CS 476 (Winter 2014 - 1141)

Numerical Computation for Financial Modeling

Prof. P.A. Forsyth
University of Waterloo

LaTeXer: W. Kong
http://wwkong.github.io
Last Revision: April 30, 2014

Table of Contents

These notes are currently a work in progress, and as such may be incomplete or contain errors.

ACKNOWLEDGMENTS:

Special thanks to *Michael Baker* and his LaTeX formatted notes. They were the inspiration for the structure of these notes.

Abstract

The purpose of these notes is to provide the reader with a secondary reference to the material covered in CS 476. The formal prerequisite to this course is CS 371/AMATH 242/CM 271 but this author believes that the overlap between the two courses is less than 30%. Readers should have a good background in linear algebra, basic statistics, and calculus before enrolling in this course.

Errata

User ID: cs476_user_2014

Pass: cs476!2014

url: www.student.cs.uwaterloo.ca/∼cs476

(Used to download the slides)

1 Introduction

The following are just notes regarding the slides in the first lecture.

- Take note of the GBM formula in slide 7 of the this lecture
- Definition of a hedged portfolio is found in slide 10 of this lecture
  - Hedinging objective: $d\Pi = \Pi(t + dt) - \Pi(t) = 0$
  - Under the Black-Scholes equation, the optimal units of the underlying asset in a hedged portfolio is $e = \frac{\partial V}{\partial S}$
- The no-arbitrage value of an option is the cost of setting up the hedged portfolio
- In reality, the bid-ask spread is composed of (profit + compensation for imperfect hedge)
- Problems with GBM?
  - Jump diffusion (1976, slide 24 of this lecture); Black Swans
- Jump diffusion adds the term $(J - 1)\ dq$ to GBM

We now can proceed to some basic terminology.

1.1 Financial Terminology

Consider a **European put/call** at some time T in the future. The holder has the right but not the obligation to buy (sell) the underlying risky asset at some strike price K. The right to buy is given for a call and the right to sell is given for a put. The payoffs of a call and put are respectively

$$\begin{aligned} \max(S - K, 0) &= (S - K)^+ \\ \max(K - S, 0) &= (K - S)^+ \end{aligned}$$

Example 1.1. Suppose a stock today has price S_0 and in 3 months it has two values $S_u = 22$ and $S_d = 18$. For a European call of $K = 21$, the value 3 months later is $V_u = 1$ and $V_d = 0$. If $p_u = 0.1$ and $p_d = 0.9$, then the EPV is

$$(0.1 \times 1 + 0.9 \times 0)e^{-r\triangle t}$$

If the risk free rate is 12% and the time unit is 1 year, then the EPV is 0.097. Suppose that someone offers to buy this option for 0.3. From your point of view, you might expect to profit 0.2, but the other person claims there is an arbitrage opportunity.

Definition 1.1. As an aside, an **arbitrage** opportunity exists if, starting from zero initial wealth, an investor can devise a strategy such that (i) there exists a zero probability of loss (ii) there is a positive probability of gain.

1.2 Arbitrage

Definition 1.2. The **risk-free rate of return** is the return earned by an investment with zero risk.

Example 1.2. An example of a risk-free asset is a government bond. Note that if a portfolio does not earn the risk-free rate then there is an arbitrage opportunity (via shorting/longing the portfolio and buying/selling at the risk free rate). Supply and demand effects (the demand for the portfolio would decrease/increase with return increasing/decreasing) would then normalize the return of the portfolio and push it towards the risk free rate.

Corollary 1.1. *(*No-arbitrage Rule #1) *A portfolio which is risk free must earn the risk free rate.*

Definition 1.3. A **short position** is constructed by borrowing an asset we don't own, selling it, but we have to give it back in the future. If you short it, then mathematically you own a negative quantity.

Continuing from the previous example, instead of examining expected values we use the following idea (1) we will construct a portfolio which has no uncertainty in 3 months (2) since the portfolio is risk free, it must earn the risk free rate. To do this, we construct the portfolio:

$$\Pi = \delta S - V$$

where S is the stock and V is the option. So we compute δ such that Π has a certain value in 3 months. The two possible values are $\Pi_u = 22\delta - 1$ and $\Pi_d = 18\delta$ and setting $\Pi_u = \Pi_d$ we get $\delta = 1/4$ and the value being 4.50. The value, discounting at the risk-free rate, is $4.50e^{0.12\times 0.25} = 4.367$. With $S_0 = 20$ and $\delta = 1/4$, the value of V_0 can be solved to be $V_0 = 0.633$.

We claim that this no-arbitrage price should be the market price. To see this, suppose the price in the market was more than 0.633. The we do the following:

1. Borrow from bank and construct the portfolio $\Pi = -V + \delta S$ (buy stock, sell one call)
2. In 3 months, sell stock, pay off option holder, and pay back the bank loan
3. Profit

Conversely, if the option sells for less than 0.633, the we do the following:

1. Deposit into bank and construct the portfolio $\Pi = V - \delta S$ (short stock, buy one call)
2. In 3 months, use the cash in the bank to buy back the stock, liquidate the short, and take the gain in the option
3. Profit

In both cases, this is a money machine and gains are only limited by borrowing. Note that the no-arbitrage price is independent of p_1 and p_2.

1.3 Hedging

Example 1.3. Here is an example of a hedge using the previous section's main example. Suppose that I sell a call and want to completely hedge the risk where we have $S_0 = 20$, $S_u = 22$, $S_d = 18$, $K = 21$, $r = 12\%$, and $T = 3$ months. This is the strategy:

1. Sell option for 0.633
2. Borrow 4.367 from the bank which means we have to pay back 4.50 in 3 months
3. Buy 1/4 of share of stock at $S_0 = 20$
4. If $S_{1/4} = S_u$ then we have 5.5 in equities so we sell the stock, pay back the option at 1 and pay back the bank loan at 4.5 to get a profit of 0
5. If $S_{1/4} = S_d$ then we have 4.5 in equities so we sell the stock, do not pay back the option (value is 0) and

Of course, no one would want to do this since there is no gain. Usually, we have the effects of the **bid-ask spread**. The ask price is usually the higher one.

2 Stochastic Calculus in Finance

In computational finance, we do not attempt to predict stock prices since they are stochastic. However, there are other applications.

2.1 Brownian Motion

Definition 2.1. A **Brownian motion with drift** is specified as follows. Suppose that X is a random variable and take t to $t+dt$ and X to $X+dX$. We model dX and dZ as

$$dX = \underbrace{\alpha}_{(1)} dt + \overbrace{\sigma}^{(2)} \times \underbrace{dZ}_{(3)}, dZ = \phi\sqrt{dt}$$

where $\phi \sim N(0,1)$, (1) is the drift (2) is the volatility, (3) is the increment of a Wiener process, and

$$\begin{aligned} E[dX] &= E[\alpha \cdot dt] + E[\sigma dZ] = \alpha\ dt + 0 \\ Var[dX] &= E[(dX - E[dx])^2] = E[\sigma^2 dZ^2] = \sigma^2\ dt \end{aligned}$$

We can model Brownian motion using something called a **lattice model** (binomial tree). Let $X(t)$ be the position of a particle at time t. At $t=0$, $X = X_0$ and after an interval $\triangle t$, $t_1 = t_0 + \triangle t$ with $X_0 \mapsto X_0 + \triangle h$ (under probability p) and $X_0 \mapsto X_0 - \triangle h$ (under probability $q = 1-p$). We are assuming that X follows a **Markov process**; that is, the probability distribution of future positions depends on where we are now. Under this structure, we have

$$\begin{aligned} E[\triangle X_i] &= (p-q)\cdot\triangle h \\ E[\triangle X_i^2] &= p(\triangle h)^2 + q(-\triangle h)^2 = (\triangle h)^2 \\ Var[\triangle X_i] &= E[\triangle X_i^2] - (E[\triangle X_i])^2 = 4pq(\triangle h)^2 \end{aligned}$$

Consider some finite time t where the total number of moves is $n = t/\triangle t$. If $X_i = X(t_i)$ then

$$E[(\triangle X_i - E[\triangle X_i])(\triangle X_j - E[\triangle X_j])] = Cov(\triangle X_i, \triangle X_j) \underset{i\neq j}{=} 0$$

We also have the property that

$$E[X_n - X_0] = E\left[\sum_{i=1}^{n-1} \triangle X_i\right] = \sum_i (p-q)\triangle h = n(p-q)\triangle h = \frac{t(p-q)\triangle h}{\triangle t}$$

$$(1)\ Var[X_n - X_0] = \sum_i Var[\triangle X_i] = \frac{t(4pq(\triangle h)^2)}{\triangle t}$$

From here, we plan to take some limits: $\triangle t \to 0$ and simultaneously $n \to \infty$, $\triangle t \to 0$. We first make a few intuitive assumptions:

$$\lim_{\triangle t \to 0} pq \neq 0, \lim_{\triangle t\to 0} pq = O(1)$$

Now if (1) is finite and non-zero as $\triangle t \to 0$ then

$$(2)\ \triangle h = C_1\sqrt{t}$$

where C_1 is independent of $\triangle t$. Substituting (2) into the $X_n - X_0$ equations, we get

$$(3)\ E[X_n - X_0] = \frac{t(p-q)}{\triangle t}\cdot C_1\sqrt{\triangle t}$$

$$(4)\ Var[X_n - X_0] = 4pqtC_1^2$$

If (3) is independent of $\triangle t$ as $\triangle t \to 0$ then

$$(5)\ (p-q) = C_2\sqrt{\triangle t}$$

where C_2 is also independent of $\triangle t$. Putting (5) into (3) gives us

$$(6)\ E[X_n - X_0] = C_1 C_2 t$$

From (4), using $p + q = 1$, we have

$$(7)\ p = \frac{1 + C_2\sqrt{\triangle t}}{2}, q = \frac{1 - C_2\sqrt{\triangle t}}{2}$$

Using (7) in (4) gives us

$$(8)\ Var[X_n - X_0] = (1 - C_2^2 \triangle t) C_1^2 t \overset{\triangle t \to 0}{=} C_1^2$$

Suppose that $[X_n - X_0] \sim dx$ and $[t - 0] \sim dt$. Recall $E[dX] = \alpha\ dt, Var[dX] = \sigma^2\ dt$ and $dx = \alpha\ dt + \sigma\ dZ$. Choose $C_1 = \sigma$ and $C_2 = \alpha/\sigma$ and remark that this causes the first two moments, (8) and (6), between the SDE and the random walk to be the same.

The solution of an SDE is the expression for the probability density of the outcome.

It can be shown that the density of the random walk converges to the solution of the SDE.

Summary 1. A discrete random walk on the lattice with the properties

$$\triangle h = \sigma\sqrt{\triangle t}, p = \frac{1 + \frac{\alpha}{\sigma}\sqrt{t}}{\sigma}, q = 1 - p$$

converges to the solution of $dx = \alpha\ dt + \sigma\ dZ$ where $dZ = \phi\sqrt{\triangle t}$ and $\phi \sim N(0, 1)$.

2.2 Properties of Brownian Motion

As $\triangle t \to 0$, the distance traveled by a particle is $|X_{i+1} - X_i| = |\triangle h|$ and the total distance is

$$n|\triangle X_i| = \frac{t}{\triangle t}|\triangle h| = \frac{t}{\triangle t}\sigma\sqrt{t} \overset{\triangle t \to 0}{=} \infty$$

To reason about this, observe that

$$\frac{\triangle X}{\triangle t} = \frac{\pm\sigma\sqrt{\triangle t}}{\triangle t} \overset{\triangle t \to 0}{=} \pm\infty$$

and hence Brownian motion is **infinitely jagged** on any time scale.

We also may want to ask the question: *Why Geometric Brownian Motion (GBM)?* Let $sd(x)$ be the standard deviation of x and remark that:

1. Asset prices cannot be negative
2. $E\left[\frac{dS}{S}\right] = \mu\ dt \approxeq E\left[\frac{S(t+dt) - S(t)}{S(t)}\right] \sim \mu\ dt$ and so we expect that the return of a stock should not depend on the units
3. $\frac{sd[S(t)]}{E[S(t)]} \sim \sqrt{e^{\sigma^2 t} - 1} \sim e^{\sigma^2 t/2}$ as $t \to \infty$
4. The density of $S(t)$ is generally log-normal and heavily right-skewed

2.3 Ito's Lemma

Suppose we have the SDE

$$(1)\ dx = \alpha(x, t)\ dt + c(x, t)\ dZ$$

We can interpret (1) as

$$(2)\ X(t) - X(0) = \int_0^t \alpha(x(s), s)\ ds + \int_0^t c(x(s), s)\ dZ(s)$$

Let

$$\begin{aligned}\triangle Z(t_j) &= Z(t_{j+1}) - Z(t_j) = Z_{j+1} - Z_j \\ \triangle t &= t_{j+1} - t_j\end{aligned}$$

and define the stochastic integral

$$(3)\ \int_0^t c(x(s), s)\ dZ(s) = \lim_{\triangle t \to 0} \sum_{j=0}^{N-1} c(x(t_j), t_j)(Z_{j+1} - Z_j)$$

where $N = t/\triangle t$ and we call this the **Ito definition**. Note in (3) that we evaluate the first multiplier in the sum on the left hand point on each interval. Alternatively,

$$(4)\ \int_0^t c(x(s), s)\ dZ(s) = \lim_{\triangle t \to 0} \sum_{j=0}^{N-1} c(x(t_{j+\frac{1}{2}}), t_{j+\frac{1}{2}})(Z_{j+1} - Z_j)$$

and we call this the **Stratonovich definition**. In stochastic calculus, (3) and (4) give different answers. In the latter definition, we are "looking forward" in time which is not what general happens in finance.

Now the latter integral in stochastic, so we need to define something called an **Ito integral**. This is the basis for using Ito's Lemma where we need to figure out

$$(5)\ \int_0^t c(x(s), s)\ (dZ(s))^2$$

We claim that

$$(6)\ \int_0^t c(x(s), s)\ (dZ(s))^2 = \int_0^t c(x(s), s)\ ds$$

which in other words,

$$(7)\ \lim_{\triangle t \to 0} \sum c_j \triangle Z_j^2 = \lim_{\triangle t \to 0} \sum c_j \triangle t$$

and in shorthand is $dZ^2 = dt$. The proof (in the course notes) is just showing

$$(8)\ \lim_{\triangle t \to 0} E\left[\left(\sum c_j \triangle Z_j^2 - \sum c_j \triangle t\right)\right] = 0$$

and in the technical proof, we show that if $h(t + \triangle t) = g(t) + O(\triangle t^n)$ as $\triangle t \to 0$ then for $\triangle t$ sufficiently small enough, there exists a c_1 such that

$$(9)\ |h(t + \triangle t) - g(t)| \leq c_1 \triangle t^n$$

Suppose that $F = F(x, t)$, $t \mapsto t\ dt$ and $F \mapsto F\ dF$. Then

$$(10)\ dF = \frac{\partial F}{\partial t} dt + \frac{\partial F}{\partial x} dx + \frac{\partial^2 F}{\partial x^2} \cdot \frac{dx^2}{2} + ...$$

and

$$\begin{aligned}(11)\ (dX)^2 &= (a\ dt + b\ dz)^2 \\ &= a^2 dt^2 + 2ab\ dt\ dZ + b^2 dZ^2 \\ &= b^2 dt + O((dt)^{3/2})\end{aligned}$$

and (11), (10), (2) give us

$$\begin{aligned}(12)\ dF &= F\ dt + F_x dx + F_{xx} \frac{dx^2}{2} + ... \\ &= F\ dt + F_x(a\ dt + b\ dZ) + F_{xx}\left(\frac{1}{2}\right) b^2 dt + O((dt)^{3/2}) \\ &= \left[aF_x + \frac{b^2}{2} F_{xx} + F_t\right] dt + bF_x dZ + O((dt)^{3/2})\end{aligned}$$

and (12) really means

$$F(t) - F(0) = \int_0^t \left[aF_x + \frac{b^2}{2} F_{xx} + F_t \right] dt + \int_0^t bF_x dZ + \underbrace{O\left[\int_0^t (dt)^{3/2} \right]}_{=0}$$

and finally

$$(15)\ F(t) - F(0) = \int_0^t \left[aF_x + \frac{b^2}{2} F_{xx} + F_t \right] dt + \int_0^t bF_x dZ$$

Lemma 2.1. *(Ito's Lemma) If* $dx = a(x,t)\ dt + b(x,t)\ dZ$ *and* $F = F(x,t)$ *then*

$$dF = \left[aF_x + \frac{b^2}{2} F_{xx} + F_t \right] dt + bF_x dZ$$

Summary 2. The following are basic Ito calculus rules, which will be all we need in this course, derived above:

1. $dZ^2 = dt$
2. $\int_0^t c(x(s), s)\ dZ(s) = \lim_{\triangle t \to 0} \sum_{j=0}^{N-1} c(x(t_j), t_j)(Z_{j+1} - Z_j)$
3. Ito's Lemma

2.4 Black-Scholes Equation

We first make a few underlying assumptions:

- Stock prices follows GBM
- Hedger can borrow / lend at the risk-free rate r
- Short selling is permitted and there are no fees

Let $V(S,t)$ be the no-arbitrage value of the claim. Let P be the hedging portfolio, V the value of the option, S the price of the underlying asset, α^h the number of shares of S in P. It is clear that

$$(1)\ P = V - \alpha^h S$$

In some interval, $t \mapsto t + dt$ and $P \mapsto P + dP$ with

$$(2)\ dP = dV - \alpha^h dS$$

and α^h being fixed in this time interval. S follows

$$(3)\ dS = \mu S\ dt + \sigma S\ dZ$$

To compute dV, use Ito's Lemma with $a = \mu S$ and $b = \sigma S$ to get

$$(4)\ dV = [\sigma S V_s]\, dZ + \left[\mu S V_s + \frac{\sigma^2 S^2}{2} V_{ss} + V_t \right] dt$$

Using (2), (3), (4), we can see that

$$(5)\ dP = \sigma S \left[V_s - \alpha^h \right] dZ + \left[\mu S V_s - \mu \alpha^h S + \frac{\sigma^2 S^2}{2} V_{ss} + V_t \right] dt$$

Note that if $\alpha^h = V_s$ then the component with dZ will disappear. So if we set $\alpha^h = V_s$, we get

$$(6)\ dP = \left[V_t + \frac{\sigma^2 S^2}{2} V_{ss} \right] dt$$

So P is risk-free over $[t, t+dt]$ and therefore

$$(7)\ dP = rP\ dt = r(V - \alpha^h S)dt = r(V - V_s S)dt$$

If we set (6) = (7) then

$$(8)\ V_t + \frac{\sigma^2 S^2}{2} V_{ss} + rSV_s - rV = 0$$

Note 1. Equation (8) above is independent of μ.

2.5 Hedging in Continuous Time

Suppose we sell an option at $t = 0$ worth V. The strategy is to sell the option for V which gives us V in cash, borrow $(SV_s - V)$ from the bank. This gives us SV_s in cash to buy V_s shares at price S. The total portfolio is then

$$\underbrace{-V}_{(A)} + \underbrace{SV_s}_{(B)} + \underbrace{V - SV_s}_{(C)} = 0$$

where (A) is the short option position, (B) is the share position and (C) is the cash in the bank. The portfolio is now hedged against small changes in S within a small time step. We now rebalance the hedge so that we own $V_s(s + ds, t + dt)$ shares, making it a dynamic hedge. At any instant in time, we liquidate P, pay off any obligations, at zero gain/loss.

Given initial cash infusion from option buyers, no further cash injection is required and the strategy is **self-financing**. So the Black-Scholes price must be the market price of the traded contracts.

Now remark that if

$$LV \equiv \frac{S^2\sigma^2}{2} V_{ss} + rSV_s - rV$$

then the Black-Scholes equation is $(1)\ V_\tau = LV$. We solve (1) backwards in time (i.e. $\tau = 0 \mapsto \tau = T$). We know that the payoff at $\tau = 0$ is

$$(2)\ V(S, \tau = 0) = \begin{cases} \max(S - K) & \text{call} \\ \max(K - S) & \text{put} \end{cases}$$

As $S \to 0$, (1) becomes $(3)\ V_\tau = -rV$ and as $S \to \infty$

$$(4)\ V(S \to \infty, \tau) = \begin{cases} S & \text{call} \\ 0 & \text{put} \end{cases}$$

(3) and (4) are **boundary conditions** for (1), and (1) to (4) are for European options. If we were examining American options,

$$V(S, \tau) \geq V(S, 0)$$

and hence

$$\begin{aligned} V_\tau - LV &\geq 0\ (5) \\ V - V(S,0) &\geq 0\ (6) \\ [V_\tau - LV]\,[V - V(S,0)] &= 0\ (7) \end{aligned}$$

where (7) implies that one of (5) and (6) are 0. More formally,

$$\min\left[V_\tau - LV, V - V(S,0)\right] = 0$$

This is called an **HJB PDE** and is related to **optimal stochastic control**.

3 Computational Methods

We describe, in this section, the various computational methods in finance.

3.1 Lattice Model for Pricing

Suppose that $dS = \mu S\ dt + \sigma S\ dZ$ and let $X = X(S,t)$ with $dX = (X_s \mu S + X_{ss}\frac{\sigma^2 S^2}{2} + X_t)dt + X_s \sigma S\ dZ$ via Ito's Lemma. Then if $X = \ln S$, we have $X_s = 1/S$, $X_{ss} = -1/S^2$ and $X_t = 0$. The above equations then give us

$$(1)\ dX = \left[\mu - \frac{\sigma^2}{2}\right] dt + \sigma\ dZ$$

Recall the results of the random walk on a lattice. This gives us

$$(2)\ \triangle h = \sigma\sqrt{\triangle t}, p = \frac{1}{2}\left[1 + \frac{\alpha}{\sigma}\sqrt{\triangle t}\right], q = 1 - p$$

This walk converges to the SDE

$$dX = \alpha\ dt + \sigma\ dZ$$

as $\triangle t \to 0$. This is the same as (1) if $\alpha = \mu - \frac{\sigma^2}{2}$. With this substitution, (2) converges to GBM as $\triangle t \to 0$. Let X at node j, time step n be X_j^n. Note that

$$\begin{aligned} X_{j+1}^{n+1} &= X_j^n + \sigma\sqrt{\triangle t} \\ X_j^{n+1} &= X_j^n - \sigma\sqrt{\triangle t} \end{aligned}$$

Since $X = \ln S$, we can write the above as

$$\begin{aligned} S_{j+1}^{n+1} &= S_j^n e^{\sigma\sqrt{\triangle t}} \\ S_j^{n+1} &= S_j^n e^{-\sigma\sqrt{\triangle t}} \end{aligned}$$

where $u = e^{\sigma\sqrt{\triangle t}}, d = e^{-\sigma\sqrt{\triangle t}}$. The hedging portfolio is constructed as

$$P_j^n = V_j^n - (\alpha^h)_j^n S_j^n$$

where $(\alpha^h)_j^n$ is the number of shares of S_j^n at (j,n). Note that we want to choose $(\alpha^h)_j^n$ such that $P_{j+1}^{n+1} = P_j^{n+1}$ and hence

$$V_j^{n+1} - (\alpha^h)_j^n S_j^{n+1} = V_{j+1}^{n+1} - (\alpha^h)_j^n S_{j+1}^{n+1} \implies \frac{V_{j+1}^{n+1} - V_j^{n+1}}{S_{j+1}^{n+1} - S_j^{n+1}}$$

Let $\triangle t \to 0$ and $S_{j+1}^{n+1} \to S_j^n$. The above equation becomes $\left(\frac{\partial V}{\partial S}\right)_j^n$ which we call out **delta**. We thus have that $P(t^{n+1})|P_j^n$ is certain (is not a random variable) and

$$P_j^n = e^{-r\triangle t} P_{j+1}^{n+1} = e^{-r\triangle t} P_j^{n+1}$$

Using the above three equations, we get

$$\begin{aligned} V_j^n &= e^{-r\triangle t}\left[p^* V_{j+1}^{n+1} + (1-p^*) V_j^{n+1}\right] \\ p^* &= \frac{e^{r\triangle t} - e^{-\sigma\sqrt{\triangle t}}}{e^{\sigma\sqrt{\triangle t}} - e^{-\sigma\sqrt{\triangle t}}} \end{aligned}$$

Note that the actual probabilities p, q do not appear in the above. If $\sigma > 0, \triangle t$ is small, then

$$0 \le p^* \le 1$$

where we call p^* the **risk-neutral probability** and not the **real probability**.

Example 3.1. (European Call) See Section 5.1. in Course Notes and Question 1 of Assignment 1.

3.2 Dynamic Programming

From any point on an optimal trajectory, the remaining trajectory is optimal for the corresponding problem initiated at that point.

Example 3.2. Determine the strategy to maximize the expected wealth of the your marriage partner.

(1) You meet at most N persons

(2) Each person has wealth $U \sim [0,1]$

(3) At step k, you can determine the wealth of person k

(4) If you ask a person to marry you, they have to accept

At step $(N-1)$, two choices: ask $(N-1)$ to marry you or marry N. We know that the expected wealth at any node is 0.5 and so if $(N-1)$ has wealth > 0.5, you stop and otherwise continue.

Let V_N be the expected partner wealth at step N and W_{N-1} be the wealth of $(N-1)$ assuming it is optimal to stop at $(N-1)$. Denote $\bar{W}_{N-1} = E\left[W_{N-1}|V_N\right]$ and $p \equiv P(W_{N-1} > V_N)$. We have:

$$\begin{aligned} V_{N-1} &= p\bar{W}_{N-1} + (1-p)V_N \\ V_N &= 0.5 \\ P(W_{N-1} > 0.5) &= p = 0.5 \\ \bar{W}_{N-1} &= 0.75 \end{aligned}$$

and hence we have

$$\begin{aligned} V_{N-1} &= (0.5)(0.75) + (0.5)(0.5) = 0.625 \\ V_k &= p\bar{W}_k + (1-p)V_{k+1} \end{aligned}$$

As a test, if $N = 10$, you should get that $V_1 \sim 0.861$.

3.3 Risk Neutral World

Suppose that we are receiving some risky cash flows in the future and the discount rate is $\rho > r$. Today, we are in the state (t, S) where S is the risky asset price. Let $\hat{V}(S,t)$ be the expected value of the option. Suppose that

$$\frac{dS}{S} = \mu\, dt + \sigma\, dZ$$

Let $\hat{V}(S+dS, t+dt)$ be the expected value at $(S+dS, t+dt)$ where

$$\hat{V}(S,t) = \frac{1}{1+\rho dt} E\left[\hat{V}(S+dS, t+dt)\right]$$

and $\frac{1}{1+\rho dt} = e^{-\rho dt} + O(dt^2)$. Now

$$E\left[\hat{V}(S+dS, t+dt)\right] = E\left[\hat{V}(S+dS, t+dt)|\hat{V} = \hat{V}(s,t)\right]$$

with $\hat{V}(S+dS, t+dt) \approx \hat{V}(S,t) + d\hat{V}$ we get

$$\begin{aligned} \hat{V}(S,t) &= \frac{1}{1+\rho\, dt}\left[E[\hat{V}(S,t) + E[d\hat{V}]\right] \\ &= \frac{1}{1+\rho\, dt}\left[\hat{V}(S,t) + E[d\hat{V}]\right] \end{aligned}$$

Simplifying this gives us

$$(1)\ \rho\, dt \cdot \hat{V}(S,t) = E[d\hat{V}]$$

and since with Ito's Lemma we have

$$d\hat{V} = \left[\hat{V}_t + \frac{\sigma^2 S^2}{2}\hat{V}_{ss} + \mu S\hat{V}_s\right] dt + \sigma S\hat{V}_s dZ$$

and hence
$$E[d\hat{V}] = \left[\hat{V}_t + \frac{\sigma^2 S^2}{2}\hat{V}_{ss} + \mu S\hat{V}_s\right] dt$$
and putting this into (1) gives us
$$(2)\ \hat{V}_t + \frac{\sigma^2 S^2}{2}\hat{V}_{ss} + \mu S\hat{V}_s - \rho\hat{V} = 0$$
Let $\tau = T - t$ and put this into (2) to get
$$\hat{V}_\tau = \frac{\sigma^2 S^2}{2}\hat{V}_{ss} + \mu S\hat{V}_s - \rho\hat{V}$$
which is the B-S equation with $\mu = r$ and $\rho = r$. The result says that we can compute the no-arbitrage by pretending that
$$\frac{dS}{S} = \underbrace{r}_{\neq\mu}\ dt + \sigma\ dZ$$
and discounting the expected payoff at r. This world is called the **risk-neutral world** where all risky assets drift at r, and all cash flows are discounted at r. Under these constraints, the **no-arbitrage price** of the option is
$$V = e^{-r(T-t)}E_Q\left[\text{Payoff}\right]$$
where Q is the probability measure under the risk-neutral world. However, the **real-expected value** is
$$\hat{V} = e^{-\rho(T-t)}E_P\left[\text{Payoff}\right]$$

3.4 Monte Carlo

We simulate the asset price forward in time in the Q measure. The simulated equation is $\frac{dS}{S} = r\ dt + \sigma\ dZ$ and we repeat this many times. To simulate a path, choose a finite timestep $\triangle t = T/N$ where in general,
$$S^{n+1} = S^n + S^n\left[r\triangle t + \sigma\phi^n\sqrt{\triangle t}\right], \phi^n \sim N(0,1)$$
Denote the payoff on the m^{th} simulation by pay^m. Then the option value is
$$e^{-r(T-t)}E_Q\left[\text{Payoff}\right] \approx e^{-r(T-t)}\sum_{m=1}^{M}\frac{pay^m}{M}$$
There are two sources of error in Monte-Carlo (MC):

- [1] Timestepping Error:
  - Consider the algorithm above, defined as
    $$\underbrace{S^{n+1} = S^n + S^n\left[r\triangle t + \sigma\phi^n\sqrt{\triangle t}\right]}_{\text{Forward Euler}}$$
    * We use a probabilistic definition of error. Let $S(T)$ be the exact solution of $\frac{dS}{S} = r\ dt + \sigma\ dZ$. In reality $S(T)$ represents a probability distribution, but using the forward Euler method generates an approximate solution $S^{\triangle t}(T)$. Suppose we would like to compute the expected value of some function of S. In particular, $E[f(S(T)]$ where f is the payoff function. We say that a numerical method for solving an SDE has **weak error of order** γ if for $\triangle t$ sufficiently small, $\exists c$ such that
    $$\left|E\left[f(S(t))\right] - E\left[f(S^{\triangle t}(T))\right]\right| \leq c(\triangle t)^\gamma$$
    * It turns out that Forward Euler has weak error order one ($\gamma = 1$).
- [2] Sampling Error

  - Here, we use an approximate expectation using a sample average. Note that if we have M samples, $(\text{Error})_{Sampling} = O\left(\frac{1}{\sqrt{M}}\right)$.

- [3] Total Error
  - We define this as a combination of the two above errors. Specifically, $\text{Error} = O\left(\max\left(\triangle t, \frac{1}{\sqrt{M}}\right)\right)$.
  - If we choose $M = \frac{C_0}{(\triangle t)^2}$ where C_0 is some constant, then the total error becomes $O(\triangle t)$ and the complexity is

$$\text{Complexity} = O\left(\frac{M}{\triangle t}\right) = O\left(\frac{1}{(\triangle t)^3}\right) \sim \frac{C_1}{(\triangle t)^3} \implies \triangle t = \frac{C_1^{1/3}}{\text{Complexity}^{1/3}} = O\left(\frac{1}{\text{Complexity}^{1/3}}\right)$$

  where C_1 is some constant. This also implies

$$\text{Error} = O(\triangle t) = O\left(\frac{1}{\text{Complexity}^{1/3}}\right)$$

  - E.g. If we were to reduce the error by 10 (one decimal digit), then it takes 1000x longer to converge (VERY slow to converge)

Alternatively, we can use a statistical estimate of the sampling error. Let

$$\hat{\mu} = \left[\frac{1}{M}\sum_{m=1}^{M} \text{pay}^m\right] e^{-r(T-t)}$$

and w be the standard deviation defined as

$$w = \left[\frac{\sum_{m=1}^{M}\left(e^{-r(T-t)}\text{Pay}^m - \hat{\mu}\right)^2}{M-1}\right]$$

The real mean V satisfies

$$\hat{\mu} - \frac{X^* w}{\sqrt{M}} \le V \le \hat{\mu} + \frac{X^* w}{\sqrt{M}}$$

with probability $(1-\alpha)$ where $X^* = \Phi^{-1}(1-\alpha)$.

If the error is so large, then why is Monte Carlo so popular?

1. It's easy to code
2. If the number of underlying stochastic factors is more than 3, then MC complexity is better than a PDE method
3. Goes wrong way (forward) in time

Example 3.3. Imagine a simple case of a Bermudan option, where we can only exercise at one time $t_1 \in [0, T]$.

The *obvious method* is to generate paths from 0 to t_1 and then at t_1, spawn many new paths from $[t_1, T]$. Take the expected terminal values of these new paths, discount them and take the max of the exercise and continuation (discounting terminal values), and discount this to time 0. Note that the complexity increases exponentially as more exercise points are created.

In the limit, we have an American option which implies that it is also exponential complexity in a Monte Carlo algorithm. There have been a few advancements in techniques: Longstaff-Schwarz algorithm, BSDE (backwards SDE).

Note 2. (Special Trick) If we have GBM with constant coefficients (i.e. $\frac{dS}{S} = r\,dt + \sigma\,dZ$ where r and σ are constant), then the exact solution is

$$S_t = S_0 \exp\left[\left(r - \frac{\sigma^2}{2}\right)t + \sigma\phi\sqrt{t}\right], \phi \sim N(0,1)$$

which will be exact for any t. In practice, $\sigma = \sigma(S,t)$ as a local volatility surface and $r = r(t)$ as yield curve. So instead, suppose

$$dS = r(t)S\,dt + \sigma(S,t)S\,dZ$$

If $r(t)S \to 0, \sigma(S,t)S \to 0, S \to 0$, then the exact solution cannot go negative. Under Forward Euler,

$$S^{n+1} = S^n + S^n \left[r(t^n)\triangle t + \sigma(S^n, t^n)\phi^n \sqrt{\triangle t} \right]$$

where this can be negative. Instead, let $X = \ln S$ where

$$dX = \left[r - \frac{\sigma^2(e^X, t)}{2} \right] dt + \sigma(e^x, t) dZ$$

and the Forward Euler is

$$X^{n+1} = X^n + \left[r - \frac{\sigma^2}{2} \right] \triangle t + \sigma\phi^n \sqrt{\triangle t}$$

We can recover S^{n+1} using $S^{n+1} = S^n e^{(X^{n+1} - X^n)}$ where using X as the dependent variable, we never get a negative value for S. A suggestion for remedying this is to use the **CIR Model** which has the form

$$dr = a(b - r)dt + \sigma\sqrt{r}dZ$$

where we cannot use $X = \ln r$.

3.5 Measuring Risk

There are different measures of risk in finance. The basic measure is standard deviation, but we also have **value-at-risk (**VaR) which is a measure of **tail risk**. Loosely speaking, it is when "y% of th time the we can lose no more than $\hat{X}$" and more formally, it is the point VAR on the x-axis where

$$\int_{VAR}^{\infty} p(t)\ dt = \frac{y}{100}$$

Typically, people quote $V\hat{A}R = -VAR$. Suppose that I have two heding portfolios $P_1(t)$ and $P_2(t)$. Then the following is NOT true:

$$V\hat{A}R(P_1(t) + P_2(t)) \leq V\hat{A}R(P_1(t)) + V\hat{A}R(P_2(t))$$

which is called the **subadditive property**. VaR is NOT subadditive but VaR IS in banking regulations. An alternative which is subadditive is CVaR which is

$$CVAR = \left(\int_{-\infty}^{VAR} tp(t)\ dt \right) / \left(\int_{-\infty}^{VAR} p(t)\ dt \right) = E[X | X \leq VAR]$$

Like VaR, people typically quote $CV\hat{A}R = -CVAR$.

3.6 Hedging Portfolios

We also have errors relating to the process of hedging:

1. We can't hedge continuously
2. When we buy/sell we lose the spread
3. Model of stochastic processes may be wrong
4. Process could have jumps

But, remember:

"All models are wrong; some are useful"

Here, we will discuss a method of hedging called **delta-gamma hedging**. Suppose you are short an option worth $V(S,t)$ and the hedging portfolio is $P(t)$ where

$$P = -V + \alpha S + B(t)$$

where S is some underlying asset and $B(t)$ is the value in the bank. Our hedging objective is to have $P(t) = 0$ for any t. Black-Scholes tells us that

$$\alpha = V_s = \text{Delta} = \triangle, B(t) = V - V_s S$$

which makes $P(t) = 0$. Let's re-derive the above using a more intuitive approach. Suppose that $t \mapsto t + \triangle t, S \mapsto S + \triangle S$ where the changes are finite. We'll ignore the changes due to simple changes in t and instead we will worry about the random changes $S \mapsto \triangle S$. We then would have

$$\triangle P = -V_s \triangle S + \alpha \triangle S + 0 + O(\triangle S^2)$$

We then choose $\alpha = V_s$ to get $\triangle P = 0$ up to $O(\triangle S^2)$. If we call

$$S_0 = S(t=0), V_0 = V(t=0), B_0 = B(t=0)$$

If we have discrete times $t_0, t_1, ...$ where we rebalance the portfolio at these times, then at t_0 we have

$$P_0 = -V(0) + (V_s)_0 S(0) + B(0) = 0 \implies B_0 = V_0 - \alpha_0 S_0, \alpha_0 = (V_s)_0$$

by the Black-Scholes equation. Define

$$t_{i-1}^+ = \lim_{\epsilon \to 0^+} t_{i-1} + \epsilon, t_{i-1}^- = \lim_{\epsilon \to 0^+} t_{i-1} - \epsilon$$

$$[B][V][S][\alpha]_{i-1} := [B][V][S][\alpha](t_{i-1}^+)$$

We have

$$P(t_{i-1}^+) = -V_{i-1} + \alpha_{i-1} S_{i-1} + B_{i-1}$$

As $t_{i-1}^+ \to t_i^-$ we have $S_{i-1} \to S_i$ (evolves randomly), $V_{i-1} \to V_i$, $B_{i-1} \to B_{i-1} e^{r \triangle t}$, and $\alpha_{i-1} \to \alpha_i$. At $t = t_i^+$ we adjust α in order to remain delta neutral. Next, since

$$\begin{aligned} P(t_i^-) &= -V_i + \alpha_{i-1} S_i + B_{i-1} e^{r \triangle t} \\ P(t_i^+) &= -V_i + \alpha_i S_i + B_i \end{aligned}$$

and the self financing condition holds, then

$$P(t_i^-) = P(t_i^+) \implies B_i = B_{i-1} e^{r \triangle t} - S_i(\alpha_i - \alpha_{i-1})$$

If $B_i > 0$, we deposit cash in the bank and earn the rate r and if $B_i < 0$, then we borrow at the rate r. It is not likely that $P(t_i) \equiv 0$ since $\triangle t$ is finite. Instead, let's examine the error in terms of the relative P/L. We define this as

$$\text{Relative P/L} = \frac{e^{-rt} P(t)}{|V(0)|}$$

where $V(0)$ is the initial option premium. In summary, with regards to the process in practice and discrete hedging error, the steps are:

1. Take out $B(0) = V(0) - \alpha_0 S_0$ to construct the portfolio.
2. Simulate a Monte-Carlo path from 0 to T. Along this simulated path, we delta hedge at intervals $\triangle t_{hedge}$ and record $B(t_i), \alpha(t_i)$, etc. [1]
3. At $t = T$, record the relative P/L as $e^{-rT} P(T)/|V(0)|$ where we convert everything to cash.
4. Repeat this many times and generate a histogram of the P/L.

[1] Make sure you remember $\triangle t_{MC} \neq \triangle t_{hedge}$ where MC is **M**onte-**C**arlo

Note that if X is the P/L and $p(X)$ is the probability density then $p(x)\ dx$ is the probability of P/L in $[x, x+dx]$,

$$p(a) \sim \frac{1}{b-a} \cdot \frac{\text{\# of occurences in } [a,b]}{\text{total \# of outcomes}}$$

Remark that under any hedging strategy, there would be more concentration in the density function around 0 P/L.

Delta-Gamma Hedging

Suppose we add an additional security to the hedging portfolio. Let the value of this security be $I(s,t)$. The new portfolio is

$$P(t) = -V + \alpha S + \beta I + B$$

where β is the number of units held in this security. As $S \mapsto S + \triangle S$,

$$P(t+\triangle t, S+\triangle S) = P(t,S) + P_t(t,S)\triangle t + \frac{\partial P}{\partial S}\triangle S + \frac{\partial^2 P}{\partial S^2} \cdot \frac{(\triangle S)^2}{2} + ...$$

and if we define $\triangle P = P(t+\triangle t, S+\triangle S) - P(t,s)$, then

$$\triangle P = P_t \triangle t + P_S \triangle S + P_{SS}\frac{(\triangle S)^2}{2} + ...$$

If we ignore the P_t term (assume it is deterministic) and set $\frac{\partial P}{\partial S} = 0$ and $\frac{\partial^2 P}{\partial S^2} = 0$. The former is called **delta** and the latter is called **gamma**. Then, when we calculate these expressions explicitly:

$$\begin{aligned} \frac{\partial P}{\partial S} &= \frac{\partial}{\partial S}\left[-V + \alpha S + \beta I + B(t)\right] = -V_s + \alpha + \beta I_s + 0 \\ \frac{\partial^2 P}{\partial S^2} &= -V_{ss} + \beta I_{ss} \end{aligned}$$

this implies that (1) $-V_s + \alpha + \beta I_s = 0$ and (2) $-V_{ss} + \beta I_{ss} = 0$. We use the previous equations to calculate β at each rebalance point Note that $\beta = V_{ss}/I_{ss}$ and we want I to have large gamma ($|I_{ss}| \gg 1$). At time $t = 0$, $P(0) = 0$, $B(0) = V_0 - \alpha_0 S_0 - \beta I_0$. At the i^{th} rebalance time,

$$(3)\ B_i = e^{r\triangle t}B_{i-1} - S_i(\alpha_i - \alpha_{i-1}) - I_i(\beta_i - \beta_{i-1})$$

Delta-gamma hedging is defined by equations (1), (2), (3). This strategy thins the tails where the right tail is thicker than the left.

Delta-Vega Hedging

Suppose instead, we want to hedge against random changes in volatility. That is, we regard (S, σ) as random variables and set $\frac{\partial P}{\partial S} = 0$ and $\frac{\partial P}{\partial \sigma} = 0$ where the latter is called **vega** and

$$\frac{\partial P}{\partial \sigma} = -V_\sigma + 0 + \beta I_\sigma + 0 = 0 \implies \beta = V_\sigma / I_\sigma$$

3.7 Random Number Generation

Standard library algorithms generate pseudo-random numbers $X \sim U[0,1]$ and until recently, library software for random number generation was very bad. Imagine sampling a hypercube in d dimensions. The standard approach is to generate numbers X_i for $i = 1, ..., d$ where $X_i \sim U[0,1]$.

For this random d dimensional vector $\mathbf{X} = [X_1 \cdots X_d]$, the sample points would lie on hyperplanes in the cube!

The “Gold Standard” method for generating $U[0,1]$ is the **Mersenne Twister** which has many public codes available online. See the notes for reference [1998]. The properties of the twister are:

- It has period $2^{19937} - 1$

  - That is, if we make 10^9 draws/second, the time to repeat a number $\gg$ the age of the universe

- Probably good properties for $d \leq 623$
- The Matlab default is the Twister

Assume we have a good method for generating $U[0,1]$ numbers. We want to generate $Y \sim N(0,1)$. To do this, we use the concept of a **transformation of probabilities**. To begin, suppose that $p(x)\ dx$ is the probability of finding $x \in [x, x+dx]$. It is obvious that

$$P(a \leq x \leq b) = \int_a^b p(x)\ dx, \int_{-\infty}^{\infty} p(x)\ dx = 1$$

In the case of a standard uniform random variable (r.v.), we have $p(x) = \chi_{x\in[0,1]}$. Assume that X is a r.v. with density $p(x)$ and let $Y = Y(X)$, assuming $\frac{dY}{dX} > 0$. That is, it is a monotone increasing function. Suppose we can solve $Y = Y(X)$ for X given Y where $X = X(Y)$. Then,

$$(1)\ P(X(\alpha) \leq X \leq X(\beta)) = \int_{X(\alpha)}^{X(\beta)} p(x)\ dx$$

$$(2)\ P(\alpha \leq Y \leq \beta) = \int_{\alpha}^{\beta} \hat{p}(y)\ dy$$

where $(1) = (2)$, $\hat{p}$ is the density of y, and

$$\int_{\alpha}^{\beta} \hat{p}(y)\ dy = \int_{X(\alpha)}^{X(\beta)} p(x)\ dx = \int_{\alpha}^{\beta} \left[p(X(y))\frac{dX}{dy}\right]\ dy$$

As we let $[\alpha,\beta] \to 0$ we get that $p(X(y))\frac{dX}{dy} = \hat{p}(y)$. We repeat the argument for $\frac{dY}{dX}$ to the get the result that given a random variable X with density $p(x)$ and a monotone function $y = y(x)$, then the density of y, $\hat{p}(y)$ is

$$\hat{p}(y) = p(X(y))\left|\frac{dX}{dy}\right|$$

In our original problem, we have $X \sim U[0,1]$ and we want to find $Y(X)$ such that $Y \sim N(0,1)$. That is, we want

$$\hat{p}(Y) = \frac{1}{\sqrt{2\pi}}e^{-y^2/2}, y \in (-\infty, +\infty)$$

Since $p(x) = \chi_{x\in[0,1]}$, then

$$\chi_{x\in[0,1]}\left|\frac{dX}{dy}\right| = \chi_{x\in[0,1]}\frac{dX}{dy} = \frac{e^{-y^2/2}}{\sqrt{2\pi}} \implies \frac{e^{-y^2/2}}{\sqrt{2\pi}}dy = \chi_{x\in[0,1]}dX$$

and integrating,

$$X = \int_{-\infty}^{y} \frac{e^{-y^2/2}}{\sqrt{2\pi}}dy + C$$

We can set the constant by noting that $X \to 0 \implies y \to -\infty$ and $X \to 1 \implies y \to \infty$. It can be shown that

$$X = \int_{-\infty}^{y} \frac{e^{-y^2/2}}{\sqrt{2\pi}}dy$$

This actually defines the transformation that we need and is the called the **Fundamental Law of Transformation of Probabilities**. So if we generate $X \sim U[0,1]$ then to generate $Y \sim N(0,1)$ we find y such that

$$X = \int_{-\infty}^{y} \frac{e^{-y^2/2}}{\sqrt{2\pi}}dy$$

or $y = F^{-1}(x)$ where F^{-1} is the inverse cumulative normal distribution function. While there is a built-in Matlab function, evaluating F^{-1} is very slow. To speed up computation, we use the **Box-Muller method** as follows.

Suppose we have a set of random points, (X_1, X_2) in 2D. Suppose the density is $p(x_1, x_2)$. Suppose there exists functions $Y_1 = Y_1(X_1, X_2)$ and $Y_2 = Y_2(X_1, X_2)$. The 2D version of the Fundamental Theorem of Transformation of Probabilities is

$$(3)\ \hat{p}(Y_1, Y_2) = p(X_1, X_2)|J|, J = \det \begin{bmatrix} \frac{\partial X_1}{\partial Y_1} & \frac{\partial X_1}{\partial Y_2} \\ \frac{\partial X_2}{\partial Y_1} & \frac{\partial X_2}{\partial Y_2} \end{bmatrix}$$

Suppose that $(X_1, X_2) \sim (U[0,1])^2$ with $X_1 \sim U[0,1]$, $X_2 \sim U[0,1]$, and $p(X_1, X_2) = \chi_{(X_1,X_2)\in[0,1]^2}$. So given this description along with $Y_1 = Y_1(X_1, X_2), Y_2 = Y_2(X_1, X_2)$ then $\hat{p}(Y_1, Y_2) = |J|$ from (3). Consider

$$\begin{aligned} Y_1 &= \sqrt{-2\log X_1}\cos(2\pi X_2) \\ Y_2 &= \sqrt{-2\log X_1}\sin(2\pi X_2) \end{aligned}$$

We then have the solutions for (X_1, X_2) being

$$\begin{aligned} X_1 &= \exp\left[-\frac{Y_1^2 + Y_2^2}{2}\right] \\ X_2 &= \frac{1}{2\pi}\arctan\left[\frac{Y_2}{Y_1}\right] \end{aligned}$$

and

$$(4)\ |J| = \frac{1}{\sqrt{2\pi}}e^{-Y_1^2/2}\frac{1}{\sqrt{2\pi}}e^{-Y_2^2/2}$$

If $(X_1, X_2) \sim U[0,1]^2$ the using (4), we generate $\hat{p}(Y_1, Y_2)$ as a product of standard normals where $Y_1 \sim N(0,1), Y_2 \sim N(0,1)$. So the general algorithm for the *standard Box-Muller* is:

1. Generate $X_1 \sim U[0,1], X_2 \sim U[0,1]$
2. Set $\rho = \sqrt{-2\log X_1}$
3. Set $Y_1 = \rho\cos(2\pi X_2)$ and $Y_2 = \rho\sin(2\pi X_2)$

We can make this even more efficient using the **Polar Form Box-Muller** where we generate a random draw in the (X_1, X_2) plane such that $(X_1, X_2) \sim D(0,1)$ where $D(0,1)$ is the unit disk. To do this, we can try the following:

1. Create $X_1 \sim U[0,1], X_2 \sim U[0,1], V_1 = 2X_1 - 1, V_2 = 2X_2 - 1, w = V_1^2 + V_2^2, \theta = \tan(V_2/V_1)$
2. If $w < 1$ then we accept
3. Otherwise go to Step 1

The probability of a successful draw is about $\pi/4$. It turns out that $\theta \sim U[0, 2\pi], w \sim U[0,1]$. Also note that

$$\begin{aligned} (5)\ \cos\theta &= V_1/\sqrt{w} \\ (6)\ \sin\theta &= V_2/\sqrt{w} \end{aligned}$$

We can now create the modified algorithm as follows:

1. Generate $X_1 \sim U[0,1], X_2 \sim U[0,1], V_1 = 2X_1 - 1, V_2 = 2X_2 - 1, w = V_1^2 + V_2^2, \theta = \arctan(V_2/V_1)$
2. Replace X_1 by $w \sim U[0,1]$ to set $\rho = \sqrt{-2\log w}$
3. Replace $2\pi X_2$ with θ to set $Y_1 = \rho\cos(\theta)$ and $Y_2 = \rho\sin(\theta)$
4. Replace the trigonometric functions with the values above

The final form, implemented in actual algorithms, is:

1. Create $X_1 \sim U[0,1], X_2 \sim U[0,1], V_1 = 2X_1 - 1, V_2 = 2X_2 - 1, w = V_1^2 + V_2^2, \theta = \arctan(V_2/V_1)$
2. If $(w < 1)$, set $Y_1 = V_1\sqrt{-\frac{2\log w}{w}}, Y_2 = V_2\sqrt{-\frac{2\log w}{w}}$
3. Otherwise go to Step 1

3.8 Efficiency of Algorithms

A basic algorithm to compute the **mean** is to do the following:

Algorithm 1 Basic summation

```
sum = 0
for i = 1,...,M
  sum = sum + Xi
end
mean = sum / M
```

For single precision, the above algorithm is prone to floating point error accumulation. Older GPUs use single precision while newer ones can do double precision. To avoid loss of precision, do recursive pairwise summation.

Next, consider the basic **integral** approximation method in one dimension, under Monte Carlo, where

$$I_1 = \int_0^1 f(x)\, dx \approx \frac{1}{M}\sum_{i=1}^{M} f(X_i)$$

and $X_i \sim U[0,1]$. Using Monte Carlo, the work is $O(M)$ and the error is $O(1/\sqrt{M})$. Alternatively, we can use the **midpoint rule** of

$$I_1 = \int_0^1 f(x)\, dx = \sum_{i=1}^{M} f(X_i)\triangle X + O(\triangle X^2)$$

where we have the same amount of complexity but the error is $O(1/M^2) = O(\triangle X^2)$. So the midpoint rule is more efficient error-wise. Now imagine a d dimensional integral $I_d = \int_0^1 \int_0^1 ... \int_0^1 f(\vec{x})d\vec{x}$. Under Monte Carlo we have,

$$I_d = \frac{1}{M}\sum_{i=1}^{M} f(\vec{x}) + O(1/\sqrt{M})$$

and under the midpoint rule, where we fix the same work as Monte Carlo (fixed number of cells), we have

$$M = \frac{1}{(\triangle X)^d} \implies \triangle X = \frac{1}{M^{1/d}}$$

$$I_d = \int_0^1 f(x)\, dx = \sum_{i=1}^{M} f(X_i)\,(\triangle X)^d + O(\triangle X^2)$$

Here, the work or complexity is the same as Monte Carlo (because we fixed it that way), but the error of the midpoint rule is $O(\triangle X^2) = O(1/M^{2/d})$. Note that this means that when the dimensionality is bigger than 4, then Monte Carlo beats the mid-point rule error-wise.

Generalizing this to finance problems, consider the problem to pricing a high dimensional American Options. We have several methods in modern research today:

- Monte Carlo Backwards Stochastic Differential Equations (BSDEs)
- Sparse Grid Methods for PDEs

3.9 Correlated Random Numbers

Here we give some examples of options that depend on a basket of stock. The general form of an option on a basket of stocks $\{S_1, S_2, ..., S_d\}$ has payoff

$$\max\left[\max\{S_1, S_2, ..., S_d\} - K, 0\right]$$

For example, a spread option with

$$\begin{aligned} dS_1 &= \mu_1 S_1 dt + \sigma_1 S_1 dZ \\ dS_2 &= \mu_2 S_2 dt + \sigma_2 S_2 dZ \end{aligned}$$

with the correlation structure $dZ_1 \cdot dZ_2 = \rho_{12} dt$ has the payoff

$$\max\left[(S_1 - S_2) - K, 0\right]$$

Suppose we have observations of asset i, S_i, at $t_k = k\triangle t$. The return of asset i in $t_k \mapsto t_k + \triangle t$ is

$$R_i(t_k) = \frac{S_i(t_k + \triangle t) - S_i(t_k)}{S_i(t_k)} \implies \bar{R}_i = \sum_{i=1}^{M} \frac{R_i(t_k)}{M}$$

with historical volatility

$$\sigma_i = \sqrt{\frac{1}{\triangle t(M-1)} \sum_{k=1}^{M} (R_i(t_k) - \bar{R}_i^2)}$$

and covariance

$$Cov(R_i, R_j) = \frac{1}{\triangle t(M-1)} \sum_{k=1}^{M} (R_i(t_k) - \bar{R}_i)(R_j(t_k) - \bar{R}_j)$$

This gives us the correlation

$$Cor(R_i, R_j) = \rho_{ij} = \frac{Cov(R_i, R_j)}{\sigma_i \sigma_j}, \rho_{ij} \in [-1, 1]$$

Suppose we use Monte Carlo (MC) to price a d dimensional basket:

$$S_i^{n+1} = S_i^n + S_i^n [r\triangle t + \sigma_i \phi_i^n \sqrt{\triangle t}]$$

In order to simulate correlated GBM, we need

$$\phi_i^n \sim N(0,1), E[\phi_i^n \phi_j^n] = \rho_{ij}$$

but algorithms seen so far only generate independent variables.

Problem 3.1. Given independent ϵ_1 to ϵ_d with $\epsilon_i \sim N(0,1)$, we have $E[\epsilon_i \epsilon_j] = \delta_{ij}$ where δ_{ij} is the **Kronecker-Delta function**. We would like to generate ϕ_1 to ϕ_d with $\phi_i \sim N(0,1)$ and $E[\phi_i \phi_j] = \rho_{ij}$.

Solution. To solve this, assume that we are given a matrix of correlation coefficients $\hat{\rho}$ with $[\hat{\rho}_{ij}] = \rho_{ij}$. Assume that $\hat{\rho}$ is **symmetric positive definite** (SPD). Symmetry is trivial while if we assume that it is not positive definite, then at least one random variable is a linear combination of the others and we can remove it with $\hat{\rho}$ being SPD. From this, we perform **Cholesky factorization** on $\hat{\rho}$ to get the form $\hat{\rho} = LL^T$ where L is lower triangular, and

$$[\hat{\rho}]_{ij} = \rho_{ij} = \sum_k L_{ik} L_{ik}^T$$

Let $\vec{\phi} = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_d \end{bmatrix}^T$, $\vec{\epsilon} = \begin{bmatrix} \epsilon_1 & \epsilon_2 & \cdots & \epsilon_d \end{bmatrix}^T$ where $\phi_i = \sum L_{ij} \epsilon_j$ and $\vec{\phi} = L\epsilon$. We then have

$$\begin{aligned} \phi_i \phi_k &= \sum_j \sum_k L_{ij} L_{kl} \epsilon_l \epsilon_j \\ &= \sum_j \sum_l L_{ij} \epsilon_j \epsilon_l L_{kl}^T \end{aligned}$$

and hence

$$\begin{aligned} E[\phi_i\phi_k] &= \sum_j \sum_l L_{ij} E[\epsilon_j \epsilon_l] L^T_{kl} \\ &= \sum_j \sum_l L_{ij} \delta_{jl} L^T_{kl} \\ &= \sum_l L_{ij} L^T_{kl} = [LL^T]_{ik} = \rho_{ik} \end{aligned}$$

Summary 3. Given independent ϵ_1 to ϵ_d with $\epsilon_i \sim N(0,1)$, we have $E[\epsilon_i\epsilon_j] = \delta_{ij}$.

(1) Cholesky factor $\hat{\rho} = LL^T$

(2) Generate ϵ_1 to ϵ_d

(3) $\vec{\phi} = L\vec{\epsilon} \implies \phi_i \sim N(0,1), E[\phi_i\phi_k] = \rho_{ij}$

Remark 3.1. For $d = 2$, $\hat{\rho} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ and $L = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix}$.

3.10 Hedging Parameters

In MC, if we compute $V(S+h,t)$ and $V(S,t)$ and try to calculate

$$\Delta = \lim_{h\to 0} \frac{V(S+h,t) - V(S,t)}{h}$$

then for really small h, this algorithm is very unstable. To improve this slightly, use common random numbers where we compute $V(S+h,t)$ and $V(S,t)$ using MC but the same seed for each t.

A better trick is to first consider the Black-Scholes equation

$$V_\tau = \frac{\sigma^2 S^2}{2} V_{SS} + rSV_S - rV$$

Let $\Delta = V_S$ and differentiate the above with respect to (w.r.t.) S to get

$$(1)\ \Delta_\tau = \frac{\sigma^2 S^2}{2} \Delta_{SS} + (r+\sigma^2) S \Delta_S$$

Recall that the expected value of the option $\hat{V}$ solves the PDE

$$(2)\ \hat{V} = \frac{\sigma^2 S^2}{2} \hat{V}_{SS} + \mu S \hat{V}_S - \rho \hat{V}$$

Comparing (1) and (2), they have the same form if $\rho = 0$ and $\mu = r + \sigma^2$. This implies that Δ is the expected value of $\Delta(S,T)$. We then can compute the payoff $V_S(S,T)$ and compute paths using

$$dS = (r+\sigma^2) S\ dt + \sigma S\ dZ$$

and $\Delta(S = S_0, s = 0) = E_\Delta[\text{Payoff}]$. Remark that for the Gamma Γ, *the payoff is undefined*.

4 Partial Differential Equations

Consider the PDE

$$(2)\ \hat{V} = \frac{\sigma^2 S^2}{2} \hat{V}_{SS} + \mu S \hat{V}_S - r\hat{V}, \tau = T - t, \tau \in [0,T]$$

Let's solvethis on a finite domain $0 \le \tau \le T$ and $0 \le S \le S_{\max}$. The approach is to discretize the PDE and seek the solution at discrete "nodes". Let

$$V(S_i, \tau^n) = (V_\tau)_i^n = \left[\frac{\sigma^2 S^2}{2}\hat{V}_{SS}\right]_i^n + \left[\mu S \hat{V}_S\right]_i^n - \left[r\hat{V}\right]_i^n$$

4.1 Finite Difference Approximations

We then define, with notation and not literal indices,

$$\Delta X_{i-1/2} := X_i - X_{i-1}, \Delta X_{i+1/2} := X_{i+1} - X_i$$

and using Taylor series

$$f_{i+1} = f_i + (f_x)_i \Delta X_{i+1/2} + (f_{xx})_i \frac{(\Delta X_{i+1/2})^2}{2} + ...$$

If we solve the above for $(f_x)_i$, then

$$(f_x)_i = \left(\frac{f_{i+1} - f_i}{\Delta X_{i+1/2}}\right) + O(\Delta X_{i+1/2})$$

with a 1st order truncation error. The above is called a **forward difference**. Similarly,

$$(f_x)_i = \left(\frac{f_i - f_{i+1}}{\Delta X_{i-1/2}}\right) + O(\Delta X_{i-1/2})$$

which we call a **backward difference**. Suppose that $\Delta X_{i+1/2} = \Delta X_{i-1/2} = \Delta X$. Then,

$$\begin{aligned} f_{i+1} &= f_i + (f_x)_i \Delta X + (f_{xx})_i \frac{\Delta X^2}{2} + O(\Delta X^3) \\ f_{i-1} &= f_i - (f_x)_i \Delta X + (f_{xx})_i \frac{\Delta X^2}{2} + O(\Delta X^3) \end{aligned}$$

If we subtract the second from the first, then

$$(f_x)_i = \left(\frac{f_{i+1} - f_{i-1}}{2\Delta x}\right) + O(\Delta x^2)$$

and we call this the **centered difference**. More generally, if $\Delta X_{i+1/2} \neq \Delta X_{i-1/2}$ then

$$(f_x)_i = \frac{f_{i+1} - f_{i-1}}{(\Delta X_{i+1/2} + \Delta X_{i-1/2})} - \frac{(f_{xx})_i}{2}(\Delta X_{i+1/2} + \Delta X_{i-1/2}) + O(\Delta X_{i+1/2}^2) + O(\Delta X_{i-1/2}^2)$$

Now remark [Note that the Taylor series (must be derived) for the below will most likely be on the final!]

$$(f_{xx})_i = \frac{\left(\frac{f_{i+1} - f_i}{\Delta X_{i+1/2}}\right) - \left(\frac{f_i - f_{i-1}}{\Delta X_{i-1/2}}\right)}{\left(\frac{\Delta X_{i+1} - \Delta X_{i-1/2}}{2}\right)} + O(\Delta X_{i+1} - \Delta X_{i-1/2}) + O\left(\frac{(X_{i+1/2}^2)}{\Delta X_{i+1} + \Delta X_{i-1/2}}\right) + O\left(\frac{(X_{i-1/2}^2)}{\Delta X_{i+1} + \Delta X_{i-1/2}}\right)$$

In the special case of $\Delta X_{i+1} = \Delta X_{i-1/2} = \Delta X$ then

$$(f_{xx})_i = \frac{f_{i+1} - 2f_i + f_{i-1}}{(\Delta X)^2} + O(\Delta X^2)$$

4.2 Discretization Methods for PDEs

Going back the the notation at the beginning of this section, recall that

$$V(S_i, \tau^n) = (V_\tau)_i^n = \left[\frac{\sigma^2 S^2}{2}\hat{V}_{SS}\right]_i^n + \left[\mu S \hat{V}_S\right]_i^n - \left[r\hat{V}\right]_i^n$$

We call $\left[\frac{\sigma^2 S^2}{2}\hat{V}_{SS}\right]_i^n$ the **diffusion term**, $\left[\mu S\hat{V}_S\right]_i^n$ the **drift term**, and $\left[r\hat{V}\right]_i^n$ the **discounting term**. We want to discretize this equation using finite difference methods.

(THE FOLLOWING NOTES ARE COMPLEMENTARY TO THE PDE SLIDES)

Definition 4.1. Let ϕ be a smooth function. The **truncation error** (T.E.) of a discretization for the B-S equation $V_\tau = \mathcal{L}V$ is

$$TE = \{(\phi_\tau)_i^n - (\mathcal{L}\phi)_i^n\} - [\phi_\tau - \mathcal{L}\phi]_i^n$$

Definition 4.2. Let $\Delta S_{max} = \max_i(S_{i+1} - S_i)$ and $\Delta\tau_{max} = \tau^{n+1} - \tau^n$. A discretization of the Black-Scholes equation is consistent if the T.E. $\to 0$ as $\Delta S_{max} \to 0$ and $\Delta\tau_{max} \to 0$.

Example 4.1. Consider the heat equation $u_\tau = u_{ss}$. Discretize to get

$$\frac{u_i^{n+1} - u_i^n}{\Delta\tau} = \frac{u_{i-1}^n - 2u_i^n + u_{i+1}^n}{(\Delta S)^2}$$

Let $\phi(S,\tau)$ be a smooth function. Then,

$$\frac{\phi_i^{n+1} - \phi_i^n}{\Delta\tau} = (\phi_\tau)_i^n + O(\Delta\tau)$$

and similarly

$$\frac{\phi_{i-1}^n - 2\phi_i^n + \phi_{i+1}^n}{(\Delta S)^2} = (\phi_{ss})_i^n + O(\Delta S^2)$$

We then have

$$\begin{aligned} TE &= \left[\left(\frac{\phi_i^{n+1} - \phi_i^n}{\Delta\tau}\right) - \left(\frac{\phi_{i-1}^n - 2\phi_i^n + \phi_{i+1}^n}{(\Delta S)^2}\right)\right]_i^n - [\phi_\tau - \phi_{ss}]_i^n \\ &= O(\Delta\tau) + O(\Delta S^2) \to 0 \end{aligned}$$

as $\Delta S \to 0$ and $\Delta\tau \to 0$.

Example 4.2. In the B-S equation, we previously derived (in the slides)

$$V_i^{n+1} = V_i^n\left(1 - (\alpha_i + \beta_i + r)\Delta\tau\right) + V_{i-1}^n\alpha_i\Delta\tau + V_{i+1}^n\beta_i\Delta\tau, \alpha_i \geq 0, \beta_i \geq 0$$

Suppose we have a mesh for our stock prices for indices $i = 0, ..., q$ and $S_q = S_{max}$. At S_0, where $i = 0$, the B-S equation becomes

$$V_\tau = -rV$$

A discretization of this DE gives us

$$\frac{V_0^{n+1} - V_0^n}{\Delta\tau} = -rV_0^n \implies V_0^{n+1} = V_0^n(1 - r\Delta\tau)$$

At $S = S_{max} = S_q$ we have

$$V_q^{n+1} \approx \begin{cases} S_q & \text{Call} \\ 0 & \text{Put} \end{cases}$$

At $\tau = 0$, we have

$$V_i^0 = \begin{cases} \max(S_i - K, 0) & \text{Call} \\ \max(K - S_i, 0) & \text{Put} \end{cases}, i = 0, ..., q-1$$

Recall that the TE is $O(\Delta\tau) + O(\Delta S^2)$. We then use the following algorithm for the explicit T.E. method:

In the Implicit method,

$$\begin{aligned} (3)\ V_i^{n+1} &= V_i^n - (\alpha_i + \beta_i + r)\Delta\tau V_i^{n+1} + V_{i-1}^{n+1}\alpha_i\Delta\tau + V_{i+1}^{n+1}\beta_i\Delta\tau \\ (4)\ V_i^n &= \left[1 + (\alpha_i + \beta_i + r)\Delta\tau\right]V_i^{n+1} - \alpha_i\Delta V_{i-1}^{n+1} - \beta_i\Delta\tau V_{i+1}^{n+1} \end{aligned}$$

Algorithm 2 Explicit T.E. Method for the B-S Equation

For n = 0,...,(N-1)
  $V_0^{n+1} = V_0^n(1 - r\Delta\tau)$
  $V_q^{n+1} = V_q^n$
  For i=1,...,(q-1)
    $V_i^{n+1} = V_i^n\left(1 - (\alpha_i + \beta_i + r)\Delta\tau\right) + V_{i-1}^n \alpha_i \Delta\tau + V_{i+1}^n \beta_i \Delta\tau, \alpha_i$
  End
End

Let $V^n = \begin{pmatrix} V_0^n & \cdots & V_q^n \end{pmatrix}^T$. Then (4) can be written as $[I+M]V^{n+1} = V^n$ where

$$M = \begin{bmatrix} r\Delta\tau & 0 & 0 & 0 & 0 & 0 \\ \ddots & \ddots & \ddots & 0 & 0 & 0 \\ 0 & -\alpha_i\Delta\tau & (\alpha_i+\beta_i+r)\Delta\tau & -\beta_i\Delta\tau & 0 & 0 \\ 0 & 0 & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The first row of M is $r\Delta\tau$ for the first entry and zero otherwise. The last row is all zeros because $V_q^{n+1} = V_q^n$. Also note that the TE is also $O(\Delta\tau) + O(\Delta S^2)$. Under Crank-Nicholson, we have

$$\frac{V_i^{n+1} - V_i^n}{\Delta\tau} = \frac{1}{2}(\mathcal{L}V)_i^{n+1} + \frac{1}{2}(\mathcal{L}V)_i^n$$

and $[I+\hat{M}]V^{n+1} = [I-\hat{M}]V^n$ where $\hat{M} = M/2$.

5 Stability

In the above algorithms, note that V^{n+1} is not an exact solution. This is because this is a finite difference approximation of derivatives which will produce truncation and round off errors.

These errors may accumulate and swamp the solution. The algorithm may be termed **unstable** in this case. A basic test for stability is to check if:

- Computed value of option should be finite for fixed T, fixed S_{max} as $\Delta S, \Delta\tau \to 0$

5.1 Positive Coefficients

Recall the explicit method

$$V_i^{n+1} = V_i^n\left(1 - (\alpha_i + \beta_i + r)\Delta\tau\right) + V_{i-1}^n \alpha_i \Delta\tau + V_{i+1}^n \beta_i \Delta\tau, \alpha_i \geq 0, \beta_i \geq 0$$

for $i = 1, ..., q-1$. We plan to bound $\|V^{n+1}\|_\infty$ in terms of $\|V^n\|_\infty$ where $\|V^n\|_\infty = \max_i |V_i^n|$. From the above

$$|V_i^{n+1}| \leq |V_i^n\left(1 - (\alpha_i + \beta_i + r)\Delta\tau\right)| + |V_{i-1}^n|\alpha_i \Delta\tau + |V_{i+1}^n|\beta_i \Delta\tau$$

We make the restrictive *assumption* that $(1 - (\alpha_i + \beta_i + r)\Delta\tau) \geq 0$ for any i. This gives us

$$\begin{aligned} |V_i^{n+1}| &\leq |V_i^n|\left(1 - (\alpha_i + \beta_i + r)\Delta\tau\right) + |V_{i-1}^n|\alpha_i \Delta\tau + |V_{i+1}^n|\beta_i \Delta\tau \\ &\leq \|V^n\|_\infty(1 - r\Delta\tau) \end{aligned}$$

and since $0 \leq 1 - r\Delta\tau \leq 1$ from our assumption above, then $|V^{n+1}| \leq \|V^n\|_\infty$. At $i = q$ we have $V_q^{n+1} = V_q^n$ and $i = 0$ gives us $V_0^{n+1} = (1 - r\Delta\tau)V_0^n$ so clearly at all the nodes $|V_i^{n+1}| \leq \|V^n\|_\infty$. Since this is true for all i, choose $i = p$, the maximal node, and get $|V_p^{n+1}| = \max_i |V_i^{n+1}| = \|V^{n+1}\|_\infty$ and

$$\|V^{n+1}\|_\infty \le \|V^n\|_\infty \implies \|V^{n+1}\|_\infty \le \|V^0\|_\infty$$

We have thus derived sufficient conditions for stability:

$$\alpha_i, \beta_i \ge 0, [1 - (\alpha_i + \beta_i + r)\Delta\tau] \ge 0$$

We can rearrange our *assumption* to get

$$\Delta\tau \le \min_i \left[\frac{1}{(\alpha_i + \beta_i + r)}\right]$$

It turns out that this is a necessary and sufficient condition. Assume that $\Delta S_{i+1/2} = \Delta S_{i-1/2} = \Delta S$ and we are using central weighting. The above equation becomes

$$\Delta\tau \le \min_i \left[\frac{1}{\frac{\sigma^2 S_i^2}{\Delta S^2} + r}\right] \approx \left[\frac{\Delta S^2}{\sigma^2 S_i^2}\right] \text{ as } \Delta S \to 0$$

Since $S_i = i\Delta S$ then $\Delta\tau \le (1/(\sigma^2 q^2))$ where $q \equiv$ # of nodes in the stock grid. This is pretty severe where the more nodes that we have, the smaller our timesteps that we need for stability.

We now examine the stability of the fully implicit method. Recall that the equation for the implicit method was given as

$$V_i^{n+1} = V_i^n - (\alpha_i + \beta_i + r)\Delta\tau V_i^{n+1} + V_{i-1}^{n+1}\alpha_i\Delta\tau + V_{i+1}^{n+1}\beta_i\Delta\tau$$

for $\alpha_i, \beta_i \ge 0$. In cases,

$$\begin{aligned} V_i^{n+1}(1 + (\alpha_i + \beta_i + r)\Delta\tau) &= V_i^n + V_{i-1}^{n+1}\alpha_i\Delta\tau + V_{i+1}^{n+1}\beta_i\Delta\tau \\ V_0^{n+1} &= (1 + r\Delta\tau)V_0^n \\ V_q^{n+1} &= V_q^n \end{aligned}$$

Using similar methods as before, we can show that

$$|V_i^{n+1}|(1 + (\alpha_i + \beta_i + r)\Delta\tau) \le \|V_i^n\|_\infty + \|V_{i-1}^{n+1}\|_\infty\alpha_i\Delta\tau + \|V_{i+1}^{n+1}\|_\infty\beta_i\Delta\tau$$

Case 1: Suppose that $p \in \{1, ..., q-1\}$ and $\|V^{n+1}\|_\infty = |V_p^{n+1}|$.

Then

$$\begin{aligned} |V_i^{n+1}|(1 + (\alpha_i + \beta_i + r)\Delta\tau) &\le \|V^n\|_\infty + \|V^{n+1}\|_\infty\alpha_i\Delta\tau + \|V^{n+1}\|_\infty\beta_i\Delta\tau \\ \implies \|V^{n+1}\|_\infty(1 + r\Delta\tau) &\le \|V^n\|_\infty \\ \implies \|V^{n+1}\|_\infty &\le \|V^n\|_\infty \end{aligned}$$

Hence, if $\|V^{n+1}\|_\infty = |V_p^{n+1}|$ for some $p \in \{1, ..., q-1\}$, then $\|V^{n+1}\|_\infty \le \|V^n\|_\infty$.

Case 2: Suppose that $\|V^{n+1}\|_\infty = |V_0^{n+1}|$.

Then

$$\|V^{n+1}\|_\infty \le \|V^n\|_\infty(1 + r\Delta\tau)^{-1} \le \|V^n\|_\infty$$

Case 3: Suppose that $\|V^{n+1}\|_\infty = |V_q^{n+1}|$.

Then trivially,

$$\|V^{n+1}\|_\infty = \|V^n\|_\infty \implies \|V^{n+1}\|_\infty \le \|V^n\|_\infty$$

Hence $\|V^{n+1}\|_\infty \le \|V^n\|_\infty$ for all $n \in \{1, ..., q\}$ and the implicit method is stable. We also then have $\|V^{n+1}\|_\infty \le \|V^0\|_\infty$. Also, there are no conditions in the implicit method for it to be stable (so it is an efficient method).

Also remark that

$$V_i^{n+1} = \frac{V_i^n + V_{i-1}^{n+1}\alpha_i\Delta\tau + V_{i+1}^{n+1}\beta_i\Delta\tau}{1 + (\alpha_i + \beta_i + r)\Delta\tau}$$

is a subconvex combination (sum of the top coefficients over the bottom is ≤ 1) and so we can infer (with a few more steps) or make intuition that that V_i^{n+1} is smaller than V_i^n, V_{i-1}^{n+1}, or V_{i+1}^{n+1}.

The explicit also has this property as well:

$$V_i^{n+1} = V_i^n\left(1 - (\alpha_i + \beta_i + r)\Delta\tau\right) + V_{i-1}^n \alpha_i \Delta\tau + V_{i+1}^n \beta_i \Delta\tau$$

5.2 Crank-Nicholson Method

Recall that

$$\hat{M} = \begin{bmatrix} \frac{r\Delta\tau}{2} & 0 & 0 & 0 & 0 & 0 \\ \ddots & \ddots & \ddots & 0 & 0 & 0 \\ 0 & \frac{-\alpha_i\Delta\tau}{2} & \frac{(\alpha_i+\beta_i+r)\Delta\tau}{2} & \frac{-\beta_i\Delta\tau}{2} & 0 & 0 \\ 0 & 0 & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and the time stepping method was $[I + \hat{M}]V^{n+1} = [I - \hat{M}]V^n$. Assume that $\hat{M}$ has p linearly independent eigenvectors. Recall that $\vec{x}$ is an eigenvector of $\hat{M}$, then

$$\hat{M}\vec{v} = \lambda\vec{x}, \lambda \in \mathbb{C}$$

Let $\vec{x}_i$ be the i^{th} eigenvector with corresponding eigenvalue λ_i. We can write the time stepping method as $V^{n+1} = [I + \hat{M}]^{-1}[I - \hat{M}]V^n$ and if $V_0 = \sum_{i=1}^p c_i \vec{x}_i$ then

$$V^{n+1} = \sum_i c_i \left[\frac{1-\lambda_i}{1+\lambda_i}\right]^{n+1} \vec{x}_i$$

In order for the above to be bounded, we must have

$$\left|\frac{1-\lambda_i}{1+\lambda_i}\right| \leq 1$$

The above will hold when $\Re(\lambda_i) \geq 0$ (will be shown below). We also note that $\hat{M}$ has the properties

$$\begin{aligned} \hat{M}_{ii} &\geq 0 \\ |\hat{M}_{ii}| &\geq \sum_{j\neq i} |\hat{M}_{ij}| \end{aligned}$$

An improvement to Crank-Nicholson is the use of **Rannacher smoothing**. This involve two fully implicit timesteps before apply Crank-Nicholson.

Theorem 5.1. *(Gershgorin Circle Theorem) Every eigenvalue of $\hat{M}$ satisfies at least one of the inequalities*

$$\left|\hat{M}_{ii} - \lambda\right| \leq \sum_{j\neq i} |\hat{M}_{ij}|$$

Corollary 5.1. *Since $|\hat{M}_{ii}| \geq \sum_j |\hat{M}_{ij}|$ then $|\hat{M}_{ii} - \lambda| \leq |\hat{M}_{ii}|$. Therefore, all eigenvalues of $\hat{M}$ has $\Re(\lambda_i) \geq 0$.*

So Crank-Nicholson (C-N) satisfies the necessary conditions for stability. In general, it is difficult to show that C-N satisfies sufficient conditions for stability. Here is the hand-wavy general approach. Suppose that

$$V^n = BV^{n-1}, B = [I+M]^{-1}[I-M] \implies \frac{\|V^n\|}{\|V^0\|} \leq \|B^n\|$$

We need to show that

$$\lim_{P\to Q} \|B^n\| \leq C$$

Where $P \to Q$ are the convergence conditions $n \to \infty$, $\Delta t \to 0$, $\Delta S \to 0$, $n\Delta t = t$.

Simple Case

Suppose B is a normal matrix where $B^T B = BB^T$. If B is normal, then there is a unitary matrix Q which diagonalizes B such that $tr(Q^{-1}BQ) = \sum_i \lambda_i$ where $Q^{-1}BQ = \Lambda$ contains eigenvalues along the diagonal and 0 elsewhere.

Theorem 5.2. *If B is normal, the if $\max_i |\lambda_i| = |\lambda_{max}| \leq 1$ as $\Delta S \to 0$, then*

$$\lim_{n\to\infty, \Delta S \to 0} \|B^{n+1}\|_2 \leq C_0, C_0 \in \mathbb{R}$$

Proof. We know that $B = Q\Lambda Q^{-1}$ and so

$$B^n = (Q\Lambda Q^{-1})(Q\Lambda Q^{-1})... = Q\Lambda^n Q^{-1}$$

$$\implies \|B^n\|_2 \leq \|Q\|_2 \|Q^{-1}\|_2 \|\Lambda^n\|_2$$

$$\implies \|B^n\|_2 \leq |\lambda_{max}|^n \leq 1$$

□

General Case

In the general case, we can put B into the Jordan normal form $B = U^{-1}JU$ where J is a matrix of diagonal blocks of generalized eigenvalues with ones on the above off diagonal. Since U is not unitary, in general, $\|U^{-1}\|_2 \|U\|_2$ may not be bounded as $\Delta S \to 0$. It also may not be possible to bound J^n.

In general, it is possible to show that C-N is unconditionally stable, but it's **not easy**. Lax-Wendroff Theorem makes some headway into this proof.

5.3 Von-Neumann Analysis

Other techniques for stability are:

- Von-Neumann Analysis
  - See notes
  - Only works for constant coefficients

6 Lattice vs. Finite Difference Methods

A lattice method is an explicit finite difference method and uses a special grid. Recall that $V_{FD} = V_{Exact} + (\Delta\tau, \Delta S^2)$. We will show that

$$|V_{Lattice} - V_{FD}| = O(\Delta\tau)$$

Now recall the properties of a *lattice method*:

- Discrete random walk on a lattice
- Parameters selected so that the walk converges to

$$\frac{dS}{S} = \mu dt + \sigma dZ$$

- If $\hat{V}_j^n$ is the lattice solution, then the basic algorithm is

$$\hat{V}_j^n = e^{-r\Delta t}\left[p^*\hat{V}_{j+1}^{n+1} + (1-p^*)\hat{V}_j^{n+1}\right]$$

where p^* is derived from a no-arbitrage argument. In particular

$$p^* = \frac{e^{\Delta t} - e^{-\sigma\sqrt{\Delta t}}}{e^{\sigma\sqrt{\Delta t}} - e^{-\sigma\sqrt{\Delta t}}}$$

Stability requires that $0 \le p^* \le 1$ and a sufficient condition for $p^* \ge 0$ is $r \ge 0$ and $\sigma > 0$. Also $p^* \le 1$ implies that

$$\begin{aligned} e^{r\Delta t} - e^{-\sigma\sqrt{\Delta t}} \le e^{\sigma\sqrt{\Delta t}} - e^{-\sigma\sqrt{\Delta t}} \quad &\Longrightarrow \quad e^{r\Delta t} \le e^{\sigma\sqrt{\Delta t}} \\ &\Longrightarrow \quad e^{r\Delta t - \sigma\sqrt{\Delta t}} \le 1 \\ &\Longrightarrow \quad \sqrt{\Delta t} \le \frac{\sigma}{r} \end{aligned}$$

In the finite difference method,

$$(0)\ V_\tau = \frac{\sigma^2 S^2}{2}V_{SS} + rSV_S - rV$$

Let $Z = \ln S$. We have

$$\begin{aligned} (1)\ V_\tau - \mathcal{L}V &= 0 \\ (2)\ \mathcal{L}V &= \left[\sigma^2 V_{ZZ} + \left(r - \frac{\sigma^2}{2}\right)V_Z - rV\right] \end{aligned}$$

We will discretize (2) using an equally space grid with central weighting and the explicit weighting. This discretization becomes

$$\begin{aligned} (3)\ \frac{V_i^{n+1} - V_i^n}{\Delta\tau} &= (\mathcal{L}_D V)_i^n \\ (4)\ (\mathcal{L}_D V)_i^n &= \left[\left(\frac{V_{i+1}^n - 2V_i^n + V_{i-1}^n}{\Delta Z^2}\right)\frac{\sigma^2}{2} + \left(\frac{V_{i+1}^n - V_{i-1}^n}{2\Delta Z}\right)\left(r - \frac{\sigma^2}{2}\right) - rV_i^{n+1}\right] \end{aligned}$$

Recall the definition of truncation error. Let ϕ be a smoothing function. The truncation error is

$$(5)\ O(\Delta\tau, \Delta Z^2) = \left(\frac{\phi_i^{n+1} - \phi_i^n}{\Delta\tau}\right) - [\mathcal{L}_D\phi]_i^n - [(\phi_\tau - \mathcal{L}\phi)]_i^n$$

We can write (4) as follows

$$(6)\ V_i^{n+1} = V_i^n + \Delta\tau\,(\mathcal{L}_D V)_i^n$$

Suppose we add a term of $O(\Delta\tau^2)$ ot the RHS of (6). Then,

$$(7)\ \left(\frac{V_i^{n+1} - V_i^n}{\Delta\tau}\right) - (\mathcal{L}_D V)_i^n + O(\Delta\tau) = 0$$

Now any approximation in which we introduce to the RHS of (6) which is $O(\Delta\tau^2)$ can be ignored. So (4) becomes

$$(8)\ V_i^{n+1} = \left(\frac{1}{1 + r\Delta\tau}\right)\left[V_i^n\left(1 - \frac{\sigma^2\Delta\tau}{\Delta Z^2}\right) + V_{i+1}^n\left(\frac{\Delta\tau\sigma^2}{2\Delta Z^2} + \left(r - \frac{\sigma^2}{2}\right)\frac{\Delta\tau}{2\Delta Z}\right) + V_{i-1}^n\left(\frac{\Delta\tau\sigma^2}{2\Delta Z^2} - \left(r - \frac{\sigma^2}{2}\right)\frac{\Delta\tau}{2\Delta Z}\right)\right]$$

Equation (8) is too complex for MBA students. Choose $\Delta Z = f(\Delta\tau)$ so that (8) becomes "simple". Choose $\Delta Z = \sigma\sqrt{\Delta\tau}$. Equation (8) becomes

$$V_i^{n+1} = \frac{1}{1 + r\Delta\tau}\left[pV_{i+1}^n + (1-p)V_{i-1}^n\right]$$

where

$$(9)\ p = \frac{1}{2}\left[1 + \sqrt{\Delta\tau}\left(\frac{r}{\sigma} - \frac{\sigma}{2}\right)\right]$$

Now

$$(10)\ e^{-r\Delta\tau} = \frac{1}{1+r\Delta\tau} + O(\Delta\tau^2)$$

Combining (9) and (10) gives us

$$V_i^{n+1} = e^{-r\Delta\tau}\left(pV_{i+1}^n + (1-p)V_{i-1}^n\right) + O(\Delta\tau^2)$$

For stability, we must have $0 \le p \le 1$ which is to say

$$(11)\ -1 \le \sqrt{\Delta\tau}\left(\frac{r}{\sigma} - \frac{\sigma}{2}\right) \le 1$$

The final FD expression then becomes

$$(12)\ V_i^{n+1} = e^{-r\Delta\tau}\left(pV_{i+1}^n + (1-p)V_{i-1}^n\right)$$

From here, we have something which we will refer to as the **domain problem**:

- The PDE domain is $[S_{\min}, S_{\max}]$ and $[\ln S_{\min}, \ln S_{\max}]$
- In the lattice domain, as $\Delta\tau \to 0$ we have $\max_j \hat{V}_j^N \to \infty$ and $\min_j \hat{V}_j^N \to \infty$
- To make the lattice method become closer to the PDE method, we bound the S_j values above and below by $S_{\max}$ and $S_{\min}$ respectively and apply the PDE boundary conditions on the $S_{\min}$ and $S_{\max}$ nodes. This is called creating a **non-exploding tree**.

We then swap some indices by taking Δt as $\Delta\tau$ and switching n and $n+1$ to get

$$(12)\ V_i^n = e^{-r\Delta t}\left(pV_{i+1}^{n+1} + (1-p)V_{i-1}^{n+1}\right)$$

Now let $Z_i = \ln S_i$, $\Delta Z = Z_{i+1} - Z_i = \ln(S_{i+1}/S_i)$ which gives us

$$\begin{aligned} S_{i+1} &= S_i e^{\Delta Z} = S_i e^{\sigma\sqrt{\Delta t}} \\ S_{i-1} &= S_i e^{-\Delta Z} = S_i e^{-\sigma\sqrt{\Delta t}} \end{aligned}$$

If $S_i(FD) = S_j^n(Lattice)$ then

$$\begin{aligned} S_{i+1} &= S_i e^{\sigma\sqrt{\Delta t}} = S_j^n e^{\sigma\sqrt{\Delta t}} = S_{j+1}^{n+1} \\ S_{i-1} &= S_i e^{-\sigma\sqrt{\Delta t}} = S_j^n e^{-\sigma\sqrt{\Delta t}} = S_j^{n+1} \end{aligned}$$

In summary,

FD Index	Lattice Index
V_i^n	V_j^n
V_{i+1}^{n+1}	V_{j+1}^{n+1}
V_{i-1}^{n+1}	V_j^{n+1}

If we replace the FD indices with lattice indices, then

$$(16)\ V_i^n = e^{-r\Delta t}\left(pV_{j+1}^{n+1} + (1-p)V_j^{n+1}\right)$$

where the lattice expression is

$$\hat{V}_i^n = e^{-r\Delta t}\left(p^* V_{i+1}^{n+1} + (1-p^*)V_{i-1}^{n+1}\right)$$

Now

$$\begin{aligned} (17)\ p^* &= \frac{e^{\Delta t} - e^{-\sigma\sqrt{\Delta t}}}{e^{\sigma\sqrt{\Delta t}} - e^{-\sigma\sqrt{\Delta t}}} \\ (18)\ p &= \frac{1}{2}\left[1 + \sqrt{\Delta\tau}\left(\frac{r}{\sigma} - \frac{\sigma}{2}\right)\right] \end{aligned}$$

You can show, after some tedious algebra,

$$(19)\ p^* = p + c_1(\Delta t)^{3/2} + O(\Delta t^2)$$

If we put (19) into (17), we get

$$(20)\ \hat{V}_i^n = e^{-r\Delta t}\left[pV_{i+1}^{n+1} + (1-p)V_{i-1}^{n+1}\right] + e^{-r\Delta t}\left[c_1(\Delta t)^{3/2}\left(\hat{V}_{j+1}^{n+1} - \hat{V}_j^{n+1}\right)\right] + O(\Delta t^2)$$

Assume that $\hat{V}(S_j^n, t^n)$ has bounded derivatives:

$$\frac{\hat{V}_{j+1}^{n+1} - \hat{V}_j^{n+1}}{\Delta Z} = \left(\frac{\partial V}{\partial Z}\right)_{j+\frac{1}{2}}^{n+1} + O(\Delta Z^2)$$

$$\begin{aligned} \implies (21)\ (\hat{V}_{j+1}^{n+1} - \hat{V}_j^{n+1}) &= (V_Z)_{j+\frac{1}{2}}^{n+1}(\Delta Z) + O(\Delta Z^3) \\ &= O(\Delta Z) = O(\sqrt{\Delta t}) \end{aligned}$$

Putting (21) into (20) gives

$$(22)\ \hat{V}_j^n = e^{-r\Delta t}\left[p\hat{V}_{j+1}^{n+1} + (1-p)V_j^{n+1}\right] + O(\Delta t)$$

The finite difference equation is

$$(23)\ V_j^n = e^{-r\Delta t}\left[p\hat{V}_{j+1}^{n+1} + (1-p)V_j^{n+1}\right]$$

Let $E_j^n = (V_j^n - \hat{V}_j^n)$ and subtract (22) from (23) to get

$$(24)\ E_j^n = e^{-r\Delta t}\left[pE_{j+1}^{n+1} + (1-p)E_j^{n+1}\right] + O(\Delta t^2)$$

Choose Δt sufficiently small so that $0 \le p \le 1$. By the triangle inequality,

$$(24A)\ |E_j^n| \le p|E_{j+1}^{n+1}| + (1-p)|E_j^{n+1}| + c_2\Delta t^2$$

where c_2 is some constant independent of Δt. Let $(25)\ \|E^n\|_\infty = \max_j |E_j^n|$. Using (25) in (24A) gives us

$$(26)\ \|E^n\|_\infty \le \|E^{n+1}\|_\infty + c_2\Delta t^2$$

Recursion gives us

$$\|E^0\|_\infty \le \|E^1\|_\infty + c_2\Delta t \le \dots \le \underbrace{\|E^N\|_\infty}_{=0} + \frac{T}{\Delta t}c_2(\Delta t)^2 = O(\Delta t)$$

The final result is that $|\hat{V}_0^0 - V_0^0| = O(\Delta t)$. Hence the lattice methods is simply a special case of an explicit finite difference method.

Summary 4. Recall that:

- Lattice Method
  - Equally spaced grid in $Z = \ln S$ space
  - Choice of timestep ensures method is stable and positive coefficient
  - We have specified the grid spacing without regard to the contract
    - For example, consider the barrier option. The barrier knock-out condition is

$$V(S, t_i) = 0 \text{ if} S \ge \beta$$

where t_i is the observation time. Let $t_i^- = t_i - \epsilon, 0 < \epsilon \ll 1$, $\tau = T - t, \tau_i^+ = T - t_i^-$ where

$$V(S, t_i^-) = 0 \text{ if } S \ge \beta \iff V(S, t_i^+) = 0 \text{ if } S \ge \beta$$

Consider a simple example with two observations $t = t_1, T$ and $0 < t_1 < T$. With a barrier of β and a strike of K, a knock-out call has payoff

$$V(S,T) = \begin{cases} 0 & S \in [0,K) \\ S-K & S \in [K,B) \\ 0 & S \in [B,\infty) \end{cases}$$

Moving the PDE from $\tau = 0$ to τ_1^- creates a smoothed curve $V(S,\tau_1^-) \geq V(S,T)$ where it increases concave up from 0 to B and decreases sharply (but continuously) from B onwards. Moving from τ_1^- to τ_1^+ is almost the same as $V(S,\tau_1^-)$ but there is a jump down at B to 0 (jump discontinuity). As τ_1^+ goes to T, the whole curve smooths over and it is continuous again. In a *lattice grid*, we don't have a choice to place nodes near K or B.

Summary 5. (Final Exam)

- 1 to 4 are trivial questions
- 5 to 6 are related to assignment or practice problems
- 7 is the most difficult question

Index