Monte Carlo Integration

- Monte Carlo integration methods are sampling methods, based on probability theory.
- They rely on trials to reveal information
- From an intuitive point of view, they rest on the central limit theorem and the law of large numbers
- Monte Carlo methods are capable of handling quite complicated and large problems
- The result of an experiment is a random number
- By the probabilistic nature, the structure of the error made has a probabilistic distribution

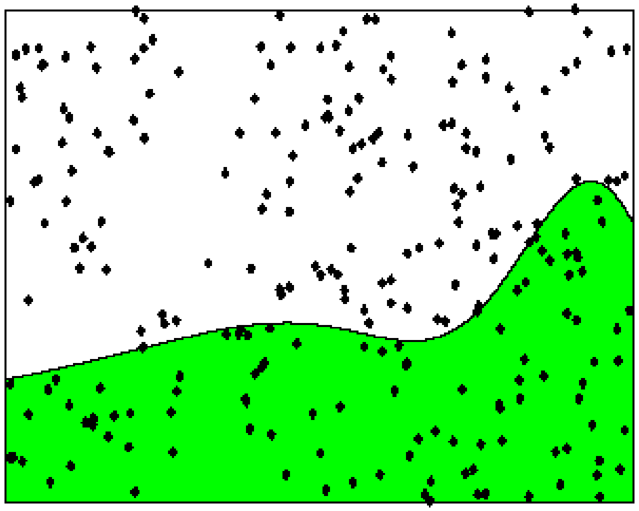
History

- Monte Carlo methods were originally practiced under more generic names such as "statistical sampling".
- The name "Monte Carlo" was popularized by physics researchers S. Ulam, E. Fermi, J. von Neumann, and N. Metropolis.
- The name is a reference to a famous casino in Monaco where Ulam's uncle would borrow money to gamble.
- Perhaps the most famous early use was by Enrico Fermi in 1930, when he used a random method to calculate the properties of the newly-discovered neutron.
- Monte Carlo methods were central to the simulations required for the Manhattan Project, though were severely limited by the computational tools at the time.

Basic Idea Monte Carlo

Integration of a Function

Figure 4.3: Basic idea of Monte–Carlo integration



- Draw random numbers in the $x - y$ plane (dots in the graph)
- Integral of function f is approximately given by the total area times the fraction of points that fall under the curve $f(x)$
- The greater the number of points the more accurate is the evaluation of this area.
- Method is only competitive for complicated and/or multi-dimensional functions

- Suppose we are trying to evaluate

$$I = \int\int\int_{\Omega} f(x,y,z)dxdydx$$

- Choose a random point inside Ω. Value $\hat{f}$ of f at that point is an unbiased estimator of f inside Ω. An unbiased estimate of I is therefore $I = \hat{f} \times \Omega$.
- Although the estimator is unbiased, it has an outrageously large expected error (so it is essentially useless).
- Reduce the error by repeating the experiment lots of times, and average the results.
- Choose N random points, then

$$I \approx (\hat{f}_1 + \hat{f}_2 + \ldots \hat{f}_N)\Omega_h/N$$

- The expected error is reduced by a factor $\sqrt{N}$, but N should be chosen very large.
- This will only be a competitive method for high dimensional quadrature, or for problems dealing with awkward volumes or surfaces.

- The evaluation of integral will be better if the points are uniformly scattered in the entire area
- Another way of thinking:

$$\int_a^b f(x)dx = (b-a)\mathsf{E}_{\mathcal{U}[a;b]}(f(x))$$

- If we draw N random numbers, $x_i,\ i = 1, \ldots, N$ from a $\mathcal{U}[a;b]$, an approximation of the integral of $f(x)$ over interval $[a, b]$ is given by

$$\frac{(b-a)}{N}\sum_{i=1}^{N} f(x_i)$$

- Accuracy attained depends on the number of trials
- The key point is to get random numbers !

Law of large numbers

- Underlying idea of Monte Carlo integration may be found in the Law of Large Numbers:
- If X_i is a collection i.i.d. random variables with density $f(x)$, then

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}X_i=\int xf(x)dx,\quad \text{almost surely}$$

- Further, we know that in this case

$$\mathsf{Var}\Big(\frac{1}{N}\sum_{i=1}^{N}X_i\Big)=\frac{\sigma^2}{N},\quad \text{where } \sigma^2=\mathsf{Var}(X_i)$$

If σ^2 is not known, it can be estimated by

$$\hat{\sigma}^2=\frac{1}{N-1}\sum_{i=1}^{N}(X_i-\overline{X})^2;\quad \overline{X}=\frac{1}{N}\sum_{i=1}^{N}X_i$$

Law of large numbers

$\Rightarrow$ Integrating a function $f(x)$ over $[0,1]$ is nothing else than computing the mean of $f(x)$ assuming that $x \sim \mathcal{U}[0,1]$. Therefore a crude application of the Monte Carlo method to compute the integral is to draw N numbers, U_i from a $\mathcal{U}[0,1]$ distribution and take:

$$\hat{I}_f = \frac{1}{N}\sum_{i=1}^{N} f(U_i)$$

as an approximation to the integral.

- This estimate of the integral is a random variable with variance

$$\sigma^2_{\hat{I}_f} = \frac{1}{N}\int_0^1 (f(x) - \hat{I}_f)^2 dx = \frac{\sigma_f^2}{N}$$

- σ_f^2 can be estimated by

$$\hat{\sigma}_f^2 = \frac{1}{N-1}\sum_{i=1}^{N}(f(U_i) - \hat{I}_f)^2$$

such that the standard error of the Monte Carlo integral is $\sigma_{\hat{I}_f} = \hat{\sigma}_f/\sqrt{N}$.

Example

- Crude application of Monte Carlo integration:

$$\int_0^1 e^x dx$$

- Table:

N	$\hat{I}_f$	$\hat{\sigma}_{\hat{I}_f}$
10	1.5490	0.1353
100	1.6995	0.0540
1000	1.7254	0.0163
10000	1.7245	0.0049
100000	1.7214	0.00156
1000000	1.7185	0.00049

- A huge number of data is needed to achieve, on average, a good enough approximation (1000000 point are needed to get an error lower than 0.5×10^{-4}.

$\Rightarrow$ Explains why Monte Carlo integration is seldom used for univariate integration, without modification.

Randomness

- Randomness is difficult to define, we usually associate randomness with unpredictability
- A sequence of numbers is random if it has no shorter description that itself
- Physical processes such as flipping a coin of tossing a dice, are deterministic if enough is known about equations governing their motion and appropriate initial conditions
- Another characteristic of True randomness is lack of repeatability
- This can make testing algorithms or debugging computer code difficult, or impossible
- Repeatability is desirable, but one must ensure independence among trials

Pseudo-random Numbers

- Computer algorithms for generating random numbers are deterministic; a sequence generated may appear random (it exhibits no apparent pattern)
- Such sequences of numbers are called Pseudo-random; They are quite predictable and reproducible !
- As only a finite number of numbers can be represented by a computer, any sequence must eventually repeat.
- Simple methods with a well-understood theoretical basis are preferable.

How we can speed up the MC?

The Monte Carlo simulation can be improved by:

- Improving the code: vectorization, removing loops, efficient memory management.
- Variance Reduction Techniques:
  - Antithetic sampling.
  - Control Variates.
  - Stratified Sampling.
  - Importance sampling.
- Taking proper random number generator:
  - Congruential Generators.
  - Fibonacci Generators.
  - Latin Hypercube sampling.
  - Sobol Sampling Algorithms.

Modifications Monte Carlo Integration

Antithetic sampling

It is well-known that if a random variable $Z \sim N(0,1)$, then also $-Z \sim N(0,1)$. We can use this property to drastically reduce the number of paths needed in the Monte Carlo simulation. Suppose that $\hat{V}$ is the approximation obtained by MC, and $\widetilde{V}$ is the one obtained by using $-Z$. By taking the average

$$V = \frac{1}{2}\left(\widetilde{V} + \hat{V}\right)$$

we obtain a new approximation. Since $\hat{V}$ and V are both random variables we aim at:

$$Var(V) < Var(\hat{V}).$$

We have:

$$Var(V) = \frac{1}{4}Var(\widetilde{V} + \hat{V}) = \frac{1}{4}Var(\widetilde{V}) + \frac{1}{4}Var(\hat{V}) + \frac{1}{2}Cov(\widetilde{V}, \hat{V}).$$

So: $Var(V) \leq \frac{1}{2}\left(Var(\hat{V}) + Var(\widetilde{V})\right)$.

Modifications Monte Carlo Integration

Stratified Sampling

- Idea: Variance of f over a subinterval of $[0,1]$ should be lower than the variance over whole interval.
- Prevent draws from clustering in a particular region of the interval; The procedure is forced to visit each subinterval. The information set used is enlarged.
- Set $\lambda \in (0,1)$ and draw $N_a = \lambda N$ data points over $[0,\lambda]$ and $N_b = N - N_a = (1-\lambda)N$ over $[\lambda, 1]$. The integral is then evaluated by

$$\hat{I}_f^s = \frac{1}{N_a}\sum_{i=1}^{N_a} f(x_i^a) + \frac{1}{N_b}\sum_{i=1}^{N_b} f(x_i^b)$$

where $x_i^a \in [0,\lambda]$ and $x_i^b \in [\lambda, 1]$.

- Variance of this estimator is:

$$\frac{\lambda^2}{N_a}\text{Var}_a(f(x)) + \frac{(1-\lambda)^2}{N_b}\text{Var}_b(f(x)) = \frac{\lambda}{N}\text{Var}_a(f(x)) + \frac{(1-\lambda)}{N}\text{Var}_b(f(x))$$

- How to select λ ? Answer: Choose it such that volatility is minimized: $\text{Var}_a((f(x)) = \text{Var}_b(f(x))$.
- This drives the overall variance to $\text{Var}_b(f(x))/N$.
- Example with $\lambda = 0.25$:

N	$\hat{I}_f$	$\hat{\sigma}_{\hat{I}_f}$
10	1.5218	0.1353
100	1.6995	0.0414
1000	1.7254	0.01188
10000	1.7245	0.00359
100000	1.7214	0.00114

Modifications

Control Variates

- This method tries to extract information from a function that approximates the function to be integrated arbitrarily well.
- Assume there exists a function φ that is similar to f, but that can be easily integrated.
- The identity
$$\int f(x)dx = \int (f(x) - \varphi(x))dx + \int \varphi(x)dx$$
restates the Monte Carlo integration of the first term plus the known integral of φ.
- The variance of $(f - \varphi)$ is given by $\sigma_f^2 + \sigma_\varphi^2 - 2\text{Cov}(f, \varphi)$. This is lower than the variance of σ_f^2 provided the covariance is high enough.

Example

Control Variates

- We use $\varphi(x) = 1 + x$, since $e^x \approx 1 + x$ in a neighborhood of zero.

N	$\hat{I}_f$	$\hat{\sigma}_{\hat{I}_f}$
10	1.6450	0.0501
100	1.7190	0.0229
1000	1.7250	0.00689
10000	1.7213	0.00210
100000	1.7198	0.00066
1000000	1.7184	0.00020

- $\int_0^1 (1 + x)dx = 1.5$ is easily computed.

Importance Sampling

- By drawing numbers for a uniform distribution in crude Monte Carlo methods, information is spread all over the interval we are sampling over.
- A simple transformation of the problem may exist for which Monte Carlo can generate a far better result in terms of variance.
- Suppose a function $g(x)$ exists such that $h(x) = f(x)/g(x)$ is almost constant over the domain of integration. Restate the problem

$$\int f(x)dx = \int \frac{f(x)}{g(x)} g(x)dx = \int h(x)g(x)dx$$

- We can now easily integrate f by instead sampling $h(x)$, but not by drawing numbers from a uniform density function, but rather from a nonuniform density $g(x)dx$.

Importance Sampling

- The approximated integral is given by

$$\hat{I}_f^{is} = \frac{1}{N}\sum_{i=1}^{N}\frac{f(x_i)}{g(x_i)}$$

- With variance

$$\begin{aligned}\sigma^2_{\hat{I}_f} &= \frac{1}{N}\Big(\int \frac{f(x)^2}{g(x)^2}g(x)dx - \big(\int \frac{f(x)}{g(x)}g(x)dx\big)^2\Big)\\ &= \frac{1}{N}\Big(\int \frac{f(x)^2}{g(x)}dx - \big(\int f(x)dx\big)^2\Big)\end{aligned}$$

- How to select $g(x)$? Answer: g should have a shape close to that of f while being simple to sample.

Importance Sampling

Example

- We use $g(x) = (1+\alpha)x^{\alpha}$, with $\alpha = 1.5$.
- The selection of the g function requires good knowledge of the function to be integrated, which may not always be the case

N	$\hat{I}_f$	$\hat{\sigma}_{\hat{I}_f}$
10	1.5490	0.0428
100	1.6995	0.0054
1000	1.7254	0.000514
10000	1.7245	0.0000495
100000	1.7214	0.0000049
1000000	1.7185	0.00000049

Usual Numerical Integration

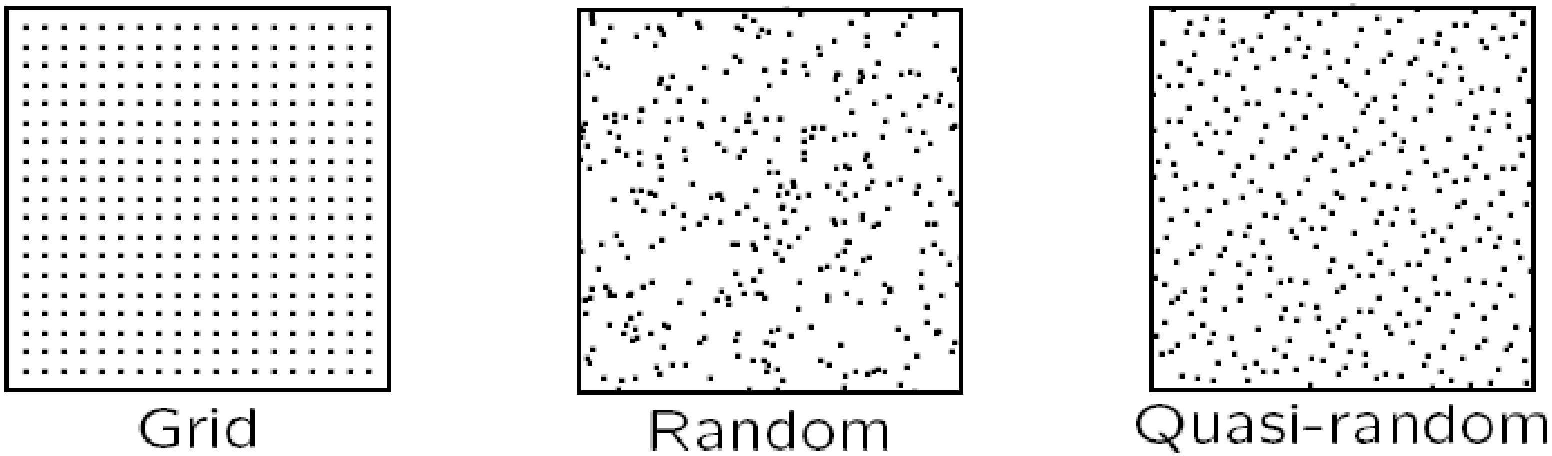
Quadrature formulas

- A classic problem of numerical analysis: Approximating the definite integral
$$\int_a^b f(x)dx$$
- Polynomial $P_m(x)$ approximates the function $f(x)$. Replace the integral by
$$\int_a^b P_m(x)dx$$
- Divide interval $[a,b]$ into n subintervals: $h = (b-a)/n$. Trapezoidal rule:
$$T(h) = h\left[\frac{f(a)}{2} + f(a+h) + \ldots + f(b-h) + \frac{f(b)}{2}\right]$$
- The error of $T(h)$ satisfies a quadratic expansion:
$$T(h) = \int_a^b f(x)dx + O(h^2)$$
depending on the differentiability of f

Quasi-Random Sequences

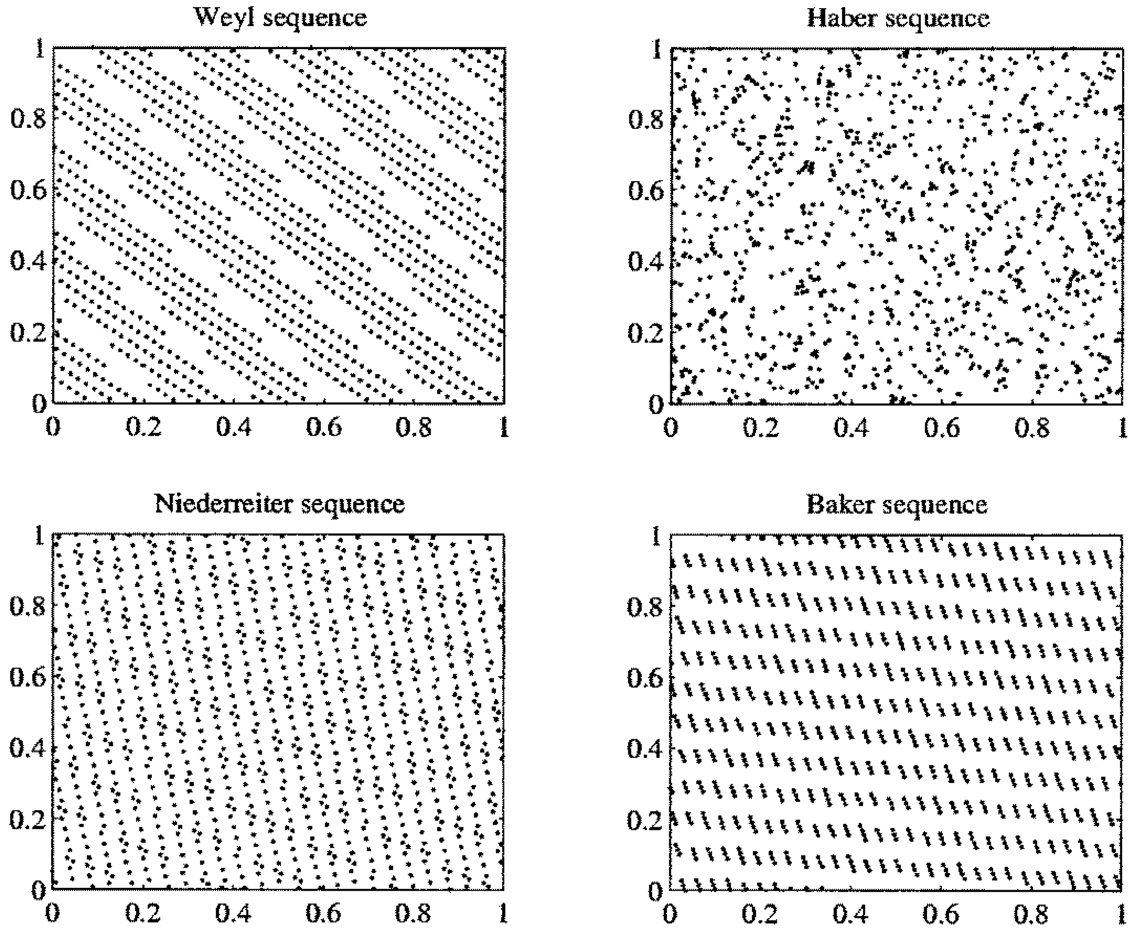
- For some applications, achieving reasonably uniform coverage of sampled data can be more important than whether the sample points are truly random.
- Truly random sequences tend to exhibit clumping, leading to an uneven coverage of sampled data for a given number of points
- A perfectly uniform coverage can be achieved by using a regular grid of sample points
- This approach does not scale well to higher dimensions
- A compromise between these extremes of coverage and randomness is provided by quasi-random sequences.
- These sequences are carefully constructed to give a uniform coverage of sampled points while maintaining a reasonable random appearance.
- By design, points tend to avoid each other, so clumping associated with true randomness is eliminated.

Quasi-Random Sequences



Quasi-Random Sequences

Figure 4.6: Quasi–Monte Carlo sequences



Low-Discrepancy Sequences

Halton Points (page 77)

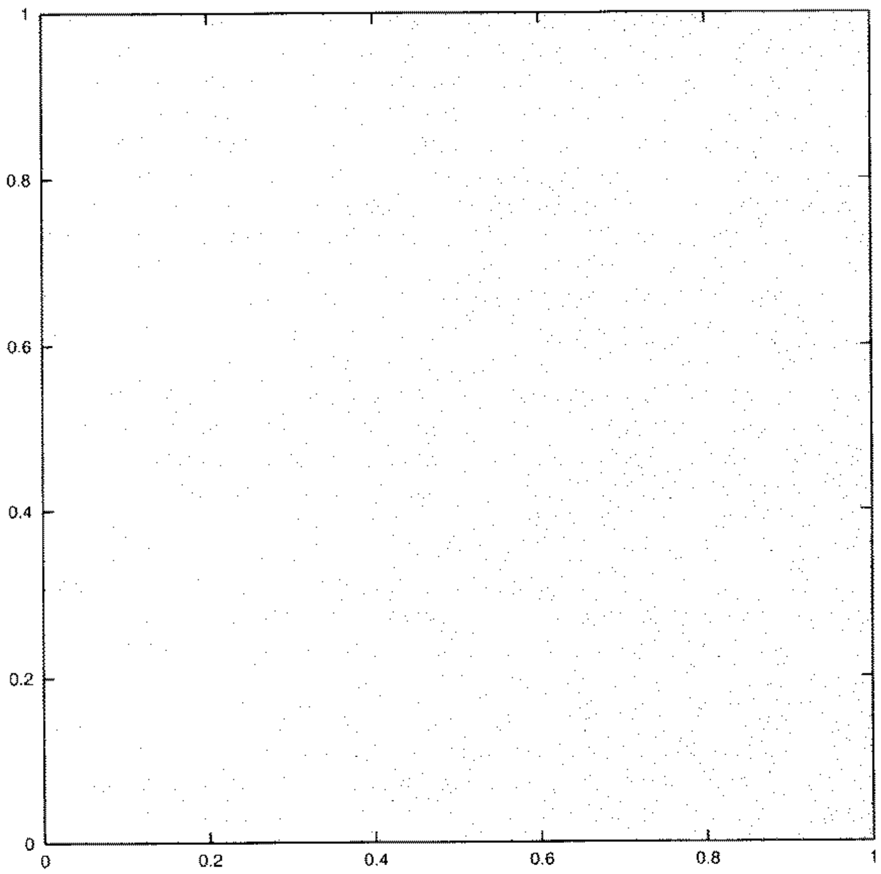


Fig. 2.7. 10000 Halton points from Definition 2.18, with $p_1 = 2$, $p_2 = 3$

Congruential Random Number Generators

- Congruential generators have the form

$$N_i = (aN_{i-1} + b) \ (\text{mod } M),$$

  with a and b given integers

- Starting integer N_0 is called seed
- Integer M is approximately (often equal to) the largest integer representable on a machine
- The quality of such generators depends on choices of a and b, and its period cannot exceed M
- To produce random floating-points numbers U_i, uniformly distributed on interval $[0, 1)$, the random integers must be divided by M: $U_i = N_i/M$.

Properties

a. $N_i \in \{0, 1, \ldots, M-1\}$

b. The N_i are periodic with period $\leq M$. $N_i = N_{i+p},\ p \leq M$.

$\Rightarrow$ Make M as large as possible; With period M, the U_i are uniformly distributed when exactly M numbers are needed. (Each grid point on $[0, 1)$ is occupied once on a grid with mesh size $1/M$.

c. $N = 0$ is ruled out in case $b = 0$.
With $a = 1$ the generator gives $N_n = (N_0 + nb) \bmod M$ which is too easily predictable

Fibonacci Generators

- Alternative methods that produce floating-point random numbers on interval $[0, 1)$ are Fibonacci Generators.
- A new value is generated as a sum, difference or product of previous values.
- Typical example:
$$N_i = N_{i-17} - N_{i-5}$$
- This generator has lags of 17 and 5. The lags must be chosen carefully
- Such a formula may produce a negative result, in which case a remedy is to add 1 to get back to interval $[0, 1)$
- These generators require more storage than the congruent generators, and a special starting procedure. Division is not required.
- Well-designed generators of this type typically have a much longer period than congruential generators, since repetition of one member of a sequence does not entail that all subsequent members will also repeat in the same order

Example

- Algorithm:

$$
\begin{aligned}
Repeat: \quad & \zeta := U_i - U_j \\
& \text{if } \zeta < 0, \ \text{set } \zeta := \zeta + 1 \\
& U_i := \zeta \\
& i := i - 1 \\
& j := j - 1 \\
& \text{if } i = 0, \ \text{set } i := 17 \\
& \text{if } j = 0, \ \text{set } j := 17
\end{aligned}
$$

- Initialization:
  Set $i = 17, j = 5$, compute $U_1, \ldots, U_{17}$ with a congruential generator $(M = 714025, a = 1366, b = 150889)$.
  Choose seed N_0 freely.

Inversion

- Suppose $U \sim \mathcal{U}[0,1]$ and F be a continuously strictly increasing distribution function. Then $F^{-1}(U)$ is a sample from F.
- Proof: $U \sim \mathcal{U}[0,1]$ means $\mathsf{P}(U \leq \xi) = \xi$ for $0 \leq \xi \leq 1$.
  Consequently, $\mathsf{P}(F^{-1}(U) \leq x) = \mathsf{P}(U \leq F(x)) = F(x)$.
- For the normal distribution there is no closed-form expression for F nor for inverse F^{-1}.

$\Rightarrow$ A possibility is "numerical inversion", i.e, finding x in $F(x) = u$, with u prescribed.

- Use Newton's method, or bisection to find x in $F(x) - u = 0$.

Transformations

Between Random Variables

- Suppose X is random variable with density $f(x)$ and distribution $F(x)$. Assume $h : \Omega \to B$, strictly monotonous with Ω, B in $\mathbb{R}$, where Ω is the support of $f(x)$ (f is zero outside Ω).

(a) Then $Y := h(X)$ is a random variable with distribution $F(h^{-1}(y))$.

(b) If h^{-1} is absolutely continuous then for almost all y the density of $h(X)$ reads

$$f(h^{-1}(y)) \left| \frac{dh^{-1}(y)}{dy} \right| .$$

– Proof: (a): $\mathsf{P}(h(X) \le y) = \mathsf{P}(X \le h^{-1}(y)) = F(h^{-1}(y))$.
(b): h^{-1} absolutely continuous $\Rightarrow$ The density of $Y = h(X)$ equals the derivative of the distribution function almost everywhere.
Evaluating $dF(h^{-1}(y))/dy$ with the chain rule implies the assertion.

Other intervals, Nonuniform Distributions

- Uniform distribution on $[a, b)$: Transform N_i from $[0, 1)$ by:

$$(b - a)N_i + a$$

- Sample nonuniform distributions: If the cumulative distribution function of desired probability density function is easily invertible, we can generate uniform random numbers and invert them.

- For example, we can take

$$-\log{(U_i)}/\lambda$$

where the U_i are uniform on $[0, 1)$. These numbers are exponentially distributed:

$$f(t) = \lambda e^{-\lambda t}, \;\; t > 0$$

Transformation in higher D

- Suppose X is a random variable in $\mathbb{R}^n$ with density $f(x) > 0$ on the support Ω. The transformation $h : \Omega \to B$ is assumed to be invertible and the inverse continuously differentiable on B.
  $Y := h(X)$ is the transformed variable. Then Y has the density
  $$f(h^{-1}(y)) \left| \frac{\partial(x_1, \ldots, x_n)}{\partial(y_1, \ldots, y_n)} \right|, \; y \in B$$
  where $x = h^{-1}(y)$ and $\partial(x_1, \ldots, x_n)/\partial(y_1, \ldots, y_n)$ is the determinant of the Jacobian matrix of all first-order derivatives of $h^{-1}(y)$.

Transformation to Normal Distribution

2D

- Start with $\Omega = [0,1]^2$ and the density of the uniform distribution $f = 1$ on Ω. Transformation

$$\left\{ \begin{array}{lcll} y_1 & = & \sqrt{-2\log x_1}\cos 2\pi x_2 & =: \; h_1(x_1,x_2) \\ y_2 & = & \sqrt{-2\log x_1}\sin 2\pi x_2 & =: \; h_2(x_1,x_2) \end{array} \right\}$$

- The inverse function h^{-1} is given by

$$\left\{ \begin{array}{lcc} x_1 & = & \exp\left(-1/2(y_1^2 + y_2^2)\right) \\ x_2 & = & \frac{1}{2\pi}\arctan\frac{y_2}{y_1} \end{array} \right\}$$

Normal Distribution

- The determinant of the Jacobian matrix is

$$\frac{\partial(x_1,x_2)}{\partial(y_1,y_2)} = det\begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{pmatrix} =$$

$$\frac{1}{2\pi}exp(-0.5(y_1^2+y_2^2))(-y_1\frac{1}{1+\frac{y_2^2}{y_1^2}}\frac{1}{y_1} - y_2\frac{1}{1+\frac{y_2^2}{y_1^2}}\frac{y_2}{y_1^2}$$

$$= -\frac{1}{2\pi}\exp\left(-0.5(y_1^2+y_2^2)\right).$$

- This shows that $|\partial(x_1,x_2)/\partial(y_1,y_2)|$ is the density of the standard normal distribution in $\mathbb{R}^2$.
- Since this density is the product of two 1D densities, the two components of y are independent.

$\mathcal{N}(0,1)$

10000 Numbers

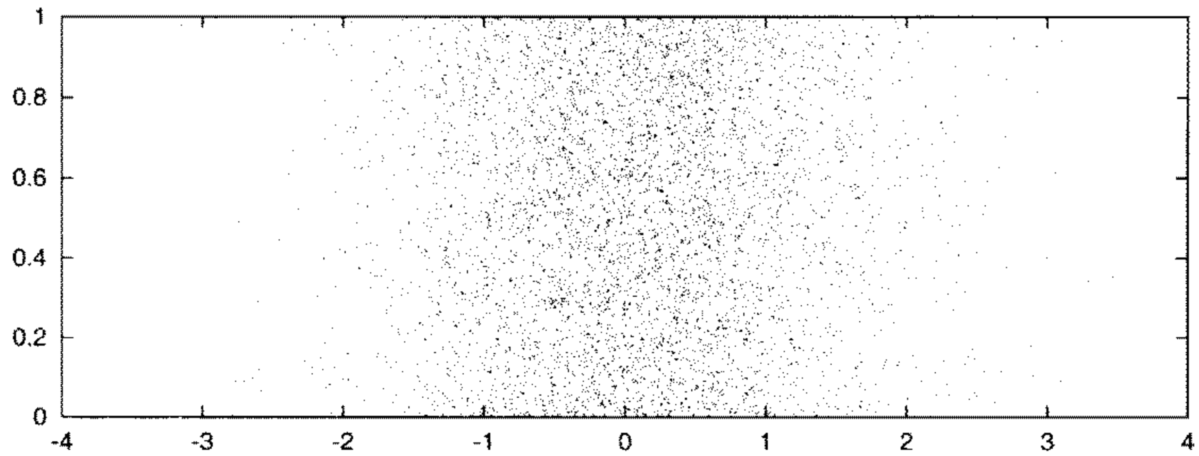


Fig. 2.5. 10000 numbers $\sim \mathcal{N}(0,1)$ (values entered horizontally and separated vertically with distance 10^{-4})

Normal Distribution

- An important example is the generation of random numbers that are normally distributed with given mean and variance.
- Available routines often assume mean of 0 and a variance of 1
- If some other mean μ and variance σ^2 are desired then each value x_i produced by a routine can be modified by a transformation $\sigma x_k + \mu$ to achieve the desired normal distribution.

Exotic Options

A Potpourri of options and equations !

- Compound, chooser, binary
- Path dependent: Barrier options, lookback, Asian
- Equations and numerics for Asians

Compound, chooser, binary

- Compound option: Call on a call: right to buy a 'call with maturity T and strike K' at time T_0 for the price K_0. Payoff:

$$CC(S, T_0, K_0, K, T) = \max\left[C(S, K, T) - K_0, 0\right]$$

- Chooser option: Gives the holder the right to choose whether the underlying option at time T_0 is a Call or a Put with the same strike K and maturity T. The payoff of a chooser option is

$$CH(S, K, T_0, T) = \max\left[C(S, K, T), P(S, K, T)\right]$$

- Binary: Cash or Nothing: Pays out Q at expiry T if option is in the money $S > K$, otherwise expires worthless. Payoff:

$$B(S, Q, K, T) = Q 1_{S \geq K}$$

  variation: Asset or nothing, where Q is the asset itself.

Binaries or Digitals

- Binary options have a discontinuous payoff at expiry.
- An example for a binary call, is:
  The contract pays 1 at T, if the asset price is then greater than the exercise price E.
- The to the binary call belonging final condition is

$$V(S,T) = \mathcal{H}(S-E)$$

  where $\mathcal{H}(\cdot)$ is the Heaviside function.

Path-dependency

- Options whose value depends on the asset history, but can still be written as $V(S,t)$ are said to be weakly path dependent.
- American options, with early exercise, are considered to be weakly path dependent. The next common reason for weak path dependence in a contract is a barrier.
- Strongly path-dependent contracts are of particular interest. These have payoff that depend on some property of the asset price path in addition to the value of the underlying at the present time. So, $V \neq V(S,t)$. The contract value is a function of at least one more independent variable, such as a 'running average' of asset prices.
- Weakly path-dependent contracts have the same number of dimensions as the non-path-dependent versions
- Strongly path dependent contracts are governed by an extra dimension. The new independent variable is a measure of the path-dependent quantity.

Path-dependent options

- Barrier options (US, 1967) are options that either come alive or die when predetermined trigger points (barriers) are reached.
  Down-and-out call: Option is knocked out if S hits a certain barrier H. Payoff:
  $$C_{DO}(S,K,T) = \max(S-K,0) \text{ if } S \geq H; \text{ else ceases to exist}$$
- Lookback options: Path dependent options whose payoffs depend on the max or min of the asset during a certain period (lookback period $[T_0,T]$).
  Payoff European fixed strike lookback call: $(\max_{[T_0,T]}(S_t) - K)^+$
  Payoff European fixed strike lookback put: $(K - \min_{[T_0,T]}(S_t))^+$

Barrier options

- Barrier options come in two main varieties, the 'in' barrier option (or knock-in) and the 'out' barrier option (knock-out). The former only have a payoff if the barrier level is reached before expiry and the latter only have a payoff if the barrier is not reached.
- Barrier options are popular for various reasons.
- Usually, a purchaser has very precise views about the direction of the market. If he wants the payoff from a call option but does not want to pay for all the upside potential, believing that the upward movement of the underlying will be limited prior to expiry, then he may choose to buy an up-and-out call. It will be cheaper than a similar vanilla call, since the upside is severely limited.
- Conversely, an 'in' option will be bought by someone who believes that the barrier level will be realized. Again the option is cheaper then the equivalent vanilla option.

Different types of barriers

- The out option only pays off if a level is not reached. If the barrier is reached the option is said to have knocked out.
- The in option pays off as long as a level is reached before expiry. If the barrier is reached then the option is said to have knocked in.
- If the barrier is above the initial asset value, we have an up option
- If the barrier is below the initial value, we have a down option
- The payoffs are the usual ones
- Barrier can be time dependent

PDE for Barrier options

- These options satisfy the Black-Scholes equation, on a special domain, with special boundary conditions
- The details of the barrier feature come in through the specification of the boundary conditions
- If the asset reaches the barrier S_u in an 'out' barrier option then the contract becomes worthless:

$$V(S_u, t) = 0 \text{ for } t < T$$

- If we have a down-and-out option with a barrier at S_d we solve for $S_d < S < \infty$ with

$$V(S_d, t) = 0$$

‘In’ Barriers

- An ‘in’ option only has a payoff if the barrier is triggered. If the barrier is not triggered we have $V(S,T) = 0$
- The value in the option is in the potential to hit the barrier. If the option is an up-and-in contract then on the upper barrier the contract must have the same value as a vanilla contract:

$$V(S_u, t) = \text{ value of vanilla contract, function of } t$$

So,

$$V(S_u, t) = V_{van}(S_u, t) \text{ for } t < T$$

Hedging barrier options

- Barrier options have discontinuous delta at the barrier

  For a knock-out, the option value is continuous, decreasing approximately continuously towards the barrier, then being zero beyond the barrier.

- A discontinuity in the delta means that the gamma is instantaneously infinite at the barrier. Delta hedging through the barrier is virtually impossible, and costly.

- There have been a number of suggestions made for ways to statically hedge barrier options. These methods try to mimic as closely as possible the value of a barrier option with vanilla calls and puts or with binary options.

Asian options

- Asian options: The payoff depends on the average of the underlying. Types of averages:
- Arithmetic average:

$$A = \frac{1}{n}\sum_{i=1}^{n} S_{t_i}$$

- Geometric average: $A = (\Pi_{i=1}^{n} S_{t_i})^{1/n}$
- Continuous average

$$A = \frac{1}{t}\int_0^t S_\tau d\tau$$

- All the above may be expressed as $A_t = \int_0^t f(S_\tau, \tau) d\tau$

Asian payoffs

- $(A-K)^+$: fixed strike call
- $(K-A)^+$: fixed strike put
- $(S_T-A)^+$: floating strike call
- $(A-S_T)^+$: floating strike put
- The PDE reads:

$$\frac{\partial V}{\partial t}+\frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2}+rS\frac{\partial V}{\partial S}+f(S,t)\frac{\partial V}{\partial A}-rV=0$$

Discrete Averaging

- A discrete averaging is most used in practice. Let sampling times $t_1, \ldots, t_N$, and define the averages by:

$$A_n = \frac{1}{n} \sum_{i=1}^{n} S(t_i)$$

- Notice that $A_n = A_{n-1} + \frac{S(t_n) - A_{n-1}}{n}$
- Denoting t^+ and t^- the times before and after the date t_n:

$$A(S, t^+) = A(S, t^-) + \frac{S - A(S, t^-)}{n}$$

Discrete Averaging (cont.)

- Simplifying notation: $A^+ = A^- + \frac{S - A^-}{n}$
- From no-arbitrage one has

$$V(S, A^+, t^+) = V(S, A^-, t^-)$$

- However, for fixed (S, A) this defines a jump across t_n
- Away from the observation dates one solves the plain BS equation

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0$$

and apply at jump times the jump condition. Summarizing:

Algorithm for fixed strike Asian put

Take a time slice $[t_N, T]$:

– Divide $[0, A_{max}]$ and $[0, S_{max}]$ with grid A_j and S_i

– Solve for each A_j the BS equation with final condition $(K - A)^+$ (say J equations simultaneously) and boundary conditions $(K - A_j)^+, S \to \infty$

– The surface obtained at time t_N (call it t_N^+) is then shifted by interpolation:

$$V(S, A, t_n^-) = V(S, A + \frac{S - A}{N}, t_N^+)$$

– The new surface $V(S, A, t_N^-)$ is used as new final condition for the new set of BS equations on time slice $[t_{N-1}, t_N]$ and so on...

Asian Options

Option depending on continuous average

- Previously details on Asian options with discrete averaging.
  Now, continuous averaging.
- The exercise price or the asset price is replaced by an average of the asset price:
- Final conditions for an arithmetic-average floating strike call:
  $u(S,T) = \max(S - \frac{1}{T}\int_0^T S(\tau)d\tau, 0)$
- With a new variable: $I(t) := \int_0^t S(\tau)d\tau$, one finds a similar Black-Scholes type equation for Asian options:

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rs\frac{\partial V}{\partial S} + S\frac{\partial V}{\partial I} - rV = 0$$

- There is no diffusion term in the I-direction.

Path dependency and the integral

- We start by assuming that the underlying asset follows the lognormal random walk:

$$dS = \mu S dt + \sigma S dW$$

  Imagine a contract that pays off at maturity T an amount that is a function of the path taken by the asset between zero and maturity

- Suppose that this path-dependent quantity can be represented by an integral of some function of the asset over the period zero to T:

$$I(T) = \int_0^T f(S, \tau) d\tau$$

- Most path-dependent quantities in exotic derivative contracts can be written in this form with a suitable choice of $f(S, t)$.

- Prior to maturity we have information about the possible final value of S (at time T) in the present value of S at time t. For example, the higher S is today, the higher it will probably end up at maturity.

  Similarly we have information about the possible final value of I in the value of the integral to date:

$$I(t) = \int_0^t f(S, \tau) d\tau$$

  As we get closer to maturity, we become more confident about the final value of I.

- The value of the option is therefore not only a function of S and t, but also a function of I; I will be our new independent variable, called a state variable

- As we will use Itô's lemma, we need to know the stochastic differential equation satisfied by I. This is simply (incrementing t by dt we find that):

$$dI = f(S,t)dt$$

- I is thus a smooth function, and the equation for dI does not contain stochastic terms

Examples

- An Asian option has a payoff that depends on the average of asset price over some period. If that period is from time zero to maturity and the average is arithmetic then:

$$I = \int_0^t S d\tau$$

- The payoff may then be, (a floating strike put), for example,

$$\max\left(\frac{I}{T} - S, 0\right)$$

- Another example: Imagine a contract that pays off a function of the square of the underlying asset, but only counts those times for which the asset is below S_u. Then

$$I = \int_0^t S^2 \mathcal{H}(S_u - S) d\tau$$

where $\mathcal{H}$ is the Heaviside function.

Continuous sampling

The pricing equation

- We derive the pricing PDE for a contract that pays some function of new variable I
- The value of the contract is now a function of the three variables: $V(S, I, t)$.
- Set up a portfolio containing one of the path-dependent options and short a number Δ of the underlying asset:

$$\Pi = V(S, I, t) - \Delta S$$

- The change in the value of this portfolio is given by

$$d\Pi = \left(\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} \right) dt + \frac{\partial V}{\partial I} dI + \left(\frac{\partial V}{\partial S} - \Delta \right) dS$$

- Choosing $\Delta = \partial V / \partial S$ to hedge the risk, we find:

$$d\Pi = \left(\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + f(S, t) \frac{\partial V}{\partial I} \right) dt$$

- This change is risk free and thus earns the risk-free rate of interest r, leading to the PDE:

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + f(S,t)\frac{\partial V}{\partial I} + rS\frac{\partial V}{\partial S} - rV = 0$$

- This is to be solved subject to

$$V(S,I,t) = \text{ payoff}$$

- This completes the formulation of the valuation problem.

Higher dimensions

- The methods outlined are not restricted to a single path-dependent quantity. Any finite number of path-dependent variables can be accomodated, theoretically.
- Imagine that a contract pays off the difference between a continuous geometric and a continuous arithmetic average. To price this one would need to introduce I_g and I_a, defined by

$$I_g = \int_0^t \log(S) d\tau, \quad I_a = \int_0^t S d\tau$$

- The solution would be a function of four variables $V(S, I_g, I_a, t)$.
- This growth in dimensionality may be difficult for solving a PDE with numerical techniques !

Similarity reductions

- Some options have a particular structure that permits a reduction in the dimensionality of the problem by use of a similarity variable.
- The dimensionality of the continuously-sampled artihmetic floating strike option can be reduced from three to two.
- The payoff for the call option is

$$\max\left(S - \frac{1}{T}\int_0^T S(\tau)d\tau, 0\right)$$

- We can write the running payoff for the call option as

$$I \max\left(R - \frac{1}{t}, 0\right), \text{ where } I = \int_0^t S(\tau)d\tau, \text{ and } R = \frac{S}{\int_0^t S(\tau)d\tau}$$

- The payoff at maturity may then be written as

$$I \max\left(R - \frac{1}{T}, 0\right)$$

- In view of the form of the payoff, it is plausible that the option value takes the form:

$$V(S,R,t) = IW(R,t). \text{ with } R = \frac{S}{I}$$

- We find that W satisfies:

$$\frac{\partial W}{\partial t} + \frac{1}{2}\sigma^2 R^2 \frac{\partial^2 W}{\partial R^2} + R(r-R)\frac{\partial W}{\partial R} - (r-R)W = 0$$

- This reduction is not possible for American variants

Put-Call Parity for European Floating Strike

- The payoff at maturity for a portfolio of one European floating strike call held long and one put held short is

$$I \max{(R - \frac{1}{T}, 0)} - I \max{(\frac{1}{T} - R, 0)}$$

- Whether R is greater or less than T at maturity, this payoff is simply:

$$S - \frac{I}{T}$$

- The value of this portfolio is identical to one consisting of one asset and a financial product whose payoff is $-I/T$

- In order to value this product find a solution of the floating strike equation of the form

$$W(R,t) = b(t) + a(t)R$$

and with $a(T) = 0$ and $b(T) = -1/T$; such a solution would have the required payoff of $-I/T$.

- Substitution and satisfying the boundary conditions, we find that

$$a(t) = \frac{-1}{rT}\left(1 - e^{-r(T-t)}\right), \quad b(t) = -\frac{1}{T}e^{-r(T-t)}$$

- We conclude that

$$V_c - V_p = S - \frac{S}{rT}(1 - e^{-r(T-t)}) - \frac{1}{T}e^{-r(T-t)}\int_0^t S(\tau)d\tau$$

where V_c and V_p are the values of the European arithmetic floating strike call and put.

- With a new variable: $A(t) := (\int_0^t S(\tau)d\tau)/t$, one obtains the following Black-Scholes type equation for Asian options:

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rs\frac{\partial V}{\partial S} + \left(\frac{S-A}{t}\right)\frac{\partial V}{\partial A} - rV = 0$$

- There is no diffusion term in the A-direction.

$\Rightarrow$ In CFD, this situation often occurs and appropriate discretizations for such terms have been developed there.

$\Rightarrow$ Ultraparabolic equation.

- Moreover, American-style Asian options exist (combination of convection dominance and free boundary aspects)
- We cannot transform the equation to a diffusion equation anymore. We need to discretization and solve the original equation.

Examples of multi-asset options

- A basket option is an option whose payoff depends on the value of a portfolio (or basket) of assets. Basket options are growing in popularity as a means of hedging the risk of a portfolio and are highly interesting for banks nowadays.
- They are attractive because an option on a basket is cheaper than buying options on the individual assets. Furthermore, their payoff profile replicates the changes in a portfolio's value more closely than any combination of options on the underlying assets.
- Basket options: $u(\mathbf{S}(T), T) = \left(\sum_{i=1}^{d} w_i S_i - K\right)^+$
- Call option on the minimum of the underlying assets $u(\mathbf{S}(T), T) = (\min_i S_i - K)^+$
- Put option on the maximum of the underlying assets $u(\mathbf{S}(T), T) = (K - \max_i S_i)^+$
- Exchange option (two-asset): $u(\mathbf{S}(T), T) = (S_1 - S_2)^+$

Multi-asset options

Problem definition

- Multi-asset options are multi-dimensional. Using numerical techniques, the number of grid points grows exponentially $\Rightarrow$ Curse of dimensionality. Problems are not solvable on nowadays machines unless advanced techniques are used.
- Sparse grid methods reduces the number of grid points per dimension, so larger problems can be computed. In finance, the sparse grid method for solving PDE's is used for the first time by C. Reisinger [4].
- Partitioning/Splitting and parallelization of the method reduces memory usage.

PDE methods

Multi-d Black-Scholes equation

The PDE-method is based on the solution of the multi-dimensional Black-Scholes equation:

$$\frac{\partial u}{\partial t} + \frac{1}{2}\sum_{i=1}^{d}\sum_{i=1}^{d}\rho_{ij}\sigma_i\sigma_j S_i S_j \frac{\partial^2 u}{\partial S_i \partial S_j} + \sum_{i=1}^{d}(r-\delta_i)\,S_i\frac{\partial u}{\partial S_i} - ru = 0 \tag{1}$$

With

- S_i, the value of underlying asset i
- σ_i volatility of asset i
- ρ_{ij} correlation between asset i and j
- r risk-free interest rate
- δ_i continuous dividend yield

TUDelft

Dividends

- The owner of a stock theoretically owns a piece of the company. This ownership can only be turned into cash, if he owns so many of the stocks that he can take over the company and keep all profits for himself, which is unrealistically (for most of us).
- To the average investor the value in holding the stock comes from the dividends and any growth in the stock's value. Dividends are the lump payments, paid out every quarter or every six months, to the holder of the stock.
- The amount of dividend varies from year to year depending on the profitability of the company. Companies like to try to keep the level of dividends about the same.
- The amount of dividend is decided by the board of directors of the company and is usually set a month or so before the dividend is actually paid.
- When the stock is bought it either comes with its entitlement to the next dividend (cum) or not (ex). There is a date at around the time of dividend payment when the stock goes from cum to ex. The original holder of a stock gets the dividend but the person who buys it obviously does not.

Options on dividend-paying equities

- A stock that is cum dividend is better than one that is ex dividend. Thus at the time that the dividend is paid there will be a drop in the value of the stock. (The jump in asset price is more complex in practice.)
- The price of an option on an dividend-paying asset is affected by these payments. Therefore we must modify the Black-Scholes analysis.
- Different structures are possible for the dividend payment
  - payments may be deterministic or stochastic
  - payments may be made continuously or at discrete times
- Here, we only consider deterministic dividends, whose amount and timing are known.
- Let's consider that in a time dt the underlying asset pays out a dividend $DSdt$ with D constant. The payment is independent of time except through the S-dependence. It represents a continuous and constant dividend yield.
- This structure is a good model for index options.

- Arbitrage considerations show that the asset price must fall by the amount of dividend payment. This is modeled as: $dS = \sigma S dw + (\mu - D) S dt$.
- The dividend payment also has its effect on the hedged portfolio: Since we receive $DSdt$ for every asset held and we hold $-\Delta$ of the underlying, the portfolio changes by an amount $-DS\Delta dt$. Therefore, we add to our $d\Pi$ from before this amount:

  $$d\Pi = dV - \Delta dS - DS\Delta dt.$$

  We find after similar reasoning as for European options that dividend is included in the following formulation:

  $$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + (r - D) S \frac{\partial V}{\partial S} - rV = 0$$
- This model is also applicable to options on foreign currencies, though only for short dated options. Since holding an amount of foreign currency yields interest at the foreign rate r_f, in this case $D = r_f$
- A nonzero dividend yield also has an effect on the boundary and final conditions.

Options on dividend-paying equities

- At the time that a dividend is paid there will be a drop in the value of the stock.
- The price of an option on an dividend-paying asset is affected by these payments.
- Different structures are possible for the dividend payment (deterministic or stochastic with payments continuously or at discrete times)
- We consider discrete deterministic dividends, whose amount and timing are known.
- Arbitrage arguments require:

$$V(S, t_d^-) = V(S - D, t_d^+)$$

Final/Boundary conditions

- European Call option: Right to buy assets at maturity $t = T$ for exercise price K.
- Final condition: $V(S, T) = \max(S - K, 0)$
- Boundary conditions $S = 0$: $V(0, t) = 0$,

  for $S \to \infty$: $V(S_{max}, t) = S_{max} - Ke^{-r(T-t)} - De^{-r(t_d - t)}$ or $V_{ss} = 0$.
- The strategy to solve the Black-Scholes equation numerically is as follows
  - Start solving from $t = T$ to $t = t_d$ with the usual pay-off.
  - Apply an interpolation to calculate the new asset and option price on the grid discounted with D.
  - Restart the numerical process with the PDE from the interpolated price as final condition from t_d to $t = 0$.

Example

- Multiple discrete dividends: analytic solution not available
- Parameters: $s_0 = K = 100,\ r = 0.06, \sigma_c = 0.25$, multiple dividends of 4 (ex-dividend date is each half year), $T = 1, 2, 3, 4, 5, 6$. Grid: $s_{max} = RK(3 \le R \le 7), \mu = 0.15$

Grid	$T=1$	Grid	$T=2$	Grid	$T=3$
20×20	10.660	20×40	15.202	20×80	18.607
40×40	10.661	40×80	15.201	40×160	18.600
Lewis (Wilmott Mag. 2003)	10.661		15.199		18.598
Grid	$T=4$	Grid	$T=5$	Grid	$T=6$
20×80	21.370	20×100	23.697	20×120	25.710
40×160	21.362	40×200	23.691	40×240	25.698
Lewis	21.364		23.697		25.710