STA2502 Stochastic Models in Investments

1 Discrete Binomial Tree Model

1.1 Financial Derivative

Informally, a derivative is a financial contract that derives its value from some underlying process. The underlying asset is often traded (e.g. stocks), but it need not be (e.g. weather).

Examples of underlying assets:

- Stocks
- Bonds
- Currencies
- Interest rates
- Indices (S&P 500, TSX, NASDAQ)
- Commodities: energy, livestock, metals and precious metals, grains, coffee, sugar, etc

Financial derivatives have two parties

- Buyer: long position w.r.t. the derivative
- Seller: short position w.r.t. the derivative

How are financial derivatives traded:

- Standard derivatives contracts are traded on exchanges
- Over-the-counter (OTC) market: derivatives traded between banks, corporations, and other major institutions.

1.2 Types of Derivatives

Definition: 1.1: Forward Contract

A forward contract is an obligation to buy or sell an underlying asset at a specified forward price on a known date

- Long position: party who agrees to purchase the underlying asset at a specified future date for a specified price
- Short position: party who agrees to sell the underlying asset at a specifed future date for a specified price
- It costs nothing to enter into a forward contract
- Forwards are traded OTC

Definition: 1.2: Futures Contract

A futures contract is a standardized forward contract, which is traded on an exchange. The standarization of the contracts allows for trading by non-institutional investors. One can purchase fractional future contracts

Definition: 1.3: Options

A **call option** gives the holder the right, but not the obligation, to purchase the underlying asset at (or by) a given date for a specified price, called the strike price.
A **put option** gives the holder the right but not the obligation, to sell the underlying asset at (or by) a given date for a specified price, called the strike price.

Main difference between options and futures/forwards: gives the holder the right to do something, but they do not have to exercise their right.

- European options: can only be exercised on the maturity date (Easier to analyze)
- American options: can be exercised at any time up to the maturity date (More popular)

Exotics

- Bermudan options: can be exercised on pre-specified dates and the maturity date
- Asian options: payoff depends on the average of the price of the underlying during the life of the option
- Binary (digital) options: pays all or nothing
- Lookback options: payoff depends on the maximum or minimum price of the underlying asset during the life of the option

Terminology:

- An option is **in the money** if the price of the underlying is at a level where the payoff would be positive
- An option is **out of the money** if the price of the underlying is at a level where the payoff would be zero

Definition: 1.4: Swap

A swap is the simultaneous selling and purchasing of cash flows involving various currencies, interest rates, and a number of other financial assets.
A swaption is an option granting its owner the right but not the obligation to enter into an underlying swap.

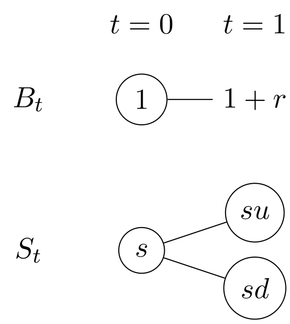
One can price swaps and swaptions by decomposing them into forwards and other options.

1.3 One-Period Binomial Model

Consider two assets: bond (B_t) and stock (S_t), and two time steps $t = 0, t = 1$. The portfolio $h = (x, y) \in \mathbb{R}^2$ has value process:

$$V_t^h = xB_t + yS_t, t = 0, 1, \tag{1}$$

x unit in bond, y unit in stock.



Proposition: 1.1:

The following are equivalent

1. The one-period binomial model is arbitrage-free
2. $d < 1 + r < u$
3. $\exists q \in (0, 1)$ s.t. $1 + r = qu + (1 - q)d$
4. There exists a martingale measure $\mathbb{Q}$ defined by $S_0 = \frac{1}{1+r}\mathbb{E}^{\mathbb{Q}}[S_1]$ s.t. $\mathbb{Q}(S_1 = su) = q$, $\mathbb{Q}(S_1 = sd) = 1 - q$

Proof. We show the equivalence of 3. and 4.

$$\begin{aligned} 1 + r &= qu + (1 - q)d \\ \Leftrightarrow s(1 + r) &= qsu + (1 - q)sd \\ \Leftrightarrow S_0 &= \frac{1}{1 + r}\mathbb{E}^{\mathbb{Q}}[S_1] \end{aligned}$$

Also we get a formula for q: $q = \frac{1+r-d}{u-d}$ □

Definition: 1.5: Complete Market

A **financial derivative** is a stochastic variable of the following form: $C_T = F(S_T)$. A financial derivative can be replicated if there exists a portfolio h s.t. $P(V_1^h = C_T) = 1$, h is the replicating portfolio.
If all derivatives can be replicated, we say the market is complete.

Proposition: 1.2:

If a derivative C_T can be replicated by a portfolio h, then at $t = 0$, any price other than V_0^h will lead to an arbitrage opportunity.

Proposition: 1.3:

If the one-period binomial model is arbitrage free, then it is complete

Proof. Let $C_1 = F(S_1)$ be an arbitrary derivative. We want to find a portfolio $h = (x, y)$ s.t. $V_1^h = \begin{cases} F(su), \text{ if } S_1 = su \\ F(sd), \text{ if } S_1 = sd \end{cases}$.

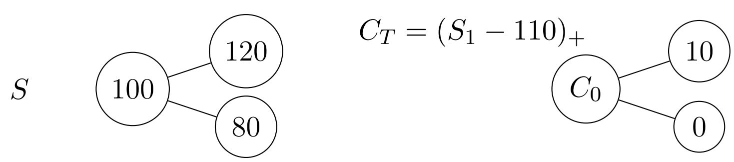
Note $V_1^h = x(1+r) + yS_1 = \begin{cases} x(1+r) + ysu = F(su), \text{ if } S_1 = su, \\ x(1+r) + ysd = F(sd), \text{ if } S_1 = sd, \end{cases}$ This gives a system of 2 equations and 2 unknowns. $\begin{cases} x = \frac{1}{1+r}\left[\frac{uF(sd)-dF(su)}{u-d}\right] \\ y = \frac{1}{S}\left[\frac{F(su)-F(sd)}{u-d}\right] \end{cases}$

$$
\begin{aligned}
C_0 = V_0^h = x + sy &= \frac{1}{1+r}\left[\frac{uF(sd) - dF(su)}{u-d}\right] + \left[\frac{F(su) - F(sd)}{u-d}\right] \\
&= \frac{1}{1+r}\left[\frac{u-(1+r)}{u-d}F(sd) + \frac{1+r-d}{u-d}F(su)\right] \\
&= \frac{1}{1+r}\left[(1-q)F(sd) + qF(su)\right] \\
&= \frac{1}{1+r}\mathbb{E}^{\mathbb{Q}}[C_1] = \frac{1}{1+r}\mathbb{E}^{\mathbb{Q}}[F(S_1)]
\end{aligned}
$$

Therefore, we can always find a replicating portfolio for any arbitrary derivative. □

Example: In the one-period binomial model, assume $s = 100, u = 1.2, d = 0.8, p_u = 0.6, r = 0.02$.

1. The market is arbitrage free, because $d < 1 + r < u$
2. Consider a European call option with $K = 110$, what is the arbitrage-free price?

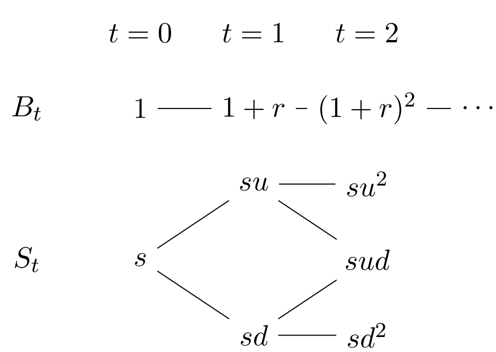


$q = \frac{1+r-d}{u-d} = \frac{1+0.02-0.8}{1.2-0.8} = 0.55$
$C_0 = \frac{1}{1+r}\mathbb{E}^{\mathbb{Q}}[C_1] = \frac{1}{1+0.02}(10 \cdot 0.55 + 0 \cdot 0.45) = \frac{5.5}{1.02} = 5.39$

3. What is the replicating portfolio?

$$
\begin{aligned}
x &= \frac{1}{1+r}\frac{uF(sd) - dF(su)}{u-d} = \frac{1}{1.02}\frac{1.2 \cdot 0 - 0.8 \cdot 10}{1.2 - 0.8} = -19.61 \\
y &= \frac{1}{S}\frac{F(su) - F(sd)}{u-d} = \frac{1}{100}\frac{10-0}{1.2-0.8} = 0.25
\end{aligned}
$$

1.4 Multi-Period Binomial Model



Assume that we have finite number of periods: Let $t \in \{0, 1, 2, ..., T\}$. Then

1. Bond price: $\begin{cases} B_0 = 1 \\ B_t = (1+r)B_{t-1} \end{cases}$, $B_t = (1+r)^t$

2. Let $Z_0, ..., Z_{T-1}$ be i.i.d. r.v.s s.t. $P(Z_t = u) = p_u$ and $P(Z_t = d) = p_d$, $p_u + p_d = 1$. Then $\begin{cases} S_0 = s \\ S_t = S_{t-1}Z_{t-1} \end{cases}$.
   Let $K_t \in \{0, ..., T\}$ be the number of up-moves of the stocks. K_t is stochastic and $K_t \sim \text{Binom}(t, p_u)$. We can write $S_t = su^{K_t}d^{t-K_t}$, $t \in \{1, ..., T\}$, where $K_t \sim \text{Binom}(t, p_u)$, $P(S_2 = sud) = P(K_2 = 1) = \binom{2}{1}p_u p_d = 2p_u p_d$.

***Definition:* 1.6: Portfolio Strategy**

A portfolio strategy is a stochastic process $h_t = (x_t, y_t)$, $t \in \{1, ..., T\}$. For a given h_t, we set $h_0 = h_1$. The process h_t depends on $S_0, ..., S_{t-1}$. The value is defined by $V_t^h = x_t(1+t) + y_t S_t$.

***Definition:* 1.7: Self-Financing Portfolio**

A self-financing portfolio (there is no exogenous cash infusions, the purchase of a new asset must be financed by the sale of an old one) satisfies the budget equation:

$$x_t(1+r) + y_t S_t = x_{t+1} + y_{t+1} S_t$$

***Definition:* 1.8: Arbitrage Portfolio**

An arbitrage portfolio is a self-financing portfolio h s.t.
- $V_t^h = 0$
- $P(V_t^h \geq 0) = 1$
- $P(V_T^h > 0) > 0$

***Proposition:* 1.4:**

The multi-period binomial model if free of arbitrage if and only if $d < 1 + r < u$.

The martingale measure $\mathbb{Q}$, and the probabilities q_u, q_d can be defined in the same way.

Let C be a derivative with payoff $C_T = F(S_T)$. It can be replicated if there exists a self-financing portfolio h s.t. $P(V_T^h = C_T) = 1$.

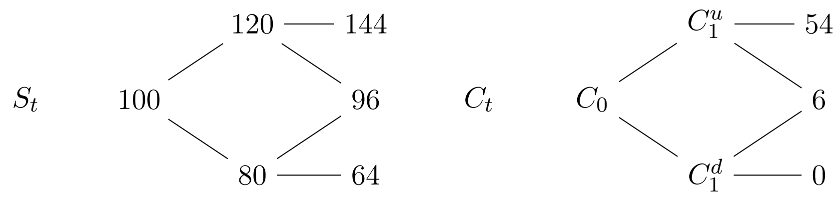
***Proposition:* 1.5:**

If C can be replicated by h, then any price other than $C_t = V_t^h$ would lead to an arbitrage. The fair price process $C_t = V_t^h$ for $t = 0, 1, ..., T$.

***Proposition:* 1.6:**

The multi-period binomial model is complete. *i.e.* all derivatives can be replicated by a self-financing portfolio.

Example: Two period model with $s = 100, u = 1.2, d = 0.8, p_u = 0.6, r = 0.02$. Find price of a European call option at $t = 2$ with $K = 90$.
It follows the one-period model that $q_u = 0.55$, $q_d = 0.45$.



$$
\begin{aligned}
C_1^u &= \frac{1}{1+r}[54q + 6(1-q)] = 31.76 \\
C_1^d &= \frac{1}{1+r}[6q + 0(1-q)] = 3.24 \\
C_0 &= \frac{1}{1+r}[31.76q + 3.24(1-q)] = 18.56
\end{aligned}
$$

Proposition: 1.7: Binomial Algorithm

The derivative C with payoff $C_T = F(S_T)$ can be replicated using a self-financing portfolio. Let $V_t(k)$ be the value of the portfolio at node (t, k). We have $V_T(k) = F(su^k d^{T-k})$. The recursion is defined as:

$$
V_t(k) = \frac{1}{1+r}[qV_{t+1}(k+1) + (1-q)V_{t+1}(k)], \text{ where } q = \frac{1+r-d}{u-d}
$$

The hedging portfolio is

$$
\begin{aligned}
x_t(k) &= \frac{1}{1+r}\frac{uV_t(k) - dV_t(k+1)}{u-d} \\
y_t(k) &= \frac{1}{S_{t-1}}\frac{V_t(k+1) - V_t(k)}{u-d}
\end{aligned}
$$

Proposition: 1.8:

The arbitrage free price of a derivative C with payoff $C_T = F(S_T)$ is

$$
C_0 = \frac{1}{(1+r)^T}\mathbb{E}^{\mathbb{Q}}[X] = \frac{1}{(1+r)^T}\sum_{k=0}^{T}\binom{T}{k} q^k (1-q)^{T-k} F(su^k d^{T-k})
$$

2 Probability

2.1 Probability Tripple

The fundational object in probability space is a probability tripple $(\Omega, \mathcal{F}, P)$.

***Definition:* 2.1: Sample Space**

The sample space Ω is the set of possible outcomes of a random event. In most cases, we consider $\Omega = \mathbb{R}$.

Example: roll a dice $\Omega = \{1, 2, 3, 4, 5, 6\}$

***Definition:* 2.2: σ-algebra**

A σ-algebra on a sample space Ω is a collection $\mathcal{F}$ of subsets of Ω s.t.

1. $\emptyset \in \mathcal{F}$
2. If $A \in \mathcal{F}$, then $A^C \in \mathcal{F}$
3. If $A_1, A_2, ... \in \mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

Example: $\Omega = \{a, b, c\}$, $\mathcal{F}_1 = \{\emptyset, \{a\}, \{b, c\}, \Omega\}$ and $\mathcal{F}_2 = \{\emptyset, \{a, b\}, \{c\}, \Omega\}$ are σ-algebras. Both are not maximal information (power set), but we can define probability measure on it.

$\mathcal{F}_1 \cup \mathcal{F}_2 = \{\emptyset, \{a\}, \{b, c\}, \{a, b\}, \{c\}, \Omega\}$ is not a σ-algebra, because $\{a, c\}$ is not included.

Example: For any Ω, the following are σ-algebra:

1. $\{\emptyset, \Omega\}$ (Trivial σ-algebra)
2. $2^\Omega = \mathcal{P}(\Omega)$ (Power set)
3. $\{\emptyset, A, A^C, \Omega\}$, where $A \in \Omega$.

***Definition:* 2.3: Sub-σ-algebra**

Let $\mathcal{F}$ be a σ-algebra, G is a sub-σ-algebra of $\mathcal{F}$ if

1. G is a σ-algebra
2. $G \subset \mathcal{F}$

Example: $\mathcal{F}_1, \mathcal{F}_2$ are sub-σ-algebra of $\mathcal{P}(\Omega)$

Terminology: $(\Omega, \mathcal{F})$ is a measurable space.

***Definition:* 2.4: Observable Events**

The σ-algebra $\mathcal{F}$ is the collection of observable events. Elements of $\mathcal{F}$ are events whose probabilities are known.

Definition: 2.5: Probability Measure

A probability measure on a measurable space $(\Omega, \mathcal{F})$ is a function $P : \mathcal{F} \to [0, 1]$ s.t.

1. $P(\Omega) = 1$
2. $P(A) \geq 0, \forall A \in \mathcal{F}$
3. For any countable set of disjoint sets $\{A_i\}_{i\in\mathbb{N}}$, $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

Some consequences:

1. $P(\emptyset) = 0$
2. $P(A^C) = 1 - P(A)$

Example: $\Omega = \{a, b, c\}$, $\mathcal{F} = \{\emptyset, \{a\}, \{b, c\}, \Omega\}$
Let P be $P(\emptyset) = 0$, $P(\{a\}) = \frac{4}{5}$, $P(\{b, c\}) = \frac{4}{5}$, $P(\Omega) = 1$. P is a probability measure.
Note: the probability measure is not uniquely defined.

Definition: 2.6: Continuous and Equivalent Probability Measures

Let P and Q be two probability measures on $(\Omega, \mathcal{F})$.

1. P is absolutely continuous w.r.t. Q if $Q(A) = 0 \Rightarrow P(A) = 0, \forall A \in \mathcal{F}$. Denote $P << Q$.
2. P and Q are equivalent, denote $P \sim Q$ if $P(A) = 0 \Leftrightarrow Q(A) = 0, \forall A \in \mathcal{F}$. *i.e.* $P << Q$ and $Q << P$.

2.2 Random Variables

Assume $(\Omega, \mathcal{F}, P)$. Let R be the smallest σ-algebra that contains all intervals in $\mathbb{R}$ (The Borel σ-algebra).

Definition: 2.7: Random Variable

A random variable is a function $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, R)$ s.t. $\forall B \in R$, $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$. We say that X is $\mathcal{F}$-measurable.

Example: Dice roll $\Omega = \{1, 2, 3, 4, 5, 6\}$. $\mathcal{F} = \{\emptyset, \{1, 2, 3\}, \{4, 5, 6\}, \Omega\}$.
Let $X : \Omega \to \mathbb{R}$, $X(\omega) = \begin{cases} 1, \omega = 1, 2 \\ 2, \omega = 3, 4 \\ 3, \omega = 5, 6 \end{cases}$ is not a random variable, because $X^{-1}((0, 1]) = \{1, 2\} \notin \mathcal{F}$. X is not $\mathcal{F}$-measurable.
If $\tilde{\mathcal{F}} = \{\emptyset, \{1, 2\}, \{3, 4\}, \{5, 6\}, \{1, 2, 3, 4\}, \{1, 2, 5, 6\}, \{3, 4, 5, 6\}, \Omega\}$, then X is $\tilde{\mathcal{F}}$-measurable.

Example: $A \in \mathcal{F}$, $X = \mathbb{1}_A$, $X(\omega) = \begin{cases} 1, \omega \in A \\ 0, \omega \notin A \end{cases}$. Let $(a, b] \subset \mathbb{R}$.

$X^{-1}((a, b]) = \begin{cases} \emptyset, 0, 1 \notin (a, b] \\ A^C, 0 \in (a, b], 1 \notin (a, b] \\ A, 0 \notin (a, b], 1 \in (a, b] \\ \Omega, 0, 1 \in (a, b] \end{cases}$. Since $A \in \mathcal{F}, A^C \in \mathcal{F}$, X is $\mathcal{F}$-measurable.

Remark 1. 1. If $\mathcal{F} = \mathcal{P}(\Omega)$, any $X : \Omega \to \mathbb{R}$ is measurable

2. If $g : \mathbb{R} \to \mathbb{R}$ is continuous and X is a random variable, then $g(X)$ is a random variable.

***Definition:* 2.8: Generated σ-algebra**

The σ-algebra generated by a r.v. X, $\sigma(X)$ is the smallest σ-algebra s.t. X is measurable.

Distribution Function $F_X(x) = P(\{\omega \in \Omega : X(\omega) \leq x\}) = P(X \leq x)$

Density $f_X(x) = \frac{dF_X(x)}{dx}$ defined only if F_X is continuously differentiable.

Expected Value $\mathbb{E}[X] = \int_\Omega X dP = \int_\mathbb{R} x f_X(x) dx.$

***Definition:* 2.9: Moment Generating Function**

The moment generating function (m.g.f.) of a random variable X is the function $M_X(t) = \mathbb{E}[e^{tX}]$, $t \in \mathbb{R}$. The m.g.f. exists if $\exists a > 0$ s.t. $M_X(s) < \infty$, $\forall s \in [-a, a]$
It can be used to find moments of X: $\mathbb{E}[X^n] = \frac{d^n}{dt^n} M_X(t)|_{t=0}$.

2.3 Conditional Expectation

Example: X: sum of 2 dices, Y: first dice.

$$\mathbb{E}[X|Y=2] = \sum_{i=3}^{8} i\frac{1}{6} = 5.5, \mathbb{E}[X|Y=y] = \sum_{i=1}^{6}(y+i)\frac{1}{6} = y + 3.5$$

$\mathbb{E}[X|Y]$ is a random variable, $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$. We condition on information (σ-algebra)

Example: S_T: stock price, $\mathcal{F}$: full information, G: limited information, G is a sub-σ-algebra of $\mathcal{F}$. We estimate the stock price using our information G, $\mathbb{E}[S_T|G]$.

***Definition:* 2.10: Conditional Expectation**

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and G be a sub-σ-algebra of $\mathcal{F}$. Let X be a integrable r.v. *i.e.* $\mathbb{E}[|X|] < \infty$ and X is $\mathcal{F}$-measurable. The conditional expectation of X given G, denoted by $\mathbb{E}[X|G]$ is a random variable satisfying:

1. $\mathbb{E}[X|G]$ is G-measurable
2. $\mathbb{E}[\mathbb{E}[X|G]\mathbb{1}_A] = \mathbb{E}[X\mathbb{1}_A]$ if $A \in G$.

Condition on a r.v.: $\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)]$, where $\sigma(Y)$ is the σ-algebra generated by Y.
Conditional Probability: $P(A|G) = \mathbb{E}[\mathbb{1}_A|G]$

Theorem: 2.1: Properties of Conditional Expectations

Let X, Y be integrable r.v., G a sub-σ-algebra of $\mathcal{F}$.

1. Measurability: If X is G-measurable, then $\mathbb{E}[X|G] = X$
2. Taking out what is known: If X is G-measurable, $\mathbb{E}[XY|G] = X\mathbb{E}[Y|G]$
3. Tower property: If H is sub-σ-algebra of G, then $\mathbb{E}[\mathbb{E}[X|G]|H] = \mathbb{E}[X|H]$. Furthermore, $\mathbb{E}[\mathbb{E}[X|G]] = \mathbb{E}[X]$.
4. Independence: X is independent of G if $P(\{x \in B\} \cap A) = P(\{x \in B\})P(A)$ for any $A \in G, B \subset \mathbb{R}$. If X is independent of G, then $\mathbb{E}[X|G] = \mathbb{E}[X]$.
5. Linearity: $\mathbb{E}[aX + bY|G] = a\mathbb{E}[X|G] + b\mathbb{E}[Y|G]$.
6. Monotonicity: If $X \geq Y$, then $\mathbb{E}[X|G] \geq \mathbb{E}[Y|G]$. Also, if $X \geq 0$, then $\mathbb{E}[X|G] \geq 0$.
7. Conditional Jenson's inequality: Let $\phi : \mathbb{R} \to \mathbb{R}$ be a convex function, $\phi(\mathbb{E}[X|G]) \leq \mathbb{E}[\phi(X)|G]$.

Proof. We prove 3 only. Consider $\tilde{\mathcal{F}} = \{\emptyset, \Omega\}$. We claim that $\mathbb{E}[X|\tilde{\mathcal{F}}] = \mathbb{E}[X]$.
Firstly, $\mathbb{E}[X]$ is constant, so $\tilde{F}$-measurable. This is because preimage of $\mathbb{E}[X]$ is either $\emptyset$ or Ω.
Secondly, the only sets in $\tilde{\mathcal{F}}$ are $\emptyset, \Omega$.

$$\mathbb{E}[\mathbb{E}[X]\mathbb{1}_\emptyset] = 0 = \mathbb{E}[X\mathbb{1}_\emptyset]$$
$$\mathbb{E}[\mathbb{E}[X]\mathbb{1}_\Omega] = \mathbb{E}[\mathbb{E}[X]] = \mathbb{E}[X] = \mathbb{E}[X\mathbb{1}_\Omega]$$

Now $\mathbb{E}[\mathbb{E}[X|G]] \underset{\text{By the claim}}{=} \mathbb{E}[\mathbb{E}[X|G]|\tilde{F}] \underset{\text{By the first part of tower property}}{=} \mathbb{E}[X|\tilde{F}] = \mathbb{E}[X]$ □

Example: Show that $\mathbb{E}[X\mathbb{E}[X|G]] = \mathbb{E}\left[(\mathbb{E}[X|G])^2\right]$.

Proof.

$$\begin{aligned}
\mathbb{E}\left[(\mathbb{E}[X|G])^2\right] &= \mathbb{E}[\mathbb{E}[X|G]\mathbb{E}[X|G]] \\
&= \mathbb{E}[\mathbb{E}[X\mathbb{E}[X|G]|G]] \text{ (By 2)} \\
&= \mathbb{E}[X\mathbb{E}[X|G]] \text{ (By 3)}
\end{aligned}$$

□

3 Stochastic Processes

Let T be set of times. For discrete, $T = \{0, 1, 2, ...\}$. For continuousm $T = [0, T]$.

Definition: 3.1: Stochastic Process

A stochastic process is a mapping $X : \Omega \times T \to \mathbb{R}$ s.t.
1. $\forall \omega \in \Omega$, $X(\omega, t)$ is a function of time.
2. $\forall t \in T$, $X(\omega, t)$ is a random variable.

Notation: $X = (X_t)_{t\in T} = \{X_t : t \in T\}$.

We can define new stochastic processes $Y = f(t, X_t) = \{f(t, X_t) : t \in T\}$.

Definition: 3.2: Filtration

A filtration on $(\Omega, \mathcal{F}, P)$ is a collection of σ-algebras indexed by time $\{\mathcal{F}_t\}_{t\in T}$ s.t.
1. $\mathcal{F}_t \subset \mathcal{F}$ is a sub-σ-algebra $\forall t \in T$
2. $\mathcal{F}_s \subset \mathcal{F}_t$, $\forall s \le t$.

It models the incoming information over time. $(\Omega, \mathcal{F}, P)$ equipped with a filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t\in T}$ is $(\Omega, \mathcal{F}, \mathbb{F}, P)$ called a *filtered probability space.*

Definition: 3.3: Adapted Process

A stochastic process is adapted to the filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t\in T}$ if X_t is $\mathcal{F}_t$-measurable for all $t \in T$.

Definition: 3.4: Adapted Filtration

Let X be a continuous stochastic process on $(\Omega, \mathcal{F}, P)$. The σ-algebra generated by X on $[0, t]$ is $\mathcal{F}_t^X = \sigma(\{X_s\} : s \le t)$. $\mathcal{F}_t^X$ is the smallest filtration that X is adapted to.

Example: Let X be a continuous stochastic process. Consider the filtration $\mathcal{F}_t^X$ generated by X, $t \in [0, \infty)$.

1. $Y_t = X_{t/2}$ is adapted to $\mathcal{F}_t^X$.
2. $Y_t = X_{t+1}$ is $\mathcal{F}_{t+1}$-measurable, but not necessarily $\mathcal{F}_t$-measurable, so it is not adapted.
3. $Y_t = \sup_{s\le t} X_s$ is adapted to $\mathcal{F}_t^X$, because it only requires information upto t.
4. $Y_t = \int_0^{2t} X_s ds$ is not adapted, but $Y_t = \int_0^t X_s ds$ is adapted.

Definition: 3.5: Martingale

A stochastic process X on $(\Omega, \mathcal{F}, \mathbb{F}, P)$ is a martingale if
1. X is $\mathbb{F}$-adapted ($\mathcal{F}_t$-measurable for all $t \in T$)
2. $\mathbb{E}^P[|X_t|] < \infty$, $\forall t \in T$
3. $\mathbb{E}^P[X_t|\mathcal{F}_s] = X_s$, $\forall s \le t, s, t \in T$

Example: Let Y be an integrable r.v. on $(\Omega, \mathcal{F}, \mathbb{F}, P)$ with $\mathbb{E}[|Y|] < \infty$. Show that $M_t = \mathbb{E}[Y|\mathcal{F}_t]$ is a martingale.

Proof. 1. M_t is $\mathcal{F}_t$ by Definition 2.10

2. $\mathbb{E}[|M_t|] = \mathbb{E}\left[|\mathbb{E}[Y|\mathcal{F}_t]|\right] \underset{\text{Jensen's Inequality}}{\leq} \mathbb{E}[\mathbb{E}[|Y|\,|\mathcal{F}_t]] = \mathbb{E}[|Y|] < \infty$

3. Let $s = t$, $\mathbb{E}[M_t|\mathcal{F}_s] = \mathbb{E}[\mathbb{E}[Y|\mathcal{F}_t]|\mathcal{F}_s] \underset{\text{Tower Property}}{=} \mathbb{E}[Y|\mathcal{F}_s] = M_s$.

Since M_t satisfies Definition 3.5, M_t is a martingale. □

Example: $C_T = f(S_T)$ price of European call option, $B_T = e^{\int_0^T r_s ds}$ bond price. $\mathbb{E}\left[\frac{C_T}{B_T}|\mathcal{F}_s\right]$ is the expected relative return.

Example Is $X_t = e^{aW_t - \frac{1}{2}a^2 t}, a \in \mathbb{R}$ a martingale?

Proof.

$$\begin{aligned}
\mathbb{E}\left[e^{aW_t - \frac{1}{2}a^2 t}|\mathcal{F}_s\right] &= e^{-\frac{1}{2}a^2 t}\mathbb{E}\left[e^{aW_t}|\mathcal{F}_s\right] = e^{-\frac{1}{2}a^2 t}\mathbb{E}\left[e^{a(W_t - W_s) + aW_s}|\mathcal{F}_s\right] \\
&= e^{-\frac{1}{2}a^2 t} e^{aW_s}\mathbb{E}\left[e^{a(W_t - W_s)}\right] \\
&= e^{aW_s - \frac{1}{2}a^2 t} e^{\frac{1}{2}a^2(t-s)} \\
&= e^{aW_s - \frac{1}{2}a^2 s} = X_s,
\end{aligned}$$

so it is a martingale. Note that $W_t - W_s \sim \mathcal{N}(0, t-s)$, and $\mathbb{E}[e^{aX}] = e^{\frac{1}{2}a^2\sigma_X^2}$. □

3.1 Brownian Motion

We construct Brownian motion using a scaled symmetric random walk.
Let $T > 0$. We divide $[0, T]$ into n intervals of size $\Delta t = \frac{T}{n}$. The kth interval is $((k-1)\Delta t, k\Delta t]$.
Define the following i.i.d. r.v.for $i = 1, 2, ..., n$:

$$Z_i = \begin{cases} 1, & \text{with probability} 0.5 \\ -1, & \text{with probability} 0.5 \end{cases}, \mathbb{E}[Z_i] = 0, \text{Var}(Z_i) = 1, \forall i$$

Define the process W as follows:

$$\begin{aligned}
W_0 &= 0 \\
W_i - W_{i-1} &= \sqrt{\Delta t} Z_i, i = 1, ..., n
\end{aligned}$$

Let $t = k\Delta t$, $W_t = \sum_{i=1}^{k} \sqrt{\Delta t} Z_i$. Then

$$\begin{aligned}
\mathbb{E}[W_t] &= \mathbb{E}\left[\sum_{i=1}^{k} \sqrt{\Delta t} Z_i\right] = \sqrt{\Delta t}\sum_{i=1}^{k} \mathbb{E}[Z_i] = 0 \\
\text{Var}(W_t) &= \text{Var}\left(\sum_{i=1}^{k} \sqrt{\Delta t} Z_i\right) = \Delta t \sum_{i=1}^{k} \text{Var}(Z_i) = \Delta t \cdot k = t
\end{aligned}$$

Let $\Delta t \to 0, k \to \infty$, by Central Limit Theorem, $W_t \sim N(0, t)$.

Now consider the increments of W: $W_{t_k} - W_{t_j}, 0 \le t_j < t_k \le T$. They are independent.

$$W_{t_k} - W_{t_j} = \sum_{i=1}^{k} \sqrt{\Delta t} Z_i - \sum_{i=1}^{j} \sqrt{\Delta t} Z_i = \sum_{i=j+1}^{k} \sqrt{\Delta t} Z_i$$

$$\mathbb{E}[W_{t_k} - W_{t_j}] = 0$$

$$\text{Var}\left(W_{t_k} - W_{t_j}\right) = (k-j)\Delta t = t_k - t_j$$

As $\Delta t \to 0$, we get $W_{t_k} - W_{t_j} \sim N(0, t_k - t_j)$.

***Definition:* 3.6: Brownian Motion**

A stochastic process $(W_t)_{t \ge 0}$ is a standard Brownian motion on $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\ge 0}, P\right)$ if

- $W_0 = 0$
- W has stationary increments with distribution $W_{t+s} - W_t \sim \mathcal{N}(0, s)$
- W has independent increments, $W_{t+s} - W_t \perp\!\!\!\perp W_t$
- W is pathwise continuous (but no-where differentiable).

Construction of Brownian motion:

1. Let $Z_1, ..., Z_n$ be i.i.d. standard normal.
2. Set $t_0 = 0$, $W(0) = 0$
3. Compute $W(t_{i+1}) = W(t_i) + \sqrt{t_{i+1} - t_i} Z_{i+1}$.

Remark 2. If $\{\mathcal{F}_t\}_{t\ge 0}$ is generated by W, then $W_{t+s} - W_t \perp\!\!\!\perp \mathcal{F}_t$. In particular, $\mathbb{E}[f(W_t - W_s)|\mathcal{F}_s] = \mathbb{E}[f(W_t - W_s)]$ if $s \le t$. Also $\mathbb{E}[W_t^2|\mathcal{F}_s] = \mathbb{E}[(W_t - W_s)^2 + 2W_tW_s - W_s^2|\mathcal{F}_s]$.

***Theorem:* 3.1: Holder's Inequality**

Let X, Y be r.v. and $p, q > 1$, $\frac{1}{p} + \frac{1}{q} = 1$. Then

$$\mathbb{E}[|XY|] \le \left(\mathbb{E}[|X|^p]\right)^{1/p} \left(\mathbb{E}[|Y|^q]\right)^{1/q}$$

If $p = q = 2$, $\mathbb{E}[|XY|] \le \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}$.

***Proposition:* 3.1:**

Brownian motion W is a martingale on $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\ge 0}, P\right)$, where $\{\mathcal{F}_t\}_{t\ge 0}$ is the filtration generated by W.

Proof.

1. By definition of $\mathcal{F}_t$
2. To show that $\mathbb{E}[|W_t|] < \infty, \forall t$, apply Theorem 3.1.

   $$\mathbb{E}[|W_t|] = \mathbb{E}[|W_t \cdot 1|] = \sqrt{\mathbb{E}[W_t^2] \cdot 1} = \sqrt{t} < \infty,$$

   because $W_t \sim \mathcal{N}(0, t)$
3. $\mathbb{E}[W_t|\mathcal{F}_s] = \mathbb{E}[W_t - W_s + W_s|\mathcal{F}_s] \underset{\text{Linearity}}{=} \mathbb{E}[W_t - W_s|\mathcal{F}_s] + \mathbb{E}[W_s|\mathcal{F}_s] = \underbrace{\mathbb{E}[W_t - W_s]}_{\text{Conditional Independence}} + W_s = W_s$

□

3.2 Stochastic Integration

Ordinary Differential Equation (ODE):

$$dX_t = a(t, X_t)dt \Leftrightarrow \frac{dX_t}{dt} = a(t, X_t) \Leftrightarrow X_t - X_t = \int_0^t a(s, X_s)ds$$

Stochastic Differential Equation (SDE):

$$dX_t = a(t, X_t)dt + b(t, X_t)dW_t \Leftrightarrow X_t - X_0 = \int_0^t a(s, X_s)ds + \int_0^t b(s, X_s)dW_s$$

- $a(t, X_t)$ is drift, the deterministic part
- $b(t, X_t)$ is diffusion

Definition: 3.7: Partition

A partition of $[0, T]$ is a set of points $\Pi = \{t_0, ..., t_n\}$ s.t. $0 = t_0 < t_1 < \cdots < t_n = T$. The norm of the partition is $\|\Pi\| = \max_{1\le k\le n}(t_k - t_{k-1})$. If we have even increments, then $\|\Pi\| = \Delta t = \frac{T}{n}$.

Definition: 3.8: Total and Quadratic Variation

Let X be a stochastic process. The *total variation* of X is $TV_X = \lim_{\|\Pi\|\to 0} \sum_{k=1}^n \left|X_{t_k} - X_{t_{k-1}}\right|$. The *quadratic variation* of X is $[X, X]_T = \lim_{\|\Pi\|\to 0} \sum_{k=1}^n (X_{t_k} - X_{t_{k-1}})^2$. TV is a proxy for differentiability.

Claim: The TV of a differentiable function is finite

Proof. Let $f(x)$ be a differentiable function.

$$\begin{aligned} TV_f &= \lim_{\|\Pi\|\to 0} \sum_{k=1}^n |f(t_k) - f(t_{k-1})| \\ &= \lim_{\|\Pi\|\to 0} \sum_{k=1}^n \left|f'(t_k^*)\right| |t_k - t_{k-1}| \text{ (By Mean Value Theorem) } t_k^* \in (t_{k-1}, t_k) \\ &\le \lim_{\|\Pi\|\to 0} \max_{1\le k\le n} \left|f'(t_k^*)\right| \sum_{k=1}^n |t_k - t_{k-1}| < \infty \end{aligned}$$

The last inequality is because $\sum_{k=1}^n |t_k - t_{k-1}| = T$. □

Claim: The TV of Brownian motion is infinite

Proof. Let $\epsilon > 0$. Since W is continuous, $\lim_{\|\Pi\|\to 0} P(|W_{t_k} - W_{t_{k-1}}| > \epsilon) = 0$.

However, $P\left(\sum_{k=1}^n (W_{t_k} - W_{t_{k-1}})^2 > 0\right) = 1$, since $\text{Var}(W) \neq 0$.

$$\sum_{k=1}^n (W_{t_k} - W_{t_{k-1}})^2 < \max_k \left|W_{t_k} - W_{t_{k-1}}\right| \sum_{k=1}^n \left|W_{t_k} - W_{t_{k-1}}\right|$$

Since LHS$\not\to 0$, but $\left|W_{t_k} - W_{t_{k-1}}\right| \to 0$, we must have $\sum_{k=1}^{n} \left|W_{t_k} - W_{t_{k-1}}\right| \to \infty$. *i.e.* TV_W is infinite. □

***Proposition:* 3.2:**

Let $(W_t)_{t\geq 0}$ be standard Brownian motion, $[W,W]_T = T$ a.s. (almost surely)

Proof. We show convergence in probability. Assume an even partition.

$$\begin{aligned}
[W,W]_T &= \lim_{\|\Pi\|\to 0} \sum_{k=1}^{n} (W_{t_k} - W_{t_{k-1}})^2 \\
&= \lim_{\|\Pi\|\to 0} \sum_{k=1}^{n} \left[(W_{t_k} - W_{t_{k-1}})^2 - (t_k - t_{k-1})\right] + \sum_{k=1}^{n} (t_k - t_{k-1})
\end{aligned}$$

By definition, $\lim_{\|\Pi\|\to 0} \sum_{k=1}^{n} (t_k - t_{k-1}) = T$. We consider the mean and variance of

$\lim_{\|\Pi\|\to 0} \sum_{k=1}^{n} \left[(W_{t_k} - W_{t_{k-1}})^2 - (t_k - t_{k-1})\right]$

$$\begin{aligned}
\mathbb{E}\left[\sum_{k=1}^{n} \left[(W_{t_k} - W_{t_{k-1}})^2 - (t_k - t_{k-1})\right]\right] &= \sum_{k=1}^{n} \mathbb{E}\left[(W_{t_k} - W_{t_{k-1}})^2\right] - \mathbb{E}\left[t_k - t_{k-1}\right] \\
&= \sum_{k=1}^{n} (t_k - t_{k-1}) - (t_k - t_{k-1}) = 0
\end{aligned}$$

Let $Z \sim \mathcal{N}(0,1)$, $W_{t_k} - W_{t_{k-1}} \sim \mathcal{N}(0, \Delta t) \sim \sqrt{\Delta t}\mathcal{N}(0,1)$.
Note that for $X \sim \mathcal{N}(0, \sigma^2)$, $\mathbb{E}[X^3] = 0$, $\mathbb{E}[X^4] = 3\sigma^4$.

$$\begin{aligned}
\operatorname{Var}\left[\sum_{k=1}^{n} \left[(W_{t_k} - W_{t_{k-1}})^2 - (t_k - t_{k-1})\right]\right] &= \sum_{k=1}^{n} \operatorname{Var}\left((W_{t_k} - W_{t_{k-1}})^2\right) \\
&= \sum_{k=1}^{n} \operatorname{Var}(\Delta t Z^2) \\
&= \sum_{k=1}^{n} (\Delta t)^2 \operatorname{Var}(Z^2) \\
&= (\Delta t)^2 \sum_{k=1}^{n} \left[\mathbb{E}[Z^4] - \mathbb{E}[Z^2]\right] \\
&= 2n(\Delta t)^2 = 2n\,(T/n)^2 = \frac{2T^2}{n}.
\end{aligned}$$

As $\|\Pi\| \to 0$, $n \to \infty$, Var $\to 0$.
Therefore, $[W,W]_T \to T$ in probability. □

Recall that the Riemann integrals is defined as

$$\int_a^b f(x)dx = \lim_{\|\Pi\|\to 0} \sum_{k=1}^{n} f(x_k^*)(x_k - x_{k-1}),$$

for some $x_k^* \in [x_{k-1}, x_k]$ and $\Pi = \{x_0, ..., x_n\}$.

However, $\int_0^t f(s,W_s)dW_s \neq \lim_{\|\Pi\|\to 0}\sum_{k=1}^n f(s^*,W_s^*)(W_{s_k}-W_{s_{k-1}})$, because TV of $W_{s_k}-W_{s_{k-1}}\to\infty$.

Example: Suppose we define $\int_0^t W_s dW_s$ by (1) $\lim_{\|\Pi\|\to 0}\sum_{k=1}^n W_{s_{k-1}}(W_{s_k}-W_{s_{k-1}})$ or (2) $\lim_{\|\Pi\|\to 0}\sum_{k=1}^n W_s(W_{s_k}-W_{s_{k-1}})$

$$
\begin{aligned}
\text{For (1), } \mathbb{E}\left[\sum_{k=1}^n W_{s_{k-1}}(W_{s_k}-W_{s_{k-1}})\right] &= \sum_{k=1}^n \mathbb{E}\left[W_{s_{k-1}}(W_{s_k}-W_{s_{k-1}})\right] \\
&= \sum_{k=1}^n \mathbb{E}[W_{s_{k-1}}]\mathbb{E}[W_{s_k}-W_{s_{k-1}}] \\
&\text{(Because } W_{s_k}-W_{s_{k-1}} \perp\!\!\!\perp W_{s_{k-1}}) \\
&= 0 \\
\text{For (2), } \mathbb{E}\left[\sum_{k=1}^n W_{s_k}(W_{s_k}-W_{s_{k-1}})\right] &= \sum_{k=1}^n \mathbb{E}\left[W_{s_k}(W_{s_k}-W_{s_{k-1}})\right] \\
&\text{(Note that } W_{s_k} \not\perp\!\!\!\perp W_{s_k}-W_{s_{k-1}}) \\
&= \sum_{k=1}^n \mathbb{E}\left[W_{s_k}^2 - W_{s_k}W_{s_{k-1}}\right] \\
&= \sum_{k=1}^n s_k - \mathbb{E}\left[W_{s_{k-1}}(W_{s_k}-W_{s_{k-1}})\right] - \mathbb{E}[W_{s_{k-1}}^2] \\
&= \sum_{k=1}^n s_k - 0 - s_{k-1} = s_n - s_0 = t
\end{aligned}
$$

The two values do not agree, so we cannot use the usual definition of Riemann integral here.

***Definition:* 3.9: L^2-Process**

A process X is $L^2[a,b]$ if

1. $\int_a^b \mathbb{E}[X_s]ds < \infty$
2. X is adapted

If $X \in L^2[0,t], \forall t > 0$, then X is L^2.
The mean squared limit (L^2-limit) of a sequence of r.v. X_n is a r.v. X s.t. $\lim_{n\to\infty}\mathbb{E}[(X_n-X)^2]=0$

***Definition:* 3.10: Ito Integral**

Let $f(t,X_t)$ be an adapted stochastic process satisfying $\mathbb{E}\left[\int_0^t f(s,X_s)ds\right] < \infty$. The integral of f w.r.t. a standard Brownian motion, called an Ito Integral, is

$$\int_0^t f(s,W_s)dW_s = L^2 - \lim_{n\to\infty}\sum_{k=1}^n f(s_{k-1},W_{s_{k-1}})[W_{s_k}-W_{s_{k-1}}]$$

Note the sum is a sequence of r.v. and the integral is a r.v.

Proposition: 3.3: Properties of Ito Integral

Let $I_t = \int_0^t f_s dW_s$, where $f_s = f(s, W_s)$.

1. I_t is a r.v. for t fixed and a stochastic process for t varied.
2. If f is deterministic, then I_t is normally distributed.
3. Ito integrals are linear, *i.e.* $\int_0^t [af_s + bg_s]dW_s = a\int_0^t f_s dW_s + b\int_0^t g_s dW_s$
4. Every Ito integral is a martingale
5. For any f_s, $\mathbb{E}\left[\int_0^t f_s dW_s\right] = 0$
6. Ito's Isometry:

$$\mathbb{E}\left[\int_0^t f_s dW_s \cdot \int_0^t g_s dW_s\right] = \mathbb{E}\left[\int_0^t f_s g_s ds\right] = \int_0^t \mathbb{E}[f_s g_s]ds$$

In particular, $\mathbb{E}\left[\left(\int_0^t f_s dW_s\right)^2\right] = \mathbb{E}\left[\int_0^t f_s^2 ds\right] = \int_0^t \mathbb{E}[f_s^2]ds$

3.3 Ito's Lemma

Consider the SDE: $dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t$, $X_0 = x_0$. Equivalenlty, $X_t - x_0 = \int_0^t \mu(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s$.

Proposition: 3.4: Uniqueness of SDE Solutions

Suppose μ, σ are adapted and there exists a constant K s.t. for all x, y, t, $|\mu(t,x) - \mu(t,y)| \leq K|x-y|$, $|\sigma(t,x) - \sigma(t,y)| \leq K|x-y|$, $|\mu(t,x)| + |\sigma(t,x)| \leq K(1+|x|)$. Then there exists a unique solution. The solution is Markov, $\mathcal{F}_t^W$-adapted with continuous trajectories.

Consider a deterministic smooth function $f(t,x)$. Its total derivative is $d(f(t,x)) = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial x}dx$. The 2nd order Taylor approximation is

$$df = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial x}dx + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(dx)^2 + \frac{1}{2}\frac{\partial^2 f}{\partial t^2}(dt)^2 + \frac{\partial^2 f}{\partial t \partial x}dtdx$$

Let X_t satisfy the SDE $dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t$. Let $f(t, X_t)$ be a $C^{1,2}$-function. Then

$$df(t, X_t) = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial x}dX_t + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(dX_t)^2 + \frac{1}{2}\frac{\partial^2 f}{\partial t^2}(dt)^2 + \frac{\partial^2 f}{\partial t \partial x}dtdX_t$$

$(dX_t)^2 = \mu^2(dt)^2 + 2\mu\sigma dtdW_t + \sigma^2(dW_t)^2$. $(dt)^2$ and $dtdW_t$ tend to 0 faster than dt and dX_t as $dt \to 0$, so we exclude them. Also $(dW_t)^2 = dt$ by the quadratic variance of W. Therefore,

$$\begin{aligned} df &= \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial x}(\mu dt + \sigma dW_t) + \frac{1}{2}\sigma^2\frac{\partial^2 f}{\partial x^2}dt \\ &= \left(\frac{\partial f}{\partial t} + \mu\frac{\partial f}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 f}{\partial x^2}\right)dt + \sigma\frac{\partial f}{\partial x}dW_t \end{aligned}$$

Example: Compute $d(te^{aW_t})$.

Proof. Here $X_t = W_t$, $dX_t = 0dt + dW_t$, $\mu = 0, \sigma = 1$, $f(t,w) = te^{aw}$. $\frac{\partial f}{\partial t} = e^{aw}$, $\frac{\partial f}{\partial w} = ate^{aw}$, $\frac{\partial^2 f}{\partial w^2} = a^2te^{aw}$.
Therefore $d(te^{aW_t}) = \left(e^{aW_t} + \frac{1}{2}a^2te^{aW_t}\right)dt + ate^{aW_t}dW_t$. □

Example: Compute $d((W_t)^2)$.

Proof. Here $\mu = 0, \sigma = 1$, $f(t,w) = w^2$. $\frac{\partial f}{\partial t} = 0$, $\frac{\partial f}{\partial w} = 2w$, $\frac{\partial^2 f}{\partial w^2} = 2$.
Therefore $d((W_t)^2) = dt + 2W_tdW_t$. □

Example: Compute $\int_0^t W_s dW_s$.

Proof. Consider the previous result $d((W_t)^2) = dt + 2W_tdW_t$. Integrate both sides and use the fact that $W_0 = 0$.

$$W_t^2 - W_0^2 = \int_0^t ds + 2\int_0^t W_s dW_s$$

Therefore, $\int_0^t W_s dW_s = \frac{1}{2}\left(W_t^2 - t\right)$ □

Example: Compute $\mathbb{E}[W_t^4]$.

Proof. $f(t,w) = w^4$, $\frac{\partial f}{\partial t} = 0$, $\frac{\partial f}{\partial w} = 4w^3$, $\frac{\partial^2 f}{\partial w^2} = 12w^2$.
Then $d((W_t)^4) = \frac{1}{2}12W_t^2dt + 4W_t^3dW_t = 6W_t^2dt + 4W_t^3dW_t$
Integrate both sides, $W_t^4 = 6\int_0^t W_s^2 ds + 4\int_0^t W_s^3 dW_s$. Note $\mathbb{E}\left[\int_0^t W_s^3 dW_s\right] = 0$ and $\mathbb{E}[W_s^2] = s$
Therefore, $\mathbb{E}[W_t^4] = 6\int_0^t \mathbb{E}[W_s^2]ds = 6\int_0^t s ds = 3t^2$. □

***Definition:* 3.11: Correlated Brownian Motion**

Correlated Brownian motion $(W_t^1, W_t^2)_{t\geq 0}$ with instantaneous correlation ρ are joint processes satisfying
1. $W_0^1 = W_0^2 = 0$
2. (W^1, W^2) has continuous sample paths
3. (W^1, W^2) has indepedent increments
4. (W^1, W^2) has stationary increments with $\begin{bmatrix} W_{t+s}^1 - W_t^1 \\ W_{t+s}^2 - W_t^2 \end{bmatrix} = \mathcal{N}\left(0, \begin{bmatrix} s & \rho s \\ \rho s & s \end{bmatrix}\right)$

***Theorem:* 3.2: 2D Ito's Lemma**

Let X^1, X^2 with SDEs: $dX_t^i = \mu_t^i dt + \sigma_t^i dW_t^i$ where (W^1, W^2) has instantaneous correlation ρ. Then

$$\begin{aligned} d(f(t, X_t^1, X_t^2)) = &\left[\frac{\partial f}{\partial t} + \sum_{i=1}^{2}\mu_t^i \frac{\partial f}{\partial x_i} + \frac{1}{2}\sum_{i=1}^{2}(\sigma_t^i)^2\frac{\partial^2 f}{\partial x_i^2} + \rho\sigma_t^1\sigma_t^2\frac{\partial^2 f}{\partial x_1 \partial x_1}\right]dt \\ &+ \sum_{i=1}^{2}\sigma_t^i\frac{\partial f}{\partial x_i}dW_t^i \end{aligned}$$

This can be extended to n-dimensional cases.

Example: Let W^1, W^2 be correlated Brownian motions with instantaneous correlation ρ. Find an SDE for $W_t^1 W_t^2$.

Proof. Let $f(t, W_t^1, W_t^2) = W_t^1 W_t^2$, $\frac{\partial f}{\partial t} = 0$, $\frac{\partial f}{\partial w^1} = w^2$, $\frac{\partial f}{\partial w^2} = w^1$, $\frac{\partial^2 f}{\partial w^1 \partial w^2} = 1$.
Then $df(t, w_t^1, w_t^2) = \rho dt + W_t^2 dW_t^1 + W_t^1 dW_t^2$. ☐

3.4 Geometric Brownian Motion

The SDE is $dX_t = \mu X_t dt + \sigma X_t dW_t$, $X_0 = x_0$. We want to use Ito's lemma to get an expansion for X_t, but we do not know X_t explicitly. Consider $Z_t = \ln X_t$.

$$\begin{aligned} dZ_t &= \left[\mu X_t \frac{1}{X_t} + \frac{1}{2}\sigma^2 X_t^2 \left(-\frac{1}{X_t^2}\right)\right] dt + \sigma X_t \frac{1}{X_t} dW_t \\ &= \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma dW_t \end{aligned}$$

Integrate both sides

$$Z_t - Z_0 = \int_0^t \left(\mu - \frac{1}{2}\sigma^2\right) ds + \sigma \int_0^t dW_s = \left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W_t$$

Therefore, $X_t = x_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W_t\right)$.

This shows that $\frac{X_t}{X_0} \sim \text{LogNormal}\left((\mu - \frac{1}{2}\sigma^2)t, \sigma^2 t\right)$

***Proposition:* 3.5:**

If X is GBM, then for $\beta \in \mathbb{R}$, $Y = X^\beta$ is GBM.

4 Continuous Time Market

Definition: 4.1: Portfolio

The financial market is defined as a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, P)$. Suppose there are N assets with price processes $X^i = (X_t^i)_{t\geq 0}$, $i = 1, ..., N$, X^i adapted. A *portfolio* $h_t = (h_t^1, ..., h_t^N)$ is an $\mathbb{F}$-adapted N-dim process, where h_t^i is the number of units of asset i held at time t. Let V_t^h be the value of the portfolio at time t.

$$V_t^h = \sum_{i=1}^N h_t^i X_t^i$$

A portfolio h is called *self-financing* if its value process satisfies

$$dV_t^h = \sum_{i=1}^N h_t^i dX_t^i$$

Define the relative portfolio weights

$$w_t^i = \frac{h_t^i X_t^i}{V_t^h}, \sum_{i=1}^N w_t^i = 1$$

The self-financing condition can be written as

$$dV_t^h = V_t^h \sum_{i=1}^N \frac{w_t^i}{X_t^i} dX_t^i$$

Definition: 4.2: Arbitrage Portfolio

An arbitrage portfolio is a self-financing portfolio h s.t.

1. $V_0^h = 0$
2. $P(V_t^h \geq 0) = 1$
3. $P(V_t^h > 0) > 0$

A market is arbitrage-free if there are no arbitrage opportunities.

Proposition: 4.1:

Suppose there is a risk-free asset B (e.g. bond) with dynamics $dB_t = r_t B_t dt$ (r is adapted). If there exists a self-financing portfolio h whose value process V_t^h has dynamics $dV_t^h = k_t V_t^h dt$, where k is adapted, then it must hold that $k_t = r_t$ for all t. Otherwise, there exists an arbitrage opportunity.

Proof. If $k > r$ constat. Borrow from bank and invest in h. Costs nothing at $t = 0$, return is positive for $t > 0$. □

Markets with 2 Assets The first asset is a risk free asset with process B. The price dynamics is

$$dB_t = r_t B_t dt,$$

where r_t is an adapted process, called *short rate* or *risk-free rate*. The dynamics is locally deterministic. Assume $B_0 = 1$, we have

$$B_t = \exp\left(\int_0^t r_s ds\right)$$

The second asset is a stock with price process S. The stock price dynamics is

$$dS_t = \mu(t, S_t)S_t dt + \sigma(t, S_t)S_t dW_t,$$

where $\mu(t, S_t)$ is the local mean rate of return, $\sigma(t, S_t)$ is the volatility.

A classical model is where (r, μ, σ) are constants.

***Definition:* 4.3: Black-Scholes Model**

The Black-Scholes model consists of 2 assets with dynamics:

$$\begin{aligned} dB_t &= rB_t dt \\ dS_t &= \mu S_t dt + \sigma S_t dW_t \end{aligned}$$

Recall that $B_t = \exp(rt)$, $S_t = s_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t\right)$

4.1 Arbitrage Pricing

We take as given the model:

$$\begin{cases} dB_t = r_t B_t dt \\ dS_t = \mu(t, S_t)S_t dt + \sigma(t, S_t)S_t dW_t \\ B_0 = 1, S_0 = s_0 \end{cases}$$

***Definition:* 4.4: Derivative**

A contingent claim (derivative) with maturity date T is, at time t, a random variable $X_T \subset \mathcal{F}_t^S$. A contingent claim X_t is simple if it is of the form $X_T = F(S_T)$, where F is the payoff function.

Example: $F(S_T) = (K - S_T)_+$ is European put, $F(S_T) = (S_T - K)_+$ is European call.

The goal is to determine the fair (arbitrage-free) price at $t \in [0, T)$. The fair price at time T is $F(S_T)$. Let X_t denote the price of the derivative. Assume that

1. The derivative can be bought and sold.
2. The price process for the deerivative is of the form $f_t = f(t, S_t)$, where f is a smooth function.
3. There are no arbitrage opportunities on the market of 3 assets: B_t, S_t, X_t.

Let w^s, w^f be the relative weights in $S_t, f(t, S_t)$. By Ito's Lemma

$$df(t, S_t) = \left(\frac{\partial f}{\partial t} + \mu S \frac{\partial f}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2}\right) dt + \sigma S \frac{\partial f}{\partial S} dW_t$$

$$\begin{aligned} dV_t &= V_t \left(\frac{w^s}{S} dS_t + \frac{w^f}{f} df(t, S_t)\right) \\ &= V_t \left(w^s \mu + \frac{w^f}{f}\left(\frac{\partial f}{\partial t} + \mu S \frac{\partial f}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2}\right)\right) dt + V_t \left(w^s \sigma + \frac{w^f}{f} \sigma S \frac{\partial f}{\partial S}\right) dW_t \end{aligned}$$

Set dW_t term to 0, we have

$$\begin{cases} 0 = w^s + \frac{w^f}{f} S \frac{\partial f}{\partial S} \\ 1 = w^s + w^f \end{cases} \Rightarrow \begin{cases} w^f = \frac{f}{f - S\frac{\partial f}{\partial S}} \\ w^s = \frac{-S\frac{\partial f}{\partial S}}{f - S\frac{\partial f}{\partial S}} \end{cases}$$

Also, by Proposition 4.1,

$$w^s \mu + \frac{w^f}{f}\left(\frac{\partial f}{\partial t} + \mu S \frac{\partial f}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2}\right) = r$$

This gives

$$0 = \frac{\partial f}{\partial t} + rS\frac{\partial f}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} - rf(t, S_t)$$
$$f(T, S_T) = F(S_T)$$

This is the *Black-Scholes equation.* S_t can take any positive values. f is the solution to the deterministic PDE. Note that the local mean return μ has no impact on the arbitrage-free price.

***Theorem:* 4.1: Feynman-Kac**

Assume f is a solution to the BVP

$$0 = \frac{\partial f}{\partial t} + a(t,x)\frac{\partial f}{\partial x} + \frac{1}{2}b^2(t,x)\frac{\partial^2 f}{\partial x^2} - rf(t,x)$$
$$f(T,x) = F(x)$$

Then f has the stochastic representation $f(t,x) = \exp(-r(T-t))\mathbb{E}[F(X_T)|X_t = x]$, where X satisfies the SDE: $dX_u = a(u, X_u)du + b(u, X_u)dW_u, X_t = x$.

Apply Theorem 4.1 to Black-Scholes equation, we have

$$f(t,S) = \exp(-r(T-t))\mathbb{E}[F(X_t)|X_t = s],$$

where $dX_u = rX_u du + \sigma X_u dW_u, X_t = s$. This looks like the SDE for S_t, but $r = \mu(t, S_t)$. They are the dynamics of S under some measure $\mathbb{Q}$. Introduce the measure $\mathbb{Q}$ for pricing. Use $\mathbb{P}$ to denote the probability measure giving the original dynamics of S.

***Theorem:* 4.2: Black-Schole Pricing**

The arbitrage-free price of a derivative with payoff $F(S_T)$ is $f(t, S_t)$, where f is given by

$$f(t,s) = \exp(-r(T-t))\mathbb{E}^{\mathbb{Q}}[F(S_t)|S_t = s],$$

and S has $\mathbb{Q}$-dynamics, $dS_u = rS_u du + \sigma(u, S_u)dW_t^{\mathbb{Q}}, S_t = s$. $\mathbb{Q}$ is the risk-neutral/martingale measure.

***Theorem:* 4.3:**

In the Black-Scholes model, the normalized process of every traded asset (including derivatives) is a $\mathbb{Q}$-martingale. *i.e.* if X_t is the price process of an asset, then $Z_t = \frac{X_t}{B_t}$ is a $\mathbb{Q}$-martingale.

4.2 The Black-Scholes Formula

Under the Black-Scholes market, where r, μ, σ are constant, derive the arbitrage-free price of a European call with strike K and maturity T.
Payoff: $F(S_T) = (S_T - K)_+ = (S_T - K)\chi_{\{S_T > K\}}$. Using the risk-neutral pricing formula (Theorem 4.2):

$$\begin{aligned} f(t,s) &= \exp(-r(T-t))\mathbb{E}^{\mathbb{Q}}[(S_T - K)_+ | S_t = s] \\ dS_u &= rS_u du + \sigma S_u dW_u, S_t = s \end{aligned}$$

Recall $S_T = S\exp\left[\left(r - \frac{1}{2}\sigma^2\right)(T-t) + \sigma(W_T^{\mathbb{Q}} - W_t^{\mathbb{Q}})\right] = S\exp(X)$, where $X \sim \mathcal{N}\left(\left(r - \frac{1}{2}\sigma^2\right)(T-t), \sigma^2(T-t)\right)$. For this normal distribution, the CDF is

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma\sqrt{T-t}}\exp\left(-\frac{1}{2}\left(\frac{x - \left(r - \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}\right)^2\right)$$

Note: since $S_T > K$, $S\exp(x) > K$, we only need to consider $x > \ln(K/S)$.

$$\begin{aligned} f(t,s) &= \exp(-r(T-t))\int_{-\infty}^{\infty} F(S\exp(X))f_X(x)dx \\ &= \exp(-r(T-t))\left(\mathbb{E}^{\mathbb{Q}}\left[S_T\chi_{S_T>K} | S_t = s\right] - K\mathbb{E}^{\mathbb{Q}}\left[\chi_{S_T>K} | S_t = s\right]\right) \end{aligned}$$

The first part:

$$\begin{aligned} &\exp(-r(T-t))\mathbb{E}^{\mathbb{Q}}\left[S_T\chi_{S_T>K} | S_t = s\right] \\ &= s\exp(-r(T-t))\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma\sqrt{T-t}}\exp(x)\exp\left(-\frac{1}{2}\left(\frac{x - \left(r - \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}\right)^2\right)dx \end{aligned}$$

Change variable by $z = \frac{x - \left(r - \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}$, $dz = \frac{dx}{\sigma\sqrt{T-t}}$, $z_0 = \frac{\ln(K/S) - (r + \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}$. Then

$$\begin{aligned} &\exp(-r(T-t))\mathbb{E}^{\mathbb{Q}}\left[S_T\chi_{S_T>K} | S_t = s\right] \\ &= s\int_{z_0}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^2\right) \\ &= s\Phi(-z_0) = s\Phi(d_+), \text{ where} \\ d_\pm &= \frac{\ln(S/K) + (r \pm \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}} \end{aligned}$$

For the second part,

$$\begin{aligned} &-K\exp(-r(T-t))\mathbb{E}^{\mathbb{Q}}\left[\chi_{S_T>K} | S_t = s\right] \\ &= -K\exp(-r(T-t))\mathbb{Q}(S_T > K | S_t = s) \\ &= -K\exp(-r(T-t))\mathbb{Q}(s\exp(X) > K) \\ &= -K\exp(-r(T-t))\mathbb{Q}(X > \ln(K/S)) \\ &= -K\exp(-r(T-t))\mathbb{Q}\left(\sigma\sqrt{T-t}Z > \ln(K/S) - (r - \frac{1}{2}\sigma^2)(T-t)\right) \\ &= -K\exp(-r(T-t))\Phi(d_-) \end{aligned}$$

Therefore, the European call price is

$$f(t, S_t) = S_t\Phi(d_+) - K\exp(-r(T-t))\Phi(d_-),$$

where $d_\pm = \frac{\ln(S/K)+(r\pm\frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$.

What is $\mathbb{Q}$?

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t^{\mathbb{P}} = rdt + \sigma dW_t^{\mathbb{P}}$$
$$\frac{\mu - r}{\sigma}dt + dW_t^{\mathbb{P}} = dW_t^{\mathbb{Q}}$$

***Theorem:* 4.4: Radon-Nikodym**

Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on the same measurable space $(\Omega, \mathcal{F})$ s.t. $\mathbb{Q} \ll \mathbb{P}$ ($\mathbb{P}(E) = 0 \Rightarrow \mathbb{Q}(E) = 0$). Then thre exists a r.v. $\frac{d\mathbb{Q}}{d\mathbb{P}}$ s.t. for any r.v. X, $\mathbb{E}^{\mathbb{Q}}[X] = \mathbb{E}^{\mathbb{P}}\left[X\frac{d\mathbb{Q}}{d\mathbb{P}}\right]$, $\frac{d\mathbb{Q}}{d\mathbb{P}}$ is the Radon-Nikodym derivative, and satisfies

1. $\frac{d\mathbb{Q}}{d\mathbb{P}} > 0$ $\mathbb{P}$-a.s.
2. $\mathbb{E}^{\mathbb{P}}\left[\frac{d\mathbb{Q}}{d\mathbb{P}}\right] = 1$

***Theorem:* 4.5: Girsanov**

Let W be a $\mathbb{P}$-Brownian motion and λ_t be an adapted process. Define a new process L_t on $[0, T]$ by

$$L_t = \exp\left(-\int_0^t \frac{\lambda_u^2}{2}du + \int_0^t \lambda_u dW_u\right)$$

Assume $\mathbb{E}^{\mathbb{P}}[L_T] = 1$. Define a new probability measure $\mathbb{Q}$ on $\mathcal{F}_T$ by $\frac{d\mathbb{Q}}{d\mathbb{P}} = L_T$. Then

$$dW_t^{\mathbb{P}} = -\lambda_t dt + dW_t^{\mathbb{Q}},$$

where $W^{\mathbb{Q}}$ is a $\mathbb{Q}$-Brownian motion. *i.e.* $W_t^{\mathbb{Q}} = W_t^{\mathbb{P}} + \int_0^t \lambda_u du$ is a Brownian motion.

Example: (Black-Schole Model) Define $\frac{d\mathbb{Q}}{d\mathbb{P}}$ as a in Theorem 4.5 for some λ_t adapted. Then

$$W_t^{\mathbb{Q}} = W_t^{\mathbb{P}} + \int_0^t \lambda_u du$$
$$dS_t = \mu S_t dt + \sigma S_t dW_t^{\mathbb{P}} = (\mu S_t - \sigma S_t \lambda_t)dt + \sigma S_t dW_t^{\mathbb{Q}}$$

For $\mathbb{Q}$ to be a martingale measure, we must have $\mu - \sigma\lambda_t = r$, or $\lambda_t = \frac{\mu - r}{\sigma}$

***Proposition:* 4.2: Linearity of Price**

Let F and G be payoff functions for derivatives $X = F(S_T)$ and $Y = G(S_T)$ with price processes f and g. Then for $\alpha, \beta \in \mathbb{R}$, the price process of the claim $\alpha F(S_T) + \beta G(S_T)$ is $\alpha f(t, S_t) + \beta g(t, S_t)$.

Proof.

$$\begin{aligned} e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}[\alpha F(S_T) + \beta G(S_T)|S_t = s] &= \alpha e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}[F(S_T)|S_t = s] + \beta e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}[G(S_T)|S_t = s] \\ &= \alpha f(t, s) + \beta g(t, s) \end{aligned}$$

□

Proposition: 4.3: Put-Call Parity

Let $p(t,s)$ be the price of a European put with strike K and maturity T. Let $c(t,s)$ be the price of a European call with strike K and maturity T. Then $p(t,s) = c(t,s) + Ke^{-r(T-t)} - s$.

Proof. $F_{put}(S_T) = (K - S_T)_+ = (S_T - K)_+ - S_T + K$. Therefore,

$$\begin{aligned} p(t,s) &= \exp(-r(T-t))\mathbb{E}^{\mathbb{Q}}[(S_T - K)_+ - S_T + K | S_T = s] \\ &= c(t,s) - \exp(-r(T-t))s\exp(r(T-t)) + \exp(r(T-t))K \\ &= c(t,s) + Ke^{-r(T-t)} - s \end{aligned}$$

□

Therefore, the European put price is

$$\begin{aligned} f^{put}(r,s) &= S_t\phi(d_+) - K\exp(-r(T-t))\Phi(d_-) - S_t + K\exp(-r(T-t)) \\ &= K\exp(-r(T-t))\Phi(-d_-) - S_t\Phi(-d_+) \end{aligned}$$

4.3 Greeks

How option prices change w.r.t. changes in underlying prices, time, and risk-free rates?

1. Delta: $\Delta = \frac{\partial f}{\partial S}$, $\Delta^{call} = \Phi(d_+)$, $\Delta^{put} = -\Phi(-d_+) = \Phi(d_+) - 1$
2. Gamma: $\Gamma = \frac{\partial^2 f}{\partial S^2}$, $\Gamma^{call} = \Gamma^{put} = \frac{\Phi(d_+)}{S\sigma\sqrt{T-t}}$
3. Rho: $\rho = \frac{\partial f}{\partial r}$, $\rho^{call} = K(T-t)\exp(-r(T-t))\Phi(d_-)$, $\rho^{put} = -K(T-t)\exp(-r(T-t))\Phi(-d_+)$
4. Theta: $\theta = \frac{\partial f}{\partial t}$
5. Vega: $\nu = \frac{\partial f}{\partial \sigma}$, $\nu^{call} = \nu^{put} = S\Phi(d_+)\sqrt{T-t}$

5 Numerical Methods

5.1 Simulate Random Variables

Let $U \sim \text{Unif}(0, 1)$. Consider the CDF, $F(x) = P(X \leq x)$ is non-decreasing and right-continuous. We can simulate discrete random variables $X = x_i, i = 0, 1, 2, ...$ with probability p_i s.t. $\sum p_i = 1, p_i \geq 0$ with

$$X = \begin{cases} x_1, u \leq p_1 \\ x_2, p_1 < u \leq p_2 \\ \vdots \\ x_i, \sum_{j=1}^{i-1} p_j < u \leq \sum_{j=1}^{i} p_j \\ \vdots \end{cases}$$

Recall: For $U \sim \text{Unif}(0, 1)$ and $0 < a < b < 1$, then $P(a < U \leq b) = b - a$.

If x_i are orderd, then $P(X = x_i) = P\left(\sum_{j=1}^{i-1} p_i < U \leq \sum_{j=1}^{i} p_i\right) = \sum_{j=1}^{i} p_j - \sum_{j=1}^{i-1} p_j = p_i$.
$P(F(x_{i-1}) < U \leq F(x_i)) = P(x_{i-1} < F^{-1}(U) < x_1)$

***Theorem:* 5.1: Inverse Transform Method**

Let $U \sim \text{Unif}(0, 1)$. For any CDF denoted F, the random variable $F^{-1}(U)$ is distributed as F, where $F^{-1}(u) = \inf \{x : F(x) \geq u\}$.

Proof. We simulate U and sample $F^{-1}(U)$. Consider $P(F^{-1}(U) \leq x)$.
Since F is monotone, we have $F^{-1}(U) \leq x \Rightarrow F(F^{-1}(U)) \leq F(x)$ and $F(F^{-1}(U)) = U$ a.e. Therefore, $P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x)$. □

Another way is to generate samples from a convenient distribution and accept/reject a subset of the generated candidates.

- Target distribution: $f(x)$ with support on D_f, difficult to generate samples
- Proposal distribution: $g(x)$ with support on $D_g \supset D_f$, easy to generate samples

We need to find a constant $c \geq 1$ s.t. $f(x) \leq cg(x), \forall x \in D_f$.

***Definition:* 5.1: Acceptance/Rejection Method**

The acceptance/rejection method consists of the following steps:

1. Find c that bounds the ratio $\frac{f}{g}$
2. Generate $Y \sim g$
3. Generate $U \sim \text{Unif}(0, 1)$, independent of Y
4. If $U \leq \frac{f(Y)}{cg(Y)}$, then $X = Y$, otherwise, return to 2.

***Theorem:* 5.2:**

The random variable generated with acceptance/rejection method is distributed as f. In addition, the number of candidates generated until one is accepted is geometrically distributed with mean c.

Proof. Let $D \subset D_f$, $P(X \in D) = P\left(Y \in D | U \leq \frac{f(Y)}{cg(Y)}\right) = \frac{P\left(Y \in D, U \leq \frac{f(Y)}{cg(Y)}\right)}{P\left(U \leq \frac{f(Y)}{cg(Y)}\right)}$.

$$
\begin{aligned}
P\left(U \leq \frac{f(Y)}{cg(Y)}\right) &= \int_{\mathbb{R}} P\left(U \leq \frac{f(y)}{cg(y)}\right) g(y)dy = \int_{\mathbb{R}} \frac{f(y)}{cg(y)} dy = \frac{1}{c} \\
P(X \in D) &= cP\left(Y \in D, U \leq \frac{f(Y)}{cg(Y)}\right) \\
&= c\int_R P\left(Y \in D, U \leq \frac{f(y)}{cg(y)}\right) g(y)dy \\
&= c\int_D \frac{f(y)}{cg(y)} g(y)dy = \int_D f(y)dy
\end{aligned}
$$

□

5.2 Simulate SDEs

Brownian Motion Given timesteps $t_0 < t_1 < \cdots < t_n$, generate $W(t_1), ..., W(t_n)$. Algorithm:

1. Set $W(0) = 0$
2. For $i = 1, 2, ..., n$, $W(t_{i+1}) = W(t_i) + \sqrt{t_{i+1} - t_i} Z_{i+1}$, where $Z_i \sim \mathcal{N}(0, 1)$ iid.

The method is exact, the joint distribution of $W(t_1), ..., W(t_n)$ is the same as Brownian motion at $t_1, ..., t_n$.

For n-dimension, $W_{t+s} - W_t \sim \mathcal{N}(0, \Sigma)$. We find B such that $BB^T = \Sigma$ by Cholesky decomposition, because Σ is PSD.

1. Set $W(0) = 0$
2. Compute B.
3. For $i = 1, 2, ..., n$, $W(t_{i+1}) = W(t_i) + \sqrt{t_{i+1} - t_i} B Z_{i+1}$, where $Z_i \sim \mathcal{N}(0, 1)$ iid.

For 2D, $\Sigma = \begin{bmatrix} s & \rho s \\ \rho s & s \end{bmatrix}$, $B = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1 - \rho^2} \end{bmatrix}$

GBM Simulation $dS_t = \mu S_t dt + \sigma S_t dW_t$, $S_0 = s_0$. $S_t = s_0 \exp\left(\left(u - \frac{1}{2}\sigma^2\right) t + \sigma W_t\right)$. If start at $u < t$, $S_t = S_u \exp\left(\left(u - \frac{1}{2}\sigma^2\right)(t - u) + \sigma(W_t - W_u)\right)$.

1. Set $S_0 = s_0$
2. For $i = 0, 1, ..., n-1$, $S(t_{i+1}) = S(t_i) \exp\left(\left(u - \frac{1}{2}\sigma^2\right)(t_{i+1} - t_i) + \sigma\sqrt{t_{i+1} - t_i} Z_{i+1}\right)$, where $Z_i \sim \mathcal{N}(0, 1)$ iid.

Discretization Scheme Consider $dX_t = a(t, X_t)dt + b(t, X_t)dW_t$, $X_0 = x_0$. The goal is to simulate X_t at $t_1 < t_2 < \cdots < d_n$. An exact scheme would be:

$$
X_{t_{i+1}} = X_{t_i} + \int_{t_i}^{t_{i+1}} a(s, X_s)ds + \int_{t_i}^{t_{i+1}} b(s, X_s)dW_s
$$

Assume time is evenly splitted $t_{i+1} - t_i = \Delta t$, and X is 1-D. Then the simplest method is Euler scheme:

$$\int_{t_i}^{t_{i+1}} a(s, X_s)ds = \int_{t_i}^{t_{i+1}} a(t_i, X_{t_i})ds = a(t_i, X_{t_i})\Delta t$$
$$\int_{t_i}^{t_{i+1}} b(s, X_s)dW_s = \int_{t_i}^{t_{i+1}} b(t_i, X_{t_i})dW_s = b(t_i, X_{t_i})(W_{t_{i+1}} - W_{t_i})$$
$$\hat{X}_{t_{i+1}} = \hat{X}_{t_i} + a(t_i, X_{t_i})\Delta t + b(t_i, X_{t_i})\sqrt{\Delta t}Z_{i+1}$$

For GBM, it is:

$$\hat{S}_{t_{i+1}} = \hat{S}_{t_i} + \mu\hat{S}_{t_i}\Delta t + \sigma\hat{S}_{t_i}\sqrt{\Delta t}Z_{i+1}$$

Milstein scheme The basic discretization is not accurate enough for the diffusion term. By Ito's lemma:

$$d(b(t, X_t)) = \left(\frac{\partial b}{\partial t} + a\frac{\partial b}{\partial x} + \frac{1}{2}b^2\frac{\partial^2 b}{\partial x^2}\right)dt + b\frac{\partial b}{\partial x}dW_t$$

Let $t \in [u, u + \Delta t]$.

$$\begin{aligned} b(t, X_t) &= b(u, X_u) + (\cdot)(t - u) + b(u, X_u)\frac{\partial b}{\partial x}(W_t - W_u) \\ &= b(u, X_u) + b(u, X_u)\frac{\partial b}{\partial x}(u, X_u)(W_t - W_u) \end{aligned}$$

Then

$$\begin{aligned} \int_{t_i}^{t_{i+1}} b(s, X_s)dW_s &= \int_{t_i}^{t_{i+1}} b(t_i, X_{t_i}) + b(t_i, X_{t_i})\frac{\partial b}{\partial x}(t_i, X_{t_i})(W_s - W_{t_i})dW_s \\ &= b(t_i, X_{t_i})(W_{t_{i+1}} - W_{t_i}) + b(t_i, X_{t_i})\frac{\partial b}{\partial x}(t_i, X_{t_i})\int_{t_i}^{t_{i+1}}(W_s - W_{t_i})dW_s \end{aligned}$$

Recall that $\int_u^t W_s ds = \frac{1}{2}\left(W_t^2 - W_u^2 - (t - u)\right)$. Therefore,

$$\begin{aligned} \int_{t_i}^{t_{i+1}} (W_s - W_{t_i})dW_s &= \int_{t_i}^{t_{i+1}} W_s dW_s - W_{t_i}\int_{t_i}^{t_{i+1}} dW_s \\ &= \frac{1}{2}\left(W_{t_{i+1}}^2 - W_{t_i}^2 - (t_{i+1} - t_i)\right) - W_{t_i}(W_{t_{i+1}} - W_{t_i}) \\ &= \frac{1}{2}\left((W_{t_{i+1}} - W_{t_i})^2 - (t_{i+1} - t_i)\right) \\ &= \frac{1}{2}\left((\Delta W_t)^2 - \Delta t\right) = \frac{1}{2}\Delta t\left(Z_{i+1}^2 - 1\right) \end{aligned}$$

Then we have:

$$\hat{X}_{t_{i+1}} = \hat{X}_{t_i} + a(t_i, X_{t_i})\Delta t + b(t_i, X_{t_i})\sqrt{\Delta t}Z_{i+1} + \frac{1}{2}b(t_i, X_{t_i})\frac{\partial b}{\partial x}(t_i, X_{t_i})\Delta t(Z_{i+1}^2 - 1)$$

2nd Order Scheme Expand $a(t, X_t)$ using Ito's lemma:

$$d(a(t, X_t)) = \left(\frac{\partial a}{\partial t} + a\frac{\partial a}{\partial x} + \frac{1}{2}b^2\frac{\partial^2 a}{\partial x^2}\right) dt + b\frac{\partial a}{\partial x}dW_t$$

Introduce operators $L^0 = \frac{\partial}{\partial t} + a\frac{\partial}{\partial x} + \frac{1}{2}b^2\frac{\partial^2}{\partial x^2}$, $L^1 = b\frac{\partial}{\partial x}$. Then $d(a(t, X_t)) = L^0 a dt + L^1 a dW_t$. Let $u < s$,

$$a(s, X_s) - a(u, X_u) = \int_u^s L^0 a(r, X_r) dr + \int_u^s L^1 a(r, X_r) dW_r$$

Apply Euler's approximation,

$$\begin{aligned}
a(s, X_s) &= a(u, X_u) + L^0 a(u, X_u) \int_u^s dr + L^1 a(u, X_u) \int_u^s dW_r \\
&= a(u, X_u) + L^0 a(u, X_u)(s - u) + L^1 a(u, X_u)(W_s - W_u) \\
\int_{t_i}^{t_{i+1}} a(s, X_s) ds &= \int_{t_i}^{t_{i+1}} a(t_i, X_{t_i}) + L^0 a(t_i, X_{t_i})(s - t_i) + L^1 a(t_i, X_{t_i})(W_s - W_{t_i}) ds \\
&= a(t_i, X_{t_i})\Delta t + L^0 a(t_i, X_{t_i})\frac{1}{2}(\Delta t)^2 + L^1 a(t_i, X_{t_i})\Delta I_t,
\end{aligned}$$

where $\Delta I_t = \int_{t_i}^{t_{i+1}} (W_s - W_{t_i}) ds$. Similary,

$$\begin{aligned}
\int_{t_i}^{t_{i+1}} b(s, X_s) dWs &= \int_{t_i}^{t_{i+1}} b(t_i, X_{t_i}) + L^0 b(t_i, X_{t_i})(s - t_i) + L^1 b(t_i, X_{t_i})(W_s - W_{t_i}) dW_s \\
\int_{t_i}^{t_{i+1}} (s - t_i) dW_s &= \int_{t_i}^{t_{i+1}} s dW_s - t_i \int_{t_i}^{t_{i+1}} dW_s \\
&= t_{i+1}W_{t_{i+1}} - t_i W_{t_i} - \int_{t_i}^{t_{i+1}} W_s ds - t_i W_{t_{i+1}} + t_i W_{t_i} \\
&= (W_{t_{i+1}} - W_{t_i})(t_{i+1} - t_i) - \int_{t_i}^{t_{i+1}} (W_s - W_{t_i}) ds \\
&= \Delta W_t \Delta t - \Delta I_t
\end{aligned}$$

To simulate ΔI_t, one can show that given W_t, ΔI_t and ΔW_t are jointly normal. $\begin{bmatrix} \Delta W_t \\ \Delta I_t \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} \Delta t & \frac{1}{2}(\Delta t)^2 \\ \frac{1}{2}(\Delta t)^2 & \frac{1}{3}(\Delta t)^3 \end{bmatrix}\right)$. Putting it all together, we have:

$$\begin{aligned}
\hat{X}_{t_{i+1}} =& \hat{X}_{t_i} + a\Delta t + b\Delta W_t + \frac{1}{2}bb_x\left((\Delta W_t)^2 - \Delta t\right) \text{ (from Euler + Milstein)} \\
&+ \frac{1}{2}\left(a_t + aa_x + \frac{1}{2}b^2 a_{xx}\right)(\Delta t)^2 + ba_x \Delta I_t \text{ (from approximation of } a\text{)} \\
&+ \left(b_t + bb_x + \frac{1}{2}b^2 b_{xx}\right)(\Delta W_t \Delta t - \Delta I_t) \text{ (from approximation of } b\text{)}
\end{aligned}$$

5.3 Variance Reduction

Motivation: we want to use Monte-Carlo methods to estimate an expectation like $\mathbb{E}\left[e^{-rT}f(S_T)\right]$. The Monte-Carlo estimator is $\frac{1}{n}\sum_{i=1}^{n} e^{-rT}f(S_{T,i})$.

Definition: 5.2: Estimator

An estimator $\hat{\theta}$ is a statistic that is used to infer the value of an unknown parameter θ from the data. The bias of an estimator is $\mathbb{E}\left[\hat{\theta}\right] - \theta$. if $\mathbb{E}\left[\hat{\theta}\right] = \theta$, then $\hat{\theta}$ is unbiased. The variance of an estimator is $\text{Var}\left(\hat{\theta}\right) = \mathbb{E}\left[\left(\hat{\theta} - \mathbb{E}\left[\hat{\theta}\right]\right)^2\right]$. A sequence of estimators $\hat{\theta}_n$ are consistent for θ if $\hat{\theta}_n \to \theta$ in probability.

Example: Sample mean $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ as an estimator for $\mathbb{E}[Y]$.

Proof. Let $Y_i, i = 1, ..., n$ be i.i.d. samples.

$$\mathbb{E}[\overline{Y}] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[Y_i] = \mathbb{E}[Y]$$

Therefore, it is unbiased.

The variance of $\overline{Y}$ is

$$\text{Var}(\overline{Y}) = \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}(Y_i) = \frac{1}{n}\text{Var}(Y)$$

It is also strongly consistent, since $\lim_{n\to\infty}\overline{Y} = \mathbb{E}[Y]$ a.s. by law of large numbers. □

5.3.1 Control Variates

Goal: estimate $\mathbb{E}[Y]$
Idea: make use of another correlated r.v. X to construct an unbiased estimator for $\mathbb{E}[Y]$ with smaller variance than $\overline{Y}$

Example: $Y = e^{-rT}(S_T - K)_+$, $X = S_T$

1. Setup: $Y_i, i = 1, ..., n$ outputs from n replications of simulation.
2. For each replication, we also calculate another output X_i.
3. Assume (X_i, Y_i) are i.i.d. and $\mathbb{E}[X]$ is known
4. For fixed $b \in \mathbb{R}$, compute $Y_i^C = Y_i - b(X_i - \mathbb{E}[X])$ (The error controls the variance).

Definition: 5.3: Control Variate Estimator

For $b \in \mathbb{R}$, the control variate estimator is

$$\hat{Y}^C = \frac{1}{n}\sum_{i=1}^{n}[Y_i - b(X_i - E[X])]$$

Theorem: 5.3:

$\hat{Y}^C$ is an unbiased and consistent estimator for $\mathbb{E}[Y]$. Its variance is

$$\frac{1}{n}\big(\text{Var}(Y) - 2b\text{Cov}(X,Y) + b^2\text{Var}(X)\big)$$

When $2b\text{Cov}(X,Y) < \text{Var}(X)$, $\text{Var}(\hat{Y}^C) < \text{Var}(\overline{Y})$.

Proof.

$$\begin{aligned}
\mathbb{E}[\hat{Y}^C] &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[Y] - b(\mathbb{E}[X] - \mathbb{E}[X]) = \mathbb{E}[Y] \\
\hat{Y}^C &= \frac{1}{n}\sum_{i=1}^{n} Y_i - b\frac{1}{n}\sum i = 1^n X_i + b\mathbb{E}[X] \to \mathbb{E}[Y] \text{ in probability} \\
\text{Var}(\hat{Y}^C) &= \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}(Y_i - bX_i + b\mathbb{E}[X]) = \frac{1}{n}\big(\text{Var}(Y) - 2b\text{Cov}(X,Y) + b^2\text{Var}(X)\big)
\end{aligned}$$

□

To choose $b \in \mathbb{R}$, we minimize $\text{Var}(\hat{Y}^C)$ over b

$$\begin{aligned}
0 &= \frac{\partial}{\partial b}\text{Var}(\hat{Y}^C) = -2\text{Cov}(X,Y) + 2b\text{Var}(X) \\
b &= \frac{\text{Cov}(X,Y)}{\text{Var}}
\end{aligned}$$

Effectiveness of the control variate estimator:
Compute ratio $\frac{\text{Var}(\hat{Y}^C)}{\text{Var}(\overline{Y})} = 1 - \rho_{XY}^2$

$$\begin{aligned}
\text{Var}(\hat{Y}^C) &= \frac{1}{n}\left[\text{Var}(Y) - 2\frac{\text{Cov}(X,Y)^2}{\text{Var}} + \left(\frac{\text{Cov}(X,Y)}{\text{Var}}\right)^2\text{Var}(X)\right] \\
&= \frac{1}{n}\left[\text{Var}(Y) - \frac{\text{Cov}(X,Y)^2}{\text{Var}}\right]
\end{aligned}$$

However, we may not know the covariance. In practice, we use the sample covariate:
$\hat{b}_n = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X}_i)(Y_i - \overline{Y}_i)}{\sum_{i=1}^{n}(X_i - \overline{X}_i)^2}$.

Theorem: 5.4:

$\hat{b}_n$ is a consistent estimator of b.

5.3.2 Importance Sampling

Goal: Estimate $\mathbb{E}[h(Y)]$
Idea: Change the probability measure from which the r.v. is generated to obtain a more convenient representation of $\mathbb{E}[h(Y)]$.

We want to estimate $\mathbb{E}[h(Y)]$ where Y has density f.

$$\mathbb{E}[h(Y)] = \int_{-\infty}^{\infty} h(y)f(y)dy$$

The ordinary Monte-Carlo estimator is $\frac{1}{n}\sum_{i=1}^{n} h(Y_i)$, $Y_i \sim f$.

We choose an importance sampling distribution $g(y)$ s.t. $f(y) > 0 \Rightarrow g(y) > 0, \forall y \in \mathbb{R}$. Then

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}[h(Y)] &= \int_{-\infty}^{\infty} h(y)f(y)dy \\ &= \int_{-\infty}^{\infty} h(y)\frac{f(y)}{g(y)}g(y)dy \\ &= \mathbb{E}^{\tilde{\mathbb{P}}}\left[h(Y)\frac{f(Y)}{g(Y)}\right],\end{aligned}$$

where $\frac{f(Y)}{g(Y)}$ is the likelihood ratio/Radon-Nikodym derivative.

***Definition:* 5.4: Importance Sampling Estimator**

The importance sampling estimator associated with g is $\hat{Y}^I = \frac{1}{n}\sum_{i=1}^{n} h(Y_i)\frac{f(Y_i)}{g(Y_i)}$, where $Y_i \sim g$.

Proof. $\hat{Y}^I$ is unbiased by construction.

$$\mathbb{E}^{\tilde{\mathbb{P}}}[\hat{Y}^I] = \mathbb{E}^{\tilde{P}}\left[h(Y)\frac{f(Y)}{g(Y)}\right] = \mathbb{E}^{\mathbb{P}}[h(Y)]$$

The variance is

$$\text{Var}^{\tilde{\mathbb{P}}}(\hat{Y}^I) = \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}^{\tilde{\mathbb{P}}}\left(h(Y)\frac{f(Y)}{g(Y)}\right)$$

Since $\mathbb{E}^{\tilde{\mathbb{P}}}\left[\left(h(Y)\frac{f(Y)}{g(Y)}\right)^2\right] = \mathbb{E}^{\mathbb{P}}\left[h^2(Y)\frac{f(Y)}{g(Y)}\right]$, then

$$\text{Var}^{\tilde{\mathbb{P}}}(\hat{Y}^I) = \frac{1}{n}\left(\mathbb{E}^{\mathbb{P}}\left[h^2(Y)\frac{f(Y)}{g(Y)}\right] - \mathbb{E}^{\mathbb{P}}[h(Y)]^2\right)$$

$$\text{Var}^{\mathbb{P}}(\overline{Y}) = \frac{1}{n}\left[\mathbb{E}^{\mathbb{P}}\left[h^2(Y)\right] - \mathbb{E}^{\mathbb{P}}[h(Y)]^2\right]$$

$\square$

Is $\mathbb{E}^{\mathbb{P}}\left[h^2(Y)\frac{f(Y)}{g(Y)}\right] < \mathbb{E}^{\mathbb{P}}\left[h^2(Y)\right]$?

Assume h is nonnegative. Choose $g(y) \propto f(y)h(y)$ *e.g.* $g(y) = cf(y)h(y)$ for c normalizing. Then

$$\text{Var}(\hat{Y}^I) = \frac{1}{n}\left[\mathbb{E}^{\tilde{\mathbb{P}}}\left[h^2(Y)\frac{f^2(Y)}{g^2(Y)}\right] - \mathbb{E}^{\tilde{\mathbb{P}}}\left[h(Y)\frac{f(Y)}{g(Y)}\right]^2\right] = \frac{1}{n}\left[\mathbb{E}^{\tilde{\mathbb{P}}}\left[\frac{1}{c^2}\right] - \mathbb{E}^{\tilde{\mathbb{P}}}\left[\frac{1}{c}\right]^2\right] = 0$$

So choosing $g \propto fh$ gives a zero variance estimation. However, $\frac{1}{c} = \int fh dy$ is what we want to compute.

Takeaways:

1. The choice of g is the key. A poor choice may increase the variance
2. There is no optimal way to choose g, but we should try to sample in proportion to $f(y)h(y)$

Example: Payoff of a European call in the Black-Scholes model. $\mathbb{E}^{\mathbb{Q}}\left[e^{-rT}(Se^Z - K)_+\right]$, where $Z \sim \mathcal{N}\left(\left(r - \frac{\sigma^2}{2}\right)T, \sigma^2 T\right)$.

If K is large, Monte-Carlo approximation will have a lot of 0s.

With the default Z, $\mathbb{E}^{\mathbb{Q}}[Se^Z] = Se^{rT}$.

Change to $Z \sim \mathcal{N}\left(\ln\frac{K}{S} + \frac{\sigma^2}{2}T, \sigma^2 T\right)$, we get

$$\mathbb{E}^{\tilde{\mathbb{Q}}}[Se^Z] = S\exp\left(\ln\frac{K}{S} - \frac{\sigma^2}{2}T + \frac{1}{2}\sigma^2 T\right) = K$$