S_t} = \ln S_{t+\Delta t} - \ln S_t = \left(r - q - \frac{\sigma^2}{2} \right) \Delta t + \sigma \epsilon \sqrt{\Delta t}$$

so that the random asset price ratio is related to the standard normal random variable ϵ via

$$\frac{S_{t+\triangle t}}{S_t} = e^{\left(r - q - \frac{\sigma^2}{2} \right) \Delta t + \sigma \epsilon \sqrt{\Delta t}}.$$

Note that the price ratio $\dfrac{S_T}{S_t}$ can be decomposed as the product of price ratios over successive time steps as follows:

$$\frac{S_T}{S_t} = \frac{S_{t+\Delta t}}{S_t} \frac{S_{t+2\Delta t}}{S_{t+\Delta t}} \cdots \frac{S_{t+N\Delta t}}{S_{t+(N-1)\Delta t}},$$

where $T = N \Delta t$.

- Suppose there are N time steps between the current time t and expiration time T, where $\Delta t = (T-t)/N$. The numerical procedure is repeated N times to simulate the price path from S_t to $S_T = S_{t+N\triangle t}$.

- The i^{th} simulated call value c_i corresponding to the i^{th} simulated terminal asset price $S_T^{(i)}$ is then computed using the discounted expectation approach under the risk neutral measure

$$\begin{aligned} c_i &= e^{-r(T-t)} \max(S_T^{(i)} - X, 0) \\ &= e^{-r(T-t)} \max\left(S_t e^{\left(r-q-\frac{\sigma^2}{2}\right)(T-t)+\sigma\sqrt{\Delta t}\sum_{j=1}^{N}\epsilon_j^{(i)}} - X, 0 \right). \end{aligned}$$

- After repeating the above simulation for a sufficiently large number of runs, the expected call value is obtained by computing the sample average of the simulated call value found in the sample simulation.

Monte Carlo convergence goes as $1/\sqrt{M}$

Suppose we have M simulations $\{c_1, c_2, \ldots, c_M\}$ of the call value, where c_i's represent independent and identically distributed random variables.

By virtue of the law of large numbers, the sample mean $\bar{c}$, tends to the "true" price μ of the call option when the number of simulations is very large. Note that $E[c_i] = \mu$. The population variance

$$\text{var}(c_i) = \sigma^2$$

comes from discretization errors. Different choices of simulation algorithms (like the choice of the Euler scheme or Milstein scheme) lead to different discretization errors, hence different population variances. This error is caused by simulation of continuous time processes by discrete time processes.

We do not know μ nor σ^2 but we can certainly estimate μ and σ^2 using a suitably chosen statistic, like sample mean and sample variance.

The sample mean $\bar{c} = \frac{1}{M}\sum_{i=1}^{M} c_i$ as an estimator of μ is unbiased since the difference between the expected value of the estimator $\bar{c}$ and the true value μ is seen to be zero, as shown by

$$E[\bar{c}] = \frac{1}{M}\sum_{i=1}^{M} E[c_i] = \frac{1}{M}\sum_{i=1}^{M} \mu = \mu.$$

Once these random variables $\{c_1, c_2, \ldots, c_M\}$ are actually observed, the sample mean $\bar{c}$ is the arithmetic average of these observations.

We can estimate the population variance σ^2 using the sample variance

$$s_V^2 = \frac{\sum_{i=1}^{M} c_i^2 - M\bar{c}^2}{M-1}.$$

The term $M-1$ in the denominator (Bessel's correction) ensures that s_V^2 is an unbiased estimator of σ^2. One can understand Bessel's correction intuitively from the degrees of freedom in the residuals vector (not errors since the population mean μ is not known):

$$(c_1 - \bar{c}, \ldots, c_M - \bar{c}),$$

where $\bar{c}$ is the sample mean. While there are M independent samples, there are only $M-1$ independent residuals, since they sum to zero.

Recall that the variance of the sample mean $\bar{c}$ (unbiased estimator of μ) and population variance var(c_i) are related by

$$\text{var}(\bar{c}) = \text{var}\left(\frac{1}{M}\sum_{i=1}^{M} c_i\right) = \frac{1}{M^2}\text{var}\left(\sum_{i=1}^{M} c_i\right) = \frac{1}{M}\text{var}(c_i) = \frac{1}{M}\sigma^2.$$

We may use $\frac{1}{M}s_V^2$ as our estimate of var($\bar{c}$). The standard error (SE) of $\bar{c}$ is taken to be the square root of var($\bar{c}$), where

$$SE = \frac{1}{\sqrt{M}}\sqrt{\frac{\sum_{i=1}^{M} c_i^2 - M\bar{c}^2}{M-1}}.$$

We expect the *SE* to decrease as $1/\sqrt{M}$.

One may use the standard error to construct a confidence interval for the estimate of the true value μ.

Central limit theorem

- Let $\bar{c}$ be an estimator of the call value. For a sufficiently large value of simulation runs M, the distribution

$$\frac{\sqrt{M}(\bar{c}-c)}{\sigma}, \quad c \text{ is the true call value,}$$

tends to the standard normal distribution. Here, σ is the population standard deviation of the simulated values of the call price. In summary, the statistical error arising from the central limit theorem decreases at the rate of $1/\sqrt{M}$.

- The confidence limits of estimation can be reduced by increasing the number of simulation runs M, though the reduction factor is $\frac{1}{\sqrt{M}}$. For example, in order to reduce the standard error by a factor of $\frac{1}{10}$, we need to increase the number of simulation runs by 100 times.

Sample calculations

$S = 100, \sigma = 0.2, r = 0.06, q = 0.03, N =$ number of time steps $=$ $10, M =$ number of simulation trials $= 100, T = 1$ so that

$$\Delta t = T/N = 0.1, \quad \sigma\sqrt{\Delta t} = 0.2\sqrt{0.1} = 0.0632,$$
$$\left(r - q - \frac{\sigma^2}{2}\right)\Delta t = \left(0.06 - 0.03 - \frac{0.04}{2}\right)0.1 = 0.001,$$
$$\ln S_0 = 4.6052.$$

Suppose the first sample value of ϵ is -0.0497, then

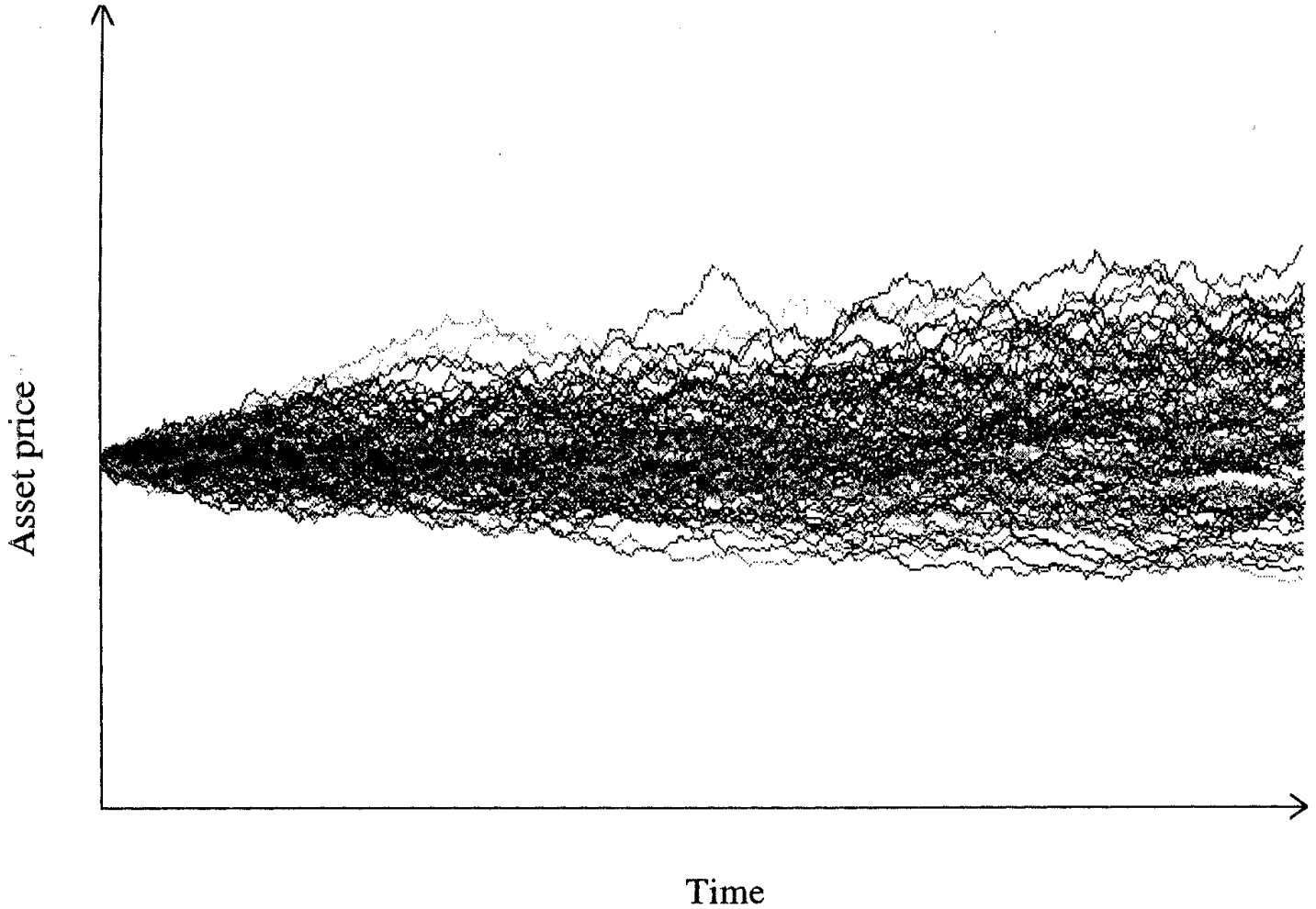
$$\ln S_{\Delta t} = 4.6052 + 0.0001 + 0.0632(-0.0497) = 4.6030.$$

- The Figure illustrates a set of $M = 100$ simulated paths repeatedly with typical parameter values for stock: $S = 100, \sigma = 20\%$, $r = 6\%$, $q = 3\%$, $T = 1$ year, $N = 10$.

- To obtain an estimate of the call price μ, based on the law of large numbers, we simply take the discounted average of these simulated terminal payoff values:

$$\overline{c} = e^{-rT} \left[\frac{1}{M} \sum_{i=1}^{M} \max(0, S_T^{(i)} - X) \right].$$

Simulated Asset Price Paths in the Black–Scholes World



Numerical Example for Monte Carlo Valuation of a European Call Option in a Black–Scholes World

K	T	S	sig	r	div	N	M	sum_CT	sum_CT2	SD
100	1	100	0.2	0.06	0.03	10	100	996.49	26610.7	12.22457

dt	nudt	sigsdt	lnS	call_value	SE
0.1	0.0010	0.0632	4.6052	9.3846	1.2225

lnSt

j i	0	1	2	3	4	5	6	7	8	9	10	ST	CT	CT*CT
1	4.6052	4.6030	4.6257	4.6738	4.6512	4.6666	4.5619	4.5864	4.5521	4.4840	4.6521	104.81	4.8070	23.11
2	4.6052	4.6862	4.6749	4.5386	4.4745	4.4546	4.4901	4.5206	4.4977	4.4618	4.4993	89.95	0.0000	0.00
3	4.6052	4.6430	4.6144	4.6184	4.6770	4.7225	4.7489	4.7402	4.7507	4.7275	4.7983	121.30	21.2996	453.67
4	4.6052	4.6188	4.6297	4.5106	4.5404	4.4814	4.4843	4.4984	4.5072	4.5060	4.4590	86.40	0.0000	0.00
5	4.6052	4.6807	4.7067	4.6915	4.7176	4.7258	4.7810	4.8131	4.8516	4.7885	4.8243	124.49	24.4939	599.95
95	4.6052	4.6121	4.6291	4.6968	4.6099	4.6949	4.5597	4.5229	4.6530	4.6761	4.6349	103.02	3.0211	9.13
96	4.6052	4.6331	4.6620	4.7152	4.7603	4.8091	4.8847	4.8218	4.7674	4.7796	4.8012	121.65	21.6545	468.92
97	4.6052	4.5729	4.7010	4.6682	4.7704	4.8236	4.9556	4.8797	4.9385	4.9607	4.9938	147.49	47.4931	2255.59
98	4.6052	4.5910	4.6111	4.6929	4.6479	4.7125	4.8040	4.9143	4.9915	4.9884	4.9925	147.30	47.3012	2237.41
99	4.6052	4.5634	4.5047	4.4652	4.4984	4.5028	4.5370	4.4913	4.5071	4.4906	4.4065	81.98	0.0000	0.00
100	4.6052	4.5189	4.5183	4.4599	4.5276	4.6812	4.6191	4.5996	4.5985	4.7130	4.7282	113.09	13.0941	171.46

ε

j i	1	2	3	4	5	6	7	8	9	10
1	−0.0497	0.3425	0.7442	−0.3723	0.2277	−1.6708	0.3709	−0.5581	−1.0924	2.6422
2	1.2660	−0.1948	−2.1717	−1.0290	−0.3296	0.5444	0.4668	−0.3777	−0.5831	0.5763
3	0.5818	−0.4677	0.0476	0.9110	0.7042	0.4014	−0.1541	0.1510	−0.3833	1.1032
4	0.1999	0.1557	−1.8976	0.4551	−0.9486	0.0294	0.2076	0.1225	−0.0347	−0.7592
5	1.1781	0.3955	−0.2564	0.3978	0.1129	0.8569	0.4924	0.5929	−1.0130	0.5489
95	0.0938	0.2533	1.0545	−1.3899	1.3276	−2.1535	−0.5969	2.0413	0.3494	−0.6672
96	0.4258	0.4411	0.8249	0.6971	0.7571	1.1791	−1.0113	−0.8750	0.1762	0.3261
97	−0.5256	2.0098	−0.5352	1.6003	0.8254	2.0709	−1.2152	0.9144	0.3343	0.5075
98	−0.2396	0.3012	1.2779	−0.7271	1.0051	1.4321	1.7281	1.2042	−0.0647	0.0485
99	−0.6757	−0.9440	−0.6416	0.5091	0.0544	0.5256	−0.7383	0.2339	−0.2779	−1.3456
100	−1.3793	−0.0259	−0.9388	1.0538	2.4134	−0.9981	−0.3233	−0.0339	1.7945	0.2252

- $N = 10$ time steps and $M = 100$ simulation runs

The sum of the values of C_T and the squares of the values of C_T are accumulated:

$$\sum_{j=1}^{M} C_{T,j} = 996.488(\text{sum\_CT}) \quad \text{and} \quad \sum_{j=1}^{M} (C_{T,j})^2 = 26610.7(\text{sum\_CT2}).$$

The estimate of the option value $\widehat{C}_0$ (call_value) is then given by

$$\widehat{C}_0 = 996.488/100 \times \exp(-0.06 \times 1) = 9.3846.$$

The standard error of the estimate is given by

$$\begin{aligned} SE &= \frac{\sqrt{\sum_{j=1}^{M} (C_{T,j})^2 - \frac{1}{M}\left(\sum_{j=1}^{M} C_{T,j}\right)^2}\exp(-rT)}{\sqrt{M-1}} \Big/ \sqrt{M} \\ &= \frac{\sqrt{26610.73 - \frac{1}{100}(996.488)^2}\exp(-0.06 \times 1)}{\sqrt{100-1}} \Big/ 10 \\ &= 12.2246/10 = 1.22246. \end{aligned}$$

The standard errors is quite sizable compared to the call value. To reduce the standard error to 0.0122246, we need to choose $M = 100 \times 100^2 = 1$ million.

Summary of important properties

1. Its ease to accommodate complicated payoff in an option model. For example, the terminal payoff of an Asian option depends on the average of the asset price over certain time interval while that of a lookback option depends on the extremum value of the asset price over some period of time. It is quite straightforward to obtain the average or extremum value in each simulated price path.

2. The demand for a large number of simulation trials in order to achieve a high level of accuracy. The reduction factor in the standard error of $\overline{c}$ is $1/\sqrt{M}$.

3. No curse of dimensionality

   The operation counts in Monte Carlo simulation of n correlated state variables is linear in n. That is, if n is doubled, the simulation time is roughly doubled. Unlike finite difference schemes where the operation counts increase in powers of n, Monte Carlo Simulation suffers no curse of dimensionality.

Multistate extension – correlated random samples

- Consider the situation where the payoff from a derivative depends on n variables $\theta_i\,(1 \leq i \leq n)$. Define σ_i as the volatility of $\theta_i, \widehat{m}_i$ as the expected growth rate of θ_i in a risk-neutral world, and ρ_{ik} as the instantaneous correlation coefficient between θ_i and θ_k.

- The life of the derivative is divided into N subintervals of length Δt. Under the Geometric Brownian motion assumption, the discrete version of the process for θ_i is

$$\theta_i(t + \Delta t) - \theta_i(t) = \widehat{m}_i \theta_i(t) \Delta t + \sigma_i \theta_i(t) \epsilon_i \sqrt{\Delta t},$$

  where ϵ_i is a random sample from a standard normal distribution. The correlation coefficient between ϵ_i and ϵ_k is $\rho_{ik}\,(1 \leq i; k \leq n)$.

- One simulation trial involves generating N samples of $\epsilon_i\,(1 \leq i \leq n)$ from a multivariate standardized normal distribution. These are substituted into the above equation to produce simulated paths for each θ_i, thereby enabling a sample value for the derivative to be calculated.

- When two correlated samples ϵ_1 and ϵ_2 from standard normal distributions are required, independent samples x_1 and x_2 from a univariate standardized normal distribution are obtained. The required samples ϵ_1 and ϵ_2 are then calculated as follows:

$$\epsilon_1 = x_1, \quad \epsilon_2 = \rho x_1 + x_2\sqrt{1-\rho^2},$$

where ρ is the coefficient of correlation.

As a verification, we consider

$$\begin{aligned} \text{cov}(\epsilon_1, \epsilon_2) &= \text{cov}(x_1, \rho x_1 + x_2\sqrt{1-\rho^2}) \\ &= \rho\text{cov}(x_1, x_1) + \sqrt{1-\rho^2}\text{cov}(x_1, x_2) = \rho \end{aligned}$$

since $\text{cov}(x_1, x_1) = \text{var}(x_1) = 1$ and $\text{cov}(x_1, x_2) = 0$.

Also, we consider

$$\begin{aligned}\text{var}(\epsilon_2) &= \text{cov}(\rho x_1 + x_2\sqrt{1-\rho^2}, \rho x_1 + x_2\sqrt{1-\rho^2}) \\ &= \rho^2\text{cov}(x_1, x_1) + 2\rho\sqrt{1-\rho^2}\text{cov}(x_1, x_2) + (1-\rho^2)\text{cov}(x_2, x_2) \\ &= \rho^2 + (1-\rho^2) = 1.\end{aligned}$$

In matrix form, we have

$$\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

where

$$\begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix}^T = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

The transformation matrix $\begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix}$ is related to the Cholesky decomposition of the correlation matrix $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$.

Extension to three samples

The required samples, $\epsilon_i\,(1 \leq i \leq 3)$, are defined as follows:

$$\begin{aligned}
\epsilon_1 &= a_{11}x_1 \\
\epsilon_2 &= a_{21}x_1 + a_{22}x_2 \\
\epsilon_3 &= a_{31}x_1 + a_{32}x_2 + a_{33}x_3
\end{aligned}$$

and so on. The matrix representation is seen to be

$$\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{pmatrix} = \begin{pmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

By equating the second order moments, we obtain

$$\begin{aligned}
\rho_{21} &= \mathrm{cov}(\epsilon_1, \epsilon_2) = a_{21}a_{11} \\
\rho_{31} &= \mathrm{cov}(\epsilon_1, \epsilon_3) = a_{31}a_{11} \\
\rho_{32} &= \mathrm{cov}(\epsilon_2, \epsilon_3) = a_{31}a_{21} + a_{32}a_{22} \\
1 &= \mathrm{var}(\epsilon_1) = a_{11}^2 \\
1 &= \mathrm{var}(\epsilon_2) = a_{21}^2 + a_{22}^2 \\
1 &= \mathrm{var}(\epsilon_3) = a_{31}^2 + a_{32}^2 + a_{33}^2.
\end{aligned}$$

Alternatively, we consider

$$
\begin{aligned}
AA^T &= \begin{pmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & 0 \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} & a_{31} \\ 0 & a_{22} & a_{32} \\ 0 & 0 & a_{33} \end{pmatrix} \\
&= \begin{pmatrix} a_{11}^2 & a_{21}a_{11} & a_{31}a_{11} \\ a_{21}a_{11} & a_{21}^2 + a_{22}^2 & a_{31}a_{21} + a_{32}a_{22} \\ a_{31}a_{11} & a_{31}a_{21} + a_{32}a_{22} & a_{31}^2 + a_{32}^2 + a_{33}^2 \end{pmatrix} \\
&= \begin{pmatrix} 1 & \rho_{21} & \rho_{31} \\ \rho_{21} & 1 & \rho_{32} \\ \rho_{31} & \rho_{32} & 1 \end{pmatrix} = \sum .
\end{aligned}
$$

Sequential steps to determine the coefficients

We choose the coefficients a_{ij} so that the correlation and variances are correct. This can be done step by step as follows. Set $a_{11} = 1$; choose a_{21} so that $a_{21}a_{11} = \rho_{21}$; choose a_{22} so that $a_{21}^2 + a_{22}^2 = 1$; choose a_{31} so that $a_{31}a_{11} = \rho_{31}$; choose a_{32} so that $a_{31}a_{21} + a_{32}a_{22} = \rho_{32}$; choose a_{33} so that $a_{31}^2 + a_{32}^2 + a_{33}^2 = 1$; and so on.

Theoretical formulation of the Cholesky decomposition

Suppose we can generate n uncorrelated normally distributed variables $x_1, x_2, \cdots, x_n$ with zero mean and unit variance. Write $\boldsymbol{x} = (x_1\ x_2 \cdots x_n)^T$ and observe that $E[\boldsymbol{x}\boldsymbol{x}^T] = I$. Let Σ be the correlation matrix, which must be symmetric and semi-positive definite. Choose a matrix A such that

$$AA^T = \Sigma$$

where A is lower triangular. This is called the Cholesky decomposition, which stems from the LU factorization. Due to the symmetric property of $\sum$, we deduce that $U = L^T$.

Define $\boldsymbol{\epsilon} = A\boldsymbol{x}$, then $\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T = A\boldsymbol{x}\boldsymbol{x}^T A^T$ so that

$$E[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T] = AE[\boldsymbol{x}\boldsymbol{x}^T]A^T = AA^T = \Sigma.$$

The semi-positive definite property of $\sum$ is revealed in the Cholesky decomposition since

$$\boldsymbol{y}^T \sum \boldsymbol{y} = \boldsymbol{y}^T AA^T \boldsymbol{y} = (A^T\boldsymbol{y})^T A^T \boldsymbol{y} \geq 0 \text{ for all } \boldsymbol{y}.$$

Cholesky factorization at work

For $d \times d$ covariance matrix Σ, we need to solve

$$\begin{aligned}\Sigma &= AA^T \\ &= \begin{pmatrix} a_{11} & & & \\ a_{21} & a_{22} & & \\ \vdots & & \ddots & \\ a_{d1} & a_{d2} & \cdots & a_{dd} \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{d1} \\ & a_{22} & \cdots & a_{d2} \\ & & \ddots & \vdots \\ & & & a_{dd} \end{pmatrix}.\end{aligned}$$

Simply by multiplying, we observe

$$\begin{aligned} a_{11}^2 &= \rho_{11} \\ a_{21}a_{11} &= \rho_{21} \\ &\vdots \\ a_{d1}a_{11} &= \rho_{d1} \\ a_{21}^2 + a_{22}^2 &= \rho_{22} \\ &\vdots \\ a_{d1}^2 + \cdots + a_{dd}^2 &= \rho_{dd}. \end{aligned}$$

Suppose we write

$$\begin{aligned}\rho_{ii} &= \sum_{l=1}^{i} a_{il}^2,\ i = 1, \ldots, d; \\ \rho_{ij} &= \sum_{l=1}^{j} a_{il} a_{jl},\ j < i.\end{aligned}$$

Taking $a_{11} = \rho_{11} = 1$, we have $a_{i1} = \rho_{i1}$. For a given i, we have computed a_{jj} for $j = 1, 2, \ldots, i-1$. We then calculate

$$a_{ij} = \left(\rho_{ij} - \sum_{l=1}^{j-1} a_{il} a_{jl} \right) / a_{jj},\ j = 2, \ldots, i-1;$$

and in particular, we calculate

$$a_{ii} = \rho_{ii} - \sum_{l=1}^{i-1} a_{il}^2.$$

The procedure is repeated for $i = 2, \ldots, d$.

Factor correlation of swap rates and 3-month LIBOR

Historical correlation of LIBOR and swap rates

ρ	3m LIBOR	5-year swap	10-year swap	30-year swap
3m LIBOR	1	0.1638	0.0817	0.0814
5y swap	0.1638	1	0.7118	0.8595
10y swap	0.0817	0.7118	1	0.6816
30y swap	0.0814	0.8595	0.6816	1

We use the Cholesky factorization to obtain

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0.1638 & 0.9865 & 0 & 0 \\ 0.0817 & 0.7080 & 0.7015 & 0 \\ 0.0814 & 0.8595 & 0.0965 & 0.4983 \end{pmatrix}.$$

The correlated unit variance normal variables $\epsilon_1, \ldots, \epsilon_4$ are related to the standard normal variables $x_1, \ldots, x_4$ by

$$\begin{aligned} \epsilon_1 &= x_1 \\ \epsilon_2 &= 0.1638x_1 + 0.9865x_2 \\ \epsilon_3 &= 0.0817x_1 + 0.7080x_2 + 0.7015x_3 \\ \epsilon_4 &= 0.0814x_1 + 0.8595x_2 + 0.0965x_3 + 0.4983x_4. \end{aligned}$$

Once sufficiently large samples of $\epsilon_1, \ldots, \epsilon_4$ are found, we can now compute the sampled correlation of them. Their correlations are displayed in the table below. Comparing correlations between the historical correlation matrix in the two tables, as expected, they are pretty close (though the percentage errors are seen to be quite significant for ρ_{13} and ρ_{14}).

Correlation of standard normal random vectors after Cholesky factorization

ρ	ϵ_1	ϵ_2	ϵ_3	ϵ_4
ϵ_1	1	0.1624	0.0646	0.0724
ϵ_2	0.1624	1	0.7087	0.8558
ϵ_3	0.0646	0.7087	1	0.6855
ϵ_4	0.0724	0.8558	0.6855	1

Computational efficiency

Suppose W_T is the total amount of computational work units available to generate an estimate of the value of an option V.

Assume that there are two methods for generating the Monte Carlo estimates for the option value, requiring W_1 and W_2 units of computation work respectively for each simulation run. For simplicity, we assume W_T to be divisible by both W_1 and W_2. For the given amount of W_T units of computer time, the number of simulation runs that can be performed for method i in $W_T/W_i, i = 1, 2$. For example, given $W_T = 100,000$ while $W_1 = 50$ and $W_2 = 100$, then the number of simulation runs that can be performed for Method 1 and Method 2 are 2,000 and 1,000, respectively.

The sample means for estimating V from the two methods using W_T amount of work are, respectively,

$$\frac{W_1}{W_T}\sum_{i=1}^{W_T/W_1} V_i^{(1)} \quad \text{and} \quad \frac{W_2}{W_T}\sum_{i=1}^{W_T/W_2} V_i^{(2)}.$$

Here, $V_i^{(1)}$ and $V_i^{(2)}$ denote the estimator of V in the i^{th} simulation using Methods 1 and 2, respectively, where each $V_i^{(1)}$ and $V_i^{(2)}$ has expectation V and their respective population standard deviations are σ_1 and σ_2.

What is the trade off between smaller standard deviation in the estimation and heavier computational work? The most fundamental criterion is to compare the standard errors achieved at a given total computation time W_T in both methods.

Their respective standard errors are

$$\sigma_1\sqrt{\frac{W_1}{W_T}} \quad \text{and} \quad \sigma_2\sqrt{\frac{W_2}{W_T}}.$$

The first method would be preferred over the second one provided that

$$\sigma_1^2 W_1 < \sigma_2^2 W_2.$$

Alternatively speaking, a lower variance estimator is preferred only if the variance ratio σ_1^2/σ_2^2 is less than the work ratio W_2/W_1, when the aspect of computational efficiency is taken into account.

Numerical instability in delta calculations

A naive approach to calculate the delta of an option using Monte Carlo simulation is to estimate the option value twice. Recall

$$\Delta \approx \frac{V(S+h,t)-V(S-h,t)}{2h}, h \text{ is small.}$$

This is an estimate of the first order derivative using the centered difference formula, with an error of $O(h^2)$. Such procedure involves subtracting two close numbers (leading to loss of significant figures) and division by a small number. For example, consider $0.123789 - 0.123456 = 0.000333$, the number of significant figures drops from 6 to 3. In general, estimating the derivative of a function numerically is unstable.

The standard errors in the estimates of the two option values at $S+h$ and $S-h$ are magnified when divided by h, resulting in an error of $O\left(\frac{1}{hM^{1/2}}\right)$, where M is the total number of simulation runs.

Calculations of the delta of the Black-Scholes call price

We would like to calculate $\Delta = \frac{\partial c}{\partial S_0}$, where c is the call option price and S_0 is the current stock price. Suppose we generate the first simulated terminal stock price

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

and a second independent terminal stock price

$$S_T(\varepsilon) = (S_0 + \varepsilon) e^{\left(r - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}Z'},$$

with Z and Z' being independent standard normal random variables.

For each simulated terminal price, the discounted call payoff is

$$\widehat{c}(S_0) = e^{-rT} \max\left(S_T - K, 0\right), \quad \widehat{c}(S_0 + \varepsilon) = e^{-rT} \max(S_T(\varepsilon) - K, 0).$$

A crude estimate of delta is given by

$$\widetilde{\Delta} = \varepsilon^{-1} \left[\widehat{c}(S_0 + \varepsilon) - \widehat{c}(S_0)\right].$$

- By generating n independent replications of S_T and $S_T(\varepsilon)$, we can calculate the sample mean of n independent copies of $\tilde{\Delta}$. As $n \to \infty$, this sample mean converges to the true finite difference ratio:

$$\varepsilon^{-1} \left[c(S_0 + \varepsilon) - c(S_0) \right].$$

- To obtain an accurate estimate of Δ, we should take ε small. Given that we have generated S_T and $S_T(\varepsilon)$ independently of each other, we have

$$\text{var}(\tilde{\Delta}) = \varepsilon^{-2} \left[\text{var}(\hat{c}(S_0 + \varepsilon)) + \text{var}(\hat{c}(S_0)) \right] = O(\varepsilon^{-2}),$$

so $\text{var}(\tilde{\Delta})$ becomes very large if we make ε small.

Method of common random numbers

Use the same normal random variable Z instead of generating 2 sets of independent Z. The variance of the new estimate of Δ is given by

$$\text{var}(\widehat{\Delta}) = \varepsilon^{-2}\left[\text{var}(\widehat{c}(S_0)) + \text{var}(\widehat{c}(S_0+\varepsilon)) - 2\text{cov}(\widehat{c}(S_0), \widehat{c}(S_0+\varepsilon))\right]$$

since $\widehat{c}(S_0)$ and $\widehat{c}(S_0+\varepsilon)$ are now no longer independent. Indeed, they are positively correlated, so $\widehat{\Delta}$ has a smaller variance than $\tilde{\Delta}$.

More precisely, we observe that

$$|\widehat{c}(S_0+\varepsilon) - \widehat{c}(S_0)| \leq |S_T(\varepsilon) - S_T| = \varepsilon e^{\left(r-\frac{\sigma^2}{2}\right)T+\sigma\sqrt{T}Z};$$

so that

$$E\left[|\widehat{c}(S_0+\varepsilon) - \widehat{c}(S_0)|^2\right] = O(\varepsilon^2).$$

This gives $\text{var}\left(\varepsilon^{-1}[\widehat{c}(S_0+\varepsilon) - \widehat{c}(S_0)]\right) = O(1)$. The variance of $\hat{\Delta}$ remains bounded as $\varepsilon \to 0$.

RMS Errors for Various Delta Estimation Methods

		Independent Z's		Common Z	
	ε	forward difference	centered difference	forward difference	centered difference
Standard Call Option	10	0.10	0.01	0.100	0.009
	1	0.18	0.09	0.012	0.006
	0.1	1.78	0.87	0.006	0.006
	0.01	7.47	8.98	0.006	0.006

Root mean square error of delta estimates for the call option using the four methods with various values of ε. The model parameters are $S_0 = 100$, $K = 100$, $\sigma = 0.40$, $r = 0.10$ and $T = 0.2$. Each entry is computed from $1,000$ delta estimates, each estimate is based on $10,000$ replications. The value of delta is 0.580 for the call option.

At large ε, the discretization error of the forward difference formula is significant.

Observations

- The RMS errors incurred in computing delta of a vanilla call using common Z are mostly independent of ϵ and the choice of difference method (forward or centered difference).

- However, the RMS errors incurred in computing delta of a vanilla call using independent Z's are highly dependent on the choice of difference method (forward difference gives higher RMS errors). Also, the RMS errors become significant with vanishingly small value of ϵ.

Governing equation for the delta under the Black-Scholes framework

Another way to calculate the delta (and gamma) is to exploit the differential equation to be satisfied by the delta. Differentiating the Black–Scholes equation with respect to S, this gives

$$\frac{\partial}{\partial t}\left(\frac{\partial V}{\partial S}\right)+\frac{\partial}{\partial S}\left(rS\frac{\partial V}{\partial S}+\frac{\sigma^2}{2}S^2\frac{\partial^2 V}{\partial S^2}\right)-r\frac{\partial V}{\partial S}=0.$$

Since

$$\frac{\partial}{\partial S}\left(rS\frac{\partial V}{\partial S}+\frac{\sigma^2}{2}S^2\frac{\partial^2 V}{\partial S^2}\right)=r\frac{\partial V}{\partial S}+rS\frac{\partial \Delta}{\partial S}+\frac{\sigma^2}{2}S^2\frac{\partial^2 \Delta}{\partial S^2}+\sigma^2 S\frac{\partial \Delta}{\partial S},$$

so

$$\frac{\partial \Delta}{\partial t}+\frac{\sigma^2}{2}S^2\frac{\partial^2 \Delta}{\partial S^2}+(r+\sigma^2)S\frac{\partial \Delta}{\partial S}=0. \qquad \text{(A)}$$

For a vanilla call option, the delta at expiry is

$$\Delta(S,T)=\mathbf{1}_{\{S>X\}}=\begin{cases}1 & \text{if } S\geq X\\ 0 & \text{otherwise}\end{cases}.$$

We can estimate the value of the delta directly without taking the finite difference of the computed option values by a Monte Carlo simulation in which we calculate the expected value of the terminal delta using the following random walk for S_t:

$$dS_t = (r + \sigma^2) S_t \, dt + \sigma S_t \, dZ_t.$$

This corresponds to

$$\frac{S_{t+\Delta t}}{S_t} = e^{\left(r + \frac{\sigma^2}{2}\right)\Delta t + \sigma\epsilon\sqrt{\Delta t}}.$$

Since there is no discounting term in the partial differential equation, there is no need to take the present value. This result does not come to a great surprise if we consider the call price formula:

$$\begin{aligned} c(S,t) &= e^{-r(T-t)} E_t \left[S_T \mathbf{1}_{\{S_T > X\}} \right] - X e^{-r(T-t)} E_t [\mathbf{1}_{\{S_T > X\}}] \\ &= S N(d_1) - X e^{-r(T-t)} N(d_2). \end{aligned}$$

Recall

$$\Delta = N(d_1) = N\left(\frac{\ln \frac{S}{X} + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}\right)$$

and

$$N(d_2) = N\left(\frac{\ln \frac{S}{X} + \left(r - \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}\right) = P[S_T > X].$$

Note that $N(d_2)$ is the solution to the undiscounted Black-Scholes equation

$$\frac{\partial V}{\partial t} + rS\frac{\partial V}{\partial S} + \frac{\sigma^2}{2}S^2\frac{\partial^2 V}{\partial S^2} = 0$$

subject to $V(S_T, T) = \mathbf{1}_{\{S_T > X\}}$.

By comparing the drift terms in Δ and $N(d_2)$, we deduce that Δ satisfies eq.(A) and the same terminal condition.

Pros and cons of Monte Carlo simulations

Advantages

- The level of mathematics required for performing a Monte Carlo simulation can be very basic.
- Correlations among underlying state variables can be easily modeled.
- Complex path dependency can often be easily incorporated.
- Suffer no curse of dimensionality since the workload increases linearly with the dimension of the pricing model.

Disadvantages

- The method is slow when compared with the lattice tree methods for solving low-dimensional option models.
- Relatively cumbersome to handle the early exercise feature of American style derivatives.

5.2 Variance reduction techniques

- Suppose $\{\epsilon^{(i)}\}$ denotes the independent samples from the standard normal distribution for the i^{th} simulation run of the asset price path so that

$$S_T^{(i)} = S_t \; e^{\left(r-\frac{\sigma^2}{2}\right)(T-t)+\sigma\sqrt{\triangle t}\sum\limits_{j=1}^{N}\epsilon_j^{(i)}}, \quad i = 1, 2, \cdots, M,$$

  where $\triangle t = \dfrac{T-t}{N}$ is the time step and M is the total number of simulation runs. Note that $\epsilon_j^{(i)}$ is randomly sampled from the standard normal distribution.

- An unbiased estimator of the price of a European call option with strike price X is given by

$$\widehat{c} = \frac{1}{M}\sum_{i=1}^{M} c_i = \frac{1}{M}\sum_{i=1}^{M} e^{-r(T-t)}\max(S_T^{(i)} - X, 0).$$

Antithetic variates method

We observe that if $\{\epsilon^{(i)}\}$ has a standard normal distribution, so does $\{-\epsilon^{(i)}\}$, and the simulated price $\widetilde{S}_T{}^{(i)}$ obtained using $\{-\epsilon^{(i)}\}$ is also a valid sample for the terminal asset price distribution. A new unbiased estimator of the call price can be obtained from

$$\widetilde{c} = \frac{1}{M}\sum_{i=1}^{M} \widetilde{c}_i = \frac{1}{M}\sum_{i=1}^{M} e^{-r(T-t)} \max(\widetilde{S}_T^{(i)} - X, 0).$$

Normally we would expect $\widehat{c}$ and $\widetilde{c}$ to be negatively correlated, that is, if one estimate overshoots the true value, the other estimate downshoots the true value. It seems sensible to take the average of these two estimates. Indeed, we take the antithetic variates estimate to be

$$\bar{c}_{AV} = \frac{\widehat{c} + \widetilde{c}}{2}.$$

Rationale for better performance

The random inputs obtained from the collection of antithetic pair $\{(Z_i, -Z_i)\}$ are more regularly distributed than a collection of $2n$ independent samples.

- A large value of $S_T^{(i)}$ resulting from a large Z_i will be paired with a small value of $\widetilde{S}_T^{(i)}$ obtained from $-Z_i$.
- The sample mean over the antithetic pairs always equals the population mean of 0, whereas the mean over finitely many independent samples is almost surely different from 0.

Example – Pricing an Asian option using antithetic variables

$$
\begin{aligned}
S^+_{t_1} &= S_{t_0} \exp\left(\left(r - q - \frac{\sigma^2}{2}\right)(t_1 - t_0) + \sigma\sqrt{t_1 - t_0}\, z_1\right), \\
S^+_{t_2} &= S^+_{t_1} \exp\left(\left(r - q - \frac{\sigma^2}{2}\right)(t_2 - t_1) + \sigma\sqrt{t_2 - t_1}\, z_2\right), \\
&\vdots \\
S^+_{t_m} &= S^+_{t_{m-1}} \exp\left(\left(r - q - \frac{\sigma^2}{2}\right)(t_m - t_{m-1}) + \sigma\sqrt{t_m - t_{m-1}}\, z_m\right); \\
S^-_{t_1} &= S_{t_0} \exp\left(\left(r - q - \frac{\sigma^2}{2}\right)(t_1 - t_0) - \sigma\sqrt{t_1 - t_0}\, z_1\right), \\
&\vdots \\
S^-_{t_m} &= S^-_{t_{m-1}} \exp\left(\left(r - q - \frac{\sigma^2}{2}\right)(t_m - t_{m-1}) - \sigma\sqrt{t_m - t_{m-1}}\, z_m\right).
\end{aligned}
$$

The antithetic variables estimator of the arithmetic Asian option value is

$$
\frac{\max\left(\frac{\sum_{i=1}^m S^+_{t_i}}{m} - X, 0\right) + \max\left(\frac{\sum_{i=1}^m S^-_{t_i}}{m} - X, 0\right)}{2}.
$$

Control variate method

- The control variate method is applicable when there are two similar options, A and B. Option A is the one whose price is desired, while option B is similar to option A in nature but its analytic price formula is available.

- Let V_A and V_B denote the true value of option A and option B respectively, and let $\widehat{V}_A$ and $\widehat{V}_B$ denote the respective estimated value of option A and option B using the Monte Carlo simulation.

- How does the knowledge of V_B and $\widehat{V}_B$ help improve the estimate of the value of option A beyond the available estimate $\widehat{V}_A$?

- The control variate method exploits information about the errors in estimates of known quantities to reduce the error of an estimate of an unknown quantity.

The error $V_B - \widehat{V}_B$ in the estimation of known quantity is used as a control in the estimation of V_A. Based on the observation $V_A - \widehat{V}_A \approx V_B - \widehat{V}_B$, we define the control variate $\widehat{V}_A^{cv}$ by adding $V_B - \widehat{V}_B$ to $\widehat{V}_A$, where

$$\widehat{V}_A^{cv} = \widehat{V}_A + (V_B - \widehat{V}_B).$$

We obtain the following relation between the variances of the above quantities

$$\text{var}\left(\widehat{V}_A^{cv}\right) = \text{ var}(\widehat{V}_A) + \text{ var}(\widehat{V}_B) - \text{ 2 cov}(\widehat{V}_A, \widehat{V}_B),$$

so that

$$\text{var}(\widehat{V}_A^{cv}) < \text{ var}(\widehat{V}_A) \text{ provided that var}(\widehat{V}_B) < 2 \text{ cov}(\widehat{V}_A, \widehat{V}_B).$$

- The control variate technique reduces the variance of the estimator of V_A when the covariance between $\widehat{V}_A$ and $\widehat{V}_B$ is large. This is true when the two options are strongly correlated.

- In terms of computational efforts, we need to compute two estimates $\widehat{V}_A$ and $\widehat{V}_B$.

- However, if the underlying asset price paths of the two options are identical, then there is only slight additional work to evaluate $\widehat{V}_B$ along with $\widehat{V}_A$ on the same set of simulated price paths.

- To facilitate the more optimal use of the control $V_B - \widehat{V}_B$, we define the control variate estimate to be

$$\widehat{V}_A^{\beta} = \widehat{V}_A + \beta(V_B - \widehat{V}_B),$$

where β is a relaxation parameter with value other than 1.

- The new relation between the variances is now given by

$$\text{var}\left(\widehat{V}_A^{\beta}\right) = \text{var}\left(\widehat{V}_A\right) + \beta^2 \text{ var}(\widehat{V}_B) - 2\beta \text{ cov}\left(\widehat{V}_A, \widehat{V}_B\right).$$

- The particular choice of β which minimizes $\text{var}(\widehat{V}_A^{\beta})$ is found to be

$$\beta^* = \frac{\text{cov}(\widehat{V}_A, \widehat{V}_B)}{\text{var}(\widehat{V}_B)}.$$

Note that when $\widehat{V}_A$ and $\widehat{V}_B$ are equal in distribution, then $\beta^* = 1$ as expected.

Remarks

- Note that

$$\begin{aligned}\text{var}(\widehat{V}_A^{\beta})\Big|_{\beta=\beta^*} &= \text{var}(\widehat{V}_A) + \frac{\text{cov}(\widetilde{V}_A, \widehat{V}_B)^2}{\text{var}(\widehat{V}_B)} - 2\frac{\text{cov}(\widehat{V}_A, \widehat{V}_B)^2}{\text{var}(\widehat{V}_B)} \\ &= \text{var}(\widehat{V}_A) - \frac{\text{cov}(\widehat{V}_A, \widehat{V}_B)^2}{\text{var}(\widehat{V}_B)}.\end{aligned}$$

  Unlike the choice of $\beta = 1$, the control variate estimate based on β^* is guaranteed to decrease variance.

- Unfortunately, the determination of β^* requires the knowledge of $\text{cov}(\widehat{V}_A, \widehat{V}_B)$, which is in general not available.

- One may estimate β^* by computing sample variance and sample convariance from the simulated option values $V_A^{(i)}$ and $V_B^{(i)}, i = 1, 2, \cdots, M$, obtained from the simulation runs.

Valuation of Asian options

- Estimation of the value of an arithmetic averaging Asian option based on the knowledge of the exact analytic formula of the corresponding geometric averaging Asian option.

- The averaging feature in the Asian options does not pose any difficulty in Monte Carlo simulation since the average of the asset prices at different observational instants in a given simulated path can be computed easily.

- Since option price formulas are readily available for the majority of geometrically averaged Asian options, the knowledge of which may be used to include a variance reduction procedure to reduce the confidence interval in the Monte Carlo simulation performed for valuation of the corresponding arithmetically averaged Asian options.

- Let V_A denote the exact price of an option whose payoff depends on the arithmetic averaging of the underlying asset price and V_G be the exact price of an option similar to the above option except that geometric averaging is taken.

- Let $\widehat{V}_A$ and $\widehat{V}_G$ denote the estimates of the option values with respect to arithmetic and geometric averaging, respectively. Recall that

$$V_A = E[\widehat{V}_A] \quad \text{and} \quad V_G = E[\widehat{V}_G].$$

  An unbiased estimator of V_A is given by

$$\widehat{V}_A^{cv} = \widehat{V}_A + (V_G - \widehat{V}_G)$$

  since the expected value of the estimator $\widehat{V}_A^{cv}$ equals the time value V_A, where

$$E[\widetilde{V}_A^{cv}] = E[\widetilde{V}_A] + V_G - E[\widetilde{V}_G] = V_A.$$

Importance sampling methods

- Importance sampling is based on using knowledge of the pricing problem to focus our sampling on critical areas of interest, values of the underlying variable which yield important results. This would achieve the goal of reducing the variance of the estimator.

- One obvious example of where this can be applied is pricing out-of-the-money options, where payoffs only occur at very large (or small) values of the underlier, and so sampling in these regions makes our estimator with less variation.

- Another example is the estimation of the probability of meltdown of a nuclear power plant. Since the occurrence of the rare event in simulation runs is rare, naive Monte Carlo simulation would require a large number of simulation runs in order to generate enough samples of meltdown.

Suppose we are interested in computing

$$\theta = \mathbb{E}_f[h(X)],$$

where X has a probability distribution function f. Let g be another probability distribution function with the property that $g(x) \neq 0$ whenever $f(x) \neq 0$. That is, g has the same support as f.

The original simulation method is to generate N samples of X from the density f and set the estimate of θ to be $\widehat{\theta}_N = \dfrac{1}{N} \displaystyle\sum_{j=1}^{N} h(X_j)$.

We performed simulation of the Monte Carlo paths using a different distribution (with a change of measure) that will give more likelihood for the simulated underlier to be located in the area of interest.

Consider the expectation calculation of θ under the two distributions f and g

$$\begin{aligned}\theta &= \mathbb{E}_f[h(X)] = \int h(x)f(x)\ dx \\ &= \int h(x)\frac{f(x)}{g(x)}g(x)\ dx = \mathbb{E}_g\left[h(X)\frac{f(X)}{g(X)}\right] = \mathbb{E}_g[h^*(X)],\end{aligned}$$

where $\dfrac{f(X)}{g(X)}$ is known as the likelihood ratio and should be chosen to be easily computable.

In the alternative estimation algorithm, however, we generate N values from $g(.)$ and set

$$\widehat{\theta}_{N,IS} = \frac{1}{N}\sum_{j=1}^{N} h(X_j)\frac{f(X_j)}{g(X_j)} = \frac{1}{N}\sum_{j=1}^{N} h^*(X_j).$$

Here, $\widehat{\theta}_{N,IS}$ is then an importance sampling estimator of θ.

Variance reduction via importance sampling

$\text{var}(\widehat{\theta}_{N,IS}) < \text{var}(\widehat{\theta})$ if and only if $\text{var}_g(h^*(X)) < \text{var}_f(h(X))$, where

$$\begin{aligned} \text{var}_g(h^*(X)) &= \int h^*(x)^2 g(x)\ dx - \theta^2 \\ &= \int \frac{h^2(x)f(x)}{g(x)} f(x)\ dx - \theta^2 \end{aligned}$$

and

$$\text{var}_f(h(X)) = \int h^2(x)f(x)\ dx - \theta^2.$$

The difference of these two variances is

$$\text{var}_f(h(X)) - \text{var}_g(h^*(X)) = \int h^2(x) \left[1 - \frac{f(x)}{g(x)}\right] f(x)\ dx.$$

In order to achieve a variance reduction, the above integral should be positive. For this to happen, we should choose

$\frac{f(x)}{g(x)} > 1$ where $h(x)f(x)$ is small, $\frac{f(x)}{g(x)} < 1$ where $h(x)f(x)$ is large.

Let us say that there is a region where $h(x)f(x)$ is large. We would like to choose g so that $f(x)/g(x)$ is small whenever x is in the region, that is, we would like a density g that puts more weight in that region. When h involves a rare event so that $h(x) \approx 0$ over most of the state space, it can then be particularly valuable to choose g so that we often sample from that part of the state space where $h(x)$ has non-negligible value.

Note that if we choose g to be

$$g(x) = h(x)f(x)/\theta$$

then we have $\text{var}_g(h^*(X)) = \theta^2 - \theta^2 = 0$. In practice, θ is not known and actually it is the quantity that we are trying to calculate. However, it tells us the closer g is to the shape of $h(x)f(x)$, the lesser the variance.

Estimating a rare normal event using importance sampling

Consider the problem of estimating

$$\theta = P[X \geq 8],$$

where $X \sim \mathcal{N}(0, 1)$. If one tries to estimate θ via simulation without doing importance sampling, we will often get zero as this event is extremely rare.

We often take g to be from the same family of distributions as f. Therefore, we try to estimate θ by doing importance sampling with a new distribution $\sim \mathcal{N}(\mu, 1)$ with some appropriate choice for μ.

Consider

$$\begin{aligned}
\theta &= P[X > 8] = \mathbb{E}_f[\mathbf{1}_{X>8}] = \int_{-\infty}^{\infty} \mathbf{1}_{z>8} \frac{1}{\sqrt{2\pi}}\ e^{-z^2/2}\ dz \\
&= \int_{-\infty}^{\infty} \mathbf{1}_{z>8} \frac{\frac{1}{\sqrt{2\pi}} e^{-z^2/2}}{\frac{1}{\sqrt{2\pi}} e^{-(z-\mu)^2/2}} \frac{1}{\sqrt{2\pi}}\ e^{-(z-\mu)^2/2}\ dz \\
&= \int_{-\infty}^{\infty} \mathbf{1}_{z>8} e^{-\mu z + \mu^2/2} \frac{1}{\sqrt{2\pi}}\ e^{-(z-\mu)^2/2}\ dz \\
&= \mathbb{E}_g[\mathbf{1}_{X>8} e^{-\mu X + \mu^2/2}].
\end{aligned}$$

Here, $g(.)$ is the density function of $\mathcal{N}(\mu, 1)$, where

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2}.$$

For a given μ, the likelihood ratio $f(x)/g(x)$ is $e^{-\mu x + \mu^2/2}$. This may be visualized as the Radon-Nikodym derivative that effects the change of measure.

Recall that the closer g to the shape of $h(x)f(x)$, the lesser is the variance. It is desirable to choose g so that $g(x)$ and $h(x)f(x)$ both take on their maximum values at the same value, say x^*. It is clear that $g(.)$ attains its maximum at $x = \mu$. To match the two values of x that give maximum value for $g(x)$ and $h(x)f(x)$, we set

$$\begin{aligned}
\mu &= \arg\max_x h(x)f(x) \\
&= \arg\max_x \mathbf{1}_{X \geq 8} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \\
&= \arg\max_{x \geq 8} e^{-x^2/2} = 8.
\end{aligned}$$

The probability of occurrence of the event $\{X > 8\}$ is much higher when we perform sampling from $\mathcal{N}(8, 1)$ instead of $\mathcal{N}(0, 1)$.

Example

Suppose we have $X \sim N(0,1)$ and one faces huge cost of $h(X)$ when $X > 10$. This may be extremely unlikely, which is almost similar to the default of the United States Treasury bills or a serious accident in a nuclear power plant.

Suppose we now use a crude Monte Carlo estimator, even for a very high number of N, we would typically not observe a single value of X_i exceeding 10 and thus estimate the mean costs $E[h(x)]$ to be zero. If we use

$$h(x) = Cx\mathbf{1}_{[10,\infty)}(x)$$

with C a typically very large constant, then it is easy to verify that we have

$$10 = \arg\max_x \left\{ Cx\mathbf{1}_{[10,\infty)}(x)\frac{1}{\sqrt{2\pi}}\exp(-x^2/2) \right\}.$$

We use

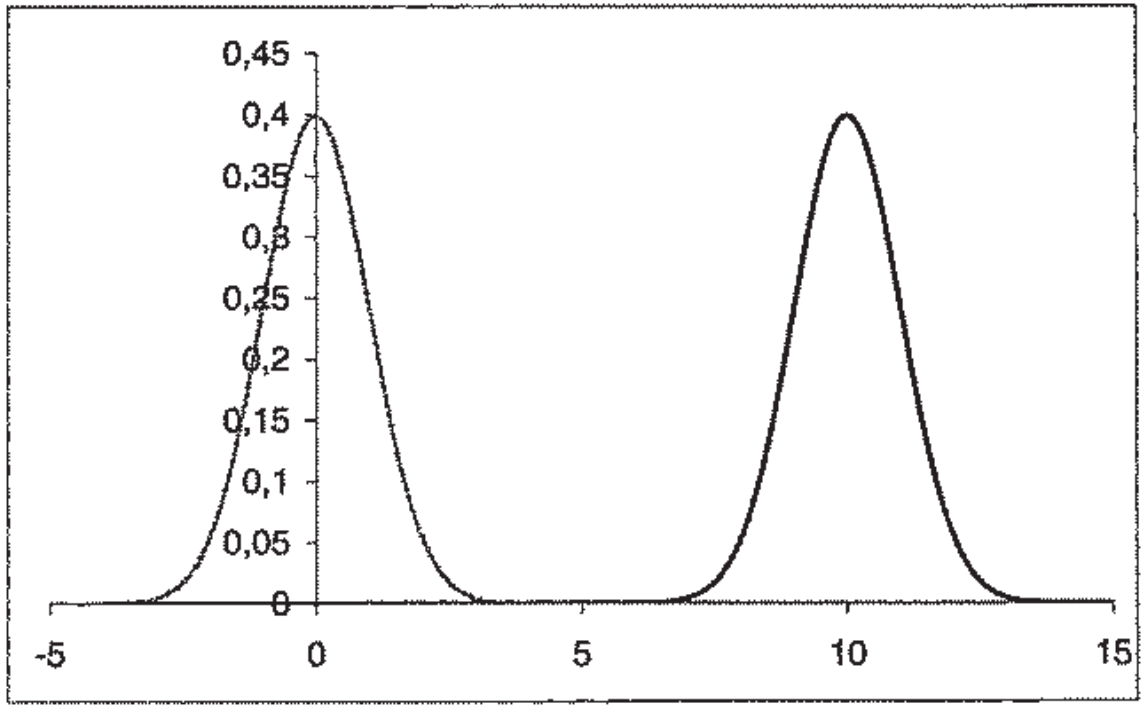
$$g(x) = \frac{1}{\sqrt{2\pi}} \exp(-(x-10)^2/2)$$

which leads to the importance sampling estimator of

$$\bar{I}_{imp,\bar{f},N}(h(X)) = \frac{1}{N} \sum_{i=1}^{N} CX_i \mathbf{1}_{X_i \geq 10} \exp(50 - 10X_i),$$

with all independent $X_i \sim N(10, 1)$.

With $N = 10,000$ and $C = 10^9$, we obtained an estimate of $7.530 \cdot 10^{-14}$ with an approximate 95%-confidence interval of $[7.029 \cdot 10^{-14}, 8.031 \cdot 10^{-14}]$. Compare this to the exact value of $C \cdot \exp(-50)/\sqrt{2\pi} = 7.695 \cdot 10^{-14}$.



Original density $f(x)$ (grey) and shifted importance sampling density $g(x)$ (black).

This shift yields sampling values in the area of importance for calculating the expectation while the likelihood ratio function assigns these samples their probability weights. In the crude Monte Carlo method this has already been done before the sampling which results in (nearly) no samples in the region of interest.

5.3 Valuation of American options

The apparent difficulties of using simulation to price American options stem from the *backward* nature of the early exercise feature since there is no way of knowing whether early exercise is optimal when a particular asset value is reached at a given time. Only with a pre-specified exercise policy, the estimated option value with respect to a given simulated path can be determined.

Backward induction procedure and dynamic programming procedure in lattice tree algorithms

Lattice tree algorithms work well in American option price calculations since they employ the backward induction procedure. The continuation value (obtained via discounted expectation procedure) can be found at each node under the backward induction procedure. The dynamic programming procedure considers taking the maximum value among exercise value and continuation value. The optimality taken at the current node is independent of the optimal decisions at earlier time points.

Parametrization of the early exercise boundary: Grant-Vora-Weeks algorithm

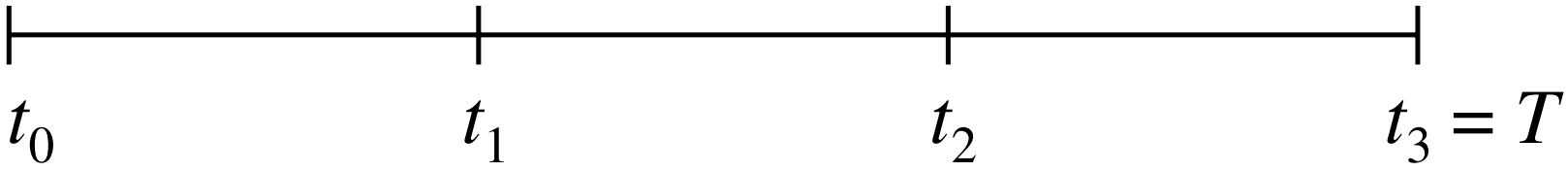
We identify the optimal exercise price $S^*_{t_i}$ at selected time instants $t_i, i = 1, 2, \cdots, N-1$ between the current t and expiration time T.

The determination of the optimal exercise prices is done by simulation at successive time steps proceeding backwards in time, starting at t_{N-1}, then $t_{N-2}, \cdots, t_1$, successively.

Once the optimal exercise boundary is identified, the option value can be estimated by the usual simulation procedure, respecting the early exercise strategy as dictated by the known optimal exercise boundary.

We illustrate the procedure by considering the valuation of an American put option and choosing only three time steps between the current time t and expiration time T, where $t_0 = t$ and $t_3 = T$.

Characterization of critical exercise prices



$$S^*(t_1) < S^*(t_2) < S^*(T^-) = \min\left(\frac{r}{q}X, X\right).$$

The American put must be deeper-in-the-money in order to induce early exercise of a longer-lived American put. In particular, when q is large, the American put must be deeper in-the-money to induce early exercise.

- Early exercise is a tradeoff between the gain on the time value of the strike price received earlier and the losses on the dividends from holding the underlying stock and the insurance value associated with holding of the American put.

Assuming a constant dividend yield q, the optimal exercise price at T is equal to $\min\left(\frac{r}{q}X, X\right)$, where X is the strike price of the put option and r is the riskless interest rate. At time δt from expiry, the insurance value associated with holding the option is negligible, early exercise occurs when the stock price S satisfies

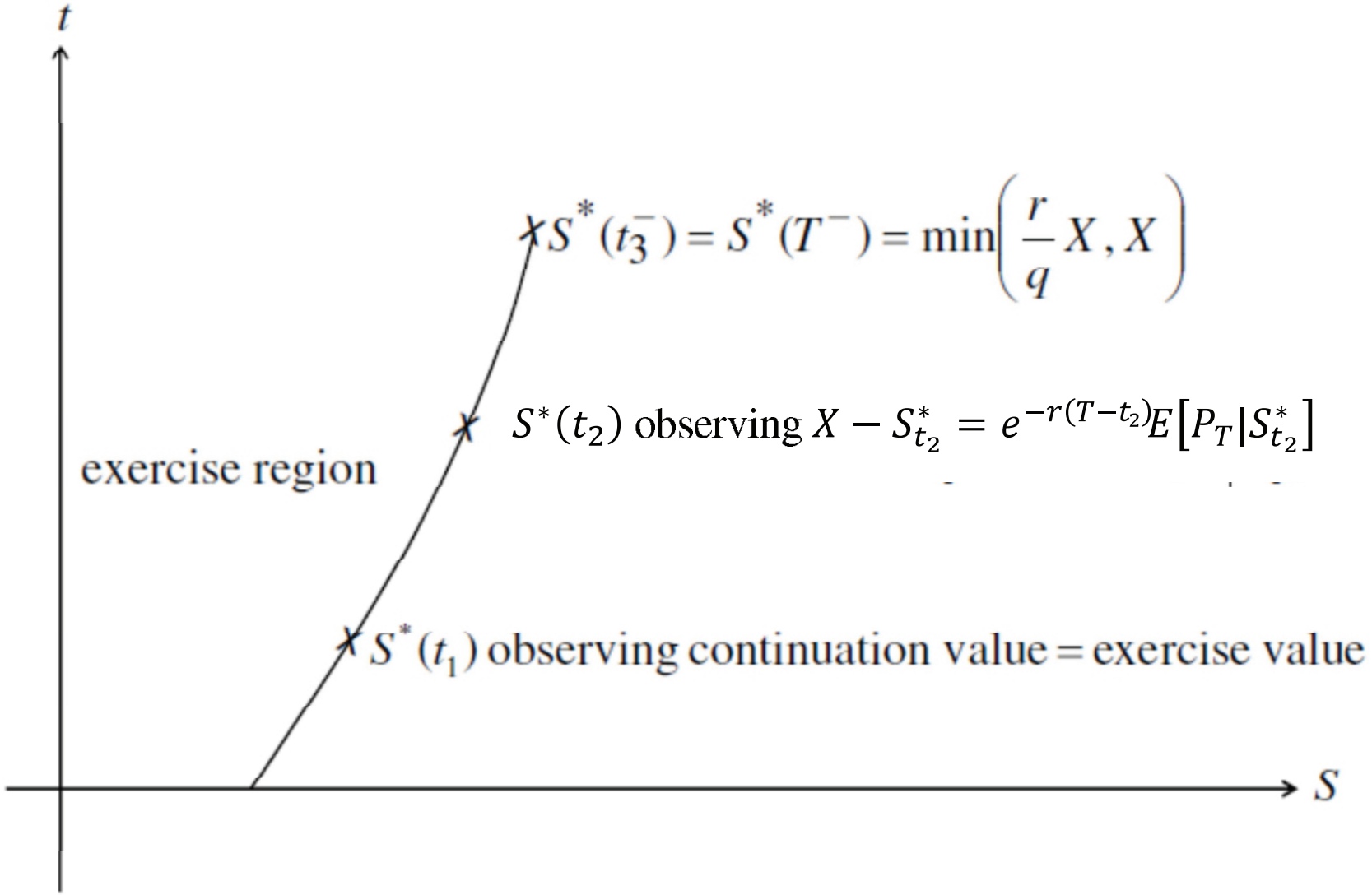
$$rX\delta t - qS\delta t \geq 0 \Leftrightarrow S \leq \frac{rX}{q}.$$

Note that $S^*(T^-)$ must be bounded above by X, otherwise the exercise payoff becomes negative. Taking these two conditions together, we have

$$S^*(T^-) = \min\left(\frac{r}{q}X, X\right).$$

At time t_2 which is one time period prior to expiration, the put value is $X - S_{t_2}$ when $S_{t_2} \leq S^*_{t_2}$, and $e^{-r(T-t_2)}E[P_T|S_{t_2}]$ when $S_{t_2} > S^*_{t_2}$. Here, $P_T = \max(X - S_T, 0)$ denotes the put option value at expiration time T. Obviously, the conditional expectation $E[P_T|S_{t_2}]$ is dependent on S_{t_2}. For a given value of S_{t_2}, one can perform a sufficient number of simulations to estimate $E[P_T|S_{t_2}]$.

Determination of the optimal exercise boundary of an American put at discrete time instants



Systematic search for $S^*_{t_2}$

- The optimal exercise price $S^*_{t_2}$ is identified by finding the appropriate value of S_{t_2} such that

$$X - S^*_{t_2} = e^{-r(T-t_2)} E[P_T | S^*_{t_2}].$$

- We find the simulation estimate of $e^{-r(T-t_2)} E[P_T | S_{t_2}]$ as a function of S_{t_2} by starting with S_{t_2} close to but smaller than $S^*_{t_3}$ (remark: $S^*_{t_3}$ is known and $S^*_{t_2}$ must be less that $S^*_{t_3}$) and repeat the simulation process for a series of S_{t_2} which decreases systematically. Note that

$$X - S_{t_2} < e^{-r(T-t_2)} E[P_T | S_{t_2}] \quad \text{when} \quad S_{t_2} > S^*_{t_2}.$$

- Once the functional dependence of the discounted expectation value $e^{-r(T-t_2)} E[P_T | S_{t_2}]$ in S_{t_2} is available, one can find a good estimate of $S^*_{t_2}$ such that the above equation is satisfied.

- Estimate $E[P_T|S_{t_2}]$ for varying S_{t_2}.

  Determine the critical value $S^*_{t_2}$ such that

$$X - S^*_{t_2} = e^{-r(T-t_2)} E[P_T|S^*_{t_2}].$$

- Estimate the continuation value for a given value of S_{t_1}, respecting $S^*_{t_2}$ obtained in the first step. Find the critical value $S^*_{t_1}$, such that continuation value equals exercise value.

- Once the optimal exercise prices at t_1 and t_2 are available, one can mimic the above numerical procedure to find the estimate of the discounted expectation value of holding the put at time t_0 by performing simulation runs with an initial asset value S_{t_0}. The put value at time t_0 for a given S_{t_0} is the maximum of the estimate of the discounted expectation value obtained from simulation (taking into account the early exercise strategy as already determined at t_1 and t_2) and the intrinsic value $X - S_{t_0}$ from early exercise.

Regression method via basis functions

- Under the discrete assumption of exercise opportunities, the option values satisfy the following dynamic programming equations

$$V_n = \max(h_n(S), H_n(S)), \quad n = 0, 1, \cdots, N-1,$$

where $S = S(t_n), H_n(S)$ is the continuation value at time t_n and $h_n(S)$ is the exercise payoff. At maturity date $t_N = T$, we have $V_N(S) = h_N(S)$ [for notational convenience, we set $H_N(S) = 0$].

- The difficulty of estimating the above conditional expectations may be resolved by considering an approximation of the continuation value $H_n(S)$ in the form

$$H_n(S) \approx \sum_{m=0}^{M} \alpha_{nm}\phi_{nm}(S),$$

for some choice of the basis functions $\phi_{nm}(S)$.

Least squares projection onto the span of basis functions

According to functional analysis, a function can be expanded in an infinite sum of basis functions, provided that the function is defined in the corresponding spanning space of the basis functions. An example is the infinite series representation for Fourier integrable functions.

Longstaff and Schwartz propose to determine the coefficients α_{nm} through the least squares projection onto the span of basis functions. Their chosen basis functions are the Laguerre polynomials defined by

$$L_m(S) = e^{-S/2}\frac{e^S}{m!}\frac{d^m}{dS^m}\left(S^m e^{-S}\right), \quad m = 0, 1, 2, \cdots .$$

The first few members are:

$$L_0(S) = e^{-S/2}, L_1(S) = e^{-S/2}(1-S), L_2(S) = e^{-S/2}\left(1 - 2S + \frac{S^2}{2}\right).$$

Pathwise approximation to the optimal stopping rule

We use $C(\omega, s; t, T)$ to denote the path of cash flows generated by the option, conditional on the option not being exercised at or prior to time t. The holder is assumed to follow the optimal stopping strategy for all subsequent time points. Here, s is the optimal stopping time within t and T, and ω indicates the randomness associated with the simulated asset price path. Recall that the value of an American option is given by maximizing the discounted cash flows from the option, where the maximum is taken over all stopping times.

- We seek for a pathwise approximation to the optimal stopping rule based on the simulated paths associated with the early exercise right in the American option.

- Like other simulation algorithms, the key is to identity the conditional expected value of continuation.

- Suppose we have chosen M basis functions, then $H_n(\omega)$ is estimated by regressing the discounted cash flow onto the basis functions for the *paths where the option is in-the-money at time* t_n. Only the in-the-money paths are used since one can better estimate the conditional expectation in the region where exercise is relevant.

- Once the functional form of the estimated continuation value $\widehat{H}_n(\omega)$ is obtained from linear regression, we can calculate the estimated continuation value from the known asset price at time t_n for that random path ω.

- When the cash flows received by the option holder for all paths are identified, we can compute an estimate of the option value by discounting each cash flow back to the issue date and averaging over all sample price paths.

Example

- Consider a 3-year American (actually Bermudan) put option on a non-dividend paying asset with strike price 1.1. The put can be exercised only at $t = 1, 2, 3$.
- We take the riskless interest rate to be 0.06. Only 8 sample paths of the asset price are generated under the risk neutral measure.

	Asset price paths				
Path	$t = 0$	$t = 1$	$t = 2$	$t = 3$	Cash flow at $t = 3$
1	1.00	1.09	1.08*	1.34	0.00
2	1.00	1.16	1.26	1.54	0.00
3	1.00	1.22	1.07*	1.03	0.07
4	1.00	0.93	0.97*	0.92	0.18
5	1.00	1.11	1.56	1.52	0.00
6	1.00	0.76	0.77*	0.90	0.20
7	1.00	0.92	0.84*	1.01	0.09
8	1.00	0.88	1.22	1.34	0.00

* Sample path for which the put is in-the-money at $t = 2$.

Note that there are 5 paths for which the put is in-the-money at $t = 2$.

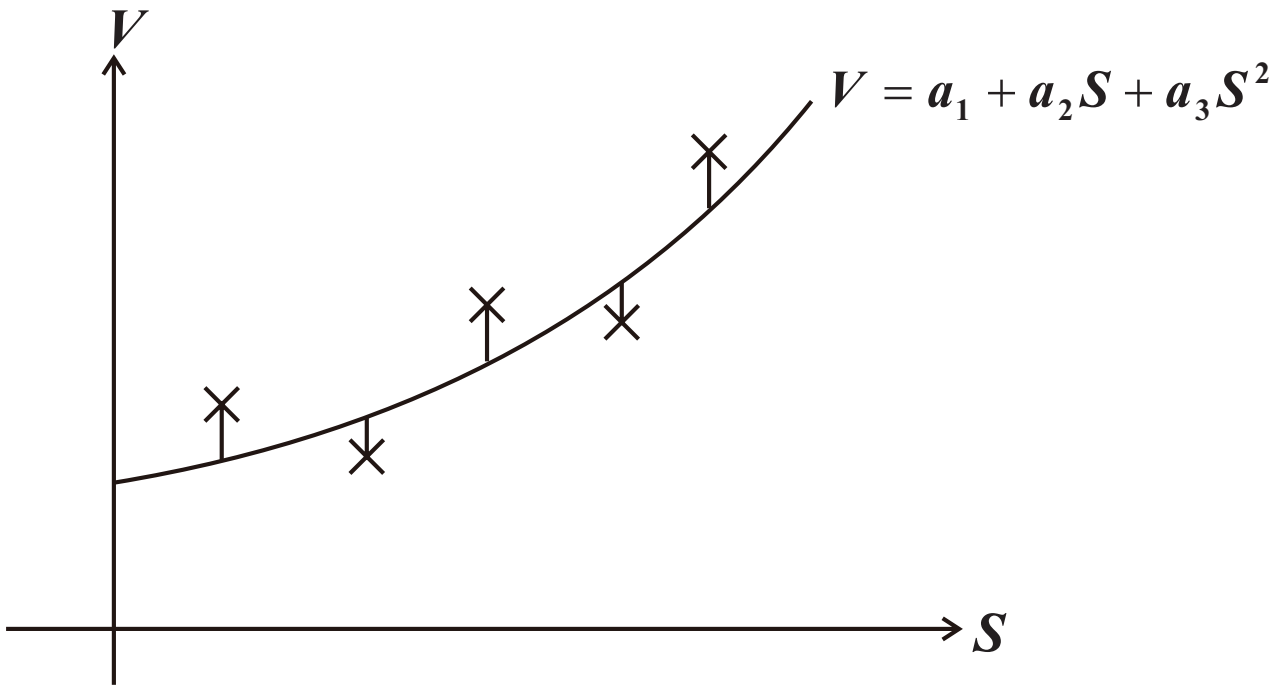
- For the 5 paths that are in-the-money at $t = 2$, we compute the corresponding discounted cash flows received at $t = 3$ if the put is not exercised at $t = 2$.

- Let X and Y denote, respectively, the asset price at $t = 2$ and the discounted cash flow conditional on no exercise at $t = 2$. The values of X and Y for those in-the-money asset price paths are listed below:

Path	Y	X	exercise value	continuation value
1	0.00×0.94176	1.08	0.02	0.0369
3	0.07×0.94176	1.07	0.03	0.0461
4	0.18×0.94176	0.97	0.13	0.1176
6	0.20×0.94176	0.77	0.33	0.1520
7	0.09×0.94176	0.84	0.26	0.1565

  * The discount factor is $e^{-0.06} = 0.94176$.

  The 5 data points of (X, Y) in the above table provide 5 observations of V and S from simulation. The conditional expectation values of continuation are computed based on the least squares fitting procedure outlined below.



We fit the observations of V_i and S_i, $i = 1, 3, 4, 6, 7$, of the 5 in-the-money paths by least squares regression using the basis functions: 1, S and S^2. Here, V is the estimate of the continuation value function in S at time t_2. We minimize the sum of squares of the errors.

We assume an approximate relationship:

$$V = a_1 + a_2 S + a_3 S^2$$

where S is the stock price at the 2-year point and V is the value of continuing, discounted back to the 2-year point. Our five observations on S are $1.08, 1.07, 0.97, 0.77$ and 0.84. The corresponding values for V are $0.00, 0.07e^{-006\times 1}, 0.18e^{-0.06\times 1}, 0.20e^{-0.06\times 1}$, and $0.09e^{-0.06\times 1}$. We use this data to calculate the values of a_1, a_2 and a_3 that minimize

$$\sum_{i=1}^{5} (V_i - a_1 - a_2 S_i - a_3 S_i^2)^2$$

where S_i and V_i are the i^{th} observation on S and V, respectively. It turns out that $a_1 = -1.070, a_2 = 2.983$ and $a_3 = -1.813$, so that the best-fit relationship is

$$V = -1.070 + 2.983S - 1.813S^2.$$

We estimate the corresponding continuation values at $t = 2$ of various simulation paths that are in-the-money at that time instant. This gives the value at the 2-year point of continuing for Paths $1, 3, 4, 6$, and 7 of $0.0369, 0.0461, 0.1176, 0.1520$, and 0.1565, respectively.

Dynamic programming procedure

- For Path 1 where $X = 1.08$, the immediate exercise value equals $1.10 - 1.08 = 0.02$ while the continuation value is

$$-1.070 + 2.983 \times 1.08 - 1.813 \times 1.08^2 = 0.0369.$$

  Since the continuation value is higher, it is not optimal to exercise the put at $t = 2$ for the first path. Note that Path 1 ends up to expire out-of-the-money at maturity. Therefore, the corresponding cash flow received by the option holder at $t = 2$ for Path 1 conditional on not exercising prior to $t = 2$ is zero.
- For Path 4, since the exercise value is higher than the continuation value, the cash flow for this path at $t = 2$ is set equal to the exercise value. One can check that it is also optimal to exercise at $t = 2$ for Paths 6 and 7.

Conditional on not exercising prior to $t = 2$, the cash flows received at $t = 2$ and $t = 3$ for the 8 simulated price paths are summarized in the following table:

Path	$t = 1$	$t = 2$	$t = 3$
1	—	0.00	0.00
2	—	0.00	0.00
3	—	0.00	0.07
4	—	0.13	0.00
5	—	0.00	0.00
6	—	0.33	0.00
7	—	0.26	0.00
8	—	0.00	0.00

Note that it is optimal to exercise the put at $t = 2$ for Paths $4, 6$ and 7. Once the option has been exercised at $t = 2$, the cash flow at $t = 3$ becomes zero.

Determination of the stopping rule at $t = 1$

- We proceed recursively to determine the stopping rule at $t = 1$. There are 5 paths (Paths $1, 4, 6, 7$ and 8) for which the put is in-the-money at $t = 1$.

- Similarly, we solve for the estimated expectation function at $t = 1$ by regressing the discounted value of subsequent option cash flow at $t = 1$ on a constant, X, and X^2, where X is the asset price at $t = 1$.

- We can compute the estimated continuation values and immediate exercise values at $t = 1$.

We compute Y at different in-the-money paths by using the cash flow information earlier based on not exercising prior to $t = 2$.

Path	Y	X	exercise value	continuation value
1	0.00×0.94176^2	1.09	0.01	0.0139
4	0.13×0.94176	0.93	0.17	0.1092
6	0.33×0.94176	0.76	0.34	0.2866
7	0.26×0.94176	0.92	0.18	0.1175
8	0.00×0.94176^2	0.88	0.22	0.1533

* The estimated conditional expectation function at $t = 1$ from least squares regression is

$$E[Y|X] = 2.038 - 3.335X + 1.356X^2.$$

Note that Path 6 would give early exercise at $t = 2$ with payoff value of 0.33 if it remains alive at $t = 1$. However, since its exercise value of 0.34 is higher than its continuation value, so it is optimal to exercise at $t = 1$ for Path 6.

Note that exercise at $t = 1$ is optimal for Paths $4, 6, 7$ and 8. The optimal stopping rules at all times are now identified.

	Stopping rule			Option cash flow matrix		
Path	$t=1$	$t=2$	$t=3$	$t=1$	$t=2$	$t=3$
1	0	0	0	0.00	0.00	0.00
2	0	0	0	0.00	0.00	0.00
3	0	0	1	0.00	0.00	0.07
4	1	0	0	0.17	0.00	0.00
5	0	0	0	0.00	0.00	0.00
6	1	0	0	0.34	0.00	0.00
7	1	0	0	0.18	0.00	0.00
8	1	0	0	0.22	0.00	0.00

* "1" represents exercise optimally at the exercise date.

- Once optimal exercise for a given path has been chosen at an earlier time, the stopping rules that have been obtained for later times in the backward induction procedure becomes immaterial. In this example, Paths $4, 6, 7$ have positive cashflow at $t = 2$ if exercise is possible only at $t = 2$ and $t = 3$. However, when exercise instant $t = 1$ is included, optimal exercise at $t = 1$ occurs for these paths. Consequently, these paths give zero cash flow at all subsequent times (corresponding to $t = 2$ and $t = 3$).

- When the cash flows generated by the put option at each time along each path have been identified, the put option value can be computed by discounting each cash flow back to the current time, and taking average value over all sample paths. This gives

$$\frac{1}{8}(0.07e^{-0.06\times 3} + 0.17e^{-0.06\times 1} + 0.34e^{-0.06\times 1} + 0.18e^{-0.06\times 1} + 0.22e^{-0.06\times 1}) = 0.1144.$$

  Note that Paths $1, 2$ and 5 survive up to maturity date and expire out-of-the-money at expiry.

Summary

1. Generate all M simulation asset price paths up to maturity T.

2. At t_{N-1}, seek the m_{N-1} paths that are in-the-money. Perform the regression procedure to determine the best-fit relationship for the continuation value V in terms of the stock price S such that

$$\sum_{i=1}^{m_{N-1}} \left[V_i - \sum_{j=1}^{n} a_j \phi_j(S) \right]^2$$

is minimized. Here, $\phi_j(S)$, $j = 1, 2, \ldots, n$ are the basis functions. Also, V_i and S_i are the observation of V and S for the i^{th} observation among the in-the-money paths. The coefficients $a_1, a_2, \ldots, a_n$ are determined using the least squares fitting procedure.

3. Find the continuation value at a given stock price at t_{N-1} for all the m_{N-1} paths using the best-fit relationship for $V = \sum_{j=1}^{n} a_j \phi_j(S)$ obtained in step 2. Compare the continuation value with the exercise value to determine the stopping rule at $t = t_{N-1}$.

4. At t_{N-2}, seek the m_{N-2} paths that are in-the-money. Compute V_i at these paths based on the discounted cash flow using the knowledge of the stopping rule obtained at t_{N-1}. Repeat the same procedure for the determination of the coefficients $a_1, a_2, \ldots, a_n$ using the least squares procedure. Again, determine the stopping rule at $t = t_{N-2}$ by comparing the continuation value and exercise payoff.

5. Repeat the same procedure until $t = t_1$ for the stopping rule.

6. Lastly, compute the sample average of the discounted cash flows from all M simulation paths, respecting all the stopping rules obtained at all time points.

Mathematical appendices

A.1 Law of Large numbers and central limit theorem

A.2 Generation of random numbers with assigned probability distributions

A.3 Numerical integration of stochastic differential equations

A.4 Quasi Monte Carlo simulation – low discrepancy sequences

A.1 Law of Large numbers and central limit theorem

Law of Large numbers

The average of the results obtained from a large number of trials should be close to the *expected value*. That is, the sample average

$$\overline{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

converges to the expected value. That is,

$$\overline{X}_n \to \mu \text{ as } n \to \infty,$$

where $X_1, X_2, \ldots$, is an infinite sequence of independent and identically distributed integrable random variables with $E[X_1] = E[X_2] = \cdots = \mu$. Integrability of X_j means that the expected value $E[X_j]$ exists and is finite.

Weak Law: The sample average converges in probability towards the expected value. That is

$$\overline{X}_n \xrightarrow{P} \mu \text{ as } n \to \infty;$$

or for any $\epsilon > 0$, we have

$$\lim_{n\to\infty} P[|\overline{X}_n - \mu| > \epsilon] = 0.$$

The weak Law states that for any non-zero tolerance ($\epsilon > 0$) specified, no matter how small, with a sufficiently large sample, there will be a very high probability that the average of the observations will be close to the expected value.

Convergence in probability is also called weak convergence of random variables.

Strong Law: The sample average converges almost surely to the expected value.

$$\overline{X}_n \xrightarrow{\text{a.s.}} \mu \text{ as } n \to \infty;$$

or

$$P[\lim_{n\to\infty} \overline{X}_n = \mu] = 1.$$

Random variables that converge strongly (almost surely) are guaranteed to converge weakly (in probability).

The strong Law states that the event $|\overline{X}_n - \mu| > \epsilon$ almost surely will not occur. With probability one, we have that for any $\epsilon > 0$, the inequality $|\overline{X}_n - \mu| < \epsilon$ holds for all sufficiently large value of n.

Central limit theorem

Let $\{X_1, X_2, \ldots, X_n\}$ be a random sample of size n, a sequence of independent and identically distributed random variables drawn from distributions of expected values given by μ and finite variances given by σ^2.

The classical central limit theorem describes the size and the distributional form of the stochastic fluctuations around the deterministic number μ during this convergence. As $n \to \infty$, the random variables $\sqrt{n}(\overline{X}_n - \mu)$ converge in distribution to $N(0, \sigma^2)$. That is,

$$\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right] \xrightarrow{d} N(0, \sigma^2).$$

In the case $\sigma > 0$, convergence in distribution means that the cumulative distribution functions of $\sqrt{n}(\overline{X}_n - \mu)$ converge pointwise to the cumulative distribution function of $N(0, \sigma^2)$ distribution. For every real number z, we have

$$\lim_{n \to \infty} P[\sqrt{n}(\overline{X}_n - \mu) \leq z] = N(z/\sigma).$$

Applications

Rolling a large number of identical and unbiased dice, the distribution of the sum (or average) of the rolled numbers will be well approximated by a normal distribution.

Why the use of the term "central limit theorem"?

The occurrence of the Gaussian density function in repeated experiments can be explained by the very same *limit theorem*, which plays a *central* role in the calculus of probability.

A.2 Generation of random numbers with assigned probability distributions

Uniform numbers generation

A congruential generator is a recursive formula returning a sequence of pseudor random numbers

1. Fix positive integers m (modulus), a (multiplier) and c (increment).
2. Set up a seed $x_0 \in \{0, 1, \ldots, m-1\}$
3. Run the recursive rule: $x_{i+1} = ax_i + c \pmod{m}$.
4. Return $u_{i+1} = \dfrac{x_{i+1}}{m} \in [0, 1]$.

Since any sample value x_i lies in the set $\{0, 1, \ldots, m-1\}$, the sequence would meet twice the same value in m stops. Typically, m is chosen to be very large, like 2^{35} or $2^{37} - 1$.

Stratified Sampling

To enhance uniformity of the generated numbers, we may divide $[0, 1]$ into M stratifying bins $\left[\frac{i}{M}, \frac{i+1}{M}\right]$, $i = 0, 1, \ldots, M-1$. We force the first sample to fall within $\left[0, \frac{1}{M}\right]$, the second sample into $\left[\frac{1}{M}, \frac{2}{M}\right]$, and so on until the M^{th} sample has been generated in the last subinterval $\left[\frac{M-1}{M}, 1\right]$. Then, the next number is generated within $\left[0, \frac{1}{M}\right]$, and so on. That is,

$$u_k \in \left[\frac{i}{M}(\text{mod } M), \frac{i+1}{M}(\text{mod } M)\right], \; k = 1, 2, \ldots, n.$$

If $n = kM$, then this method ensures that K samples fall into each interval $\left[\frac{i}{M}, \frac{i+1}{M}\right]$.

Pseudo-code for stratified sampling

```
h[1] = h[2] = ... = h[m] = 0;
M = number of stratifying bins; /* bin length=1/M */
n = number of samples /* of the form k*M, k = integer */;
m = number of histogram bins /* of form p*M, bin length=1/m */
for (j = 1, j <= k, j++){
      for (i = 1, i <= M, i++){
            v = Uniform[0,1];/* sampling a uniform in [0,1]*/
            u = (v+i-1)/M;/* zoom [0,1] into [(i-1)/M,i/M]*/
            b = integerPart[(u*m)+1];
            h[b]+=1;
      }
};
Plot[(h[i]/n)/(1/m)] over i = 1,...,m;
```

Inverse transformation method

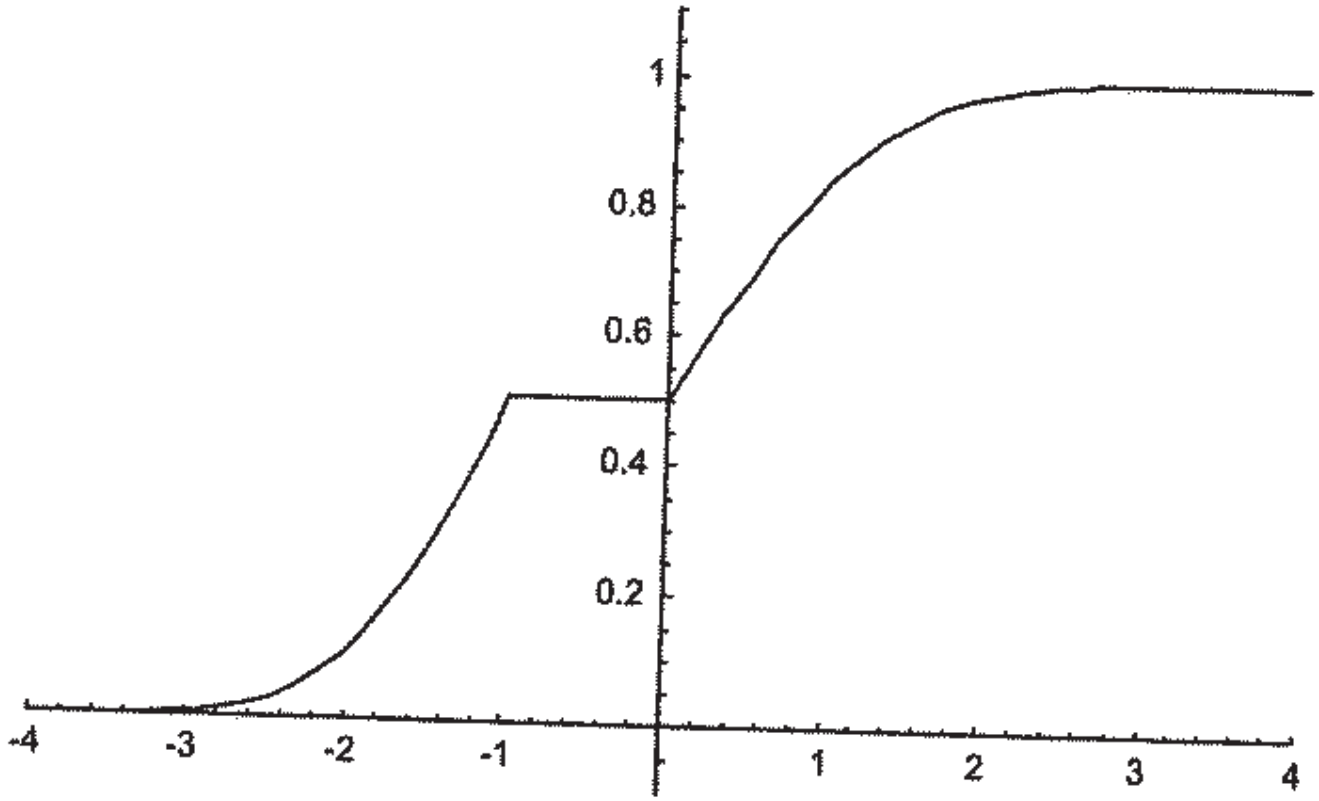
Given a uniform random variable U on $[0, 1]$, we seek a transformation f of U such that $f(U)$ has a cumulative distribution function given by F, that is,

$$P[f(U) \leq x] = F(x).$$

If f is bijective and monotonically increasing, the inverse function f^{-1} is well defined. We may write

$$P[f(U) \leq x] = P[U \leq f^{-1}(x)] = f^{-1}(x).$$

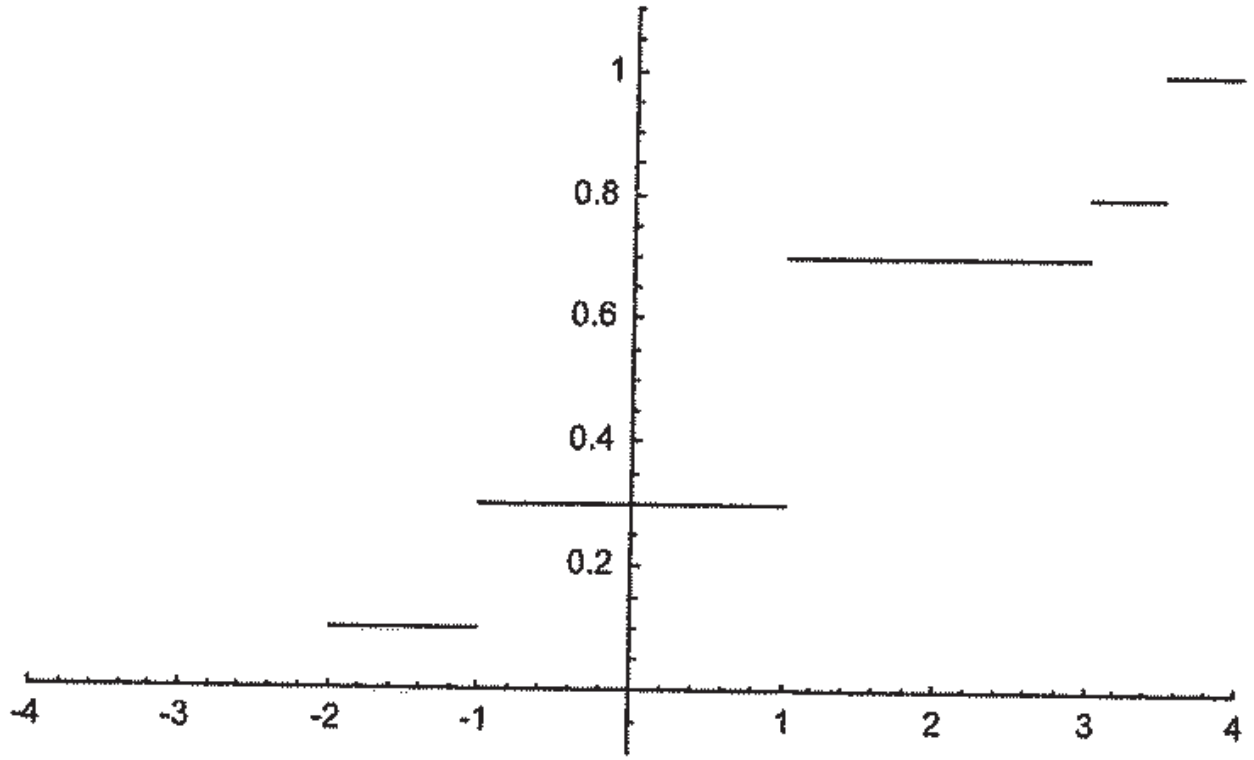
Comparing the two results, we set f such that $f^{-1} = F$. If F is continuous and strictly increasing, then F is bijective and $f = F^{-1}$ satisfies the required properties.



Suppose F is continuous but not injective, then we define the generalized inverse function

$$F^{-1}(y) = \min\{x : F(x) = y\};$$

which always exists since F is right-continuous.



Define $F^{-1}(y) = \min\{x : F(x) \geq y\}$, which always exists since F is right continuous. For any $u \in [0,1]$, $\{x : u \leq F(x) < F \circ F^{-1}(U)\}$ is always an empty set. We then have

$$\begin{aligned} P[F^{-1}(U) \leq x] &= P[F \circ F^{-1}(U) \leq F(x)] \\ &= P[F \circ F^{-1}(U) \leq F(x)] + P[U \leq F(x) < F \circ F^{-1}(U)] \\ &= P[U \leq F(x)] = F(x). \end{aligned}$$

Box-Muller method for generating normal variables

It takes uniformly distributed variables and turn them into normal. Let u_1 and u_2 be two uniform random numbers between zero and one. We combine them to give 2 numbers x_1 and x_2 that are close to be normally distributed (zero mean and unit variance):

$$x_1 = \sqrt{-2 \ln u_1} \cos(2\pi u_2) \quad \text{and} \quad x_2 = \sqrt{-2 \ln u_1} \sin(2\pi u_2).$$

Let X_1 and X_2 be independent and identically distributed standard normal random variable $N(0, 1)$ with joint pdf $f(x, y)$, where

$$f(x_1, x_2) = \frac{1}{2\pi} e^{-(x_1^2 + x_2^2)/2}.$$

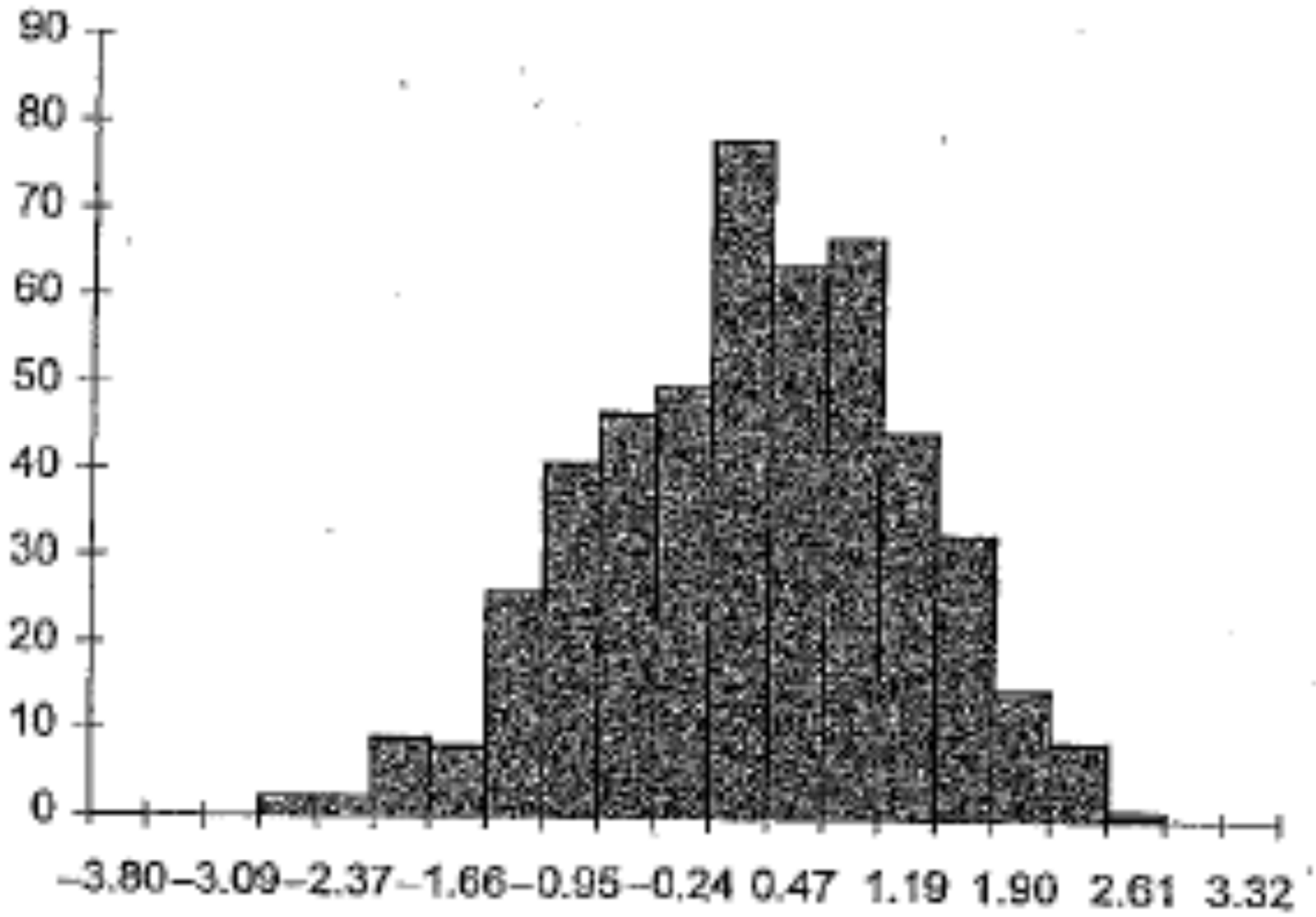
Define $X_1 = \sqrt{R}\cos\Ⓗ$ and $X_2 = \sqrt{R}\sin\Ⓗ$, then

$$f_{R,\Ⓗ}(r,\theta) = \frac{1}{2}\frac{e^{-r/2}}{2\pi} = f_R(r) f_{\Ⓗ}(\theta).$$

Here, we choose $f_R(r)$ to be the pdf of the exponential distribution while $f_{\Ⓗ}(\theta)$ to be the pdf of the uniform distribution on $[0, 2\pi]$. Suppose we generate U_1 and U_2 to be iid $U(0,1)$, then

$$X_1 = \sqrt{-2\ln U_1}\cos 2\pi U_2 \text{ and } X_2 = \sqrt{-2\ln U_1}\sin 2\pi U_2.$$

This is because $-2\ln U_1$ is a sample from an exponential with mean 2 and $2\pi U_2$ is a sample from $U(0, 2\pi)$.



The approximation to the Normal distribution using 500 uniformly distributed points and the Box-Muller method.

A.3 Numerical integration of stochastic differential equation

Consider the following generic one-dimensional stochastic differential equation (SDE):

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t,$$

where $W = \{W_t, 0 \leq t \leq T\}$ is a one-dimensional standard Wiener process, μ and σ are the measurable drift and diffusion coefficients, respectively.

The Ito-Taylor expansion of the SDE is

$$\begin{aligned} X_t &= X_{t_0} + \mu(X_{t_0}) \int_{t_0}^{t} ds + \sigma(X_{t_0}) \int_{t_0}^{t} dW(s) \\ &\quad + \frac{\sigma(X_{t_0})\sigma'(X_{t_0}, t_0)}{2}\{[W(t) - W(t_0)]^2 - (t - t_0)\} + R, \end{aligned}$$

where $\sigma'(x, t) = \dfrac{\partial \sigma}{\partial x}(x, t)$, R is the remainder of higher order.

After dropping R, the discretized version is

$$X_{t_{j+1}} = X_{t_j} + \mu(X_{t_j})\Delta t_j + \sigma(X_{t_j})\Delta W_j + \frac{\sigma(X_{t_j})\sigma'(X_{t_j})}{2}[(\Delta W_j)^2 - \Delta t_j],$$

where $\Delta t_j = t_{j+1} - t_j$ and $\Delta W_j = W(t_{j+1}) - W(t_j)$. Note that $\Delta W_j = O(\sqrt{t_j})$.

Euler Scheme

This is the simplest discretization scheme, expanding the diffusion term only up to $O(\sqrt{\Delta t})$. Let Z_j be iid $N(0,1)$, the explicit Euler scheme is

$$X_{t_{j+1}} = X_{t_j} + \mu(X_{t_j}, t_j)\Delta t_j + \sigma(X_{t_j}, t_j)\sqrt{\Delta t_j} Z_j.$$

Milstein Scheme

The Milstein scheme improves upon the Euler scheme by expanding the diffusion term to $O(\Delta t)$, which gives

$$\begin{aligned} X_{t_{j+1}} &= X_j + \mu(X_j, t_j)\Delta t_j + \sigma(X_j, t_j)\sqrt{\Delta t_j} Z_j \\ &\quad + \frac{\sigma(X_{t_j})\sigma'(X_{t_j})}{2}\Delta t_j (Z_j^2 - 1). \end{aligned}$$

The analytic form of the derivative of the volatility function is required.

When σ is constant, the Milstein scheme reduces to the Euler scheme.

Runge-Kutta scheme

We would like to avoid the inclusion of $\sigma'(X_{t_j})$ in the discretized scheme. By observing

$$\Delta X_i = \mu(X_i)\Delta t + \sigma(X_i)\Delta W_i,$$

we obtain

$$\begin{aligned}
\sigma(X_i + \Delta X_i) - \sigma(X_i) &= \sigma'(X_i)\Delta X_i + O((\Delta X)^2) \\
&= \sigma'(X_i)[\mu(X_i)\Delta t + \sigma(X_i)\Delta W_i] + O((\Delta X)^2) \\
&= \sigma'(X_i)\sigma(X_i)\Delta W_i + O(\Delta t) \quad \text{as } (\Delta X)^2 \sim O(\Delta t).
\end{aligned}$$

Performing the following Taylor expansions, we deduce that

$$\begin{aligned}
\sigma(X_i + \Delta X_i) = \sigma(&X_i + \mu(X_i)\Delta t + \sigma(X_i)\sqrt{\Delta t}) \\
&+ \sigma'(X_i + \mu(X_i)\Delta t + \sigma(X_i)\sqrt{\Delta t})\sigma(X_i)(\Delta W_i - \sqrt{\Delta t}) \\
&+ O(\Delta t),
\end{aligned}$$

and

$$\sigma'(X_i + \mu(X_i)\Delta t + \sigma(X_i)\sqrt{\Delta t}) = \sigma'(X_i) + O(\sqrt{\Delta t}).$$

Combining the results, we obtain

$$\sigma(X_i + \mu(X_i)\Delta t + \sigma(X_i)\sqrt{\Delta t}) - \sigma(X_i) = \sigma'(X_i)\sigma(X_i)\sqrt{\Delta t} + O(\Delta t)$$

giving

$$\sigma'(X_i)\sigma(X_i) = \frac{1}{\sqrt{\Delta t}}[\sigma(X_i + \mu(X_i)\Delta t + \sigma(X_i)\sqrt{\Delta t}) - \sigma(X_i)] + O(\Delta t).$$

Finally, the Runge-Kutta scheme can be represented in the predictor-corrector form:

$$\begin{aligned} \widehat{X_i} &= X_i + \mu(X_i)\Delta t + \sigma(X_i)\sqrt{\Delta t} \\ X_{i+1} &= X_i + \mu(X_i)\Delta t + \sigma(X_i)\Delta W_i \\ &\quad + \frac{1}{2\sqrt{\Delta t}}[\sigma(\widehat{X_i}) - \sigma(X_i)][(\Delta W_i)^2 - \Delta t]. \end{aligned}$$

Note that when $\sigma(X_i)$ becomes constant, both the Milstein and Runge-Kutta schemes reduce to the Euler scheme.

A.4 Quasi Monte Carlo simulation method

Quasi Monte Carlo methods are commonly used to evaluate multi-dimensional integrals. Consider

$$\int_{I_n} f(x)\ dx$$

over the n-dimensional unit cube, $I_n = [0,1] \times \cdots \times [0,1]$. We generate a sequence of pseudo random numbers $x_1, x_2, \ldots, x_N$, over the unit cube, and approximate the integral by

$$\theta_N = \frac{1}{N} \sum_{i=1}^{N} f(x_i).$$

In Quasi Monte Carlo methods, we produce a deterministic sequence of points (called low-discrepancy sequence) that provides the best possible spread in I_n.

Halton sequence

The Halton sequence is a sequence of numbers $h(i, b)$ for $i = 1, 2, ...$ The integer b is the base. The numbers all lie between zero and one. The numbers are constructed as follows. First choose your base. Let us choose 2. Now write the positive integers in ascending order in base 2, i.e. 1, 10, 11, 100, 101, 110, 111, etc. The Halton sequence base 2 is the reflection of the positive integers in the decimal point, i.e.

Integer base 10	Integers base 2	Halton sequence base 2	Halton number base 10
1	1	$1 \times \frac{1}{2}$	0.5
2	10	$0 \times \frac{1}{2} + 1 \times \frac{1}{4}$	0.25
3	11	$1 \times \frac{1}{2} + 1 \times \frac{1}{4}$	0.75
4	100	$0 \times \frac{1}{2} + 0 \times \frac{1}{4} + 1 \times \frac{1}{8}$	0.125
...	...	...	...

This has been called reflecting the numbers about the decimal point. If you plot the Halton points successively you will see that the next number in the sequence is always as far as possible from the previous point. Generally, the integer i can be written as

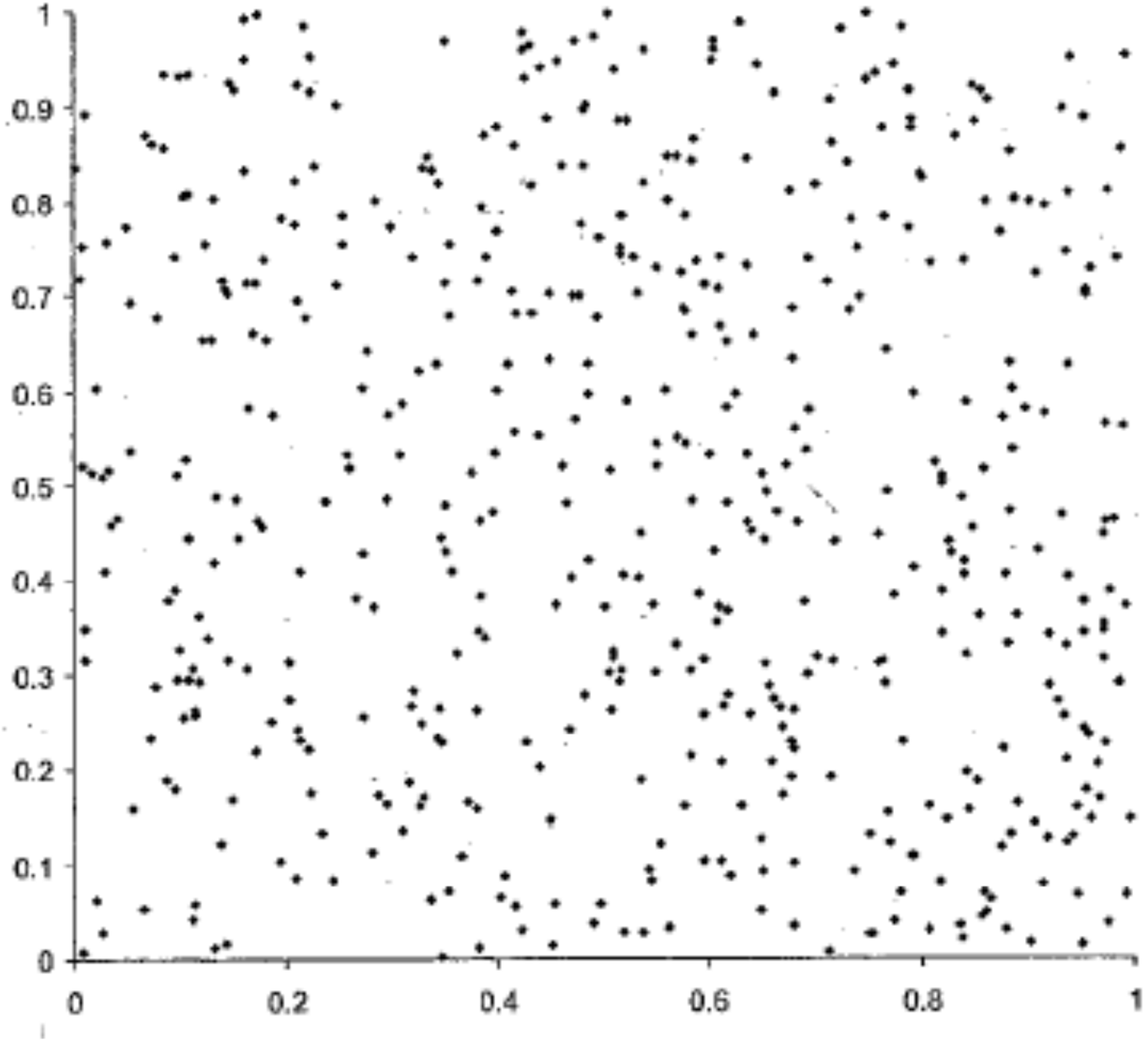
$$i = \sum_{j=1}^{m} a_j b^j$$

in base b, where $0 \leq a_j < b$. The Halton numbers are then given by

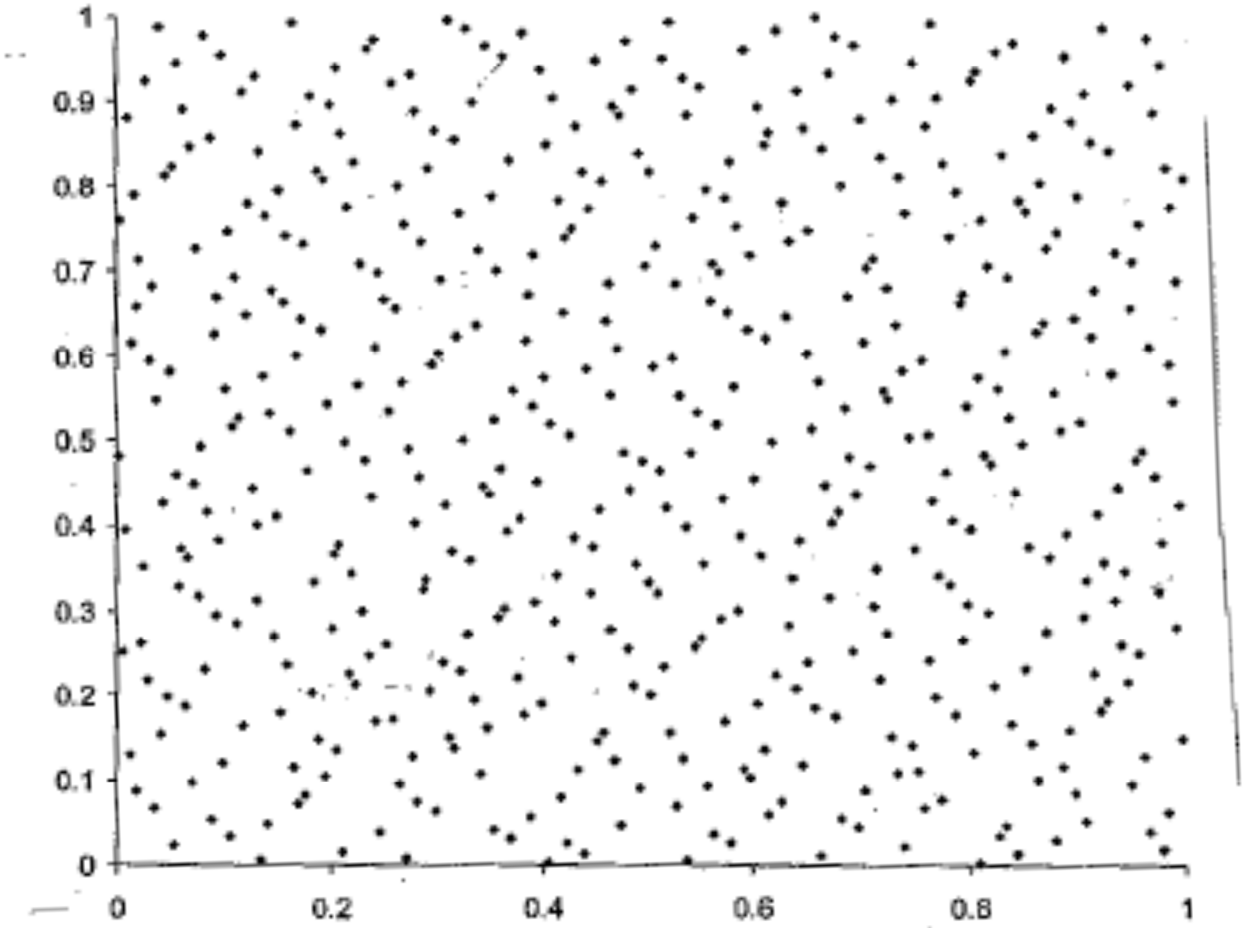
$$h(i; b) = \sum_{j=1}^{m} a_j b^{-j-1}.$$

Here is an algorithm for calculating Halton numbers of arbitrary base; the nth term in a Halton sequence of base b is given by Halton(n,b).

```
Function Halton(n,b)
Dim n0, n1, r, As Integer
Dim h As Double
Dim f As Double
n0 = n
h = 0
f = 1/b
While (n0 > 0)
n1 = Int (n0 / b)
r = n0 - n1 * b
h = h + f * r
f = f / b
n0 = n1
Wend
Halton = h
End Function
```



A Monte Carlo sample in two dimensions.



Halton points in two dimensions.

It is better to use a non-random series of points with better distributional properties.

- Quasi-random sequence seek to fill space uniformly, though they fail many statistical tests for randomness.

The estimate of the d-dimensional integral

$$\int_0^1 \cdots \int_0^1 f(x_1, \cdots, x_d)\ dx_1 \cdots dx_d$$

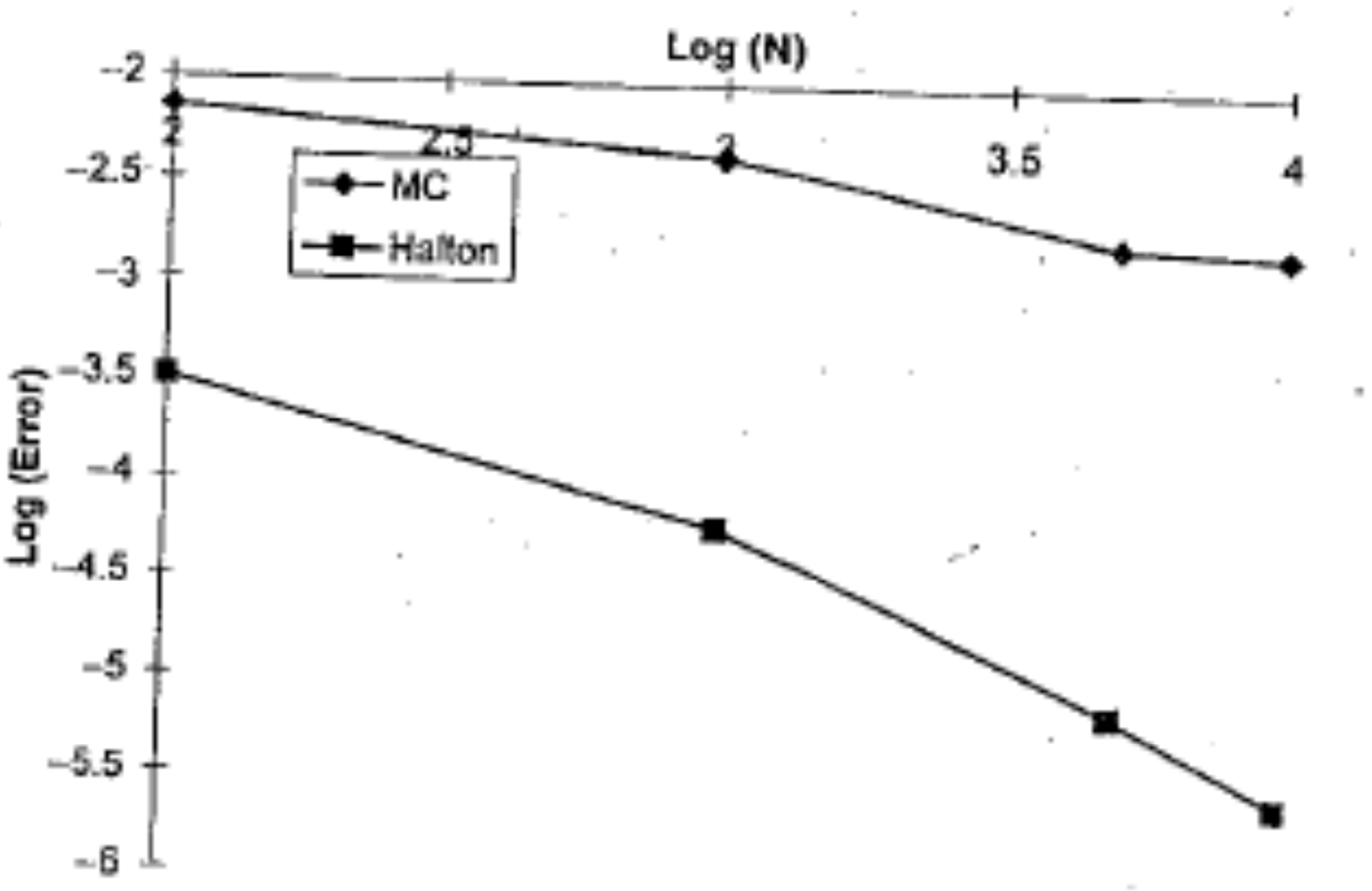
is given by

$$\frac{1}{N} \sum_{i=1}^{N} f\left(h(i, b_1), \cdots, h(i, b_n)\right),$$

where b_j are distinct prime numbers. The error in these quasi-random methods is

$$O\left((\log N)^d / N\right),$$

and it is better than Monte Carlo at all dimensions.



Estimate of the error in the value of a five-dimensional contract using the basic Monte Carlo simulation and a low discrepancy sequence.