ECON 765 - Model for Financial Economics

Pr. Russell Davidson

Course notes by
Léo Raymond-Belzile
Leo.Raymond-Belzile@mail.mcgill.ca

The current version is that of June 19, 2013
Summer 2013, McGill University

Please signal the author by email if you find any typo.

These notes have not been revised and should be read carefully.

Contents

The book for the course will be **Stochastic Calculus for Finance II - Continuous-Time models** by Steven E. Shreve.

Section 1
Binomial Asset Pricing Model

Any process that can be modeled as a **diffusion**, that is the evolution over time of financial assets or returns in the case which is of interest to us. In this section, we consider a simple model, named binomial asset pricing. Before describing it, we begin with some terminology.

Options (which we call **contracts**) are **derivative securities**, derivative in the sense that there is an **underlying asset** (bond, stock, etc.) The evolution of options is random, and so difficult to forecast. We consider **hedging** strategies, that is investment positions taken as to reduce loss incurred in financial or money markets.

Suppose for simplicity a two periods model, with period 0 and 1. At time zero, denote the price of the asset by S_0. At period 1, one of two things can happen: in one of these, the price increases, which we denote $S(H)$ or the price go down $S(T)$ for H, T denoting head and tail. We can go from one to the other, with some probabilities associated with this random walk. These probabilities do not have any impact on the pricing of the derivatives, based on the principle of **no arbitrage**. An **arbitrage** is defined as being a strategy for which there is a positive probability of a gain, and no probability of loss. The market would be drawn back to equilibrium if there was positive gain to be made. Here, we are assuming that the markets are at equilibrium to begin with, there is no arbitrage.[1]

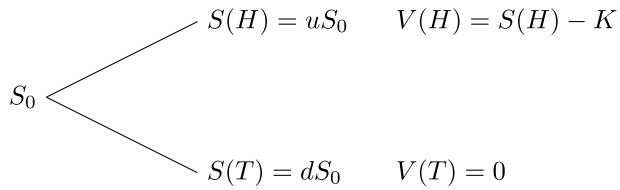
A **portfolio** is a collection of assets owned by an individual. There are two underlying assets considered in our settings, a risky asset, denoted S, and a risk-free asset, which we could think of as money. The number of units of the risky asset (the number of shares), will form the portfolio.

We will consider the **European call options**, an option which gives the owner the possibility to buy one share at t_1, time one, for a **strike price** K; this is what you pay to get one unit of the underlying asset if you exercise the option. Thus, in an European call option, one can **exercise** the option in period 1 if the result is H and pay K in exchange of the price of the stock $S_1(H)$.[2]

We speak about the exercise, that is do what the options is, i.e. if call, then buy. We base our decision on and choose to exercise at expiration (maturity) sometimes denoted T.

[1] There is no problem for an horizon in a discrete model.

[2] Here, **call** is defined for buy, **put** for sell. This implies that exercising an option (in an European call option) yields profit $(S_1 - K)_+$, while in an European put option, we can sell the option at t_0 to get profit $(K - S_1)_+$.



Let $S(H) = uS_0$ and $S(T) = dS_0$. If $S_1 - K > 0$, we exercise the option. The value thus is $(S_1 - K)_+ = \max\{0, S_1 - K\}$.

The other parameter needed in this model, apart from the parameters u, d and K is the interest rate on money, denoted r. Obviously, we are interested in strike prices within the range (dS_0, uS_0). Can we duplicate the effect of the option; that is can we get the same payoff, by building a portfolio by making things in all states of the world, getting the same payoff with risky asset and risk-free assets? [3]

We would have arbitrage if the models did not have equal values.

We have the inequality $d < 1 + r < u$ where r is the return from time t_0 to t_1.

To constitute a hedging portfolio, we need to check availability and get what the portfolio looks like. Set $V(H) = S(H) - K$ and $V(T) = 0$. Assume the portfolio contains $\varDelta_0$ of the risky asset and M_0 units of the risk free asset. We can then say that X_0 is the cost of acquiring this portfolio, given by

$$\begin{aligned} X_0 &= M_0 + S_0\varDelta_0 \\ X_1(H) &= \varDelta_0 S(H) + M_0(1+r) \\ X_1(T) &= \varDelta_0 S(T) + M_0(1+r) \end{aligned}$$

with conditions $X_1(H) = V_1(H)$ and $X_1(T) = V_1(T)$. The equations we have to solve are thus simply

$$\begin{aligned} V_1(H) &= \varDelta_0 S(H) + M_0(1+r) \\ V_1(T) &= \varDelta_0 S(T) + M_0(1+r) \end{aligned}$$

$\varDelta_0$ and M_0 tells us the quantity we have to buy and X_0 tells us the value to get no arbitrage. We then get the simple equation, equating the above, is

$$\varDelta_0 = \frac{V_1(H) - V_1(T)}{S_1(H) - S_1(T)}$$

referred to as the **delta-hedging rule**.

[3] This is not always possible, for example with the trinomial model.

Once we have this, further simplification gives

$$\begin{aligned}
M_0 &= \frac{1}{1+r}\left(V_1(H) - \frac{V_1(G)-V_1(T)}{S_1(H)-S_1(T)}S_1(H)\right)\\
&= \frac{1}{1+r}\left(\frac{V_1(H)S_1(H)-V_1(H)S_1(T)}{S_1(H)-S_1(T)} - V_1(H)S_1(H)+V_1(T)S_1(H)\right)\\
&= \frac{1}{1+r}\left(\frac{V_1(T)S_1(H)-V_1(H)S_1(T)}{S_1(H)-S_1(T)}\right)
\end{aligned}$$

and the denominator must be positive, although the numerator needs not be. We need thus to introduce long and short sales, and we will assume that one can borrow money or lend money at the same interest rate r. If Δ_0 is negative, you sell short, *i.e.* you sell something you don't have.

The quantity that we are really interest in is X_0, which is the value of the option.

$$\begin{aligned}
X_0 &= \frac{1}{1+r}\left(\frac{V_1(T)S_1(H)-V_1(H)S_1(T)}{S_1(H)-S_1(T)}\right) + \frac{S_0(V_1(H)-V_1(T))}{S_1(H)-S_1(T)}\\
&= \frac{V_1(T)S_1(H)-V_1(H)S_1(T)+(1+r)S_0(V_1(H)-V_1(T))}{(1+r)(S_1(H)-S_1(T))}
\end{aligned}$$

since the denominators are common and thus joining the terms together

$$= \frac{1}{(1+r)(S_1(H)-S_1(T))}\Big[\{(1+r)S_0 - S_1(T)\}\,V_1(H) + \{S_1(H)-(1+r)S_0\}\,V_1(T)\Big]$$

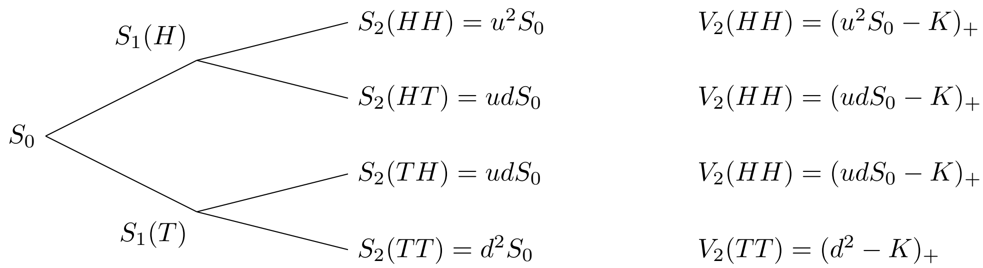
we will see shortly that both Δ_0 and V_0 cannot be both simultaneously negative, otherwise X_0 would be negative – this case is ruled out. If $V_1(T) = V_1(H)$, then the above reduces to $(1+r)^{-1}$. If we consider $S_1(H) = uS_0$ and $S_1(T) = dS_0$, then

$$\begin{aligned}
S_1(H) - S_1(T) &= S_0(u-d)\\
(1+r)S_0 - S_1(T) &= S_0(1+r-d)\\
S_1(H) - (1+r)S_0 &= S_0(u-(1+r))
\end{aligned}$$

We have then

$$X_0 = \frac{1}{1+r}\left(\frac{u-(1+r)}{u-d}V_1(T) + \frac{1+r-d}{u-d}V_1(H)\right) \tag{1.1}$$

and $\widetilde{p} \equiv \frac{u-(1+r)}{u-d}$, $\widetilde{q} = \frac{1+r-d}{u-d}$ means that $\widetilde{p}$ and $\widetilde{q}$ satisfies the requirements probabilities, they are both positive and sum to one, *i.e.* $\widetilde{q} + \widetilde{p} = 1$. They are what are called **risk-neutral probabilities**, which are fundamental to the pricing of derivative security. X_0 is the convex combination of the two options and the current value, in the usual sense of



present discounted values.[4]

The states of the world, which is the collection of states at a given period in our graph, in our case is two period and two possible outcomes. We can however very easily extend this to more periods. The graph for the state space is thus again very simple, but now the state in period $t+1$ depends on the previous period t. where now

$$\begin{aligned} X_1(H) &= \frac{1}{1+r}\left(\widetilde{p}V_2(HH) + \widetilde{q}V_2(HT)\right) \\ X_1(T) &= \frac{1}{1+r}\left(\widetilde{p}V_2(TH) + \widetilde{q}V_2(TT)\right) \\ X_0 &= \frac{1}{1+r}\left(\widetilde{p}X_1(H) + \widetilde{q}X_1(T)\right) \end{aligned}$$

where we ascribe the risk neutral probability, ascribe and discount in order to get what we need to form the portfolio S_0. Since these are risk-neutral probabilities, we have a "pseudo-independence" structure and we can use the the delta-hedging rule to determine what $S_1(H)$ and $S_1(T)$ are. We abstract of transaction costs in this framework.

In more periods, with n time period, a possible outcome would be *e.g.* $HTTHHHTHHTTT$ where the cardinality is 2^n for the set of possible outcomes. If n_H is the number of H, n_T corresponding and $n_H + n_T = n$, then

$$\mathsf{P}\left(n_H, n_T\right) = \widetilde{p}^{n_H}\widetilde{q}^{n_T}$$

We can still ascribe the risk-neutral probabilities to the model with associated risk neutral probabilities here $\widetilde{p}_1, \widetilde{q}_2, \widetilde{q}_3, \widetilde{p}_4, \widetilde{p}_5$. In the limit, any outcome is probability zero, as $n \to \infty$. This is one example of discontinuity of what one gets, as now every event has mass zero.

[4] This discounting is perfectly normal.

Section 2
General Probability Theory

An **outcome** space could be more general, either discrete or continuous. We done the outcome space Ω and corresponding outcome (events) $\omega \in \Omega$.

We need to impose a structure on this space, namely that of a σ-algebra, denoted $\mathcal{F}$ the set of subsets of Ω. The set of all powers, is 2^{Ω} is the power set, that is the set of all subsets of Ω and in general $\mathcal{F} \subseteq 2^{\Omega}$. In order for it to be σ-algebra, we require the following to hold:

1. $\emptyset \in \mathcal{F}$ and $\Omega \in \mathcal{F}$
2. $A \in \mathcal{F} \Rightarrow A^{\complement} \in \mathcal{F}$
3. If $A_i \in \mathcal{F}$, for $i = 1, 2, \ldots$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

Suppose we consider this in the context of a three period binomial model. We can enumerate in this simple case Ω as

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

and all collections of the above, the triples correspond to Ω. Suppose we want to consider the set of outcomes where the first two periods, are known. The proper subsets of Ω would then be the collection

$$\{\{HHH, HHT\}, \{HTH, HTT\}, \{THH, THT\}, \{TTH, TTT\}\}$$

This gives thus a notion of available information. Denote the pair $(\Omega, \mathcal{F})$, the measurable space. We will be interested mostly in probability measures P, where $\mathsf{P} : \mathcal{F} \to [0, 1]$ is the underlying mapping. We will have the usual constraints, the Kolmogorov axioms, with

1. $\mathsf{P}(\emptyset) = 0, \mathsf{P}(\Omega) = 1$
2. If $A_1, \ldots, A_i$ disjoint $(A_i \cap A_j = \emptyset)$, whenever $i \neq j$, then

$$\mathsf{P}\left(\bigsqcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{n} \mathsf{P}(A_i)$$

This implies that $A \in \mathcal{F}$, then $A^{\complement} \in \mathcal{F}$ and as such $\mathsf{P}(A) = 1 - \mathsf{P}\left(A^{\complement}\right)$ as $A \cup A^{\complement} = \Omega$.

The triple $(\Omega, \mathcal{F}, \mathsf{P})$ is termed **probability space**.

Denote by X a random variable. X is a measurable mapping $X : \Omega \to (\mathbb{S}, \mathcal{S})$, where usually $\mathbb{S} = \mathbb{R}$ or an appropriate space for vectors, say $\mathbb{R}^{n+}$. We need to be able to define a a

σ-algebra on it. If $X : (\Omega, \mathcal{F}) \to (\mathbb{S}, \mathcal{S})$, then if $B \in \mathcal{S} \Rightarrow X^{-1}(B) \in \mathcal{F}$. The set

$$X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\}$$

We have $\mu_X(B) = \mathsf{P}\left(X^{-1}(B)\right)$ where μ_X is the measure induced on $\mathcal{S}$.

The canonical σ-algebra on $\mathbb{R}$ is the Borel σ-algebra, denoted by $\mathcal{B}$ that is the smallest σ-algebra that contains $[a, b]$. We generate this way (a, b) open intervals, or rays (a, ∞), etc. Indeed, if we consider

$$\left[a + \frac{1}{n}, b + \frac{1}{n}\right] = B_n \; \forall \, n \in \mathbb{N}$$

the points a, b are not in the interval. One can construct non-measurable sets, using the axiom of choice and construct a set that is not Borel set. We will be typically be interested in the pair $(\mathbb{R}, \mathcal{B})$.

In general, we are interested in CDFs of the real valued random variables, the so-called cumulative distribution function, which we denote F_X. This F_X is a function mapping $\mathbb{R} \to [0, 1]$, and $F_X(x) = \mathsf{P}\left(X \leq x\right) = \mu_X(X \leq x) = \mu_X((-\infty, x])$.

Example 2.1

If $\Omega = \{1, 2, 3, \ldots, N\}$ is the sample space corresponding to the number of people in classroom, then $\mathcal{F} = 2^\Omega$, $X(i) = i$ and $\mathsf{P}\left(A\right) = \frac{n}{N}$ if $A = \{n \text{ points}\}$ and $\mathsf{P}\left(i\right) = \frac{1}{N}$ and $X(i) = i$.

Definition 2.1 (Expectation)

We define the expectation of a random variable to be

$$\mathsf{E}\left(X\right) = \sum_{i=1}^{N} \mathsf{P}\left(i\right) X(i) = \sum_{\omega \in \Omega} X(\omega) \mathsf{P}\left(\omega\right)$$

in the discrete setting. More generally, we want

$$\mathsf{E}\left(X\right) = \int_\Omega X(\omega) \, \mathrm{d}\mathsf{P}\left(\omega\right)$$

If $\Omega = \mathbb{R}$ and X has a density, we could write

$$\int_{-\infty}^{\infty} X(x) f_X(x) \, \mathrm{d}x = \int_{-\infty}^{\infty} X(x) \, \mathrm{d}F_X x$$

the **Stieltjes** integral. Here $X(x)$ is as usual the **integrand** and $\mathrm{d}F_X(x)$the integrator. Provided the derivative exist, $f_X(x) = F'_X(x)$. We can add another condition, that of bounded variation. Since $F_X(x)$ is bounded, we can always use this as a candidate integrator.

Consider the simple case where $X(x) \sim \mathcal{U}(0,1)$. Then, the Riemann integral is defined as follow, $\int_0^1 X(x)\,\mathrm{d}x$ is defined as follows. Let $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = 1$, a partition of the unit interval where $t_i = i/n$ with $I_i = [t_{i-1}, t_i]$

The corresponding lower Riemann sum is $\frac{1}{n}\sum_{i=1}^n \inf_{x\in I_i} X(x) \to \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X(x)$ provided it exist and the corresponding upper Riemann sum where inf is replaced with sup. If both exist and the limit coincide, then the limit is the Riemann sum.

We now construct a lower sum $\sum_{i=1}^n \inf_{x\in I_i} X(x)(F_X(t_i) - F_X(t_{i-1}))$ and analogous to the Riemann sum, this defines in the limit with both lower and upper Riemann-Stieltjes sums. Then, if $x_1, \ldots, x_N$ is a set of values for a discrete random variable with associated probability $p_1, \ldots, p_N$ the the corresponding $F_X(x) = \sum_{X_i \le x} p_i = \sum_{i=1}^N \mathrm{I}\,(X_i \le x)p_i$ where $\mathrm{I}\,(\cdot)$ is the indicator function. If $X_i = X(x_i)$, what we want to get is $\sum_i X_i p_i$.

For an interval $a < x_i < b$, $\forall\, i$, then F_X is a step function and notice that F_X is continuous at the points of discontinuity of X. If $\varepsilon, \eta > 0$ arbitrarily small, then $F(X_i+\varepsilon) - F(X_i-\eta) = p_i$.

Recall that we wanted to define expectation by $\int_\Omega X(\omega)\,\mathrm{dP}\,(\omega)$. We can answer two problems in one step. Consider the function $g(x)$ for $x \in [0,1]$, which is defined as

$$g(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{otherwise} \end{cases}$$

and recall that the rationals are countable and are dense in the interval $[0,1]$. This function is not Riemann integrable. For any $x \in I_i$, then $\inf_{x\in I_i} g(x) = 0$ and $\sup_{x\in I_i} g(x) = 1$.

However, according to Lebesgue theory, we would want to assign the value of zero for the integral. The Lebesgue measure, countably finite, on $(\mathbb{R}, \mathcal{B})$ which denote by μ, $\mu(a,b) = |b-a|$ and $\mu(\{a\})$. Of course, this is not a probability measure as $\mu(\mathbb{R}) = \infty$, yet for any interval the measure is finite. Here $\mu(\mathbb{Q}) = 0$, since this is a countable union of points, and so this is a correct conclusion to draw from this.

The Lebesgue integral considers the partition of the ordinates and look at the corresponding measure of the interval. The corresponding Lebesgue lower sum is $\sum_{i=1}^N y_{i-1}\mu(f^{-1}(y_{i-1}, y_i))$ for $y = f(x)$ for f measurable. Correspondingly, we have $\sum_{i=1}^N y_i\mu(f^{-1}(y_{i-1}, y_i))$ for the upper sum and as usual, in the limit we define the Lebesgue integral to be this function provided the two limits agree. In this case, we need only to partition from zero to one.

This allows us to define $\int_\Omega X(\omega)\,\mathrm{dP}\,(\omega)$ as $\sum_{i=1}^N y_i \mathrm{P}\left(X^{-1}(y_{i-1}, y_i)\right)$.

Remark (Partitions of the Riemann and Lebesgue integral)

We define the sum to be $S = \sum_{i=0}^{n-1} y_i\mu(y_i < f(x) < y_{i+1})$ where $\mu(\cdot)$ is the Lebesgue measure.

Recall that a probability space is the tuple $(\Omega, \mathcal{F}, \mathrm{P})$ where $\mathcal{F} \subseteq 2^\Omega$, the power set, and

$\mathsf{P} : \mathcal{F} \to [0, 1]$ is a mapping onto the unit interval. Then the random variable (whether real valued, complex valued, vector or matrix) is said to be $\mathcal{F}$-measurable if for the pair $(\mathbb{R}, \mathcal{B})$ where we want for each $B \in \mathcal{B}$, $X^{-1}(B) = \{\omega \in \Omega | X(\omega) \in \mathcal{B}\} \in \mathcal{F}$. Thus, our measure is $\mu_X(B) = \mathsf{P}\left(X^{-1}(B)\right) = \Pr(X \in B)$.

We are interested in $\mathsf{E}(X)$, the expectation of the random variable, defined as being

$$\mathsf{E}(X) = \int_{\Omega} X(\omega \, \mathrm{d}\mathsf{P}(\omega) = \lim_{n \to \infty} \sum_{i=0}^{n-1} y_i \mathsf{P}(X \in (y_i, y_{i+1}])$$

where

$$\{\omega \in \Omega : X(\omega) \in (y_i, y_{i+1}]\}$$

For discrete valued random variables, this boils down to $\mathsf{E}(X) = \sum_{i=1}^{m} p_i x_i$ if $x_i, i = 1, \ldots, m$ is the set of all possible values with associated measure p_i.

Assume now that $X(\omega)$ is non-negative, wlog, since we can define the above to be $X = X^+ - X^-$ with the convention that $X^+ = \max\{0, X\}$ and $X^- = \max\{0, -X\}$. If $\mathsf{E}(X^+) = \infty$ and $\mathsf{E}(X^-) < \infty$, then we define $\mathsf{E}(X) \coloneqq \infty$. Now, if $\mathsf{E}(X^+) < \infty$ and $\mathsf{E}(X^-) = \infty$, then we define $\mathsf{E}(X) \coloneqq -\infty$. If both $\mathsf{E}(X^+) < \infty, \mathsf{E}(X^-) < \infty$, then X is integrable

We have the following proposition:

Proposition 2.2

1. X is integrable if and only if $\int_{\Omega} |X(\omega)| d\mathsf{P}(\omega) < \infty$. Here $|X(\omega)| = X^+ + X^-$
2. Linearity: $\int_{\mathbb{R}} (f + g) \, \mathrm{d}\mu + \int f \, \mathrm{d}\mu + \int g \, \mathrm{d}\mu$ and as a result $\mathsf{E}(X + Y) = \mathsf{E}(X) + \mathsf{E}(Y)$ and $\int_R (\alpha f) \, \mathrm{d}\mu = \alpha \int_{\mathbb{R}} f \, \mathrm{d}\mu$ and so $\mathsf{E}(aX + bY) = a\mathsf{E}(X) + b\mathsf{E}(Y)$.
3. Comparison of $X \le Y$ almost surely if $\mathsf{P}(X \le Y) = 1$, then $\mathsf{E}(X) \le \mathsf{E}(Y)$.

Theorem 2.3 (Jensen's inequality)

If $\phi : \mathbb{R} \to \mathbb{R}$ is convex, then $\mathsf{E}(|X|) < \infty$, then

$$\phi(\mathsf{E}(X)) \le \mathsf{E}(\phi(X))$$

A **deterministic function** of a random variable is defined with $X : \Omega \to \mathbb{R}$ and $\phi(X) : \Omega \to \mathbb{R}$ with $[\phi(X)](\omega) = \phi(X(\omega))$ (since the inverse image under $\phi(X)$ is trivially measurable).

Convergence of integrals

If we denote $\phi(X)$ and $\Phi(X)$ respectively the PDF and CDF of the standard Normal random variable. Scale family If $W \sim \mathcal{N}(0, 1)$ and we take $\sigma > 0$, then $Z = \sigma W \sim \mathcal{N}(0, \sigma^2)$. Indeed

$$\Pr(Z \le z) = \Pr(\sigma W \le z) = \Pr(W \le z/\sigma) = \Phi(z/\sigma)$$

Consider a sequence of functions which shrinks the variance of a $\mathcal{N}(0,1)$ random variable with corresponding n^{th} element

$$f_n(x) = \sqrt{\frac{n}{2\pi}} \exp\left(-\frac{nx^2}{2}\right)$$

the limiting distribution of which is a generalized, the Dirac δ function, $\delta(x)$, defined as $\delta(x) = \mathrm{I}\,(x = 0)$

$$\int_a^b f(x)\delta(x)\,\mathrm{d}x = f(0)$$

provided $a \leq 0 \leq b$. We can also consider shifts, namely we have $\int_a^b f(x)\delta(x-c)\,\mathrm{d}x = f(x)$ if $a \leq c \leq b$. Since for each n, we have by standard requirements for densities that $\int_{-\infty}^{\infty} f_n(x)\,\mathrm{d}x = 1$, but for $x \neq 0$, we have $\lim_{n\to\infty} f_n(x) = 0$. However, at $x = 0$, we have $\lim_{n\to\infty} f_n(x) = \infty$. Thus $\lim_{n\to\infty} f_n(x) = 0$ a.s.

This however imply that

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f_n(x)\,\mathrm{d}x = 1 \neq \int_{-\infty}^{\infty} \left(\lim_{n\to\infty} f_n(x)\right)\mathrm{d}x$$

We thus need some convergence theorem and regularity criterion for interchanging integral and limit operator.

Theorem 2.4 (Monotone convergence)
Let $\{f_n\}$ be a monotone sequence of non-negative functions. We impose the condition $f_n(x) \leq f_{n+1}(x)$ a.e. Then if $\lim_{n\to\infty} f_n(x) = f(x)$ a.e. exists, then

$$\lim_{n\to\infty} \int f_n(x)\,\mathrm{d}x = \int f(x)\,\mathrm{d}x$$

Another result is as follow

Theorem 2.5 (Dominated convergence)
If $\exists g \geq 0$, with $\int g(x)\,\mathrm{d}x < \infty$ and $\forall\, n, f_n(x) \leq g(x)$ a.e., then

$$\lim_{n\to\infty} \int f_n(x)\,\mathrm{d}x = \int f(x)\,\mathrm{d}x$$

Theorem 2.6
Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space and $g : \mathbb{R} \to \mathbb{R}$ be Borel measurable with $\mathsf{E}\left(|g(X)|\right)$ and $X : \Omega \to \mathbb{R}$. Then

$$\mathsf{E}\left(|g(X)|\right) = \int g(X(\omega))\,\mathrm{d}\mathsf{P}\left(\omega\right) = \int_{\mathbb{R}} |g(x)|\,\mathrm{d}\mu_X(x).$$

If $\mathsf{E}\left(|g(X)|\right) < \infty$, then

$$\mathsf{E}\left(g(X)\right) = \int_{\mathbb{R}} g(x)\,\mathrm{d}\mu_X(x)$$

Proof The technique uses the **standard machine**. We first prove the result for

1. indicator functions
2. non-negative simple functions
3. non-negative measurable functions
4. arbitrary measurable functions.

and usually, we simply need to invoke linearity, then monotone convergence theorem and density for $\mathcal{L}^1$ functions, and finally appeal to decomposition of functions.

Let

$$g(x) = \mathrm{I}_B\left(x\right) = \begin{cases} 1 & \text{if } x \in B \\ 0 & \text{otherwise} \end{cases}$$

then

$$\begin{aligned}\mathsf{E}\left(|g(X)|\right) &= \mathsf{E}\left(\mathrm{I}_B\left(X\right)\right) = 0 \cdot \mathsf{P}\left(\mathrm{I}_B\left(X\right) = 0\right) + 1 \cdot \mathsf{P}\left(\mathrm{I}_B\left(X\right) = 1\right) \\ &= \mathsf{P}\left(X \in B\right)\end{aligned}$$

Now consider $\mu_X(A) \coloneqq \mathsf{P}\left(X^{-1}(A)\right) = \mathsf{P}\left(\mathrm{I}_B\left(A\right)^{-1}\right)$, [5]. Then

$$\mu_X(A) = \mathsf{P}\left(\omega : X(\omega) \in A\right).$$

Thus $\int_{\mathbb{R}} |g(x)|\,\mathrm{d}\mu_X(x) = \mu_X(g(X) = 1) = \mathsf{P}\left(X = B\right)$.

For the second step, consider $g(x) = \sum_{i=1}^{n} a_i \mathrm{I}_{B_i}\left(x\right)$ with $a_i > 0$. Then, by linearity of expectation,

$$\mathsf{E}\left(g(X)\right) = \sum_{i=1}^{n} a_i \mathsf{E}\left(\mathrm{I}_{B_i}\left(X\right)\right) = \sum_{i=1}^{n} a_i \mathsf{P}\left(B_i\right)$$

For the RHS, using linearity of the integral, we have

$$\sum_{i=1}^{n} a_i \int_{\mathbb{R}} \mathrm{I}_{B_i}\left(x\right)\mathrm{d}\mu_X(x) = \sum_{i=1}^{n} a_i \mathsf{P}\left(B_i\right)$$

[5] Remark that by definition of the induced measure, if $Y : \omega \mapsto g(X(\omega)) = \mathrm{I}_B\left(X(\omega)\right)$, then $\mu_Y(A) = \mathsf{P}\left(\omega : \mathrm{I}_B\left(X(\omega)\right) \in A\right)$

For the third step, consider the Borel sets

$$B_{k,n} = \left\{x : k2^{-n} \leq g(x) < (k+1)2^{-n}\right\}$$

for $k = 0, 1, \ldots, 4^n - 1$.[6]

Then

$$g_n(x) = \sum_{k=0}^{4^n-1} \frac{k}{2^n} \mathrm{I}_{B_{k,n}}(x)$$

As n increases, then this forms a non-decreasing sequence. As $n \to \infty$, then $g_n(x) \to g(x)$. We get the result applying monotone convergence theorem.

For the fourth part, take $g = g^+ - g^-$. If $|g| = g^+ + g^-$ is integrable, then the result follows. ■

Suppose we want to have to take expectation of a random variable. We have $\int_{\mathbb{R}} g(x)\,\mathrm{d}\mu_X(x)$, and we want to have this for continuous random variables. We have $f_X(x)$ and define for random variables with density to have $\mu_X(B) = \int_B f_X(x)\,\mathrm{d}x$. We then would have

$$\mathsf{E}\left(g(X)\right) = \int_{\mathbb{R}} g(x)f(x)\,\mathrm{d}x$$

We then define the **cumulative distribution** $F_X(x) = \mu_X((-\infty, x])$, which will be cadlag, that is *continue à droite, limite à gauche*. Thus, if we take $g(x) = \mathrm{I}_B(x)$, then $\mathsf{E}\left(g(X)\right) = \mathsf{P}\left(X \in B\right)$. Now $\int_{\mathbb{R}} \mathrm{I}_B(x)f(x)\,\mathrm{d}x = \int_B f(x)\,\mathrm{d}x = \mu_X(B) = \mathsf{P}\left(X \in B\right)$. Getting the CDF or the induced measure is equivalent, *i.e.* having either is enough to get the measure for intervals, as $\mu_X((a,b]) = F_X(b) - F_X(a)$ for $a \leq b$.

Change of measure

Suppose we want to make a change of measure from P to $\widetilde{\mathsf{P}}$, thus $(\Omega, \mathcal{F}, \mathsf{P}) \to (\Omega, \mathcal{F}, \widetilde{\mathsf{P}})$ with $\mathsf{E}\left(Z\right) = 1$ and $Z \geq 0$ a.s. Then, the $\widetilde{\mathsf{P}}$ measure of A, $\widetilde{\mathsf{P}}(A) = \int_A Z\,\mathrm{d}\mathsf{P} = \int_\Omega \mathrm{I}_A(Z)\,\mathrm{d}\mathsf{P}$. Thus, $\mathsf{E}\left(\mathrm{I}_A(Z)\right) = \int_\Omega \mathrm{I}_A(Z)\,\mathrm{d}\mathsf{P}$. For this to be a valid change, we need that $\widetilde{\mathsf{P}}$ be a valid probability measure.

1. $\widetilde{\mathsf{P}}(\Omega) = \mathsf{E}\left(\mathrm{I}_\Omega(Z)\right) = \mathsf{E}\left(Z\right) = 1$
2. $\widetilde{\mathsf{P}}\left(\bigsqcup_{i=1}^\infty A_i\right) = \sum_{i=1}^\infty \widetilde{\mathsf{P}}(A_i)$; consider $\bigsqcup_{i=1}^n A_i \coloneqq B_n$. We may decompose $\mathrm{I}_{B_n}(x) =$

[6] So given n, all values from $g(x) \in [0, 2^n)$ are covered.

$\sum_{i=1}^{n} \mathrm{I}_{A_i}(x)$ as to get

$$\widetilde{\mathsf{P}}(B_n) = \mathsf{E}\left(\mathrm{I}_{B_n}(Z)\right) = \sum_{i=1}^{n} \mathsf{E}\left(\mathrm{I}_{A_i}(Z)\right) = \sum_{i=1}^{n} \widetilde{\mathsf{P}}(A_i)$$

and we can apply monotone convergence theorem to get the result.

Given $\mathsf{P}, \widetilde{\mathsf{P}}$ two measures, does there exist a random variable X that satisfies the above conditions. To make a change of variable in the other direction, we need $\widetilde{\mathsf{P}}$ be **absolutely continuous** with respect to P. That is, if $\mathsf{P}(A) = 0 \Rightarrow \widetilde{\mathsf{P}}(A) = 0$. If the relation holds both way, then the measure $\mathsf{P}, \widetilde{\mathsf{P}}$ are equivalent. If this hold, then by the Radon-Nikodym theorem says that the random variable is given by

$$Z(\omega) = \frac{\mathrm{d}\widetilde{\mathsf{P}}}{\mathrm{d}\mathsf{P}}(\omega)$$

and $\mathrm{d}\widetilde{\mathsf{P}} = Z\,\mathrm{d}\mathsf{P}$ where the corresponding $\widetilde{\mathsf{E}}(X) = \mathsf{E}(ZX)$.

This concept will allow us to define more generally the concept of a density of P with respect to $\widetilde{\mathsf{P}}$. The one we were discussing before was the Radon-Nikodym with respect to the Lebesgue measure (relaxing the conditions of Z).[7]

Consider $X \sim \mathcal{N}(0,1)$ and introduce $Y = X + \theta$ for $\theta \in \mathbb{R}^+$. We want to redefine Y to be standard normal and shift the probability mass to the right. Consider $Z = \exp\left(-\theta X - \frac{1}{2}\theta^2\right)$. We want to show that in the $\widetilde{\mathsf{P}}$ measure, that Y is standard normal. We need to show that $Z \geq 0$ (immediate) and that $\mathsf{E}(Z) = 1$. Indeed,

$$\begin{aligned}\mathsf{E}(Z) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\theta x - \frac{1}{2}\theta^2\right) \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(x+\theta)^2\right) \mathrm{d}x\end{aligned}$$

Make the change of variable $y = x + \theta$, we get then

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}y^2\right) \mathrm{d}y$$

[7] For discrete choice model, we may derive maximum likelihood estimation, then we define the likelihood with a density. The change of measure thus yields a density with mass at points.

Thus

$$
\begin{aligned}
\widetilde{\mathsf{P}}(Y \in A) &= \widetilde{\mathsf{E}}\left(\mathrm{I}_A\left(Y\right)\right) \\
&= \mathsf{E}\left(Z \mathrm{I}_A\left(Y\right)\right) \\
&= \mathsf{E}\left(e^{-\theta X - \frac{1}{2}\theta^2} \mathrm{I}_A\left(X + \theta\right)\right) \\
&= \int_{-\infty}^{\infty} e^{-\theta X - \frac{1}{2}\theta^2} \mathrm{I}_A\left(X + \theta\right) \phi(x)\,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(x+\theta)^2\right) \mathrm{I}_A\left(X + \theta\right)\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} \mathrm{I}_A\left(y\right)\mathrm{d}y \\
&= \mathsf{P}\left(A\right) = \mathsf{P}\left(X \in A\right).
\end{aligned}
$$

We say that $\widetilde{\mathsf{P}}$ is absolutely continuous w.r.t. P if and only if $\mathsf{P}\left(A\right) = 0 \Rightarrow \widetilde{\mathsf{P}}(A) = 0$.

Section 3
Information and conditioning

The idea of information is captured formally in this context by the σ-algebra. An **information set** is a description of what is known to agents at a particular point in time.

This can be considered in the case of the binomial asset pricing model, consider splitting up events depending of the current period, in terms of σ-algebra. We can define $\mathcal{F}_0$ to be the information at time zero, consisting of the trivial σ -algebra (the collection $\{\emptyset, \Omega\}$) and for time 1, we have $\mathcal{F}_1$ with $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \cdots$ with $\{\emptyset, H_{\bullet\bullet}, T_{\bullet\bullet}, \Omega\}$. $\mathcal{F}_3$ would be the set of all subsets with all 8 possible outcomes. This increasing sequence of σ-algebra is called a **filtration**. This implies that we have information accumulation, with a history.

We have random variables on $(\Omega, \mathcal{F})$ or in most cases is $(\mathbb{R}, \mathcal{B})$. Most often, we will deal with call a **stochastic process**

Definition 3.1 (Stochastic process)

An indexed set of random variables, indexed by time, denoted $\mathcal{F}(t)$. For $s \leq t$, this imply a filtration $\mathcal{F}(s) \subseteq \mathcal{F}(t)$; we denote the **stochastic process** by $X(t)$.

We say that $X(t)$ is an **adapted process to the filtration** $\{\mathcal{F}(t)\}$, *i.e.* $X(t)$ is $\mathcal{F}(t)$-measurable.[8]

If we specify a probability measure, (with possibly same measurable space), $(\Omega, \mathcal{F}, \mathsf{P})$,

We define for measurable sets **independence.**

Definition 3.2 (Independence)

For two events $A, B \in \mathcal{F}$, we say $A \perp\!\!\!\perp B$ if $\mathsf{P}\left(A \cap B\right) = \mathsf{P}\left(A\right)\mathsf{P}\left(B\right)$.

If X, Y are random variables and $\sigma(X)$ is the σ-algebra generated by X, a subalgebra of $\mathcal{B}$, that is $\forall\ B \in \mathcal{B}$, $X(B) \in \sigma(X)$., We can define the product of Borel algebras as $(\mathbb{R} \times \mathbb{R}, \mathcal{B} \times \mathcal{B})$, corresponding to rectangles in 2d space, of the form $B_1 \times B_2 \equiv \{(x, y) : x \in B_1, y \in B_2\}$.

We have $X \perp\!\!\!\perp Y$ if $\forall\ B_X \in \sigma(X)$, $\forall\ B(y) \in \sigma(Y)$, then $B_X \perp\!\!\!\perp B_Y$ and thus independence of σ-algebra is $\mathcal{G} \perp\!\!\!\perp \mathcal{H}$. We can also have $X \perp\!\!\!\perp \mathcal{G}$, which simply means $\sigma(X) \perp\!\!\!\perp \mathcal{G}$.

We now are interested in conditioning, e.g. weather, and conditional expectation $\mathsf{E}\left(X|\mathcal{G}\right)$. Recall that we have $X = \int_\Omega X(\omega)\, \mathrm{d}\mathsf{P}\left(\omega\right)$. To say X is $\mathcal{F}_0$-measurable. The inverse image of any Borel set under X must be either the null set or everything, meaning all realizations are the same, so $X(\omega) = x$. We will want that $Y = \mathsf{E}\left(X|\mathcal{G}\right)$ be a random variable that is $\mathcal{G}$-measurable; we want $\forall\ G \in \mathcal{G}, \int_G X(\omega)\, \mathrm{d}\mathsf{P}\left(\omega\right) = \int_G Y(\omega)\, \mathrm{d}\mathsf{P}\left(\omega\right)$. This is referred to as the **partial averaging** property. The unconditional expectation can be view as the conditional expectation on $\mathcal{F}_0$ (we cannot split Y over G).

[8] That is, $\forall\ B \in \mathcal{B}$, $\{\omega : X(t\omega) \in B\}$

We can then write Y as a version of conditional expectation of X, that is almost surely.

If $X \perp\!\!\!\perp Y$ and $f, g : \mathbb{R} \to \mathbb{R}$ are measurable functions, then $f(X) \perp\!\!\!\perp g(Y)$. This follows since $\sigma(f(X)) \subseteq \sigma(X)$. Indeed, if we assume $B \in \sigma(f(X))$, then $\exists D \in \mathcal{B}$ such that $f(X(D)) = B$ which implies $X(D) = f^{-1}(B) \in \sigma(X)$.

If we consider $\mathsf{E}\left(X|\mathcal{G}\right)$ and $X \perp\!\!\!\perp \mathcal{G}$, then we would expect the former to be equal to $\mathsf{E}\left(X\right)$, which has to be $\mathcal{G}$-measurable, but its trivial since this is generating $\mathcal{F}_0$. What about partial averaging? We would like to have $\forall\, G \in \mathcal{G}$, $\int_G X \,\mathrm{d}\mathsf{P} = \mathsf{E}\left(X\right) \int_G \mathrm{d}\mathsf{P} = \mathsf{E}\left(X\right) \mathsf{P}\left(G\right)$. We can write $\int_\Omega X \mathrm{I}\left(G\right) \mathrm{d}\mathsf{P} = \mathsf{E}\left(X \mathrm{I}\left(G\right)\right) = \mathsf{E}\left(X\right) \mathsf{E}\left(\mathrm{I}\left(G\right)\right)$.

This uses properties of expectations which are described in Theorem 2.2.7. Recall we had $\mu_X(B) = \mathsf{P}\left(X^{-1}(B)\right)$. For two random variables X, Y, we want

$$\mu_{X,Y}(B_1 \times B_2) = \mathsf{P}\left(X^{-1}(B_1) \cap Y^{-1}(B_2)\right) = \mathsf{P}\left(\{\omega : (X(\omega), Y(\omega)) \in C = B_1 \times B_2\}\right)$$

The joint distribution function is then

$$F_{X,Y}(x, y) = \Pr(X \leq x \quad \text{and} \quad Y \leq y) = \mu_{X,Y}((-\infty, x] \times (-\infty, y]).$$

The **marginal measures**, μ_X, μ_Y on a single copy of the real line by the random variable; then we have $\mu_X(B)$ is the induced measure $\mu_{X,Y}(B \times \mathbb{R})$ and correspondingly, $\mu_Y(D) = \mu_{X,Y}(\mathbb{R} \times D)$. We also define a joint density, $f_{X,Y}(x, y)$ such that $\mu_{X,Y}(C) = \iint_C f_{X,Y}(x, y)\,\mathrm{d}x\,\mathrm{d}y$. We have

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \mathrm{d}x' \int_{-\infty}^{y} \mathrm{d}y' f(x', y')$$

and

$$\frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x, y) = f_{X,Y}(x, y).$$

If we have independence, then those factorize into a product of marginal or CDFs, *i.e.* $F_{X,Y}(x, y) = F_X(x) F_Y(y)$ and $f_{X,Y}(x, y) = f_X(x) f_Y(y)$. We formalize this in the following

Theorem 3.3

Let X, Y be independent random variables. Then if $A \in \sigma(X)$ and $B \in \sigma(Y)$, $A \perp\!\!\!\perp B$. We consider the measure induced by both random variables:

$$\begin{aligned}\mu_{X,Y}(A \times B) &= \mathsf{P}\left((X(\omega) \in A) \cap (Y(\omega) \in B)\right) \\ &\mathsf{P}\left(X(\omega) \in A\right) \mathsf{P}\left(Y(\omega) \in B\right) \\ &= \mu_X(A) \mu_Y(B)\end{aligned}$$

Start now with

$$F_{X,Y} = F_X(x)F_Y(y)$$

and assume the mixed partials exist, so we have

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x,y) = f_X(x)f_Y(y)$$

We define the moment generating function (MGF), $M_X(u) = \mathsf{E}\left(e^{uX}\right)$ provided it exists. We need to care about the domain of integration; notice that e^{uX} can be Taylor-expanded into

$$\sum_{n=0}^{\infty} \frac{u^n X^n}{n!}$$

with $0! = 1$, since $n! = \Gamma(n+1)$. This series converges for all $x \in \mathbb{C}$, thus we can take the expectation inside and get

$$\sum_{n=0}^{\infty} \frac{u^n}{n!} \mathsf{E}\left(X_n\right)$$

If $\psi_X(t) = \mathsf{E}\left(e^{itX}\right)$, then this function always exists,[9] giving rise to the characteristic function, of which $\log(\psi_X(t)$ is the cumulant generating function. Note that $\psi_X(0) = 1$.[10]

The density, provided it exists, is the inverse Fourier transform of the characteristic function, that is

$$\psi_X(t) = \int_{-\infty}^{\infty} f(x)e^{itx}\,\mathrm{d}x$$

thus the density $f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \psi_X(t)e^{-itx}\,\mathrm{d}t$. This means we could also factor the characteristic function, and as such the MGF.

For an arbitrary Borel measurable function $h(x,y)$, we have

$$\mathsf{E}\left(h(X,Y)\right) = \int h(x,y)\,\mathrm{d}\mu_{X\times Y}(x,y)$$

and if $X \perp\!\!\!\perp Y$, this measure $\mu_{X\times Y}$ becomes $\mu_X(x)\mu_Y(y)$

$$= \iint h(x,y)\,\mathrm{d}\mu_X(x)\,\mathrm{d}\mu_Y(y)$$

[9]Since $|e^{itX}| = 1$ since $e^{ix} = \cos(x) + i\sin(x)$ and the absolute value of a complex number of is the square root of the Euclidean distance – recall that $\sin^2(x) + \cos^2(x) = 1$, the result follows then immediately.

[10]The cumulants are functions of the centered moments, but they are not the centered moments.

Proposition 3.4

1. The joint moment generating function is $h = e^{ux}e^{vy}$. This means that we also have factorization of the moment generating function for independent random variables,

$$\mathsf{E}\left(e^{ux}e^{vy}\right) = \mathsf{E}\left(e^{ux+vy}\right) = M_{X,Y}(u,v)$$

and if $X \perp\!\!\!\perp Y$, $M_{X,Y}(u,v) = M_X(u) \cdot M_Y(v)$.

Proof

$$\mathsf{E}\left(h(X,Y)\right) = \iint e^{ux}\,\mathrm{d}\mu_X(x) e^{vy}\,\mathrm{d}\mu_Y(y) = \int e^{ux}\,\mathrm{d}\mu_X(x) \int e^{vy}\,\mathrm{d}\mu_Y(y)$$

■

Taking $h(x,y) = xy$, for independent random variables $X \perp\!\!\!\perp Y$ we have

$$A = \iint xy\,\mathrm{d}\mu_X(x) v\,\mathrm{d}\mu_Y(y) = \int x\,\mathrm{d}\mu_X(x) \int y\,\mathrm{d}\mu_Y(y) = \mathsf{E}\left(X\right)\mathsf{E}\left(Y\right) = \mathsf{E}\left(XY\right)$$

2. More generally, if $X \perp\!\!\!\perp \mathcal{G}$, then $\mathsf{E}\left(X \mid \mathcal{G}\right) = \mathsf{E}\left(X\right)$. For all $G \subset \mathcal{G}$,

$$\begin{aligned}\int_G x\,\mathrm{d}\mathsf{P} &= \int_G \mathsf{E}\left(X\right)\,\mathrm{d}\mathsf{P} = \mathsf{E}\left(X\right)\mathsf{P}\left(G\right) \\ &= \int_\Omega x\mathrm{I}\left(G\right)\mathrm{d}\mathsf{P} \overset{\perp\!\!\!\perp}{=} \mathsf{E}\left(X\right)\mathsf{E}\left(\mathrm{I}\left(G\right)\right) = \mathsf{E}\left(X\right)\mathsf{P}\left(G\right)\end{aligned}$$

3. If X is $\mathcal{G}$ measurable, $\mathsf{E}\left(X \mid \mathcal{G}\right) = X$. For all $G \in \mathcal{G}$

$$\int_G x\,\mathrm{d}\mathsf{P} = \int_G \mathsf{E}\left(X \mid \mathcal{G}\right)\,\mathrm{d}\mathsf{P} \qquad \Rightarrow \mathsf{E}\left(X \mid \mathcal{G}\right) = X$$

4. **Taking out what is known:** Consider $\mathsf{E}\left(XY \mid \mathcal{G}\right)$; since X is $\mathcal{G}$ measurable, X is known and $\mathsf{E}\left(XY \mid \mathcal{G}\right) = X\mathsf{E}\left(Y \mid \mathcal{G}\right)$

Proof We use the standard machine. In step 1, consider $X = \mathrm{I}\left(B\right)$ for $B \in \mathcal{G}$; then

$$\mathsf{E}\left(\mathrm{I}\left(B\right)Y \mid \mathcal{G}\right) = \mathrm{I}\left(B\right)\mathsf{E}\left(Y \mid \mathcal{G}\right)$$

thus

$$\int_G \mathrm{I}\left(B\right)\mathsf{E}\left(Y \mid \mathcal{G}\right)\,\mathrm{d}\mathsf{P} = \int_{G\cap B} \mathsf{E}\left(Y \mid \mathcal{G}\right)\,\mathrm{d}\mathsf{P} = \int_{G\cap B} \mathsf{E}\left(Y \mid \mathcal{G}\right)\,\mathrm{d}\mathsf{P}$$

and so we have

$$\int_G \mathrm{I}\left(B\right)Y\,\mathrm{d}\mathsf{P} = \int_{G\cap B} Y\,\mathrm{d}\mathsf{P} \quad \Leftrightarrow \quad \int_{G\cap B} \mathsf{E}\left(Y \mid \mathcal{G}\right)\,\mathrm{d}\mathsf{P}$$

by definition of $\mathsf{E}(Y \mid \mathcal{G})$, since $G \cap B \subset \mathcal{G}$. ■

Note

Anything which is $\mathcal{G}$ measurable can be treated as deterministic.

5. **Law of iterated expectation**, sometimes referred to as the Tower property. If $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$, then

$$\mathsf{E}(\mathsf{E}(X \mid \mathcal{G}) \mid \mathcal{H}) = \mathsf{E}(X \mid \mathcal{H})$$

following the principle of more information, less average.
6. **Partial averaging**: If $Y = \mathsf{E}(X \mid \mathcal{G})$, we claim $\mathsf{E}(Y \mid \mathcal{H}) = \mathsf{E}(X \mid \mathcal{H})$. Indeed, let $h \in \mathcal{H}$. Then we want to assess whether $\int_h x \, \mathsf{dP} \stackrel{?}{=} \int_h \mathsf{E}(X \mid \mathcal{G}) \, \mathsf{dP}$.
Since $\mathcal{H} \subseteq \mathcal{G}$, we have for all $g \in \mathcal{G}$ that

$$\int_h \mathsf{E}(X \mid \mathcal{G}) \, \mathsf{dP} = \int_h x \, \mathsf{dP}$$

therefore $h \subseteq g$,

$$\int_G \mathsf{E}(X \mid \mathcal{G}) \, \mathsf{dP} = \int_G x \, \mathsf{dP}$$

using the definition of $\mathsf{E}(X \mid \mathcal{G})$.
7. If X is $\mathcal{G}$-measurable and $Y \perp\!\!\!\perp \mathcal{G}$, then

$$\mathsf{E}(f(X,Y) \mid \mathcal{G}) = g(x) = \mathsf{E}_Y(f(X,Y)) = \mathsf{E}_Y(f(X,Y) \mid \mathcal{G})$$

which is a deterministic function of X; thus $\mathsf{E}(f(x,y) \mid \mathcal{G}) = g(X)$ implies $g(X)$ is $\mathcal{G}$-measurable.
8. **Partial averaging**: for all $G \in \mathcal{G}$,

$$\int_G f(x,y) \, \mathsf{dP} = \int_G g(x) \, \mathsf{dP}.$$

Before presenting some properties of the multinormal distribution, we go on with a few definitions.

Definition 3.5 (Martingale)

Let $X(t)$ be a stochastic process adapted to $\mathcal{F}(t)$. $X(t)$ is a **martingale** if $\forall\, t \geq s$,

$$\mathsf{E}(X(t) \mid \mathcal{F}(s)) = X(s).$$

Similarly, if $\mathsf{E}(X(t) \mid \mathcal{F}(s)) \geq X(s)$, we term $X(t)$ a submartingale, while we say that $X(t)$

is a supermartingale if $\mathsf{E}\left(X(t) \mid \mathcal{F}(s)\right) \leq X(s)$.

An example of example of such is gain over time of a biding game. Then, if the game is fair (for example a coin toss with gain/loss of 1), then the expected gain in later periods is the current amount collected.

Recall the binomial asset pricing model: we were able to price derivative securities using risk-neutral measure; we had risk-neutral probabilities, which were the expected values at the later time. The value of the asset in this model is thus a martingale.

We have also the following concept, sometimes referred to as a "forgetting property", that of

Definition 3.6 (Markov process)
$X(t)$ is Markov for $s \leq t$ if there exists a measurable function g such that

$$\mathsf{E}\left(f(X(t)) \mid \mathcal{F}(s)\right) = g(X(s))$$

for f arbitrary measurable.

There are distinctions between the two concepts: f is arbitrary in the definition of the Markov process, while in the definition of martingale, we have the identity function.

Properties of the multinormal Normal distribution

We have $\Phi(x)$ the CDF of the Normal distribution and the density is

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

The normal distribution is a location-scale family, that is the location denotes the expectation and the scale is the square-root of the variance. If we denote those quantities respectively by μ, σ^2, then $\Phi\left(\frac{x-\mu}{\sigma}\right)$ is standard Normal.

In the case of the exponential distribution, this is different since changing the scale, we have CDF $1 - e^{-x}\mathrm{I}\left(x \geq 0\right)$, then the location shift would be in PDF form equivalent to $e^{-\frac{x}{\lambda}}$.

Let $\boldsymbol{W} = (W_1, \ldots, W_n)^\top$ where $W_i \sim \mathcal{N}(0,1)$ and $W_i \perp\!\!\!\perp W_j$. We would say that the vector $\boldsymbol{W} \sim \mathcal{N}_n(\mathbf{0}_n, \mathbf{I}_n)$ where in general

$$[\mathsf{Var}\left(\boldsymbol{X}\right)]_{ij} = \mathsf{E}\left((X_i - \mathsf{E}\left(X_i\right))(X_j - \mathsf{E}\left(X_j\right))\right)$$

Consider $\boldsymbol{Z} = \boldsymbol{A}\boldsymbol{W} + \boldsymbol{E}$ respectively of dimension $(m \times 1)$, $(m \times n)$, $(n \times 1)$ and $(m \times 1)$.

If $\boldsymbol{W}$ is as above, then $\mathsf{E}\left(\boldsymbol{Z}\right) = \boldsymbol{E}$ and

$$\mathsf{Var}\left(\boldsymbol{Z}\right) = \mathsf{Var}\left(\boldsymbol{A}\boldsymbol{W}\right) = \mathsf{E}\left(\boldsymbol{A}\boldsymbol{W}\boldsymbol{W}^\top\boldsymbol{A}^\top\right) = \boldsymbol{A}\mathsf{E}\left(\boldsymbol{W}\boldsymbol{W}^\top\right)\boldsymbol{A}^\top = \boldsymbol{A}\boldsymbol{A}^\top$$

and so $\boldsymbol{Z} \sim \mathcal{N}_m(\boldsymbol{E}, \boldsymbol{A}\boldsymbol{A}^\top)$. Now, wlog let $\mathsf{E}\left(\boldsymbol{Z}\right) = 0$ Now if $\boldsymbol{Z} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma}$ must be **symmetric** and **non-negative definitive.** Recall that non-negative definite means that for any $(1 \times n)$ vector $\boldsymbol{a}$, we have $\boldsymbol{a}^\top \geq 0$. Indeed, we have $\boldsymbol{a}^\top \mathsf{E}\left(\boldsymbol{X}\boldsymbol{X}^\top\right)\boldsymbol{a} = \mathsf{E}\left(\boldsymbol{a}^\top\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{a}\right) = \mathsf{E}\left((\boldsymbol{a}^\top\boldsymbol{X})^2\right)$.

We then get $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}^\top$, which is usually taken to be lower-triangular or upper-triangular. One can use the above as a rule at least for the simulation of the multinormal variates.

The density for our independent random variables will be of the form

$$f_{\boldsymbol{W}}(\boldsymbol{w}) = \prod_{i=1}^{n} \phi(w_i) = (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\boldsymbol{w}^\top\boldsymbol{w}\right)$$

If we perform a location transform, we would get $(\boldsymbol{w} - \boldsymbol{\mu})^\top(\boldsymbol{w} - \boldsymbol{\mu})$ in the expression above. For scaling, recall in the univariate case we have $\Phi\left(\frac{x-\mu}{\sigma}\right)$ gives $\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)$. We suppose $\boldsymbol{X}$ is $(n \times 1)$ and $\boldsymbol{Y}$ is a deterministic function of $\boldsymbol{X}$ of the form $Y_i = y_i(X_1, \ldots, X_n)$, then comparing the probabilities over the two, one would get the transformation

$$f_{\boldsymbol{Y}}(y_1, \ldots, y_n)\,\mathrm{d}y_1 \ldots,\ \mathrm{d}y_n = f_{\boldsymbol{X}}(x_1, \ldots, x_n)\,\mathrm{d}x_1 \ldots,\ \mathrm{d}x_n$$

then we could have dividing through and taking absolute value of the derivative

$$f_{\boldsymbol{Y}}(y_1, \ldots, y_n) = f_{\boldsymbol{X}}(x_1, \ldots, x_n)\frac{\mathrm{d}x_1 \ldots,\ \mathrm{d}x_n}{\mathrm{d}y_1 \ldots,\ \mathrm{d}y_n}$$

This leads to the **Jacobian**, which is a matrix function [of change of variable], with element

$$\left[\frac{\partial(x_1, \ldots, x_n)}{\partial(y_1, \ldots, y_n)}\right]_{ij} = \frac{\partial x_i}{\partial y_j}(y_1, \ldots, y_n)$$

with the Jacobian determinant[11]

Now if $\boldsymbol{Z} = \boldsymbol{A}\boldsymbol{W} + \boldsymbol{\mu}$, then $\frac{\partial \boldsymbol{Z}}{\partial \boldsymbol{W}} = \boldsymbol{A}$. Provided $\boldsymbol{A}$ is non-singular, that is invertible, we may write $\boldsymbol{W} = \boldsymbol{A}^{-1}(\boldsymbol{Z} - \boldsymbol{\mu})$. Recall that singularity would imply linear dependence between the columns of $\boldsymbol{A}$, so we could express one as a linear combination of other variables; this would be happen if we had more. We can also take the inverse of the determinant since this transformation is an homomorphism. Thus $\det(\boldsymbol{\Sigma}) = \det(\boldsymbol{A})\det(\boldsymbol{A}^\top) = \left(\det(\boldsymbol{A})^2\right)$

[11] In French, we avoid this frequent confusion, as "Jacobien" is used for determinant (which is masculine), while matrix is feminine, so is termed "Jacobienne"

The formula for change of variable boils down to

$$f(\boldsymbol{z}) = f(\boldsymbol{w}) \left|\det \boldsymbol{A}\right|^{-1}$$

Now in our specific case, we get

$$(2\pi)^{-\frac{n}{2}} (\det(\boldsymbol{\Sigma}))^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu})^\top (\boldsymbol{A}^\top)^{-1}\boldsymbol{A}^{-1}(\boldsymbol{z}-\boldsymbol{\mu})\right)$$

where $\boldsymbol{A}^{\top^{-1}}\boldsymbol{A}^{-1} = (\boldsymbol{A}\boldsymbol{A}^\top)^{-1} = \boldsymbol{\Sigma}^{-1}$ and if we had this, yields $(\boldsymbol{z}-\boldsymbol{\mu})^\top (\boldsymbol{A}^\top)^{-1}\boldsymbol{A}^{-1}(\boldsymbol{z}-\boldsymbol{\mu}) \sim \chi^2_n$.

Consider now the case of the bivariate Normal distribution; we can parametrize the covariance matrix in terms of three parameters, namely $\mathsf{Var}\,(X) = \sigma_1^2, \mathsf{Var}\,(Y) = \sigma_2^2$ and the Pearson linear correlation $\rho = \mathsf{Cor}(X,Y)$. Recall that

$$\mathsf{Cor}(X,Y) = \frac{\mathsf{Cov}\,(X,Y)}{\sqrt{\mathsf{Var}\,(X)\,\mathsf{Var}\,(Y)}}$$

We have

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \cdot \begin{bmatrix} 1 & \rho \\ 1 & \rho \end{bmatrix} \cdot \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}$$

and correspondingly the inverse gives

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{1-\rho^2} \begin{bmatrix} 1/\sigma_1 & 0 \\ 0 & 1/\sigma_2 \end{bmatrix} \cdot \begin{bmatrix} 1 & -\rho \\ 1 & -\rho \end{bmatrix} \cdot \begin{bmatrix} 1/\sigma_1 & 0 \\ 0 & 1/\sigma_2 \end{bmatrix}$$

We want to find $\boldsymbol{B}\boldsymbol{B}^\top$ and suppose that $\boldsymbol{B}$ is lower-triangular; we get

$$\begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \cdot \boldsymbol{B}\boldsymbol{B}^\top \cdot \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} = \boldsymbol{A}\boldsymbol{A}^\top$$

The procedure is termed **Choleski** decomposition, or sometimes referred to as the Choleski decomposition. Let

$$\boldsymbol{B} = \begin{bmatrix} b_11 & 0 \\ b_{21} & b_{22} \end{bmatrix} \qquad \boldsymbol{B}\boldsymbol{B}^\top = \begin{bmatrix} b_11^2 & b_{11}b_{21} \\ b_{11}b_{21} & b_{22}^2 + b_{21}^2 \end{bmatrix}$$

and choosing $b_{11} = 1$, then this restriction forces $b_{21} = \rho$ and $b_{22} = \sqrt{1-\rho^2}$. This leaves us with

$$\boldsymbol{A} = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix} = \begin{bmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_1\sqrt{1-\rho^2} \end{bmatrix}$$

In the case of the bivariate Normal distribution with zero mean vector, we have

$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_1\sqrt{1-\rho^2} \end{bmatrix} \cdot \begin{bmatrix} W_1 \\ W_2 \end{bmatrix} = \begin{bmatrix} \sigma_1 W_1 \\ \sigma_2(\rho W_1 + rW_2) \end{bmatrix}$$

where $r = \sqrt{1-\rho^2}$. From this construction, we can also derive the conditional distribution of $Z_2|Z_1$. Now

$$\mathsf{E}\left(Z_2 \mid Z_1\right) = \sigma^2 \mathsf{E}\left(\rho W_1 + rW_2\right) \mid W_1) = \rho\sigma_2 W_1 = \frac{\rho\sigma_2 Z_1}{\sigma_1}$$

and accordingly,[12] we have

$$\begin{aligned} \mathsf{Var}\left(Z_2|Z_1\right) &= \mathsf{E}\left(Z_2 - \rho\frac{\sigma_2 Z_1}{\sigma_1} \middle| Z_1 \right) \\ &= \mathsf{E}\left((\sigma_2(\rho W_1 + rW_2) - \rho\sigma_2 W_1)^2 \middle| Z_1 \right) \\ &= \mathsf{E}\left((r\sigma_2 W_2)^2 | W_1 \right) = r^2\sigma_2^2 = \sigma_2^2(1-\rho^2) \end{aligned}$$

This leads to the conclusion that $Z_2|Z_1 \sim \mathcal{N}(\rho\sigma_2 Z_1/\sigma_1, \sigma_2^2(1-\rho^2))$. Conditioning leads to using information from the σ-algebra and allows to improve the forecast (reduce the variance) and change the expectation.

[12]To remember, think of different currency change

Section 4
Brownian motion

We will treat the concept of random walk in one dimension. Consider the simple case with

$$X_1 = \begin{cases} +1 & \text{with probability } \frac{1}{2} \\ -1 & \text{with probability } \frac{1}{2} \end{cases}$$

with $M_0 = 0$ and $M_t = \sum_{j=1}^{t} X_j$. We could also consider the drunken walk, that is walking on an integer lattice in $\mathbb{Z} \times \mathbb{Z}$. We consider **increments** $M_t - M_s$ for $s < t$ and write

$$M_t = M_s + (M_t - M_s);$$

the process is characterized by independent increments, if $0 < t_1 < t_2 < \ldots, t_n$. Then $M_{t_n} = M_0 + (M_{t_1} - M_{t_0}) + (M_{t_2} - M_{t_1}) + \cdots + (M_{t_n} - M_{t_{n-1}})$. Now M_t is a martingale, since $\mathsf{E}\left(M_t \mid \mathcal{F}_s\right) = \mathsf{E}\left((M_t - M_s) + M_s \mid \mathcal{F}_s\right) = M_s$.

The **quadratic variation** of the process, $[M, M]_t = \sum_{j=1}^{t} (M_j - M_{j-1})^2 = \sum_{j=1}^{t} X_j^2 = t$, as opposed to the variance; we know $M_t = \sum_{j=1}^{t} X_j$; since the increments are independent, we have

$$\mathsf{Var}\left(M_t\right) = \sum_{j=1}^{t} \mathsf{Var}\left(X_j\right) = t$$

again, because we have a symmetric random walk with Rademacher distribution. We have $\mathsf{E}\left(X_j\right) = 0$, with $\mathsf{Var}\left(X_j\right) = 1$. This is just a coincidence.

Define now the **scaled random walk** to be

$$W_{nt}^{(n)} = \frac{1}{\sqrt{n}} \sum_{j=1}^{nt} X_j = \frac{1}{\sqrt{n}} M_{nt}$$

with $\mathsf{E}\left(W_{nt}^{(n)}\right) = 0$ and $\mathsf{Var}\left(W_{nt}^{(n)}\right) = \frac{tn}{n} = t$. From there, we generalize to continuous time process; now $W^{(n)}(t) = W_{\lfloor nt \rfloor}^{(n)}$ for $t \in [0, 1]$. Again, it is clear that the expectation is zero, but now the quadratic variation now is

$$\left[W^{(n)}, W^{(n)}\right]_t = \sum_{j=1}^{\lfloor nt \rfloor} \left(\frac{1}{\sqrt{n}}\right)^2 = \frac{\lfloor nt \rfloor}{n}$$

We can establish $W_t^{(n)} \xrightarrow{d} \mathcal{N}(0, t)$ as $n \to \infty$, known as convergence in distribution or weak convergence. That is $\{X_t\} \xrightarrow{d} F$ and

$$\lim_{t\to\infty} \int g(X_t)\,\mathrm{d}\mathsf{P}_t = \int g(x)\,\mathrm{d}F(x)$$

The result as stated is simply a consequence of the Central Limit Theorem, which we use here since $\mathsf{E}\,(X_t) = 0$ and $\mathsf{Var}\,(X_t) = \sigma^2$, and X_t are IID random variables, then

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

that is the partial sum converges in law to a scaled Normal distribution. Recall we had defined

$$W_t^{(n)} = \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor nt \rfloor} X_j$$

and since $\sqrt{t}^{-1} W_t^{(n)} \xrightarrow{d} \mathcal{N}(0,1)$. Thus $W_t^{(n)} \xrightarrow{d} \mathcal{N}(0,t)$.

Denote the partial sum by Z_n; we are interested in

$$\begin{aligned}
M_{Z_n}(u) &= \mathsf{E}\left(e^{uZ_n}\right) \\
&= \mathsf{E}\left(\exp\left(u\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_t\right)\right) \\
&= \mathsf{E}\left(\prod_{t=1}^{n} \exp\left(\frac{uX_t}{\sqrt{n}}\right)\right) \\
&= \prod_{i=1}^{n} \mathsf{E}\left(\exp\left(\frac{uX_t}{\sqrt{n}}\right)\right)
\end{aligned}$$

and further denote $M_X(y) = \mathsf{E}\left(e^{uX_t}\right)$

$$\prod_{i=1}^{n} M_X\left(\frac{u}{\sqrt{n}}\right) = \left(M_X\left(\frac{u}{\sqrt{n}}\right)\right)^n$$

In the case of interest to us, the MGF for a single term is $\frac{1}{2}(e^u + e^{-u}) \equiv \cosh(u)$. We then are left with

$$\frac{1}{2^n}\left(e^{\frac{u}{\sqrt{n}}} + e^{-\frac{u}{\sqrt{n}}}\right)^n$$

Take logs, we get

$$-n\log(2)+n\log\left(e^{\frac{u}{\sqrt{n}}}+e^{-\frac{u}{\sqrt{n}}}\right)=n\log\left(1+\left(\frac{e^{\frac{u}{\sqrt{n}}}+e^{-\frac{u}{\sqrt{n}}}}{2}-1\right)\right).$$

We now want to perform a Taylor series expansion around 1; recall $\log(1+x)=x-\frac{x^2}{2}+\frac{x^3}{3}-\frac{x^4}{4}\ldots$. Now we have

$$x\mapsto\frac{1}{2}\left(e^{\frac{u}{\sqrt{n}}}+e^{-\frac{u}{\sqrt{n}}}\right)-1$$

and this is

$$\frac{1}{2}\left(e^{\frac{u}{\sqrt{n}}}+e^{-\frac{u}{\sqrt{n}}}\right)-1=\frac{u^2}{n}+O\left(n^{-2}\right)$$

Now multiply both sides by n to get $\frac{u^2}{2}+O\left(n^{-1}\right)$. The quantity tends to $e^{\frac{u^2}{2}}$ as $n\to\infty$; this corresponds to the MGF of a standard normal. Indeed, if $X\sim\mathcal{N}(0,1)$, then

$$M_X(y)=\mathsf{E}\left(e^{uX}\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{ux}e^{-\frac{x^2}{2}}\,\mathrm{d}x$$

where now we complete the square; the terms in the exponent are

$$-\frac{x^2}{2}+ux=-\frac{1}{2}(x^2-2ux+u^2)+\frac{1}{2}u^2$$

as to get

$$\frac{e^{\frac{u^2}{2}}}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}(x-\mu)^2\right)\mathrm{d}x=e^{\frac{u^2}{2}}$$

We do not need much extra regularity conditions before to finish the proof. We have

$$\psi(u)=\sum_{j=1}^{\infty}\frac{iu^j}{j!}\kappa_i=\frac{u^2}{2n}-\kappa_3\frac{u^3}{3!n^{\frac{3}{2}}}$$

as to get $n\psi(u/\sqrt{n})=-\frac{u^2}{2}+O(n^{-\frac{1}{2}})$.[13] The cumulants are the same as those of the characteristic function; we need regularity conditions that are κ_3 existing, which is restrictive.

What we want is $W^{(n)}(t)\xrightarrow{d}W(t)$ for $t\in[0,1]$. We need a functional CLT, which is beyond the scope of the course. We will be interested in $\mathsf{Cov}\left(W(s),W(t)\right)$ for $s\le t$; we approximate

[13] In the earlier case, the cumulants were that of the symmetric random variable.

the continuous version with the discrete analog, with

$$
\begin{aligned}
\mathsf{Cov}\left(W^{(n)}(s), W^{(n)}(t)\right) &= \mathsf{Cov}\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor ns\rfloor} X_j, \frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor nt\rfloor} X_j\right) \\
&= \frac{1}{n}\sum_{j=1}^{\lfloor ns\rfloor}\sum_{k=1}^{\lfloor nt\rfloor} \mathsf{E}\left(X_j X_k\right) \\
&= \sum_{j=1}^{\lfloor ns\rfloor} 1 \\
&= \frac{\lfloor ns\rfloor}{n} \to s
\end{aligned}
$$

We thus establish $\mathsf{Cov}\left(W(s), W(t)\right) = \min(s,t)$ and we have independent increments. From the functional CLT, we establish

$$
\mathsf{Cov}\left(W(s), W(t) - W(s)\right) = \mathsf{E}\left(W(s)W(t)\right) - \mathsf{E}\left(W^2(s)\right) = s - s = 0
$$

Conditions for the Brownian motion; see Shreve.

We now look at transition from the Binomial asset pricing model to infinite horizon when the period between two coin toss goes to zero. Suppose that the events "up" and "down" with constant probability p, q. Now, we express

$$
u = 1 + \frac{\sigma}{\sqrt{n}} \qquad d = 1 - \frac{\sigma}{\sqrt{n}}
$$

There are a total of $nt = H_t + T_t$ periods, where H_t is the number of up transitions before time t and T_t the number of down transition. Let $M_t = H_t - T_t$ corresponding to the gain; we require M_t to be a martingale.

$$
\begin{aligned}
\lfloor nt\rfloor &= H_t + T_t \\
M_t &= H_t - T_t \\
H_t &= \frac{1}{2}\left(\lfloor nt\rfloor + M_t\right) \\
T_t &= \frac{1}{2}(\lfloor nt\rfloor - M_t)
\end{aligned}
$$

so that, using earlier notation,

$$
M_t = \frac{1}{\sqrt{n}} W_{\lfloor nt\rfloor}
$$

We then have

$$\begin{aligned} S_t &= S_0 u^{H_t} d^{T_t} \\ &= S_0 \left(1+\frac{\sigma}{\sqrt{n}}\right)^{\frac{1}{2}(\lfloor nt \rfloor + M_t)} \left(1-\frac{\sigma}{\sqrt{n}}\right)^{\frac{1}{2}(\lfloor nt \rfloor - M_t)} \end{aligned}$$

and this is largely intractable. Take logarithm on both sides to get

$$\log(S_t) = \log(S_0) + \frac{1}{2}(\lfloor nt \rfloor + \sqrt{n} W_{\lfloor nt \rfloor} \log\left(1+\frac{\sigma}{\sqrt{n}}\right) + \frac{1}{2}(\lfloor nt \rfloor - \sqrt{n} W_{\lfloor nt \rfloor} \log\left(1-\frac{\sigma}{\sqrt{n}}\right)$$

which we Taylor expand, setting $n^{-\frac{1}{2}} W_{\lfloor nt \rfloor} \equiv W(t)$

$$\begin{aligned} &= \log(S_0) + \frac{1}{2}\left(\lfloor nt \rfloor + \sqrt{n} W_{\lfloor nt \rfloor}\right)\left(\frac{\sigma}{\sqrt{n}} - \frac{\sigma^2}{2n} + \frac{\sigma^3}{3n\sqrt{n}} + O(n^{-2})\right) \\ &\qquad + \frac{1}{2}\left(\lfloor nt \rfloor - \sqrt{n} W_{\lfloor nt \rfloor}\right)\left(-\frac{\sigma}{\sqrt{n}} - \frac{\sigma^2}{2n} - \frac{\sigma^3}{3n\sqrt{n}} + O(n^{-2})\right) \\ &= \log(S_0) + \frac{1}{2}\left[\sigma t\sqrt{n} - \frac{\sigma^2 t}{2} + O(n^{-\frac{1}{2}}) + \sigma W_{\lfloor nt \rfloor} - \sigma t \sqrt{n} - \frac{\sigma^2 t}{2} + \sigma W_{\lfloor nt \rfloor}\right] \\ &= \log(S_0) - \frac{\sigma^2 t}{2} + \sigma W_n(t) + O(n^{-\frac{1}{2}}) \\ &\xrightarrow{n\to\infty} \quad \log(S_0) - \frac{\sigma^2 t}{2} + \sigma W(t) \end{aligned}$$

as to get

$$S_t = S_0 \exp\left(\sigma W(t) - \frac{\sigma^2 t}{2}\right)$$

which has lognormal distribution. This construction gives rise to **geometric Brownian motion**. Clearly, it is positive and is moreover is martingale. We have independent increments which follows since $\mathsf{Cov}\left(W(t), W(s)\right) = \min(t,s)$ and the joint distribution is multivariate Normal.

At this stage, we will use the σ-algebra (the filtration) such that the Brownian motion at time t is adapted at time t to $\mathcal{F}(t)$. We want to show that for $s \leq t$,

$$\mathsf{E}\left(W(t) \mid \mathcal{F}(s)\right) \stackrel{?}{=} W(s)$$

Starting by splitting into increments, as to get

$$\mathsf{E}\left(W(s) + (W(t) - W(s)) \mid \mathcal{F}(s)\right) = W(s)$$

To show that the Brownian motion is a Markov process,

$$\mathsf{E}\left(f(W(t)) \mid \mathcal{F}(s)\right) \stackrel{?}{=} g(W(s))$$

We go about proving this the same way as before namely

$$\mathsf{E}\left(f(W(t) + (W(t) - W(s))) \mid \mathcal{F}(s)\right)$$

and by the independence lemma.

$$g(x) = \mathsf{E}\left(f(x + (W(t) - W(s)))\right)$$

which yields the result. We now aim at establishing that

$$\begin{aligned}
\mathsf{E}\left(S(t) \mid \mathcal{F}(s)\right) &= S_0 \mathsf{E}\left(\exp\left(-\frac{\sigma^2 t}{2}\right) \exp\left(\sigma(W(s) + (W(t) - W(s))\right) \middle| \mathcal{F}(s)\right) \\
&= S_0 \exp\left(-\frac{\sigma^2 t}{2}\right) \exp(\sigma W(s)) \mathsf{E}\left(\exp(\sigma(W(t) - W(s)))\right)
\end{aligned}$$

where we established that the inside is $\mathcal{N}(0, \sigma^2(t-s))$; now use MGF

$$\begin{aligned}
&= \mathsf{E}\left(\exp\left(\sigma\sqrt{t-s}Z\right)\right) \\
&= \exp\left(\frac{\sigma^2(t-s)}{2}\right)
\end{aligned}$$

Coming back to the equation $\mathsf{E}\left(S(t) \mid \mathcal{F}(s)\right)$

$$= S_0 \exp\left(-\frac{\sigma^2 s}{2} + \sigma W(s)\right) = S(s)$$

since the expression on the left is the definition of the geometric Brownian motion.

All these properties could hold for filtrations larger than the one generated by the Brownian motion itself. For it to be adaptive to the "arbitrary " filtration denoted $\mathcal{F}(t)$ we require $W(t)$ is $\mathcal{F}(t)$-measurable. We also need $W(t) - W(s) \perp\!\!\!\perp \mathcal{F}(s)$.

The next step is to look at the quadratic variation of the process itself. Again, in our case we will still get the quadratic variation to be non-random (and equal to the variance once again); we will need to dig harder to get a stochastic process to get something stochastic. We consider the sum of all variations, either positive or negative, added for the decomposition. An example of such function is $\sin(x^{-1})$, since it will oscillate infinitely often at the origin. We may be interested in **first-order variation**, the oscillation, which is the function f is continuously differentiable is equal to $\int_a^b |f'(t)| \,\mathrm{d}t$. It is standard fact of life that Brownian motion has infinite point of non-differentiability; almost nowhere differentiable.

Definition 4.1 (Quadratic-variation)

We have for an interval $[0, T]$ the quadratic variation QV for Π a partition, where the norm is just the biggest distance between two points in the partition, namely $\|\Pi\| = \max_{i=1}^n (t_i - t_{i-1})$.

$$\mathrm{QV} = \lim_{\|\Pi\|\to 0} \sum_{i=1}^n (f(t_i) - f(t_{i-1}))^2$$

If f was differentiable, by the mean-value theorem, then

$$f(t_i) - f(t_{i-1}) = (t_i - t_{i-1}) f'(t_i^*)$$

which now yields

$$\sum_{i=1}^n (t_i - t_{i-1})^2 (f'(t_i^*))^2 \leq \|\Pi\| \sum_{i=1}^n (t_i - t_{i-1})(f'(t_i^*))^2 \to 0$$

as $\|\Pi\| \to 0$. Here, the partition yields $\int_0^T (f'(t))^2 \,\mathrm{d}t$ and the quadratic variation goes to zero if the integral is finite.

For a Brownian motion, we have

$$\mathsf{E}\left(\sum_{i=1}^n (W(t_i) - W(t_{i-1})^2\right) = \sum_{i=1}^n (t_i - t_{i-1}) = T$$

and we show that this is deterministic in the limit by showing that it's variance goes to zero. Indeed,

$$\begin{aligned}\mathsf{Var}\left(\sum_{i=1}^n (W(t_i) - W(t_{i-1}))^2\right) &= \sum_{i=1}^n \mathsf{Var}\left((W(t_i) - W(t_{i-1}))^2\right) \\ &= 2\sum_{i=1}^n (t_i - t_{i-1}) \\ &\leq 2\|\Pi\| \sum_{i=1}^n (t_i - t_{i-1}) \to 0\end{aligned}$$

where $W(t_i) - W(t_{i-1}) \sim \mathcal{N}(0, t_i - t_{i-1})$. If $Z \sim \mathcal{N}(0,1)$ and $\mathsf{Var}\left(Z^2\right) = \mathsf{E}\left(Z^4\right) - (\mathsf{E}\left(Z^2\right))^2 = 3-1 = 2$; we could get the fourth moment using the formula$(2n-1)(2n-3)\cdots 1$ to recover the moments of higher order (even); all odd moments are zero by symmetry. Or you could just try brute force, or MGF.

We say that the Brownian motion accumulates quadratic variation of amount T. The notation is akin to what we had before, that is $[W, W](T) = T$.

Some notation, we could consider $(W(t_i) - W(t_{i-1}))^2$, we could express this as $\mathrm{d}W(t)\,\mathrm{d}W(t) = \mathrm{d}t$ using non-rigorous notation; this will be useful for formulas and calculations. We could similarly look at $\mathrm{d}t\,\mathrm{d}t = 0$. Now, we could also write $\mathrm{d}W(t)\,\mathrm{d}t$, the cross-variation, defined as

$$\lim_{\|\Pi\|\to 0} \sum_{i=0}^{n} (W(t_i) - W(t_{i-1}))(t_i - t_{i-1}) \leq \|\Pi\| \sum_{i=0}^{n} (W(t_i) - W(t_{i-1})) = \|\Pi\| W(T) \to 0$$

so the rule is $\mathrm{d}W(t)\,\mathrm{d}t = 0$. These will be the rules of stochastic calculus.

Volatility of geometric Brownian Motion

We consider now $S(t) = S(0)\exp\left(\sigma W(t) + \left(\alpha - \frac{\sigma^2}{2}\right)t\right)$, where we can interpret α as the **drift**. Unless we had a term, the expectation would be equal to $S(0)$ at all time. We would like an exponential growth for the rate of return of money. We can develop the concept of log return, which would be

$$\begin{aligned} \log\left(\frac{S(t_i)}{S(t_{i-1})}\right) &= \log\left(1 + \frac{S(t_i) - S(t_{i-1})}{S(t_{i-1})}\right) \\ &\approx \frac{S(t_i) - S(t_{i-1})}{S(t_{i-1})} \end{aligned}$$

using a Taylor-series expansion.

If we look at the squared log returns, that is $\sum_{i=1}^{n}\left(\log\left(\frac{S(t_i)}{S(t_{i-1})}\right)^2\right)$; before, we can look at the asymptotic behavior of

$$\log\left(\frac{S(t_i)}{S(t_{i-1})}\right) = \sigma(W(t_i) - W(t_{i-1})) + \left(\alpha - \frac{\sigma^2}{2}\right)(t_i - t_{i-1})$$

so as to get

$$\begin{aligned} \sum_{i=1}^{n} \sigma^2 (W(t_i) - W(t_{i-1}))^2 + 2\left(\alpha - \frac{\sigma^2}{2}\right)(W(t_i) - W(t_{i-1}))(t_i - t_{i-1}) \\ + \left(\alpha - \frac{\sigma^2}{2}\right)^2 (t_i - t_{i-1})^2 \end{aligned}$$

$\to \sigma^2 T$ where

$$\sigma^2 = \frac{1}{T} \lim_n \sum_{i=1}^n \left(\log \left(\frac{S(t_i)}{S(t_{i-1})} \right) \right)^2$$

termed the **realized volatility**.

If we are interested in passage to a point m, we have the **first-passage time** of $W(t)$,

$$\tau_m = \min \{ t \mid W(t) = m \}$$

We will look at stopped martingale, denoted $W(t \wedge \tau_m)$, where we stop at τ_m and so the argument will be $\min(t, \tau_m)$. We make use of the properties of the exponential martingale

$$\begin{aligned} 1 &= \mathsf{E}\left(\exp\left(\sigma W(t) - \frac{\sigma^2 t}{2} \right) \right) \\ &= \mathsf{E}\left(\exp\left(\sigma W(t \wedge \tau_m) - \frac{\sigma^2 (t \wedge \tau_m)}{2} \right) \right) \end{aligned}$$

since the term in the exponential is a martingale, and is zero at $t = 0$. By properties of the martingale, this is zero, so $e^0 = 1$ gives the result. Wlog, suppose $m > 0$; if $\tau_m < \infty$, then

$$1 = \mathsf{E}\left(\mathrm{I}\left(\tau_m < \infty \right) \exp\left(\sigma m - \frac{1}{2} \sigma^2 \tau_m \right) \right)$$

We can let $\sigma \to 0$ with monotone convergence. We also take $t \to 0$, then by MCT, we get in the limit

$$1 = \mathsf{E}\left(\mathrm{I}\left(\tau_m < \infty \right) \right) = \mathsf{P}\left(\tau_m < \infty \right).$$

We get

$$1 = \mathsf{E}\left(\exp\left(\sigma m - \frac{1}{2} \sigma^2 \tau_m \right) \right)$$

and rearrangement yields $\mathsf{E}\left(-\frac{1}{2} \sigma^2 \tau_m \right) = e^{-\sigma m}$ Now if $\alpha = \sigma^2/2$, then $\sigma = \sqrt{2\alpha}$ and

$$\mathsf{E}\left(e^{-\alpha \tau_m} \right) = e^{m\sqrt{2\alpha}}$$

and $\mathsf{E}\left(\tau_m \right) = \infty$ as differentiating with respect to α gives

$$\mathsf{E}\left(\tau_m e^{-\alpha \tau_m} \right) = \frac{m\sqrt{2}}{2\sqrt{\alpha}} e^{m\sqrt{2\alpha}}$$

This is yet a paradoxical result, moment generating function $e^{m\sqrt{2\alpha}}$, $\mathsf{P}\left(\tau_m < \infty\right) = 1$ yet $\mathsf{E}\left(\tau_m\right) = \infty$.

We now study the **reflexion principle**; suppose we have hitting time τ_m for m; then the continuation of the Brownian motion for times $t > \tau_m$. Suppose we reach $w < m$ at t_i; the path has length $m - w$ and the reflexion principle states that this has same probability as drift to $2m - w$ positively reflected around m, *i.e.*

$$\mathsf{P}\left(W(t) < w \wedge \tau_m < t\right) = \mathsf{P}\left(W(t) > 2m - w\right)$$

If $w = m$, then

$$\mathsf{P}\left(W(t) < m \wedge \tau_m < t\right) = \mathsf{P}\left(W(t) > m\right) = \mathsf{P}\left(Z > \frac{m}{\sqrt{t}}\right) = 1 - \Phi\left(\frac{m}{\sqrt{t}}\right)$$

and similarly for $\mathsf{P}\left(W(t) > m \wedge \tau_m < t\right)$, so by the law of total probability, we get

$$\mathsf{P}\left(\tau_m < t\right) = 2 - 2\Phi\left(\frac{m}{\sqrt{t}}\right)$$

We will now consider the **maximum to date** of the Brownian motion [14]

Definition 4.2 (Maximum-to-date)
We denote **maximum to date** $M(t) = \max_{s \leq t} W(s)$. For $\tau_m \leq t$, then $M(t) \geq w$.

What we are interested in is to look at the joint distribution

$$\begin{aligned}
\mathsf{P}\left(M(t) \geq m \wedge W(t) \leq w\right) &= \mathsf{P}\left(\tau_m \leq t \wedge W(t) \leq w\right) \\
&= \mathsf{P}\left(W(t) \geq 2m - w\right) \\
&= 1 - \Phi\left(\frac{2m - w}{\sqrt{t}}\right)
\end{aligned}$$

If we have the joint density of $M(t), W(t)$, this would be

$$\int_m^{\infty} \mathrm{d}y \int_{-\infty}^{w} \mathrm{d}x f_{M(t),W(t)}(y, x)$$

First, differentiate both expressions with respect to m, to get

$$\frac{2}{\sqrt{t}}\phi\left(\frac{2m - w}{\sqrt{t}}\right) = \int_{-\infty}^{w} f_{M(t),W(t)}(m, x)\,\mathrm{d}x$$

[14] This is not appealing for financial economic purposes, since this is allowed to take negative values. Certainly, the exponential martingale is a more interesting model.

and further differentiating with respect to w, we get

$$f_{M(t),W(t)}(m,w) = \frac{2}{t}\frac{2m-w}{\sqrt{t}}\phi\left(\frac{2m-w}{\sqrt{t}}\right)$$

since $\phi'(x) = -x\phi(x)$.

We saw that the Brownian motion is a Markov process, which told us that if we consider $f(W(t))$ any function, then we have

$$\mathsf{E}\left(f(W(t)) \mid \mathcal{F}(s)\right) = g(W(s)) = \int_{-\infty}^{\infty} f(y)p(\tau,x,y)\,\mathrm{d}y$$

with $t-s=\tau$ and $p(\tau,x,y)$ is a density function of y for time increment τ from x. Again, we have $g(x) = \mathsf{E}\left(f(x+W(t)-W(s))\right)$. We want to express this knowing that $W(t)-W(s) \sim \mathcal{N}(0,t-s)$. We can write this as

$$\frac{1}{\sqrt{2\pi(t-s)}}\int_{-\infty}^{\infty} \mathrm{d}z f(x+z)\exp\left(-\frac{z^2}{2(t-s)}\right)$$

and so make the change of variable $x+z=y$ as to get

$$= \frac{1}{\sqrt{2\pi\tau}}\int_{-\infty}^{\infty} \mathrm{d}y f(y)\exp\left(-\frac{(y-x)^2}{2\tau}\right)$$

and so the termed **transition density** is then

$$p(\tau,x,y) = \frac{1}{\sqrt{2\pi\tau}}\exp\left(-\frac{(y-x)^2}{2\tau}\right)$$

Section 5
Stochastic calculus

We will consider $\int_0^\tau W(t)\,\mathrm{d}t$, but also integrals $\int_0^T \Delta(t)\,\mathrm{d}W(t)$, but non-differentiability of $W(t)$ implies we need careful rules. We will show that $\mathrm{d}W(t)\,\mathrm{d}W(t) = \mathrm{d}t$ and $\mathrm{d}t\,\mathrm{d}t = 0$ and $\mathrm{d}W(t)\,\mathrm{d}t$.

We will be interested in the Itō integral, for $\int_a^b \Delta(t)\,\mathrm{d}W(t)$, *i.e.* where the integrator is a Brownian motion. This is simply an example of Stieltjes integral; if we had a differentiable integrator, we would have

$$\int f\,\mathrm{d}g = \int fg'\,\mathrm{d}t$$

with $\mathrm{d}g = g'\,\mathrm{d}t$, with those being interpreted as one form differential on a manifold. Incidentally, this allows us to write $\mathrm{d}(fg) = f\,\mathrm{d}g + g\,\mathrm{d}f$, we get a formula for integration by part

$$\int_a^b f\,\mathrm{d}g = f(b)g(b) - f(a)g(a) - \int_a^b g\,\mathrm{d}f$$

Recall when we had something not integrable, we would consider partitions of the form $\sum_{i=1}^n f(t_i)(g(t_i) - g(t_{i-1}))$ and in the limit, we would have no difference using $f(t_i^*)$ where t_i^* is between t_i and t_{i-1}, not necessarily a convex combination. If we have a random integrator and a (potentially random) integrand, as long as we had some properties, such as bounded quadratic variation.

We will impose the condition that $\Delta(t)$ is adapted, that is there is a filtration $\mathcal{F}(t)$, and such that we have independent increments, that is $W(t+s) - W(t) \perp\!\!\!\perp \mathcal{F}(t)$. Suppose $\Delta(t)$ is a simple function, or piecewise-constant. For $t_{i-1} \leq t < t_i$, $\Delta(t) = \Delta(t_{i-1})$. If, in the ordinary circumstances have a piecewise constant, then

$$\int_a^b \Delta(t)\,\mathrm{d}W(t) = \sum_{i=1}^n \Delta(t_{i-1})(W(t_i) - W(t_{i-1})), \qquad t_0 = a, t_n = b$$

This is then perfectly well-defined and each factor in the product are independent. Since $\Delta(t_{i-1})$ is $\mathcal{F}(t)$ measurable since adapted and so is independent of $W(t_i) - W(t_{i-1})$.[15]

We will use a limiting procedure to develop the Itō integral for arbitrary measurable func-

[15] There is also another sort of stochastic integral, the Stratonovich integral, with $\Delta(t_{i-1})$, the lower end, replaced by $\Delta((t_i - t_{i-1})/2)$. Since we are interested in the value of a portfolio; since Δ is the number of units of assets, with $W(t_i)$ the evolution of the price. The Itō integral would thus give the profit in a given period on the stock market. These integrals were developed for chemical diffusion models.

tions, not necessarily piecewise constant. We will first look at some properties.

Proposition 5.1 (Properties of the Itō integral)
Wlog, suppose $I(t) = \int_0^t \Delta(S)\,\mathrm{d}W(s)$

1. $I(t)$ is a martingale
Proof We want to show $\mathsf{E}\left(I(t+s) \mid \mathcal{F}(t)\right) = I(t)$. Clearly, conditional on $\mathcal{F}(t)$ we have measurability. Now let $t_0 = t$ and consider

$$I(t+s) = I(t) + \sum_{i=1}^{n_{t,s}} \Delta(t_{i-1})(W(t_i) - W(t_{i-1}))$$

We have $\mathsf{E}\left(I(t+s) \mid \mathcal{F}(t)\right)$ is by linearity

$$\begin{aligned}
\mathsf{E}\left(I(t+s) \mid \mathcal{F}(t)\right) &= I(t) + \sum_{i=1}^{n_{t,s}} \mathsf{E}\left(\Delta(t_{i-1})(W(t_i) - W(t_{i-1})) \mid \mathcal{F}(t)\right) \\
&= I(t) + \sum_{i=1}^{n_{t,s}} \Delta(t_{i-1})\mathsf{E}\left(W(t_i) - W(t_{i-1})\right) \\
&= I(t) + \sum_{i=1}^{n_{t,s}} \Delta(t_{i-1})\left(\mathsf{E}\left(W(t_i)\right) - \mathsf{E}\left(W(t_{i-1})\right)\right) \\
&= I(t)
\end{aligned}$$

using the fact that $\Delta(t_{i-1})$ is $\mathcal{F}(t)$ measurable and that $W(t_i) - W(t_{i-1})$ is independent of $\mathcal{F}(t)$; then the first term is zero in expectation and we can use linearity. Since the expectation of $W(t)$ is zero, the result follows. ■

2. **Itō isometry**: $\mathsf{E}\left(I^2(t)\right) = \mathsf{E}\left(\int_0^t \Delta^2(u)\,\mathrm{d}u\right)$
Proof Because $I(t)$ is defined as a sum, we have

$$I^2(t) = \sum_{i=1}^{n}\sum_{j=1}^{n} \Delta(t_{i-1})\Delta(t_{j-1})\left(W(t_i) - W(t_{i-1})\right)\left(W(t_j) - W(t_{j-1})\right)$$

Take expectation as to get (by linearity)

$$\begin{aligned}
\mathsf{E}\left(I^2(t)\right) &= \sum_{i=1}^{n}\sum_{j=1}^{n}\mathsf{E}\left(\Delta(t_{i-1})\Delta(t_{j-1})\left(W(t_i)-W(t_{i-1})\right)\left(W(t_j)-W(t_{j-1})\right)\right)\\
&= \sum_{i=1}^{n}\mathsf{E}\left(\Delta^2(t_{i-1})(W(t_i)-W(t_{i-1}))^2\right)\\
&= \sum_{i=1}^{n}\mathsf{E}\left(\mathsf{E}\left(\Delta^2(t_{i-1})\left(W(t_i)-W(t_{i-1})\right)^2\middle|\mathcal{F}(t_{i-1})\right)\right)\\
&= \sum_{i=1}^{n}\mathsf{E}\left(\Delta^2(t_{i-1})\mathsf{E}\left((W(t_i)-W(t_{i-1}))^2\middle|\mathcal{F}(t_{i-1})\right)\right)\\
&= \sum_{i=1}^{n}\mathsf{E}\left(\Delta^2(t_{i-1})(t_i-t_{i-1})\right)\\
&= \mathsf{E}\left(\int_0^t \Delta^2(u)\,\mathrm{d}u\right)
\end{aligned}$$

which is some form of Parseval identity. Since the cross terms have cross-terms, one of those i, j has to come earlier and iterating on $\mathcal{F}(\min(t_{j-1}, t_{i-1}))$, one of the terms has expectation zero and all off-diagonal terms cancel. ■

3. $I(t) = \int_0^t \Delta(s)\,\mathrm{d}W(s)$ and $\mathrm{d}I(t) = \Delta(t)\,\mathrm{d}W(t)$. The relation

$$\int_a^b \mathrm{d}I(t) = I(b) - I(a) = \int_a^b \Delta(t)\,\mathrm{d}W(t)$$

and $I(a)$ is the constant of integration, which is unknown.[16] If we have an adapted integrand which is not simple, we can approximate (by density) by simple functions and as the mesh gets finer, then $\mathsf{E}\left((\Delta(t)-\Delta_n(t))^2\right) \to 0 \xrightarrow{\text{a.s.}} n \to \infty$ for partition with n points of discontinuity. We will further require $\mathsf{E}\left(\int_0^T \Delta^2(t)\,\mathrm{d}t\right) < \infty$.
4. $I(t)$ is continuous with respect to t;

Proof Obvious; this follows since $W(t)$ is almost surely continuous, since we only care about points of discontinuity (since otherwise the function is piecewise constant). ■

5. $I(t)$ is adapted and $\mathcal{F}(t)$–measurable
6. $I(t)$ is linear
7. The quadratic variation is given by $[I, I]_t = \int_0^t \Delta^2(u)\,\mathrm{d}u$

[16] We need to fix the level of the function we are integrating, so that when we combine with the differential we are integrating, we get the correct answer.

Example 5.1

Consider $\int_0^T W(t)\,\mathrm{d}W(t)$; this is adapted. We will calculate this as

$$\sum_{i=1}^{n} W(t_{i-1})(W(t_i) - W(t_{i-1}))$$

Consider instead

$$\begin{aligned}
&\frac{1}{2}\sum_{i=1}^{n}(W(t_i) - W(t_{i-1}))^2 \\
&= \frac{1}{2}\sum_{i=1}^{n} W^2(t_i) + \frac{1}{2}\sum_{i=1}^{n} W^2(t_{i-1}) - \sum_{i=1}^{n} W(t_i)W(t_{i-1}) \\
&= \frac{1}{2}\sum_{i=1}^{n} W^2(t_i) + \frac{1}{2}\sum_{i=1}^{n-1} W^2(t_i) - \sum_{i=1}^{n} W(t_i)W(t_{i-1}) \\
&= \frac{1}{2}W^2(T) + \sum_{i=1}^{n-1} W^2(t_i) - \sum_{i=1}^{n} W(t_{i-1})\left(W(t_{i-1}) + [W(t_i) - W(t_{i-1})]\right)
\end{aligned}$$

Now we get

$$\sum_{i=1}^{n} W(t_{i-1})(W(t_i) - W(t_{i-1})) = \frac{1}{2}W^2(T) - \frac{1}{2}\sum_{i=1}^{n}(W(t_i) - W(t_{i-1}))^2$$

and so

$$\int_0^T W(t)\,\mathrm{d}W(t) = \frac{1}{2}W^2(T) - \frac{1}{2}T$$

where we have as opposed to regular integral, with

$$\int_0^T g(t)\,\mathrm{d}g(t) = \frac{1}{2}\int_0^T \mathrm{d}g^2(t) = \frac{1}{2}g^2(T)$$

if $g(0) = 0$. The extra $-\frac{1}{2}T$ term comes from the non-zero quadratic variation. We have seen that the Brownian motion accumulated quadratic variation at linear rate. The rules of stochastic calculus we have are thus

$$\begin{aligned}
\mathrm{d}W(t)\,\mathrm{d}W(t) &= \mathrm{d}t \\
\mathrm{d}W(t)\,\mathrm{d}t &= 0 \\
\mathrm{d}t\,\mathrm{d}t &= 0
\end{aligned}$$

We have $\mathrm{d}W^2(t) = 2W(t)\,\mathrm{d}W(t) + \text{term}$ where we want now to think of this term as a Taylor series expansion of the form for f some function

$$f(W(t+s)) = f(W(t)) + f'(W(t))\,\mathrm{d}W(t) + \frac{1}{2}f''(W(t))\,\mathrm{d}W(t)\,\mathrm{d}W(t)$$

stopping knowing the cubic variation of the Brownian motion is zero. Thus

$$W^2(t+s) = W^2(t) + 2W(t)\,\mathrm{d}W(t) + \mathrm{d}W(t)\,\mathrm{d}W(t)$$

and now integrating this from 0 to T, this is

$$W^2(T) - W^2(0) = 2\int_0^T W(t)\,\mathrm{d}W(t) + T$$

There is something in this intuition that we can formalize.

Theorem 5.2 (Itō-Doeblin formula)
If $f(W(t))$ is a measurable function and $f \in \mathcal{C}^2(\mathbb{R})$, then

$$\mathrm{d}f(W(t)) = f'(W(t))\,\mathrm{d}W(t) + \frac{1}{2}f''(W(t))\,\mathrm{d}t$$

using the fact $\mathrm{d}W(t)\,\mathrm{d}W(t) = \mathrm{d}t$.[17] We have furthermore

$$\mathrm{d}f(t, W(t)) = f_t(t, W(t))\,\mathrm{d}t + f_w(t, W(t))\,\mathrm{d}W(t) + \frac{1}{2}f_{ww}(t, W(t))\,\mathrm{d}t$$

proved in exactly the same way.

Definition 5.3 (Itō process)
We define the **Itō process** as

$$X(t) = X(0) + \int_0^t \Delta(u)\,\mathrm{d}W(u) + \int_0^t \Theta(u)\,\mathrm{d}u$$

where $\int_0^t \Delta(u)\,\mathrm{d}W(u) \equiv I(t)$ and $\int_0^t \Theta(u)\,\mathrm{d}u \equiv R(t)$ where $\Theta(u)$ is an adapted process, with regularity conditions $\int_0^t |\Theta(u)|\,\mathrm{d}u < \infty$ and $\mathsf{E}\left(\int_0^t \Delta^2(u)\,\mathrm{d}u\right) < \infty$. The quadratic variation of $X(t)$ is

$$[X, X]_t = \int_0^t \Delta^2(u)\,\mathrm{d}u$$

[17]This is argued by looking simply at sample path, since all we are interested in using Taylor theorem, we require differentiability of f; this corresponds to $\lim_{s\downarrow 0}(f(W(t+s)) - f(W(t)))/s$.

To remember this result, we can rewrite it in differential form, as

$$\begin{aligned}
\mathrm{d}X(t) &= \Delta(t)\,\mathrm{d}W(t) + \Theta(t)\,\mathrm{d}t \\
\mathrm{d}X(t)\,\mathrm{d}X(t) &= \Delta^2(t)\,\mathrm{d}t
\end{aligned}$$

since $R(t)$ is continuous with respect to t. Let $\Gamma(u)$ be an adapted stochastic process. We define

$$\int_0^t \Gamma(u)\,\mathrm{d}X(y) \coloneqq \int_0^t \Gamma(u)\Delta(u)\,\mathrm{d}W(u) + \int_0^t \Gamma(u)\Theta(u)\,\mathrm{d}u$$

$$\begin{aligned}
f(T, X(T)) = f(0, X(0)) + \int_0^T [f_t(t, X(t)) + f_x(t, X(t)) + \Theta(t) + \frac{1}{2} f_{xx}(t, X(t))\Delta^2(t)]\,\mathrm{d}t \\
+ \int_0^T f_x(t, X(t))\Delta(t)\,\mathrm{d}W(t)
\end{aligned}$$

We would expect that

$$\begin{aligned}
\mathrm{d}f(t, X(t)) &= f_t(t, X(t))\,\mathrm{d}t + f_x(t, X(t))\,\mathrm{d}X(t) + \frac{1}{2} f_{xx}(t, X(t))\,\mathrm{d}X(t)\,\mathrm{d}X(t) \\
&= f_t(t, X(t))\,\mathrm{d}t + f_x(t, X(t))(\Delta(t)\,\mathrm{d}W(t) + \Theta(t)\,\mathrm{d}t) + \frac{1}{2} f_{xx}(t, X(t))\Delta^2(t)\,\mathrm{d}t \\
&= (f_t + f_x\Theta(t) + \frac{1}{2} f_{xx}\Delta^2(t))\,\mathrm{d}t + f_x\Delta(t)\,\mathrm{d}W(t)
\end{aligned}$$

which yields the above.

Generalization of the geometric Brownian motion

If we have formally have the differential

$$\begin{aligned}
\mathrm{d}X(t) &= \sigma(t)\,\mathrm{d}W(t) + (\alpha(t) - \frac{1}{2}\sigma^2(t))\,\mathrm{d}t \\
\mathrm{d}X(t)\,\mathrm{d}X(t) &= \sigma^2(t)\,\mathrm{d}t
\end{aligned}$$

Consider now $S(t)$; we promoted $\sigma(t)$ and $\alpha(t)$ as adapted process and now have

$$S(t) = S(0)e^{X(t)}$$

and with $X(0) = 0$, we get

$$\begin{aligned}
\mathrm{d}S(t) &= S(0)\,\mathrm{d}e^{X(t)} \\
&= S(0)\left[e^{X(t)}\,\mathrm{d}X(t) + \frac{1}{2}e^{X(t)}\,\mathrm{d}X(t)\,\mathrm{d}X(t)\right] \\
&= S(0)e^{X(t)}\left[\sigma(t)\,\mathrm{d}W(t) + (\alpha(t) - \frac{1}{2}\sigma^2(t))\,\mathrm{d}t + \frac{1}{2}\sigma^2(t)\,\mathrm{d}t\right] \\
&= S(t)\left[\alpha(t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}W(t)\right]
\end{aligned}$$

and if $\alpha(t)$ was absent, then $S(t)$ being adapted, and $\sigma(t)$ adapted, we would have a martingale. The $\alpha(t)$ term corresponds to a **drift** and $\sigma(t)$ to the **diffusion**. In applications, $\sigma(t)$ could be an Itō process; these two Brownian motions would drive the process; these stochastic volatility models are popular because they fit well the data. If we have correlation between the volatility process $\sigma(t)$ and the price process $\mathrm{d}W(t)$, then we will term the latter occurrence **leverage**.

Almost surely positive (at all times) stocks, it has no jumps and is driven by a single Brownian motion can be modeled with the geometric Brownian motion. The beta is the correlation between Brownian motions describing the price of the various assets in a market.

When we have our deterministic integrand, we will consider the usual Itō integral which is just

$$I(t) = \int_0^t \mathit{\Delta}(s)\,\mathrm{d}W(s)$$

then $I(t) \sim \mathcal{N}(0, \int_0^t \mathit{\Delta}^2(s)\,\mathrm{d}s)$. The moments follow from the Itō isometry property and from usual arguments. To show the distributional claim, we argue by MGF. We have

$$\mathsf{E}\left(\exp(uI(t))\right) \stackrel{?}{=} \exp\left(\frac{1}{2}u^2\int_0^t \mathit{\Delta}^2(s)\,\mathrm{d}s\right)$$

and then

$$\mathsf{E}\left(\exp\left(uI(t) - \frac{1}{2}u^2\int_0^t \mathit{\Delta}^2(s)\,\mathrm{d}s\right)\right) = 1$$

If we can show that $X(t) = \exp\left(uI(t) - \frac{1}{2}u^2\int_0^t \mathit{\Delta}^2(s)\,\mathrm{d}s\right)$ is a martingale with respect to t, then we will get that the expectation is 1 with respect to t. This is always a martingale, but if $\mathit{\Delta}$ is non-random, then we can take it outside the integral. To show that it is a martingale, we use the Itō-Doeblin formula with $\mathrm{d}I(t) = \Delta(t)\,\mathrm{d}W(t)$:

$$\begin{aligned}\mathrm{d}X(t) &= X(t)\left[-\frac{1}{2}u^2\Delta^2(t)\,\mathrm{d}t + u\,\mathrm{d}I(t) + \frac{1}{2}u^2\,\mathrm{d}I(t)\,\mathrm{d}I(t)\right]\\ &= X(t)\left[-\frac{1}{2}u^2\Delta^2(t)\,\mathrm{d}t + u\Delta(t)\,\mathrm{d}W(t) + \frac{1}{2}u^2\Delta^2(t)\,\mathrm{d}t\right]\\ &= X(t)u\Delta(t)\,\mathrm{d}W(t)\end{aligned}$$

and $\Delta(t), X(t)$ being adapted, $W(t)$ being a martingale, then the result is a martingale.

The next example is actually drawn from finance.

Example 5.2 (Vasicek interest-rate model)
We define

$$\mathrm{d}R(t) = (\alpha - \beta R(t))\,\mathrm{d}t + \sigma\,\mathrm{d}W(t)$$

This is (implicitly) the first instance of a **stochastic differential equation** which we encounter. We can draw tools from ODE to get a closed-form solution of the SDE. If $\sigma = 0$, we have

$$\mathrm{d}R = (\alpha - \beta R)\,\mathrm{d}t \Rightarrow \frac{\mathrm{d}R}{\alpha - \beta R} = \mathrm{d}t$$

using separation of variables and integrate, from R_0 to $R(t)$,

$$-\frac{1}{\beta}\log(\alpha - \beta R)\Big|_{R_0}^{R(t)} = t$$

and so

$$\log\left(\frac{\alpha - \beta R(t)}{\alpha - \beta R_0}\right) = -\beta t$$

exponentiate the result

$$(\alpha - \beta R(t))e^{\beta t} = (\alpha - \beta R_0)$$

which implies

$$R_0 = \frac{1}{\beta}\left(\alpha - (\alpha - \beta R)e^{\beta t}\right)$$

and now $X(t) = (\alpha - \beta R)e^{\beta t}$. Now

$$\begin{aligned}\mathrm{d}X &= \alpha\beta(\alpha - \beta R)e^{\beta t}\,\mathrm{d}t - \beta e^{\beta t}\,\mathrm{d}R \\ &= \beta e^{\beta t}\left[(\alpha - \beta R)\,\mathrm{d}t - (\alpha - \beta R)\,\mathrm{d}t - \sigma\,\mathrm{d}W(t)\right]\end{aligned}$$

and

$$\begin{aligned}\mathrm{d}X &= -\beta e^{\beta t}\sigma\,\mathrm{d}W \\ X(t) &= X(0) - \beta\sigma\int_0^t e^{\beta s}\,\mathrm{d}W(s)\end{aligned}$$

and now, since we are interested in $R(t)$ and not $X(t)$, we can replace out the terms to get

$$(\alpha - \beta R)e^{\beta t} - (\alpha - \beta R_0) = -\beta\sigma\int_0^t e^{\beta s}\,\mathrm{d}W(s)$$

which we can simplify to, isolating $R(t)$,

$$R(t) = \frac{\alpha}{\beta} - \frac{1}{\beta}(\alpha - \beta R_0)e^{-\beta t} + \sigma\int_0^t e^{\beta(t-s)}\,\mathrm{d}W(s)$$

which gives as final answer

$$R(t) = e^{-\beta t}R_0 + \frac{\alpha}{\beta}(1 - e^{-\beta t}) + \mathcal{N}\left(0, \sigma^2\int_0^t e^{-2\beta s}\,\mathrm{d}s\right)$$

with variance $\frac{1}{2\beta}\left(1 - e^{-2\beta t}\right)$, so

$$R(t) \xrightarrow{d} \mathcal{N}\left(\frac{\alpha}{\beta}, \frac{1}{2\beta}\right)$$

Example 5.3 (Cox-Ingersoll-Ross interest rate model)
Starting with something which is almost surely positive, this model

$$\mathrm{d}R(t) = (\alpha - \beta R(t))\,\mathrm{d}t + \sigma\sqrt{R(t)}\,\mathrm{d}W(t)$$

which avoids negative, since the randomness vanishes. Unfortunately, the solution will not be available in closed form, but we can nevertheless characterize it.

We will again look at the evolution $e^{\beta t}R(t)$ and so the differential

$$
\begin{aligned}
\mathrm{d}\left(e^{\beta t}R(t)\right) &= e^{\beta t}\left[\beta R(t)\,\mathrm{d}t + (\alpha - \beta R(t))\,\mathrm{d}t + \sigma\sqrt{R(t)}\,\mathrm{d}W(t)\right] \\
\Rightarrow \qquad e^{\beta t}R(t) &= R(0) + \alpha\int_0^t e^{\beta s}\,\mathrm{d}s + \sigma\int_0^t e^{\beta s}\sqrt{R(s)}\,\mathrm{d}W(s)
\end{aligned}
$$

since the equation is linear in $R(t)$, the second derivative vanishes. Instead of deriving the full stochastic solution to the SDE, we will consider the expected value

$$
\mathsf{E}\left(e^{\beta t}R(t)\right) = R(0) + \frac{\alpha}{\beta}\left(e^{\beta t} - 1\right).
$$

and we get

$$
\mathsf{E}\left(R(t)\right) = R(0)e^{-\beta t} + \frac{\alpha}{\beta}\left(1 - e^{-\beta t}\right)
$$

which is a steady state. We see that the process forget where it came from at an exponential rate. We have now for

$$
\begin{aligned}
\mathrm{d}\left(e^{2\beta t}R^2(t)\right) &= e^{2\beta t}\left[2\beta R^2(t)\,\mathrm{d}t + 2R(t)\,\mathrm{d}R(t) + \,\mathrm{d}R(t)\,\mathrm{d}R(t)\right] \\
&= e^{2\beta t}\left[2\beta R^2(t)\,\mathrm{d}t + 2R(t)(\alpha - \beta R)\,\mathrm{d}t + 2R^{\frac{3}{2}}\sigma\,\mathrm{d}W + \sigma^2 R\,\mathrm{d}t\right] \\
&= e^{2\beta t}\left[R(t)(2\alpha + \sigma^2)\,\mathrm{d}t + 2R^{\frac{3}{2}}\sigma\,\mathrm{d}W\right] \\
\Rightarrow \quad e^{2\beta t}R^2(t) &= R^2(0) + (2\alpha + \sigma^2)\int_0^t e^{2\beta s}R(s)\,\mathrm{d}s + 2\sigma\int_0^t e^{2\beta s}R^{\frac{3}{2}}(s)\,\mathrm{d}W(s)
\end{aligned}
$$

and now taking expectation,

$$
e^{2\beta t}\mathsf{E}\left(R^2(t)\right) = R^2(0) + (2\alpha + \sigma^2)\int_0^t e^{2\beta s}\left[R(0)e^{-\beta s} + \frac{\alpha}{\beta}\left(1 - e^{-\beta s}\right)\right]\,\mathrm{d}s
$$

and some painful calculations yields

$$
\mathsf{Var}\left(R(t)\right) = \frac{\sigma^2}{\beta}R(0)\left(e^{-\beta t} - e^{-2\beta t}\right) + \frac{\alpha\sigma^2}{2\beta^2}\left(1 - 2e^{-\beta t} + e^{-2\beta t}\right)
$$

and taking the limit as $t \to \infty$,

$$
\lim_{t\to\infty}\mathsf{Var}\left(R(t)\right) = \frac{\alpha\sigma^2}{2\beta^2}
$$

Black-Scholes-Merton model

We are looking at the price of an asset which satisfies a SDE and the parameters of which are assumed to be constant, of the form

$$\mathrm{d}S(t) = \alpha S\,\mathrm{d}t + \sigma S\,\mathrm{d}W(t)$$

which gives the random time evolution of the asset. We may ask what

$$\begin{aligned}\mathrm{d}\log(S(t)) &= \frac{1}{S(t)}\,\mathrm{d}S(t) - \frac{1}{2S^2(t)}\sigma^2 S^2(t)\,\mathrm{d}t \\ &= \left(\alpha - \frac{1}{2}\sigma^2\right)\,\mathrm{d}t + \sigma\,\mathrm{d}W(t)\end{aligned}$$

is; if we were to exclude the term $\sigma\,\mathrm{d}W(t)$ for the moment, we see that the solution would then be $S(t) = S(0)e^{\left(\alpha - \frac{\sigma^2}{2}\right)t}$. This would yield

$$\begin{aligned}\mathrm{d}\left(e^{\left(\alpha-\frac{\sigma^2}{2}\right)t}S\right) &= e^{\left(\alpha-\frac{\sigma^2}{2}\right)t}\left[-\frac{\sigma^2}{2}S\,\mathrm{d}t + \sigma S\,\mathrm{d}W\right] \\ &= \frac{\sigma^2}{2}e^{\left(\alpha-\frac{\sigma^2}{2}\right)t}S\,\mathrm{d}t + e^{\left(\alpha-\frac{\sigma^2}{2}\right)t}\sigma S\,\mathrm{d}W\end{aligned}$$

We begin by considering the European call, coming at maturity at time T and the payoff at maturity is $(S(T)-K)_+$ for strike price K. Let $S(t)$ be the price of the option, distinguished from the value of a general portfolio generated by holding at each period $\Delta(t)$ units of the underlying asset for unit time t. Consider the differential of the value of the portfolio

$$\mathrm{d}X(t) = \Delta(t)\,\mathrm{d}S(t) + r(X(t) - \Delta(t)S(t))\,\mathrm{d}t \tag{5.2}$$

with X_0 the starting point. This is making absurd assumptions about the market. We are interested mostly in the discounted value

$$\mathrm{d}e^{-rt}X(t) = -re^{-rt}X(t)\,\mathrm{d}t + e^{-rt}\,\mathrm{d}X(t)$$

Let us take (5.2) and substitute it in our geometric Brownian motion model. We then get

$$\begin{aligned}\mathrm{d}X(t) &= \Delta\alpha S\,\mathrm{d}t + \Delta\sigma S\,\mathrm{d}W + rX\,\mathrm{d}t - r\Delta S\,\mathrm{d}t \\ &= \Delta S(\alpha - r)\,\mathrm{d}t + rX\,\mathrm{d}t + \sigma\Delta S\,\mathrm{d}W\end{aligned}$$

and now

$$\begin{aligned}\mathrm{d}e^{-rt}X(t) &= e^{-rt}\left[-rX\,\mathrm{d}t + \mathit{\Delta} S(\alpha - r)\,\mathrm{d}t + rX\,\mathrm{d}t + \sigma \mathit{\Delta} S\,\mathrm{d}W\right] \\ &= e^{-rt}\mathit{\Delta} S\left[(\alpha - r)\,\mathrm{d}t + \sigma\,\mathrm{d}W\right]\end{aligned}$$

a Brownian motion with drift. Suppose some mechanism tells us what the value of the option is at time t and of the price, $c(t,x)$ such that $c(t,S(t))$ is the price of the option at time t conditional on the price at time t for the stock. We want a Markov process property for such function to exist. We will delay this until the next section.[18] We want

$$e^{-rt}X(t) = e^{-rt}c(t,S(t)) \qquad \Rightarrow \qquad \mathrm{d}\left(e^{-rt}X(t)\right) = \mathrm{d}\left(e^{-rt}c(t,S(t))\right)$$

and we also need the initial value condition be satisfied, that is $X(0) = c(0,S(0))$. This then is the principle of no-arbitrage pricing. We will evaluate the right hand side, using Itō-Doeblin, assuming the function $c(t,x)$ for x some price variable

$$\begin{aligned}&e^{-rt}\mathit{\Delta} S\left[(\alpha - r)\,\mathrm{d}t + \sigma\,\mathrm{d}W\right] \\ &\quad = e^{-rt}\left[-rc(t,S)\,\mathrm{d}t + c_t\,\mathrm{d}t + c_x S(\alpha\,\mathrm{d}t + \sigma\,\mathrm{d}W) + \frac{1}{2}c_{xx}S^2\sigma^2\,\mathrm{d}t\right] \\ &\quad = e^{-rt}\left[\mathrm{d}t\left(-rc + c_t + \frac{1}{2}c_{xx}S^2\sigma^2 + \alpha c_x S\right) + \sigma c_x S\,\mathrm{d}W\right]\end{aligned}$$

which implies that $\mathit{\Delta} S = \mathit{\Delta} c_x S$ and $\mathit{\Delta}(t) = c_x(t,S(t))$ which is the **delta-hedging rule**, an extension (analogous) to continuous price of the concept introduced for the Binomial asset pricing model.

We thus want

$$(\alpha - r)c_x s = -rc + c_t + \frac{1}{2}\sigma^2 S^2 c_{xx} + \alpha c_x S$$

and canceling common factors, we get

$$c_t + rxc_x + \frac{1}{2}\sigma^2 x^2 c_{xx} = rc \tag{5.3}$$

which is a second order PDE (partial differential equation), which is backward parabolic equation. This equation, (5.3), is referred to as the **Black-Scholes-Merton equation**. In order to get a unique solution, we need boundary conditions.[19] This has not specified the call price or the nature of the option (call or put); the only condition we required was the Markov property. We have $c(T,x) = (x-K)_+$. If we look at the space for (t,x), we have a

[18] We could have prices being a function of $M(t)$, the maximum and so to get such function $c(t,S(t))$ would be impossible

[19] This requires a 1d object in a two dimensional space, the same way we needed some initial condition and smoothness for an ODE.

line at T and another at $x = 0$. $c(t, 0) = 0$; for our model, recall we had $\mathrm{d}S = S(\alpha\,\mathrm{d}t + \sigma\,\mathrm{d}W)$ to get $c_t(t, 0) = rc(t, 0)$; we have furthermore $c_t(t, 0) = 0$, which is enough conditions for uniqueness of the PDE solution, given by

$$c(t, x) = x\Phi(d_+(T - t, x)) - Ke^{-(T-t)}\Phi(d_-(T - t, x))$$

where

$$d_\pm(\tau, x) = \frac{1}{\sigma\sqrt{\tau}}\left[\log\left(\frac{x}{K}\right) + \left(r \pm \frac{\sigma^2}{2}\right)\tau\right]$$

It is easy to verify for $c(t, 0)$ since x multiplies the first argument and $\log(x) = -\infty$ when $x \to 0^+$. Also, the condition holds for $c(T, x)$; we now have the three partial derivative, the **delta** (c_x), the **theta** (c_t) and the **gamma** (or c_{xx}). Now

$$\frac{\partial}{\partial x}d_\pm(\tau, x) = \frac{1}{\sigma x\sqrt{\tau}}.$$

Note that τ here is interpreted as the **time to maturity**; we get

$$\begin{aligned} c_x &= \Phi(d_+(\tau, x)) + \phi(d_+(\tau, x))\frac{1}{\sigma\sqrt{\tau}} - Ke^{-r\tau}\phi(d_-(\tau, x))\frac{1}{\sigma x\sqrt{\tau}} \\ &= \Phi(d_+(\tau, x)) > 0. \end{aligned}$$

`SOMETHING WRONG IN THE ABOVE`

It will turn out theta is positive and gamma is negative. See Figure 4.5.1 on p.161.

The portfolio is such that $\Delta = c_x$ and this is termed **delta-neutral**. We also say the portfolio, which has convex shape, that gives it the terminology **long-gamma**; it is profitable in times of high stock volatility. The explanation gives intuition about why the PDE might work, economically speaking, for the case of interest.

We will now look at European put option, recall call was to buy the stock at a strike price, while the put is the other way around and to sell at the strike price $(K - S(T))_+$. This leads to **put-call parity**. We introduce a **forward contract** $S(T) - K$; it pays off if the strike price is high then the market price. We use this as intermediate to relate put and call. For arbitrary K, we can relate it to

$$x - K = (x - K)_+ - (K - x)_+$$

which is long one call and short one put. We can thus get the value of the put as a function of the value of the call and the forward contract. The answer is then

$$f(t,x) = x - e^{-r(T-t)}K$$

where again $\tau = T - t$ is the time to maturity. We have $S(0)$ is one unit of the asset. One gets $S(0) - e^{-rT}K$, the latter has to be borrowed on the money market to buy one unit of the asset. What is the hedging portfolio then worth? In the money market, we have $-K$ and one unit of the asset, worth $S(T)$ at maturity; this yields a **static edge**. $f(t,x) = c(t,x) - p(t,x)$. The two, put and call, do not vary independently.

This needs not be confounded with the **forward price**, the price paid now to get later delivery, $e^{-r\tau}S(t)$. This is used for commodities, for example with airlines and fuel; a lot of companies went bankrupt during the oil price shock.

Multivariate Stochastic Calculus

Suppose we start with d independent and identical Brownian motions, We denote

$$\boldsymbol{W}(t) = [W_1(t) \cdots W_d(t)]^\top$$

for this d-dimensional Brownian motion adapted to a filtration $\mathcal{F}(t)$. We will want to discuss the possibility of correlated Brownian motion. Further extensions to the rules are:

$$\mathrm{d}W_i(t)\,\mathrm{d}W_i(t) = \mathrm{d}t, \quad \mathrm{d}W_i(t)\,\mathrm{d}t = 0 \quad \mathrm{d}t\,\mathrm{d}t = 0 \quad \mathrm{d}W_i(t)\,\mathrm{d}W_j(t) = 0$$

for $i \neq j$. This means the cross-quadratic variation $[W_i, W_j](t) = 0$. We thus want to show this to consider a sum over a partition

$$\sum_{k=1}^{n} (W_i(t_k) - W_i(t_{k-1}))(W_j(t_k) - W_j(t_{k-1}))$$

and all increments are independent with mean zero, therefore for any partition are zero. We can reasonably claim then that the expectation is zero, for partition. We now want

$$
\begin{aligned}
&\mathsf{E}\left(\sum_{k=1}^{n}\sum_{l=1}^{n}(W_i(t_k)-W_i(t_{k-1}))(W_i(t_l)-W_i(t_{l-1}))(W_j(t_k)-W_j(t_{k-1}))(W_j(t_l)-W_j(t_{l-1}))\right)\\
&\qquad=\sum_{k=1}^{n}\mathsf{E}\left(((W_i(t_k)-W_i(t_{k-1}))^2\,((W_j(t_k)-W_j(t_{k-1}))^2\right)\\
&\qquad=\sum_{k=1}^{n}(t_k-t_{k-1})^2\\
&\qquad\le\|\Pi\|\sum_{k=1}^{n}(t_k-t_{k-1})\to 0
\end{aligned}
$$

et cetera. We now give a generalization of Itō formula for more than two processes, but we will consider for now the simple case for two process

$$
\begin{aligned}
X(t)&=X(0)+\int_0^t \Theta_1(s)\,\mathrm{d}s+\int_0^t \sigma_{11}(s)\,\mathrm{d}W_1(s)+\int_0^t \sigma_{12}(s)\,\mathrm{d}W_2(s)\\
Y(T)&=Y(0)+\int_0^t \Theta_2(s)\,\mathrm{d}s+\int_0^t \sigma_{21}(s)\,\mathrm{d}W_1(s)+\int_0^t \sigma_{22}(s)\,\mathrm{d}W_2(s)
\end{aligned}
$$

and in differential notation

$$
\begin{aligned}
\mathrm{d}X(t)&=\Theta_1(t)\,\mathrm{d}t+\sigma_{11}(t)\,\mathrm{d}W)1(t)+\sigma_{12}(t)\,\mathrm{d}W_2(t)\\
\mathrm{d}Y(t)&=\Theta_2(t)+\sigma_{21}(t)\,\mathrm{d}W_1(t)+\sigma_{22}(t)\,\mathrm{d}W_2(t).
\end{aligned}
$$

Theorem 5.4 (Itō-Doeblin formula in two variables)

Consider a function $f(t,X(t),Y(t))$ with $f_{xx}, f_{yy}, f_{yx}=f_{xy}$ and f_t all exist. Then

$$
\begin{aligned}
\mathrm{d}f(t,X(t),Y(t))&=f_t\,\mathrm{d}t+f_x\,\mathrm{d}X(t)+f_y\,\mathrm{d}Y(t)\\
&\qquad+\frac{1}{2}f_{xx}\,\mathrm{d}X(t)\,\mathrm{d}X(t)+\frac{1}{2}f_{yy}\,\mathrm{d}Y(t)\,\mathrm{d}Y(t)+f_{xy}\,\mathrm{d}X(t)\,\mathrm{d}Y(t)
\end{aligned}
$$

where

$$
\begin{aligned}
\mathrm{d}X(t)\,\mathrm{d}X(t)&=(\sigma_{11}^2(t)+\sigma_{12}^2(t))\,\mathrm{d}t\\
\mathrm{d}Y(t)\,\mathrm{d}Y(t)&=(\sigma_{21}^2(t)+\sigma_{22}^2(t))\,\mathrm{d}t\\
\mathrm{d}X(t)\,\mathrm{d}Y(t)&=(\sigma_{11}\sigma_{21}+\sigma_{12}\sigma_{22})\,\mathrm{d}t
\end{aligned}
$$

This gives an interesting corollary, taking $f(t,x,y) = xy$. By the previous theorem, taking the corresponding derivatives, this is then

Corollary 5.5 (Itō product rule)

$$\mathrm{d}(X(t)Y(t)) = Y(t)\,\mathrm{d}X(t) + X(t)\,\mathrm{d}Y(t) + \,\mathrm{d}X(t)\,\mathrm{d}Y(t)$$

Recognizing Brownian Motion

This will be used when we will change measure to get a risk-neutral measure, and using it and its properties will give the solution of the Black-Scholes-Merton PDE. Thus a characterization of the Brownian motion will be necessary in the upcoming chapter.

Theorem 5.6 (Lévy)

Let $M(t)$ be a process which is $\mathcal{F}(t)$ measurable and is a martingale relative to the filtration $\mathcal{F}(t)$, $t \geq 0$, such that $\mathsf{E}\left(M(t+s) \mid \mathcal{F}(t)\right) = M(t)$. If furthermore $M(0) = 0, M(t)$ should have almost surely continuous path and its quadratic variation accumulated at rate t, *i.e.* $[M,M](t) = t$. Then $M(t)$ is a Brownian motion.

Note here we do not need to check Gaussianity and it has the correct covariance structure; the last part is easy from the martingale property and the quadratic variation. Remark that the Itō-Doeblin formula does not depend on the Normality; we use Taylor theorem to expand a function and having done so looked at the resulting quadratic variation. The above is thus enough to use Itō-Doeblin formula. We look again at the moment generating function, being interested in $e^{uM(t)-\frac{1}{2}u^2t}$. The fact that this function is a martingale follows from the quadratic variation. Now define $f(t,M) = e^{uM-\frac{1}{2}u^2t}$, so that by Itō-Doeblin formula, we get

$$\begin{aligned}\mathrm{d}f(t,M) &= e^{uM-\frac{1}{2}u^2t}\left[-\frac{u^2}{2}\,\mathrm{d}t + u\,\mathrm{d}M + \frac{1}{2}u^2\,\mathrm{d}t\right] \\ &= e^{uM-\frac{1}{2}u^2t}u\,\mathrm{d}M\end{aligned}$$

and when we integrate this up, we get a martingale whose value at zero is 1, and so this establishes the Normality assumption. We can extend this result to more than one dimension:

Theorem 5.7 (Lévy, two dimensions)

Let $W_1(t)$ and $W_2(t), t \geq 0$ be martingales adapted to the filtration $\mathcal{F}(t)$. Assume that $W_i(0)$=0, $W_i(t)$ has continuous paths and $[W_i, W_i](t) = t$, $\forall\ t \geq 0, i = 1, 2$. If in addition $[W_1, W_2](t) = 0$ for all $t \geq 0$ then $W_1(t)$ and $W_2(t)$ are **independent Brownian motions.**

We need to establish the full covariance structure.

The first application is to correlated stock prices. The model considered in Shreve is

$$\frac{\mathrm{d}S_1(t)}{S_1(t)} = \alpha_1\,\mathrm{d}t + \sigma_1\,\mathrm{d}W(t), \frac{\mathrm{d}S_2(t)}{S_2(t)} \qquad = \alpha_2\,\mathrm{d}t + \sigma_2\left[\rho\,\mathrm{d}W_1(t) + r\,\mathrm{d}W_2(t)\right]$$

the SDE used in Black-Scholes-Merton, and one using correlation Brownian motions, with $r = \sqrt{1-\rho^2}$ as to get $\rho^2 + r^2 = 1$. We have

$$(\rho\,\mathrm{d}W_1 + r\,\mathrm{d}W_2)(\rho\,\mathrm{d}W_1 + r\,\mathrm{d}W_2) = \rho^2\,\mathrm{d}t + r^2\,\mathrm{d}t$$

since the cross-term by assumption, using Itō-Doeblin formula, are zero. We could then express $[\rho\,\mathrm{d}W_1(t) + r\,\mathrm{d}W_2(t)] = \mathrm{d}W_3(t)$. Thus

$$\mathrm{d}W_1\,\mathrm{d}W_3 = \mathrm{d}W_1(\rho\,\mathrm{d}W_1 + r\,\mathrm{d}W_2) = \rho\,\mathrm{d}t$$

In the next chapter, all of these will be time-dependent, so as to get generalized geometric Brownian motions.

Brownian bridge

Definition 5.8 (Brownian bridge)

A **Brownian bridge** is a Gaussian stochastic process denoted B and characterized by the conditions

- $B(0) = B(1) = 0$, zero at the endpoints.
- $\mathsf{Cov}\left(B(t), B(s)\right) = s(1-t)$ for $s \leq t$; notice in passing the variance is lower than the one from Brownian motion $W(t)$.

One can easily build a Brownian bridge from a standard Brownian motion, $B(t) = W(t) - tW(1)$. We could compute as $\mathsf{E}\left(B(t)\right) = 0$ that

$$\begin{aligned}\mathsf{Cov}\left(B(t), B(s)\right) = \mathsf{E}\left(B(t)B(s)\right)&\\ &= \mathsf{E}\left((W(t) - tW(1))(W(s) - sW(1))\right)\\ &= s - s\mathsf{E}\left(W(t)W(1)\right) - t\mathsf{E}\left(W(s)W(1)\right) + ts\mathsf{E}\left(W(1)W(1)\right)\\ &= s - st = s(1-t)\end{aligned}$$

Now, we have $B(t)$ and $W(1)$ are independent; from the multinormality of $W(t)$ and by standard computations

$$\mathsf{Cov}\left(B(t), W(1)\right) = \mathsf{E}\left((W(t) - tW(1))W(1)\right) = t - t = 0$$

What is the interest of this Brownian bridge. Well, we can define an **empirical process**, which will have a Brownian bridge as asymptotic distribution. Indeed, start from an IID sample $X_i, i = 1, \ldots, n$ and consider the empirical distribution function (EDF) of the sample defined to be

$$\widehat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}\left(X_i \leq x\right) \xrightarrow{\text{a.s.}} \mathsf{E}\left(\mathrm{I}\left(X \leq x\right)\right) = \mathsf{P}\left(X \leq x\right) = F(x)$$

holds except at points of discontinuity of F. This result is sometimes referred to in the literature as the fundamental theorem of statistics and establishes consistency via the strong law of large numbers. If we consider the uniform distribution on the unit interval, with CDF given by

$$F(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

We have $n^{\frac{1}{2}}(\widehat{F}(x) - F(x)) \to B(x)$ for $x \in [0,1]$. Write

$$\widehat{F}(x) - F(x) = \frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{I}\left(X_i \leq x\right) - F(x)\right).$$

Consider as usual $0 \leq s \leq t < 1$, as

$$\begin{aligned}
&\mathsf{E}\left((\widehat{F}(t) - F(t))(\widehat{F}(s) - F(s))\right) \\
&\quad= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathsf{E}\left((\mathrm{I}\left(X_i \leq t\right) - F(t))(\mathrm{I}\left(X_j \leq s\right) - F(s))\right) \\
&\quad= \frac{1}{n^2}\sum_{i=1}^{n}\mathsf{E}\left((\mathrm{I}\left(X_i \leq t\right) - t)(\mathrm{I}\left(X_i \leq s\right) - s)\right) \\
&\quad= \frac{1}{n^2}\sum_{i=1}^{n}(s - st - st + st) \\
&\quad= \frac{1}{n}s(1-t)
\end{aligned}$$

which yields correct covariance structure. If we go back to X_i and use the **probability integral transform**, then $F(X_i) \sim \mathcal{U}(0,1)$ provided F is continuous. We can extend the result for continuous random variables and assert that

$$\sqrt{n}(\widehat{F}(x) - F(x)) \to B(F(x))$$

The left hand side is the **empirical process**.

Other properties of the Brownian bridge: we have

$$(1-t)W\left(\frac{t}{1-t}\right) = B(t)$$

and it is only a matter of determining appropriate constants, since W ranges in $[0, \infty]$ while $B(t)$ has support $t \in [0, 1]$; to get covariance. We can invert this relation, set $\frac{t}{1-t} = s$ so $t = \frac{s}{1+s}$ and $(1-t) = (1-s)^{-1}$; thus

$$W(s) = (1+s)B\left(\frac{s}{1+s}\right)$$

What now about simulation? If we were interested in a finite set of values of t, we could use Cramer-Wold device, otherwise we need to use the Functional central limit theorem, as to get

$$\sqrt{n}(\widehat{F}(t) - t) \xrightarrow{d} B(t)$$

one can show that it is impossible to get convergence in probability. If instead of looking at the whole stochastic process, we looked at some functional of it, we can use the result to get convergence in probability.

We can simulate one sample path of the Brownian bridge, and simulate longer and longer paths. Consider a Brownian motion, starting at zero and ending at $W(1)$, known to be Normal, and interpolate linearly. Then, take distribution at $\frac{1}{2}$ and use conditional distribution and multinormality. We know

$$\begin{bmatrix} W\left(\frac{1}{2}\right) \\ W(1) \end{bmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}\right)$$
$$\rho\sigma_1\sigma_2 = \frac{1}{2}, \quad \sigma_1 = \rho = \frac{1}{\sqrt{2}}, \quad \sigma_2 = 1$$

as to get $W_2 \mid W_1 \sim \mathcal{N}(\frac{1}{2}W(1), \frac{1}{4})$ since $\mathsf{E}\left(W_2 \mid W_1\right) = \rho\sigma_x y/\sigma_y$ and similarly for the variance. We linearly interpolate and fill, conditioning successively for points between the different dyadic expansions. Note that the variance decreases as powers of two. This is referred to as **infill asymptotics** or **cts-record**.

Another property, dating back from fractals, of the Brownian motion is that it is **self-similar**. We could look at stretched interval, and zooming in we get exactly same behavior. This procedure would allow to get the asymptotic distribution of say the augmented Dickey-Fuller statistic.

Remark
If we consider Brownian motion $W(t) \mid W(1) = 0$, we get the covariance structure of the Brownian bridge

We had found earlier a joint distribution for the pair $M(t), W(t)$, given by

$$f_{M,W}(m, w, t) = \frac{2(2m-w)}{t\sqrt{2\pi t}} \exp\left(-\frac{(2m-w)^2}{2t}\right)$$
$$f_M(m, t \mid W(t) = w) = \frac{2(2m-w)}{t} \exp\left(-\frac{2m(m-2)}{t}\right)$$

Proposition 5.9 (Kolmogorov-Smirnov test)
This is a measure of goodness of fit, $\widehat{F}(x) - F(x)$, and look at

$$\sup_{x \in [0,1]} \sqrt{n}(\widehat{F}(x) - F(x))$$

or correspondingly

$$\sup_{t \in [0,1]} B(t)$$

which is the Kolmogorov-Smirnov test statistic. We want to find the asymptotic distribution of this. Write now

$$f_M(m, 1 \mid W(1) = 0) = 4m \exp(-2m^2)$$

for $m \geq 0$, we can integrate to get

$$\int_0^\infty f_m(m, 1 \mid W(1) = 0)\,\mathrm{d}m = \int_0^\infty -\frac{\mathrm{d}}{\mathrm{d}m} \exp(-2m^2) = 1$$

so as to verify this is a well-define density on $[0, \infty)$ interval. The CDF is $F(m) = 1 - e^{-2m^2}$, the distribution of the supremum of the KS for a one-sided test. For the two-sided test, however, if we look $\sup_{t \in [0,1]} |B(t)|$, one gets

$$F(m) = 1 - 2\sum_{k=1}^\infty (-1)^{k-1} e^{-2k^2m^2}$$

Section 6
Risk-Neutral measure

We will cover in this section the fundamental theorem of asset pricing; if a risk-measure exists, we can always edge a security (replicate the payoff of the derivative securities using assets). The other theorem says that arbitrage can still be possible even if we edge all securities; we need uniqueness of the risk-neutral measure to guarantee no-arbitrage. We then lead to Girsanov's theorem.

Recall the Radon-Nikodym theorem, which deals with two probability measure, $\mathsf{P}, \widetilde{\mathsf{P}}$ where $\widetilde{p}$ is absolutely continuous under p, *i.e.* $\mathsf{P}(A) = 0 \Rightarrow \widetilde{\mathsf{P}}(A) = 0$. If the other inclusion holds, we say the measures are equivalent. We then have the existence of a measurable function $Z = \frac{d\widetilde{p}}{d\mathsf{P}}$, or $d\widetilde{\mathsf{P}} = Z\, d\mathsf{P}$. If Z is strictly positive, then we have

$$\frac{1}{Z} = \frac{d\mathsf{P}}{d\widetilde{\mathsf{P}}} \qquad\qquad d\mathsf{P} = \frac{1}{Z}\, d\widetilde{\mathsf{P}}.$$

We also have

$$1 = \int_\Omega d\widetilde{\mathsf{P}} = \int_\Omega Z\, d\mathsf{P} \quad \Rightarrow \quad \mathsf{E}(Z) = 1, \widetilde{\mathsf{E}}\left(\frac{1}{Z}\right) = 1$$

If $X \overset{\mathsf{P}}{\sim} \mathcal{N}(0,1)$, then $Z = \exp\left(-\theta X - \frac{1}{2}X^2\right)$, and $Y = X + \theta$, then $Y \overset{\widetilde{\mathsf{P}}}{\sim} \mathcal{N}(0,1)$. This was for a single random variable; we want to extend this to a stochastic process.

Suppose we have a function $Z = Z(1)$ and a filtration $\mathcal{F}(t)$; now $Z(t) = \mathsf{E}(Z \mid \mathcal{F}(t))$. We want Z to be $\mathcal{F}(1)$-measurable. For $s \leq t$, then

$$\mathsf{E}(Z(t) \mid \mathcal{F}(s)) = \mathsf{E}(\mathsf{E}(Z \mid \mathcal{F}(t)) \mid \mathcal{F}(s)) = \mathsf{E}(Z \mid \mathcal{F}(s)) = Z(s)$$

using the definition of Z and the tower property. Recall $\widetilde{\mathsf{E}}(Y) = \int_\Omega Y\, d\widetilde{\mathsf{P}} = \int_\Omega ZY\, d\mathsf{P}$.

Lemma 6.1
Fix t and let Y be $\mathcal{F}(t)$ measurable. Then $\widetilde{\mathsf{E}}(Y) = \mathsf{E}(YZ(t))$.

Proof

$$\widetilde{\mathsf{E}}(Y) = \mathsf{E}(YZ) = \mathsf{E}(Y\mathsf{E}(Z \mid \mathcal{F}(t))) = \mathsf{E}(YZ(t))$$

using iterated expectations and "taking out what is known". ■

Lemma 6.2

Fix $s \leq t$ let Y be $\mathcal{F}(t)$ measurable. Then

$$\widetilde{\mathsf{E}}(Y \mid \mathcal{F}(s)) = \frac{1}{Z(s)} \mathsf{E}\left(YZ(t) \mid \mathcal{F}(s)\right)$$

Proof We need measurability and the above $\widetilde{\mathsf{E}}$ partial averaging property. Measurability for $\mathcal{F}(s)$ is directly satisfied; we then have to show that for $A \in \mathcal{F}(s)$,

$$\begin{aligned}
\int_\Omega \mathrm{I}_A(\omega) ZY \,\mathrm{d}\mathsf{P} &= \int_A YZ \,\mathrm{d}\widetilde{\mathsf{P}} \\
&\stackrel{?}{=} \int_A \widetilde{\mathsf{E}}(Y \mid \mathcal{F}(s)) \,\mathrm{d}\widetilde{\mathsf{P}} \\
&= \int_A \frac{1}{Z(s)} \mathsf{E}\left(YZ(t) \mid \mathcal{F}(s)\right) \,\mathrm{d}\widetilde{\mathsf{P}} \\
&= \int_\Omega \frac{\mathrm{I}_A(\omega)}{Z(s)} \mathsf{E}\left(YZ(t) \mid \mathcal{F}(s)\right) \,\mathrm{d}\widetilde{\mathsf{P}} \\
&= \widetilde{\mathsf{E}}\left(\frac{\mathrm{I}_A(\omega)}{Z(s)} \mathsf{E}\left(YZ(T) \mid \mathcal{F}(s)\right)\right) \\
&= \mathsf{E}\left(\frac{\mathrm{I}_A(\omega)}{Z(s)} Z(s) \mathsf{E}\left(YZ(t) \mid \mathcal{F}(s)\right)\right) \\
&= \mathsf{E}\left(\mathsf{E}\left(\mathrm{I}_A(\omega) YZ(t) \mid \mathcal{F}(s)\right)\right) \\
&= \mathsf{E}\left(\mathrm{I}_A(\omega) YZ(t)\right)
\end{aligned}$$

The first result follows from taking $\mathcal{F}(1)$ since $Z(0) = 1$.

Theorem 6.3 (Girsanov, 1-dimension)

Let $W(t)$ be a Brownian motion on $(\Omega, \mathcal{F}, \mathsf{P})$ and $\mathcal{F}(t)$ be a filtration for $W(t)$. Let $\Theta(t)$ be an adapted process to $\mathcal{F}(t)$ and define[20]

$$\begin{aligned}
Z(t) &= \exp\left(-\int_0^t \Theta(u) \,\mathrm{d}W(u) - \frac{1}{2}\int_0^t \Theta^2(u) \,\mathrm{d}u\right) \\
\widetilde{W}(t) &= W(t) + \int_0^t \Theta(u) \,\mathrm{d}u
\end{aligned}$$

Assume $\Theta(t)$ is square-integrable. We have $\mathsf{E}\left(Z\right) = 1$ and under the probability measure $\widetilde{\mathsf{P}}$, $\widetilde{W}(t)$ is a Brownian motion.

Looking at differential notation, we establish from definition the martingale property. Let

[20]Analogous to $Z = \exp(-\theta X - \frac{1}{2}\theta^2)$ and $Y = X + \theta$

$X(t)$ be the argument in the exponent:

$$\begin{aligned}
\mathrm{d}Z(t) &= Z(t)\,\mathrm{d}X(t) + \frac{1}{2}\,\mathrm{d}X(t)\,\mathrm{d}X(t) \\
\mathrm{d}X(t) &= -\Theta(t)\,\mathrm{d}W(t) - \frac{1}{2}\Theta^2(t)\,\mathrm{d}t \\
\mathrm{d}X(t)\,\mathrm{d}X(t) &= \Theta^2(t)\,\mathrm{d}t
\end{aligned}$$

so we can rewrite

$$\mathrm{d}Z(t) = Z(t)\left(-\Theta\,\mathrm{d}W - \frac{1}{2}\Theta^2(t)\,\mathrm{d}t + \frac{1}{2}\Theta^2(t)\,\mathrm{d}t\right) = -Z(t)\Theta(t)\,\mathrm{d}W(t)$$

and so is a martingale, we get the result for $\mathsf{E}\left(Z(1) \mid \mathcal{F}(t)\right) = \mathsf{E}\left(Z(t)\right)$. For Lévy theorem to apply, we need accumulation of quadratic variation at time t, since $W(t)$ is continuous and the integral is continuous as a function of the limit of integration, $\widetilde{W}(t)$ is also continuous. We want

$$\widetilde{\mathsf{E}}(\widetilde{W}(t) \mid \mathcal{F}(s)) = \widetilde{W}(s) \tag{6.4}$$

Clearly, $\widetilde{W}(t)$ is $\mathcal{F}(t)$ measurable since $\Theta(t)$ is adapted and $W(t)$ is $\mathcal{F}(t)$ measurable. By the previous lemma, we have (6.4) equal to

$$\begin{aligned}
&= \frac{1}{Z(s)}\mathsf{E}\left(\widetilde{W}(t)Z(t) \mid \mathcal{F}(s)\right) \\
&= \mathsf{E}\left(W + \int_0^t \Theta(u)\,\mathrm{d}u \frac{Z(t)}{Z(s)} \;\middle|\; \mathcal{F}(s)\right) \\
&\stackrel{?}{=} W(s) + \int_0^s \Theta(u)\,\mathrm{d}u
\end{aligned}$$

so it suffices to show that the term $\widetilde{W}(t)Z(t)$ is a regular martingale with respect to P. Using Itō product rule, we have

$$\mathrm{d}(\widetilde{W}(t)Z(t)) = \widetilde{W}(t)\,\mathrm{d}Z(t) + Z(t)\,\mathrm{d}\widetilde{W}(t) + \,\mathrm{d}Z(t)\,\mathrm{d}\widetilde{W}(t)$$

and recall that

$$\mathrm{d}\widetilde{W} = \,\mathrm{d}W + \Theta\,\mathrm{d}t \qquad \mathrm{d}Z = -Z\Theta\,\mathrm{d}W$$

and we then get

$$-\widetilde{W}(t)Z(t)\Theta(t)\,\mathrm{d}W(t) + Z(\,\mathrm{d}W(t) + \Theta(t)\,\mathrm{d}t) - Z(t)\Theta(t)\,\mathrm{d}t$$

and so this establishes the martingale property. The quadratic variation has nothing to do with the probability measure, so it is trivial to establish, and $\mathrm{d}\widetilde{W}(t)\,\mathrm{d}\widetilde{W}(t) = \mathrm{d}W(t)\,\mathrm{d}W(t) = \mathrm{d}t$. ■

Example 6.1

We can apply this to the geometric Brownian motion of the form

$$\mathrm{d}S(t) = \alpha(t)S(t)\,\mathrm{d}t + \sigma(t)S(t)\,\mathrm{d}W(t)$$

for a stock price. We introduce the discounted version, the discount process as

$$D(t) = \exp\left(-\int_0^t R(s)\,\mathrm{d}s\right)$$

where $R(s)$ is though of as the instantaneous interest rate (thus $D(t)$ is continuous discounting with time varying interest rate). Notice that $D(t)$ has zero quadratic variation; the differential $\mathrm{d}D(t)$ is

$$\mathrm{d}D(t) = -D(t)R(t)\,\mathrm{d}t$$

What we are interested in is $\mathrm{d}(DS)$; using Itō product rule, this is

$$\begin{aligned}
\mathrm{d}(DS) &= D\,\mathrm{d}S + D\,\mathrm{d}D + \mathrm{d}S\,\mathrm{d}D \\
&= \alpha(t)D(t)S(t)\,\mathrm{d}t + \sigma(t)D(t)S(t)\,\mathrm{d}W(t) - R(t)D(t)S(t)\,\mathrm{d}t \\
&= D(t)S(t)((\alpha(t) - R(t))\,\mathrm{d}t + \sigma\,\mathrm{d}W(t)) \\
&= \sigma DS\left(\frac{\alpha - R}{\sigma}\,\mathrm{d}t + \mathrm{d}W\right) \\
&\equiv \sigma D(t)S(t)\,\mathrm{d}\widetilde{W}(t)
\end{aligned}$$

most of the time, we expect $\alpha - R$ to be positive, so that the return be greater than the standard return on the money market. Obviously, since the quantities is random, nothing could prevent but the average return should be greater.

We can apply Girsanov theorem on the latter expression: recall we have $\mathrm{d}S = \Theta\,\mathrm{d}t + \mathrm{d}W$ and change of measure $\widetilde{W} = W + \Theta$ with associated differential $\mathrm{d}\widetilde{W} = \mathrm{d}W + \Theta\,\mathrm{d}t$. In our example, we set $\Theta = \frac{\alpha - R}{\sigma}$ where we choose wlog $\sigma > 0$; here Θ is termed **market price of risk**.

We are going to look at a portfolio; denote by $X(t)$ the value of the portfolio at any time and the balance from the money market by $\Delta(t)$ (the investment in the risky asset); $\Delta(t)S(t)$ is then the value in the asset. Then

$$\mathrm{d}X(t) = \Delta(t)\,\mathrm{d}S(t) + R(t)(X(t) - \Delta(t)S(t))\,\mathrm{d}t$$

What we will be interested in is once again $\mathrm{d}(D(t)X(t))$, using the product rule and noticing $\mathrm{d}D(t)\,\mathrm{d}X(t) = 0$, we have

$$\begin{aligned}
\mathrm{d}(D(t)X(t)) &= D(t)\,\mathrm{d}X(t) + X(t)\,\mathrm{d}D(t) \\
&= D\Delta\,\mathrm{d}S + DR(X - \Delta S)\,\mathrm{d}t - DRX\,\mathrm{d}t \\
&= \alpha D\Delta S\,\mathrm{d}t + \sigma D\Delta S\,\mathrm{d}W - D\Delta RS\,\mathrm{d}t \\
&= D\Delta S((\alpha - R)\,\mathrm{d}t + \sigma\,\mathrm{d}W) \\
&= \Delta(t)\,\mathrm{d}(D(t)S(t))
\end{aligned}$$

Suppose we have some derivative security $V(T)$ and we want to edge it; we need to determine $X(t), \Delta(t)$ for the hedging portfolio. In order to avoid arbitrage we want $V(t) = X(t)$; we know that under the risk neutral measure , anything involving $\mathrm{d}\widetilde{W}(t)$ is a martingale. We want thus

$$\widetilde{\mathsf{E}}(D(T)V(T)) = V(0) = X(0)$$

whatever the value of $\Delta(t)$ might be. We use the tilde-martingale property to get

$$D(t)V(t) = \widetilde{\mathsf{E}}(D(T)V(T) \mid \mathcal{F}(t))$$

We consider an European call, so $V(T) = (S(T) - K)_+$.[21] Let us now derive the Black-Scholes-Merton equation:

Recall $D(t) = \exp\left(-\int_0^t R(s)\,\mathrm{d}s\right)$. We can write

$$\begin{aligned}
V(t) &= \widetilde{\mathsf{E}}\left(\frac{D(T)}{D(t)}V(T)\middle|\,\mathcal{F}(t)\right) \\
&= \widetilde{\mathsf{E}}\left(\exp\left(-\int_t^T R(s)\,\mathrm{d}s\right)(S(T) - K)_+\,\middle|\,\mathcal{F}(t)\right)
\end{aligned}$$

[21] An American option can be exercised at any time between 0 and T included; this is the most complicated to deal with mathematically, at least for the American put.

Remember what we have

$$
\begin{aligned}
\mathrm{d}S(t) &= RS\,\mathrm{d}t + \sigma S\,\mathrm{d}\widetilde{W} \\
S(t) &= S_0 \exp\left[\int_0^t \sigma(s)\,\mathrm{d}\widetilde{W}(s) + \int_0^t (R(s) - \frac{1}{2}\sigma^2(s)\,\mathrm{d}s\right]
\end{aligned}
$$

so that

$$
\begin{aligned}
\frac{S(T)}{S(t)} &= \exp\left(\int_t^T \sigma(s)\,\mathrm{d}\widetilde{W}(s) + \int_t^T \left(R(s) - \frac{1}{2}\sigma^2(s)\right)\,\mathrm{d}s\right) \\
&= \exp\left(\sigma(\widetilde{W}(T) - \widetilde{W}(t)) + \left(\tau - \frac{1}{2}\sigma^2\right)(T-t)\right)
\end{aligned}
$$

The earlier tilde-expectation may thus be written as

$$
\begin{aligned}
&\widetilde{\mathsf{E}}\left(e^{-r(T-t)}\left(e^{r\tau}S(t)e^{\left(r-\frac{1}{2}\sigma^2\right)(T-t)}\exp\left(\sigma(\widetilde{W}(T)-\widetilde{W}(t))\right) - K\right)_+ \,\middle|\, \mathcal{F}(t)\right) \\
&\quad= \widetilde{\mathsf{E}}\left(\left(e^{r\tau}S(t)e^{-\frac{\sigma^2}{2}(T-t)}e^{\sigma(\widetilde{W}(T)-\widetilde{W}(t))} - K\right)_+ \,\middle|\, \mathcal{F}(t)\right) \\
&\quad= \widetilde{\mathsf{E}}\left(\left(e^{r\tau}xe^{-\frac{\sigma^2\tau}{2}}e^{\sigma\mathcal{N}(0,\tau)} - K\right)_+\right) \\
&\quad= \mathsf{E}\left(\left(e^{r\tau}xe^{-\frac{\sigma^2\tau}{2}}e^{\sigma\mathcal{N}(0,\tau)} - K\right)_+\right) \\
&\quad= \mathsf{E}\left(\left(e^{r\tau}xe^{-\frac{\sigma^2\tau}{2}}e^{\sigma\sqrt{\tau}Y} - K\right)_+\right)
\end{aligned}
$$

using the independence lemma; we have thus demonstrated the existence of the function $c(t,x) \coloneqq \left(e^{r\tau}xe^{-\frac{\sigma^2\tau}{2}}e^{\sigma\sqrt{\tau}Y} - K\right)_+$.[22] Further denote

$$
Y = \frac{\widetilde{W}(T) - \widetilde{W}(t)}{\sqrt{\tau}} \sim \mathcal{N}(0,1)
$$

and before using the density of the standard Normal distribution, we want to determine the

[22] K here is not discounted by convention, while the payoff is.

conditions for which this is positive,[23] that is

$$
\begin{aligned}
& K \leq x e^{\left(r-\frac{\sigma^2}{2}\right)\tau} e^{\sigma\sqrt{\tau} Y} \geq K \\
\Leftrightarrow \quad & Y \geq \frac{1}{\sigma\sqrt{\tau}} \log\left(\frac{K}{x} + \frac{1}{\sigma\sqrt{\tau}}\left(r - \frac{\sigma^2}{2}\right)\tau\right). \\
& \quad = d_-(\tau, x)
\end{aligned}
$$

and

$$
d_\pm(\tau, x) = \frac{1}{\sigma\sqrt{\tau}}\left[\log\left(\frac{x}{K} + \left(r \pm \frac{\sigma^2}{2}\right)\tau\right)\right]
$$

$$
= \frac{1}{\sqrt{2\pi}} \int_{\frac{1}{\sigma\sqrt{\tau}}\log\left(\frac{K}{x} + \frac{\sigma\sqrt{\tau}}{2}\right)}^{\infty} e^{-\frac{y^2}{2}} x
$$

and to complete the derivation, we complete the square to get $-\frac{1}{2}\left(y - \sigma\sqrt{\tau}\right)^2$ making the change of variable $z = y - \sigma\sqrt{\tau}, z = d_+(\tau, x)$

$$
\begin{aligned}
&= \frac{1}{\sqrt{2\pi}} \int_{d_-(\tau,x)}^{\infty} x e^{r\tau} \exp\left(-\frac{1}{2}(y - \sigma\sqrt{\tau})^2\right) \mathrm{d}y \\
&= \frac{1}{\sqrt{2\pi}} \int_{d_+(\tau,x)}^{\infty} x e^{r\tau} e^{-\frac{z^2}{2}} \mathrm{d}z \\
&= \int_{d_+(\tau,x)}^{\infty} x e^{r\tau} \phi(z) \, \mathrm{d}z
\end{aligned}
$$

and we have taken a different road from Shreve; the book final result is

$$
c(t, x) = x\Phi(d_+(\tau, x)) - e^{r\tau} K \Phi(d_-(\tau, x))
$$

Martingale Representation Theorem

This is a restrictive result, yet useful for our needs.

Theorem 6.4 (Martingale representation)

Let $W(t)$ is Brownian motion and corresponding $\mathcal{F}(t)$ filtration,$W(t) \to \mathcal{F}(t)$. Before, we allowed $\mathcal{F}(t)$ to contain more information, yet not to be giving any more information for forecasting. Now, $\mathcal{F}(t)$ is the smallest σ-algebra so that $W(t)$ is adapted. Suppose $M(t)$ is

[23] To compute the conditional expectation –integrating w.r.t y –recall the independence lemma: if we want $\mathsf{E}\left(f(X,Y) \mid \mathcal{F}\right)$ and if $X \sim \mathcal{F}$-measurable and $Y \perp\!\!\!\perp \mathcal{F}$, then the expectation is $g(X)$ where $g(x) = \mathsf{E}\left(f(x,Y)\right)$.

a martingale and is adapted to $\mathcal{F}(t)$. Then there is $\Gamma(t)$ adapted such that

$$M(t) = M(0) + \int_0^t \Gamma(s)\,\mathrm{d}W(s)$$

Corollary 6.5

If $\Theta(t)$ is adapted and we define as usual the Radon-Nikodym derivative process as

$$\begin{aligned} Z(t) &= \exp\left(-\int_0^t \Theta(s)\,\mathrm{d}W(s) - \frac{1}{2}\int_0^t \Theta^2(s)\,\mathrm{d}s\right) \\ \widetilde{W}(t) &= W(t) + \int_0^t \Theta(s)\,\mathrm{d}s \\ \mathrm{d}\widetilde{W}(t) &= \mathrm{d}W(t) + \Theta(t)\,\mathrm{d}t \end{aligned}$$

Then, there must exist an adapted process $\widetilde{\Gamma}(t)$ such that

$$\widetilde{M}(t) = \widetilde{M}(0) + \int_0^t \widetilde{\Gamma}(s)\,\mathrm{d}\widetilde{W}(s)$$

Any portfolio discounted and the portfolio value is also a martingale. We used this for BSM equation. We investigate the details of hedging a portfolio: $R(t), \alpha(t), \sigma(t)$ are adapted if random and we have

$$\frac{\mathrm{d}S}{S} = \alpha\,\mathrm{d}t + \sigma\,\mathrm{d}\widetilde{W}$$

and $D(t)V(t)$ is a $\widetilde{\mathsf{P}}$-martingale. We want to combine this to the fact that and equate drift and diffusion; using the Martingale Representation theorem

$$D(t)V(t) = V(0) + \int_0^t \widetilde{\Gamma}(s)\,\mathrm{d}\widetilde{W}(s)$$

and in differential form, we want $\mathrm{d}(DV) = \widetilde{\Gamma}\,\mathrm{d}\widetilde{W}$. We had earlier $\mathrm{d}(DS) = DS\sigma\,\mathrm{d}\widetilde{W}$ and

$$\mathrm{d}(D(t)X(t)) = \Delta(t)\,\mathrm{d}(DS) = DS\sigma\Delta\,\mathrm{d}\widetilde{W}$$

and so for the hedging, Δ has to satisfy

$$\widetilde{\Gamma} = DS\Delta\sigma \quad \Leftrightarrow \quad \Delta = \frac{\widetilde{\Gamma}}{\sigma DS}$$

provided $\sigma \neq 0$, to define the hedging portfolio (since in such case we have something equivalent to the money market or we have arbitrage). It is also not observed in practical applications. What is $\widetilde{\Gamma}$ will be investigated in the next chapter using PDEs. We first need to extend Girsanov theorem and the Martingale Representation theorem.

Let $\boldsymbol{W}(t)$ be a $d \times 1$ vector of Brownian motion and $\mathcal{F}(t)$ be the filtration generated by the d independent Brownian motion.

Theorem 6.6 (Girsanov)
Let $\boldsymbol{\Theta}(t)$ be $d \times 1$ adapted and

$$\begin{aligned} \mathrm{d}\widetilde{\boldsymbol{W}} &= \mathrm{d}\boldsymbol{W} + \boldsymbol{\Theta}\,\mathrm{d}t \\ \widetilde{\boldsymbol{W}}(t) &= \boldsymbol{W}(t) + \int_0^t \boldsymbol{\Theta}(s)\,\mathrm{d}s \end{aligned}$$

with

$$Z(t) = \exp\left(-\int_0^t \boldsymbol{\Theta}(s)\cdot\,\mathrm{d}\boldsymbol{W}(s) - \frac{1}{2}\int_0^t \|\boldsymbol{\Theta}^2(s)\|^2\,\mathrm{d}s\right)$$

Theorem 6.7 (Martingale Representation)

$$M(t) = M(0) + \sum_{i=1}^{d}\int_0^t \Gamma_i(s)\,\mathrm{d}W(s)$$

and transformed, with $\widetilde{\Gamma}$ adapted to the original filtration $\mathcal{F}(t)$ giving

$$\widetilde{M}(t) = \widetilde{M}(0) + \sum_{i=1}^{d}\int_0^t \widetilde{\Gamma}_i(s)\,\mathrm{d}\widetilde{W}(s)$$

Multidimensional market model

Suppose we have m assets and d Brownian motions. We have

$$\mathrm{d}S_i(t) = \alpha_i(t)S_i(t)\,\mathrm{d}t + S_i(t)\sum_{j=1}^{d}\sigma_{ij}(t)\,\mathrm{d}W_j(t)$$

for $i = 1, \ldots, m$. Then the quadratic variation

$$\mathrm{d}\left(\sum_{j=1}^{d}\sigma_{ij}(t)\,\mathrm{d}W_j(t)\right)\cdot\sum_{k=1}^{d}\sigma_{ij}(t)\,\mathrm{d}W_j(t) = \left(\sum_{j=1}^{d}\sigma_{ij}^2\right)\,\mathrm{d}t$$

and we can define σ_i as

$$\sigma_i := \sqrt{\left(\sum_{j=1}^{d} \sigma_{ij}^2\right)}$$

and let $B_i(t)$ be another Brownian motion. We can write then

$$\mathrm{d}S_i(t) = \alpha_i(t)S_i(t)\,\mathrm{d}t + S_i(t)\sigma_i\,\mathrm{d}B_i(t)$$

so that

$$\mathrm{d}B_i(T) = \frac{\sum_{j=1}^{d} \sigma_{ij}(t)\,\mathrm{d}W_j(t)}{\sigma_i(t)}$$

This allows us to look at the cross quadratic variation

$$\mathrm{d}B_i\,\mathrm{d}B_j = \frac{1}{\sigma_i\sigma_j}\sum_k\sum_l \sigma_{ik}\sigma_{jl}\,\mathrm{d}W_k\,\mathrm{d}W_l$$

and by the extended Itō-Doeblin rule, only when $k = l$ does those differ from zero and we have

$$= \frac{1}{\sigma_i\sigma_j}\sum_k \sigma_{ik}\sigma_{jk}\,\mathrm{d}t \equiv \rho_{ij}(t)$$

for i, j index of assets. In terms of cross quadratic variations of the prices,

$$\mathrm{d}S_i\,\mathrm{d}S_j = S_iS_j\rho_{ij}\sigma_i\sigma_j\,\mathrm{d}t \qquad \Rightarrow \qquad \frac{\mathrm{d}S_i}{S_i}\frac{\mathrm{d}S_j}{S_j} = \rho_{ij}\sigma_i\sigma_j\,\mathrm{d}t$$

where we can think of ρ_{ij} as the instantaneous correlation. This will be the setup of the multidimensional model. Now

$$\begin{aligned}\mathrm{d}(D(t)S_i(t)) &= (\alpha_i - R)DS_i\,\mathrm{d}t + \sigma_i DS_i\,\mathrm{d}B_i \\ &= \sigma_i DS_i\left(\frac{(\alpha_i - R)}{\sigma_i}\,\mathrm{d}t + \,\mathrm{d}B_i\right)\end{aligned}$$

A risk neutral measure will make all discounted prices martingales. If we have such risk neutral measure, we want to have

$$\begin{aligned} \mathrm{d}(D(t)S_i(t)) &= D(t)S_i(t)\sum_{j=1}^{d}\sigma_{ij}(t)\,\mathrm{d}\widetilde{W}_j(t) \\ &= D(t)S_i(t)\sum_{j=1}^{d}\sigma_{ij}(t)(\Theta_j(t)\,\mathrm{d}t + \mathrm{d}W_j(t)) \end{aligned}$$

To get $\boldsymbol{\Theta}$, we write

$$\begin{aligned} \mathrm{d}(DS_i) &= (\alpha_i - R)DS_i\,\mathrm{d}t + DS_i\sum_{j=1}^{d}\sigma_{ij}\,\mathrm{d}W_j \\ &= DS_i\left((\alpha_i - R)\,\mathrm{d}t + \sum_{j=1}^{d}\sigma_{ij}\,\mathrm{d}W_j\right) \\ &= DS_i\sum_{j=1}^{d}\sigma_{ij}\left(\mathrm{d}W_j + \frac{\alpha_i - R}{\sigma_{ij}}\,\mathrm{d}t\right) \end{aligned}$$

so that

$$\sum_{j=1}^{d}\sigma_{ij}\Theta_j = \alpha_i - R$$

a system of linear equation for $i = 1, \ldots, m$; we would thus expect to have as many Brownian motions as there are assets since we have d unknowns and m equations. If $d > m$, we expect an infinity of solutions, if $d < m$, we expect no solution.

If we put more randomness in the model, we have more information than needed to solving the asset process. We don't expect the risk neutral measure to be unique in that, since we can fiddle around to change it on the non-relevant randomness. If $d < m$, we could well have arbitrage.

Suppose in the context of a very simple problem we have two assets and the underlying processes are

$$\begin{aligned} \alpha_1 - r &= \sigma_1\theta \\ \alpha_2 - r &= \sigma_2\theta \end{aligned}$$

is true if and only if

$$\frac{\alpha_1 - r}{\sigma_1} = \frac{\alpha_2 - r}{\sigma_2} = \theta$$

and either they are the same or we have arbitrage, go short on one and long on another and make infinite profit.

Let us look formally at **arbitrage**. Let $X(t)$ be the portfolio value and $X(0) = 0$; we have arbitrage if

$$\mathsf{P}\left(X(T) > 0\right) > 0 \quad \text{and} \quad \mathsf{P}\left(X(T) < 0\right) = 0$$

Theorem 6.8 (First fundamental theorem of asset pricing)
If there exists a risk-neutral measure $\widetilde{\mathsf{P}}$, then there is no arbitrage.

We have by the existence of a risk-neutral measure that

$$\widetilde{\mathsf{P}}(X(T) < 0) = 0$$

since the $\widetilde{\mathsf{P}}$ measure agrees with P for sets of measure zero and

$$\widetilde{\mathsf{E}}(D(T)X(T)) = 0.$$

Indeed,

$$\int D(T)X(T)\,\mathrm{d}\widetilde{\mathsf{P}} = 0$$

and since $D(T)$ is zero. For this to be zero, we need $\widetilde{\mathsf{P}}(X(T) > 0) = 0$ and this implies that $\mathsf{P}\left(X(T) > 0\right) = 0$. This is a contradiction, hence there cannot be such arbitrage.

We have inferred the interest rate process $R(t)$ allows for discounting; with

$$D(t) = \exp\left(-\int_0^t R(s)\,\mathrm{d}s\right).$$

The risk neutral measure has to be such that the discounted price measure $D(t)S_i(t)$ from $i = 1, \ldots, m$, is a martingale, so that

$$\widetilde{\mathsf{E}}(D(t)S_i(t) \mid \mathcal{F}(s)) = D(s)S_i(s)$$

for $i = 1, \ldots, m$ for $s \leq t$. If we consider a portfolio $X(t)$ with $\varDelta_i(t)$ units of asset i; then we see that the process $D(t)X(t)$ is also a martingale. Arbitrage would correspond to having a portfolio where $X(0) = 0$, $\mathsf{P}\,(X(T) < 0) = 0$ and $\mathsf{P}\,(X(T) > 0) > 0$. If $\widetilde{\mathsf{P}}$ is the risk-neutral measure, then $\widetilde{\mathsf{E}}(D(T)X(T)) = \int D(T)X(T)\,\mathrm{d}\widetilde{\mathsf{P}} = 0$ under $\widetilde{\mathsf{P}}$. Since both measures agree and D, X are positive, then

$$\int D(T)X(T)\,\mathrm{d}\widetilde{\mathsf{P}} > 0$$

but then this implies $\widetilde{\mathsf{P}}(D(T)X(T) > 0) = 0$, the $\widetilde{\mathsf{P}}(X(T) > 0) = 0$ and as such $\mathsf{P}\,(X(T) > 0) = 0$.

We can say then that the existence of a risk-neutral measure rules out arbitrage. If we allow out-of-equilibrium events, for short time, they certainly can exist until we are back to economic rules prevail again.

We introduce the notion of hedging; a model is said to be **complete** if every derivative security can be hedged. Consider the process

$$\begin{aligned} \mathrm{d}X(t) &= \sum_{i=1}^{m} \mathit{\Delta}_i(t)\,\mathrm{d}S_i(t) + R(t)\left(X(t) - \sum_{i=1}^{m} \mathit{\Delta}_i(t)S_i(t)\right)\,\mathrm{d}t\,\mathrm{d}(D(t)X(t)) \\ &= D\,\mathrm{d}X + X\,\mathrm{d}D + \,\mathrm{d}X\,\mathrm{d}D \end{aligned}$$

but from before we have that

$$\mathrm{d}D(t) = -D(t)R(t)\,\mathrm{d}t.$$

so we can ignore $\mathrm{d}X\,\mathrm{d}D$. We thus get

$$\begin{aligned} &= D\sum_{i=1}^{m} \mathit{\Delta}_i\,\mathrm{d}S_i + DR\left(X - \sum_{i=1}^{d} \mathit{\Delta}_i S_i\right)\,\mathrm{d}t - XDR\,\mathrm{d}t \\ &= D\sum_{i=1}^{m} \mathit{\Delta}_i\,\mathrm{d}S_i - DR\sum_{i=1}^{d} \mathit{\Delta}_i S_i\,\mathrm{d}t \\ &= \sum_{i=1}^{m} \mathit{\Delta}_i\,\mathrm{d}(DS_i) \end{aligned} \tag{6.5}$$

recalling that

$$\mathrm{d}(DS_i) = D\,\mathrm{d}S_i - DRS_i\,\mathrm{d}t.$$

Since each term in the sum above in (6.5) is a martingale, the term

$$D(t)X(t) = X(0) + \sum_{j=1}^{d} \int_0^t \widetilde{\Gamma}_j(s)\,\mathrm{d}\widetilde{W}_j(s)$$

and in differential notation

$$d(D(t)X(t)) = \sum_j \widetilde{\Gamma}_j(t)\,\mathrm{d}\widetilde{W}_j(t) \sum_i \Delta_i\,\mathrm{d}(DS_i).$$

We now assume in our differential market model, that

$$\mathrm{d}(DS_i) = DS_i \sum_i \sigma_{ij}\,\mathrm{d}\widetilde{W}_j$$

where σ_{ij} could be though of as the "covariance" of the Brownian motions. This tell us that

$$\widetilde{\Gamma}_j = D \sum_i \sigma_{ij} \Delta_i S_i$$

or rearranging

$$\frac{\widetilde{\Gamma}_j}{D} = \sum_i \sigma_{ij} \Delta_i S_i \tag{6.6}$$

where (6.6) are the so-called **hedging equations**.

Theorem 6.9 (Second Fundamental Theorem of Asset Pricing)
Under the assumption of existence of a risk-neutral measure, a market model is complete if and only if the risk neutral measure associated with it is unique.

Assume that we have hedged a market and there are (for a contradiction) associated measures $\widetilde{\mathsf{P}}_1$ and $\widetilde{\mathsf{P}}_2$ are risk-neutral measures. Then, if

$$D(T)V(T) = \mathrm{I}_A$$

for some event A. If we can edge it, then there is $D(t)X(t)$ a martingale under any risk-neutral measure for which

$$\widetilde{\mathsf{E}}_1(\mathrm{I}_A) = X_0 = \widetilde{\mathsf{E}}_2(\mathrm{I}_A)$$

which means since I_A is an indicator that $\widetilde{\mathsf{P}}_1(A) = \widetilde{\mathsf{P}}_2(A)$. Conversely, we want to solve a

$d \times m$ system of equation for arbitrary adapted process $\widetilde{\Gamma}_j$; we had

$$\begin{aligned} \mathrm{d}S_i &= \alpha_i S_i \, \mathrm{d}t + S_i \sum_j \sigma_{ij} \, \mathrm{d}W_j \\ &= R S_i \, \mathrm{d}t + S_i \sum_j \sigma_{ij} \, \mathrm{d}\widetilde{W}_j \end{aligned}$$

and

$$\alpha_i - R = \sum_j \sigma_{ij} \Theta_j$$

Consider now $\mathbf{A}$, the $m \times d$ matrix $[\sigma_{ij}]$ and we want to find

$$\mathbf{A}^\top \begin{bmatrix} \Delta_1 S_1 \\ \vdots \\ \Delta_m S_m \end{bmatrix} = \begin{bmatrix} \widetilde{\Gamma}_i / D \\ \vdots \\ \widetilde{\Gamma}_d / D \end{bmatrix}$$

Then $\mathbf{A}\Theta = \alpha - R$ can be found if $\mathbf{A}$ is non-singular and invertible, we have a unique solution to the system of linear equation and we can solve for arbitrary $\widetilde{\Gamma}_i$.

Dividend paying stocks

We assume that there is a rate A at which dividends are paid, which mean they are paid according to $AS \, \mathrm{d}t$. Here, the essential insight into this analysis is that if we reinvest it into the security to get larger unit of share; this will have under the risk-neutral measure the properties of a martingale. With a little more details, what we see is

$$\begin{aligned} \mathrm{d}S &= -\alpha S \, \mathrm{d}t + \sigma S \, \mathrm{d}W - AS \, \mathrm{d}t \\ \mathrm{d}X &= \Delta \, \mathrm{d}S + \Delta A S \, \mathrm{d}t + R[X - \Delta S] \, \mathrm{d}t \end{aligned}$$

and substituting $\mathrm{d}S$ in $\mathrm{d}X$, one obtains

$$\begin{aligned} &= \Delta \alpha S \, \mathrm{d}t + \Delta \sigma S \, \mathrm{d}W - \Delta A S \, \mathrm{d}t + R[X - \Delta S] \, \mathrm{d}t + \Delta A S \, \mathrm{d}t \\ &= \Delta \alpha S \, \mathrm{d}t + \Delta \sigma S \, \mathrm{d}W + R[X - \Delta S] \, \mathrm{d}t \\ &= \Delta S(\alpha - R) \, \mathrm{d}t + \Delta \sigma S \, \mathrm{d}W + R X \, \mathrm{d}t \end{aligned}$$

and we get as usual

$$\begin{aligned} \mathrm{d}(DX) &= D\Delta S(\alpha - R) \, \mathrm{d}t + DRX \, \mathrm{d}t + D\Delta\sigma S \, \mathrm{d}W - XRD \, \mathrm{d}t \\ &= D\Delta S(\alpha - R) \, \mathrm{d}t + D\Delta\sigma S \, \mathrm{d}W \end{aligned}$$

and setting

$$\Theta = \frac{\alpha - R}{\sigma} \qquad \mathrm{d}\widetilde{W} = \mathrm{d}W + \Theta\,\mathrm{d}t$$

and

$$\mathrm{d}(DX) + D\Delta\sigma S\,\mathrm{d}\widetilde{W}$$

which leads to the formula with a constant

$$c(\tau, x) = xe^{-a\tau}\Phi(d_+(\tau, x)) - e^{-r\tau}K\Phi(d_-(\tau, x))$$

where

$$d_\pm(\tau, x) = \frac{1}{\sigma\sqrt{\tau}}\left[\log\left(\frac{x}{K}\right) + \left(r - a \pm \frac{\sigma^2}{2}\right)\tau\right]$$

Suppose now that we set dividends paid at discrete time, with $0 < t_1 < t_2 < \ldots < t_d < T$ and at time t_j, dividend of $a_j S(t_j^-)$, which is cadlag – we shall also assume it is reinvested in the market. We get

$$S(t_j) - S(t_j^-) - a_j S(t_j^-) = (1 - a_j)S(t_j^-) \tag{6.7}$$

and we assume that in between periods, the stock follows a geometric Brownian motion. Our derivative is

$$\begin{aligned} \mathrm{d}S &= \alpha S\,\mathrm{d}t + \sigma S\,\mathrm{d}W \\ \mathrm{d}X &= RX\,\mathrm{d}t + \Delta\sigma S(\Theta\,\mathrm{d}t + \mathrm{d}W) \end{aligned}$$

Solving with the recurrence relationship with the jump in (6.7), we get

$$S(T) = S(0)\prod_{j=0}^{n-1}(1 - a_{j+1})\exp\left(\sigma\widetilde{W}(T) + \left(r - \frac{\sigma^2}{2}\right)T\right)$$

so the only difference lie in the change from $S(0)$ to $S(0)\prod_{j=0}^{n-1}(1 - a_{j+1})$.

Section 7
Connections with Partial Differential Equations

We will generalize the ideas with Black-Scholes-Merton and somewhat more general SDEs and how they can be linked with PDEs and the associated solutions. We start with a definition

Definition 7.1

A stochastic differential equation (SDE) is an equation satisfying $\mathrm{d}X(u) = \beta(u, X(u))\,\mathrm{d}u + \gamma(u, X(u))\,\mathrm{d}W(u)$

A geometric Brownian motion is clearly an example of the form. Here, we term the function β the drift and γ the diffusion. We are in the presence of a linear SDE if it is of the form

$$\mathrm{d}X(u) = [a(u) + b(u)X(u)]\,\mathrm{d}u + [\gamma(u) + \sigma(u)X(u)]\,\mathrm{d}W(u)$$

which can be solved explicitly if they are one-dimensional if the functions are adapted and a, b, γ, σ are nonrandom.

In integral form, we have

$$X(t) = X(0) + \int_0^t \beta(u, X(u))\,\mathrm{d}u + \int_0^t \gamma(u, X(u))\,\mathrm{d}W(u)$$

to solve numerically, one can use Euler method of discretization to get some tractability with step size

$$X(t+\delta) = X(t) + \beta(t, X(t))\delta + \gamma(t, X(t))\sqrt{\delta}\epsilon_1$$

where $\epsilon_1 \sim \mathcal{N}(0,1)$. This is by no mean the only way to discretize, but it is the easiest for ODEs, PDEs or SDEs. This procedure is recursive; one gets $X(t+\delta)$ and replace $X(t)$; we take an independent standard normal variable there, *i.e.*

$$X(t+2\delta) = X(t+\delta) + \beta(t+\delta, X(t+\delta))\delta + \gamma(t+\delta, X(t+\delta))\sqrt{\delta}\epsilon_2$$

and so on for $\epsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$.

This could be used for **stochastic volatility** models, with

$$\begin{aligned}\mathrm{d}S(t) &= \beta(t, S(t))\,\mathrm{d}t + \sigma(t)\,\mathrm{d}W_S(t)\\ \mathrm{d}\sigma(t) &= \alpha(t, \sigma(t))\,\mathrm{d}t + V(t, \sigma(t))\,\mathrm{d}W_\sigma(t)\end{aligned}$$

where W_S, W_σ are allowed to be correlated; in such case, the correlation is termed the leverage in the model. We could use Euler method to generate that: add the step

$$\sigma(t+\delta) - \sigma(t) = a(t, \sigma(t)) + \delta + V(t, \sigma(t))\sqrt{\delta}\eta_1$$

where η_1 is standard normal.

We want to establish the Markov properties of the solutions of SDEs. If we have a process X governed by a SDE, then

Theorem 7.2

If $X(u), 0 \leq u \leq T$ be a solution to the SDE with given initial conditions. Then

$$g(t,x) = \mathsf{E}^{t,x}(h(X(T)))$$

with $X(t) = x$.

Recall we say $Y(t)$ is a Markov process if $\mathsf{E}\left(h(Y(T)) \mid \mathcal{F}(s)\right) = g(t, Y(s))$ for $s < t$. Once we establish the existence and uniqueness of the solution to the SDE, then the result

$$\mathsf{E}\left(h(X(T)) \mid \mathcal{F}(t)\right) = g(t, X(t))$$

follows from the independence lemma, as

$$X(T) = X(t) + \int_t^T \cdots \,\mathrm{d}t + \int_t^T \cdots \,\mathrm{d}W$$

Theorem 7.3 (Feynman-Kac)

Consider the stochastic differential equation

$$\mathrm{d}X(u) = \beta(u, X(u))\,\mathrm{d}u + \gamma(u, X(u))\,\mathrm{d}W(u)$$

and let the measurable function h such that $g(t,x) = \mathsf{E}^{t,x}(h(X(T))$. [24] Then the function $g(t,x)$ satisfies the partial differential equation

$$g_t(t,x) + \beta(t,x)g_x(t,x) + \frac{1}{2}\gamma^2(t,x)g_{xx}(t,x) = 0$$

and the terminal condition $g(T,x) = h(x)$.

The two physicist after whom the theorem is named have made great contributions.

Moreover, $g(t, X(t))$ is a martingale. We have to show that

[24]The previous theorem gives us the Markov property and the solution to the stochastic differential equation g.

$$\mathsf{E}\left(g(t, X(t) \mid \mathcal{F}(s)\right)$$

Using the Markov property explicitly, we have

$$\mathsf{E}\left(h(X(T)) \mid \mathcal{F}(t)\right) = g(t, X(t))$$

thus

$$\begin{aligned}
\mathsf{E}\left(h(X(T)) \mid \mathcal{F}(s)\right) &= g(s, X(s)) \\
&= \mathsf{E}\left(\mathsf{E}\left(h(X(T)) \mid \mathcal{F}(t)\right) \mid \mathcal{F}(s)\right) \\
&= \mathsf{E}\left(g(t, X(t)) \mid \mathcal{F}(s)\right)
\end{aligned}$$

which is what we wanted to prove. Since it is a martingale, its differential will involve no $\mathrm{d}t$ term. If we look at the differential of g using Itō-Doeblin formula, one gets

$$\begin{aligned}
\mathrm{d}g(t, X(t)) = g_t\,\mathrm{d}t + g_x\,\mathrm{d}X + \frac{1}{2} g_{xx}\,\mathrm{d}X\,\mathrm{d}X \\
&= g_t\,\mathrm{d}t + g_x(\beta\,\mathrm{d}t + \gamma\,\mathrm{d}W) + \frac{1}{2} g_{xx}\gamma^2\,\mathrm{d}t \\
&= (g_t + \beta g_x + \frac{1}{2}\gamma^2 g_x x)\,\mathrm{d}t + \gamma g_x\,\mathrm{d}W
\end{aligned}$$

but the martingale property implies that $(g_t + \beta g_x + \frac{1}{2}\gamma^2 g_x x) = 0$. The function h comes in for the terminal condition $g(T, x) = h(x)$.

If we introduce the following discounted SDE along with the function

$$f(t, x) = \mathsf{E}^{t,x}\left(e^{-r(T-t)} h(X(T))\right)$$

and for the martingale property, we have analogous to the previous theorem

$$\begin{aligned}
\mathsf{E}\left(e^{-rT} h(X(T))\right) = e^{-rt} f(t, X(t)) \\
\mathsf{E}\left(e^{-rT} h(X(T))\right) = e^{-rs} f(t, X(s))
\end{aligned}$$

and we can use the discounted process and iterated conditioning to get as before

$$\mathsf{E}\left(e^{-rt} f(t, X(t)) \mid \mathcal{F}(t)\right) = \mathsf{E}\left(e^{-rs} f(s, X(s)) \mid \mathcal{F}(s)\right)$$

Then, looking at the differential, where $\mathrm{d}X = \beta\,\mathrm{d}t + \gamma\,\mathrm{d}W$, we have

$$\begin{aligned}\mathrm{d}(e^{-rt}f(t,X(t))) &= e^{-rt}\left[(--rf+f_t)\,\mathrm{d}t + f_x\,\mathrm{d}X + \frac{1}{2}f_{xx}\,\mathrm{d}X\,\mathrm{d}X\right] \\ &= e^{-rt}\left[(-rf+f_t+\beta f_x+\frac{1}{2}\gamma^2 f_{xx})\,\mathrm{d}t + f_x\gamma\,\mathrm{d}W\right]\end{aligned}$$

so as to get

$$f_t + \beta f_x + \frac{1}{2}\gamma^2 f_{xx} = rf$$

now that we discount the Feynman-Kac equation, we have an additional term on the right.

Let us work with Black-Scholes-Merton equation: recall we have

$$\begin{aligned}\mathrm{d}S &= \alpha S\,\mathrm{d}t + \sigma S\,\mathrm{d}W \\ \mathrm{d}S &= rS\,\mathrm{d}t + \sigma S\,\mathrm{d}\widetilde{W}\end{aligned}$$

and assume $h(S(T)) = V(T)$ is the payoff. Thus

$$\begin{aligned}V(t) &= \widetilde{\mathsf{E}}\left(e^{-r(T-t)}h(S(T)) \mid \mathcal{F}(t)\right) \\ v(t,S(t)) = V(t) &= v(t,x)\end{aligned}$$

In the context of the discounted Feynman-Kac theorem, we have

$$v_t + rxv_x + \frac{1}{2}\sigma^2 x^2 v_{xx} = rv$$

with terminal condition $v(T,x) = h(x) = (x-K)_+$. For any given strike price, fixing the interest rate we can get a solution and 'fiddle' σ^2 such that it fits the market price of the asset; the resulting $\widehat{\sigma}$ is termed the **implied volatility**, which is a function of the strike price and is typically convex (referred to as the volatility smile) function.

Let $[W_1\ W_2]$ be independent standardized Brownian motions and two processes

$$\begin{aligned}\mathrm{d}X_1 &= \beta_1(t,X_1,X_2)\,\mathrm{d}t + \gamma_{11}(t,X_1,X_2)\,\mathrm{d}W_1 + \gamma_{12}(t,X_1,X_2)\,\mathrm{d}W_2 \\ \mathrm{d}X_2 &= \beta_2\,\mathrm{d}t + \gamma_{21}\,\mathrm{d}W_1 + \gamma_{22}\,\mathrm{d}W_2\end{aligned}$$

Let $h(x_1,x_2)$ be a function for which

$$\begin{aligned}g(t,x_1,x_2) &= \mathsf{E}^{t,x_1,x_2}\left(h(X_1(T),X_2(T))\right) \\ f(t,x_1,x_2) &= \mathsf{E}^{t,x_1,x_2}\left(e^{-r(T-t)}h(X_1(T),X_2(T))\right)\end{aligned}$$

its discounted counterpart. Then

$$g_t + \beta_1 g_{x_1} + \beta_2 g_{x_2} + \frac{1}{2}(\gamma_{11}^2 + \gamma_{12}^2) g_{x_1 x_1} \\ + (\gamma_{11}\gamma_{21} + \gamma_{12}\gamma_{22})\, g_{x_1 x_2} + \frac{1}{2}(\gamma_{21}^2 + \gamma_{22}^2) g_{x_2 x_2} = 0$$

One application of the Feynman-Kac equation is the **Asian option**, which is a call option with payoff at maturity given by

$$V(T) = \left(\frac{1}{T}\int_0^T S(t)\,\mathrm{d}t - K\right)_+$$

which is the average over time. We now have something which depends on the trajectory of the asset price rather than the simple value at T. We can get around with this by introducing a new stochastic process,

$$Y(t) = \int_0^t S(s)\,\mathrm{d}s$$

where now

$$V(T) = \left(\frac{1}{T}Y(T) - K\right)_+$$

Now the pair $S(t), Y(t)$ pair satisfies a system of differential equation

$$\begin{aligned} \mathrm{d}S &= rS\,\mathrm{d}t + \sigma S\,\mathrm{d}\widetilde{W} \\ \mathrm{d}Y &= S\,\mathrm{d}t \end{aligned}$$

which satisfy the requirements for a stochastic differential equation we laid out, so $(S(t), Y(t))$ is a Markov process. Y is not a Markov process; this is what we use in order to price the option. As usual, we write for intermediate time $0 \le t < T$

$$e^{-rt}V(t) = \widetilde{\mathsf{E}}\left(e^{-rT}\left(\frac{1}{T}Y(T) - K\right)_+\right)$$

We assume the existence of a function of the time to maturity $e^{-rt}v(t,x,y)$ where x stands for the asset price and y for the corresponding variable (from the Markov property). Using the two-dimensional Feynman-Kac with $\sigma_{11} = \sigma S$ and all other corresponding terms

$\sigma_{12}, \sigma_{21}, \sigma_{22}$=0, we have

$$\begin{aligned}
\mathrm{d}(e^{-rt}v) &= e^{-rt}\left[(v_t - rv)\,\mathrm{d}t + v_x\,\mathrm{d}S + v_y\,\mathrm{d}Y + \frac{1}{2}v_{xx}\,\mathrm{d}S\,\mathrm{d}S\right] \\
&= e^{-rt}\left[\left(v_t - rv + rxv_x + xv_y + \frac{1}{2}\sigma^2x^2v_{xx}\right)\mathrm{d}t + \sigma x v_x\,\mathrm{d}\widetilde{W}\right]
\end{aligned}$$

so the corresponding PDE is

$$v_t - rv + rxv_x + xv_y + \frac{1}{2}\sigma^2x^2v_{xx} = 0$$

a function of three independent variables. This is very similar to Black-Scholes-Merton except for the additional term; we further require the boundary conditions $v(T,x,y) = \left(\frac{y}{T} - K\right)_+$; we need a two-dimensional cut of the three-dimensional plane specified to get a unique solution.

We next look at interest rate models. We had varying interest rates in previous models, but this time the volatility will involve accumulating quadratic variation and the volatility will be governed by a Brownian motion. Which of those two is more appropriate in the real world is almost an empirical question.

Example 7.1 (Zero-coupon interest rate model)
Let

$$\mathrm{d}R(t) = \beta(t,R(t))\,\mathrm{d}t + \gamma(t,R(t))\,\mathrm{d}\widetilde{W}(t)$$

As usual, we define the discount process to be

$$D(t) = \exp\left(-\int_0^t R(s)\,\mathrm{d}s\right)\quad \mathrm{d}D(t) = -R(t)D(t)\,\mathrm{d}t$$

We will define $B(t,T)$ for a bond, called the zero-coupon bond. This bound only pays at maturity, and we assume for simplicity it delivers one unit of money, so that $B(T,T) = 1$. We will want to discount the bond since it has value before maturity. We will also want $D(t)B(t,T)$ to be a martingale, so that

$$D(t)B(t,T) = \widetilde{\mathsf{E}}\left(D(T) \mid \mathcal{F}(t)\right)$$

We can then write that

$$B(t,T) = \widetilde{\mathsf{E}}\left(\exp\left(-\int_t^T R(s)\,\mathrm{d}s\right)\,\middle|\,\mathcal{F}(t)\right)$$

We may ask the question: what is the value of the discounted bond at T, denoted

$$e^{-Y(T,t)(T-t)} = B(t,T)$$

Here Y is called the **yield**; taking logs, we get

$$Y(T,t) = -\frac{1}{T-t}\log\left(B(t,T)\right)$$

Thus $D(t)f(t,R(t))$ has to be a martingale; such function f exists by the martingale property. We proceed to finding the differential, using Itō product rule

$$\begin{aligned}
\mathrm{d}(D(t), f(t,R(t))) &= D(t)\,\mathrm{d}f(t,R(t)) + f(t,R(t))\,\mathrm{d}D(t) \\
&= D(t)(f_t\,\mathrm{d}t + f_r\,\mathrm{d}R + \frac{1}{2}f_{rr}\,\mathrm{d}R\,\mathrm{d}R) - RDf\,\mathrm{d}t \\
&= D(t)\left[\left(f_t - Rf + f_r\beta + \frac{1}{2}f_{rr}\gamma^2\right)\mathrm{d}t + \gamma f_r\,\mathrm{d}\widetilde{W}\right]
\end{aligned}$$

which once again gives a partial differential equation for f,

$$f_t + f_r\beta + \frac{1}{2}f_{rr}\gamma^2 = Rf$$

and we need a terminal condition, $f(T,r) = 1$ (the payoff at maturity is the value of the bond).

Example 7.2 (Hull-White interest rate)
We have

$$\mathrm{d}R(t) = (a(t) - b(t)R(t))\,\mathrm{d}t + \sigma(t)\,\mathrm{d}\widetilde{W}(t)$$

where a, b, σ are non-random positive adapted functions of time. This corresponds to $\beta \equiv (a(t) - b(t)R(t))$ and $\gamma \equiv \sigma(t)$. We can substitute this into the equation for the zero-coupon model. We have

$$f_t + (a - br)f_r + \frac{1}{2}\sigma^2 f_{rr} = rf$$

We guess the solution has the form

$$e^{-r(T-t)}B(t,T) = f(t,r) = e^{-rC(t,T)-A(t,T)}$$

for non-random C, A functions. Using this and the definition of the yields, we can see that

$$\log(B(t,T)) = -rC(t,T) - A(t,T) + r(T-t)$$

taking logs, in which case the yield

$$Y(T,t) = -r + r\frac{C(t,T)}{T-t} + \frac{A(t,T)}{T-t}$$

which is **affine** in r. All such models are termed **affine yield model**. Dropping the (explicit) dependence on T, we have $f = e^{-rC-A}$, with differentials

$$f_t = f(-rC' - A') \qquad f_r = -fC \qquad f_{rr} = fC^2$$

Since f is positive and proportional to this, we get

$$-rC' - A' - (a - br)C + \frac{1}{2}\sigma^2 C^2 = r$$

or correspondingly

$$-A' - aC + \frac{1}{2}\sigma^2 C^2 = r\left[1 + C' - bC\right]$$

Since this must hold for all values of r, t. Setting $r = 0$, we get that

$$\begin{aligned} -A' - aC + \frac{1}{2}\sigma^2 C^2 &= 0 \\ 1 + C' - bC &= 0 \end{aligned}$$

two ordinary differential equations. We have $C' = b(t)C - 1$, so

$$\frac{\mathrm{d}C}{\mathrm{d}t} - b(t)C = -1$$

with integrating factor $\exp\left(t^T(s)\,\mathrm{d}s\right)$. Now,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\left(C\exp\int_t^T b(s)\,\mathrm{d}s\right) &= \exp\left(\int_t^T b(s)\,\mathrm{d}s\right)\left[\frac{\mathrm{d}C}{\mathrm{d}t} - b(t)C\right] \\ &= -\exp\left(\int_t^T b(s)\,\mathrm{d}s\right) \end{aligned}$$

and so

$$C(t,T) = \int_t^T \exp\left(-\int_t^s b(u)\,\mathrm{d}u\right)$$

Example 7.3

Suppose we have a stationary economic environment with varying interest rate, positive, so that

$$\mathrm{d}R(t) = (a - bR(t))\,\mathrm{d}t + \sigma\sqrt{R(t)}\,\mathrm{d}\widetilde{W}$$

and we get again two ordinary differential equations, with solutions given in the book.

Example 7.4 (Options on a bond)

Let $t \leq T_1 < T_2$ and $B(T_2, T_2) = 1$ and from the Markov property, with maturity at T_1 and then discount onward

$$c(t, R(t)) = \widetilde{\mathsf{E}} = \left(e^{-\int_t^{T_1} R(s)\,\mathrm{d}s} \left(f(T_1, R(T_1)) - K\right)_+ \,\middle|\, \mathcal{F}(t) \right)$$

with $e^{-\int_t^{T_1} R(s)\,\mathrm{d}s} = D(T_1)/D(t)$ and $D(t)C(t, R(t))$ is a martingale. The usual business gives the same partial differential equation as before, given as

$$c_t + \beta(t, r)c_r + \frac{1}{2}\gamma^2(t, r)c_{rr} = rc$$

with this time a different terminal condition $c(T, r) = (f(T, r) - K)_+$.

Section 8
American Derivative Securities

What distinguishes these securities from other is that they can be exercised at any time, worth $(V(t_0)-K)_+$ for call options and similarly for put options. It will turn out that with call option, the American call provides no advantage over the European option. It is advantageous for the put option however. This gives rise to all sorts of complication; if one wishes to hedge a position where one sell short an option, then it is not just having a question of having a portfolio that meets the value of the option at T, if it is desired to hedge the short term. We need to have a portfolio to have a value at least that of the stock.

Will the holder exercise at any particular time; this depends on the person who is long on the option; for the person who is short, this implies more randomness. This means the portfolio may be higher in order for the person to be able to meet its requirement.

This gives rise to the **perpetual American put**, exercised at any time. In order to deal with this maturity, we introduce the concept of a **stopping time**, and we will look at the optimal stopping time (the time at which the holder would exercise the option). Our generic notation will be τ; the event $\tau = t$ as an event is $\mathcal{F}(t)$ measurable. This is not a suitable way to define, however we will require this condition in our definition. We want

$$(\tau \leq t) \in \mathcal{F}(t), \quad \text{and so} \quad (\tau > t) \in \mathcal{F}(t)$$

and the requirement to get

$$(\tau = t) = (\tau \leq t) \cap \left(\bigcap_{n=1}^{\infty} \left(\tau > t - \frac{1}{n} \right) \right)$$

is also in $\mathcal{F}(t)$ by properties of σ-algebras. We came across first-passing time for a Brownian motion, and looked the distribution and the joint distribution of the current value and the maximum. We have

$$\tau_m = \min\{t \geq 0 \mid X(t) = m\}$$

where $X(t)$ is adapted and has continuous path. If $t=0$, then $(\tau \leq t) = (\tau = 0) \in \mathcal{F}_0$. We consider the following event:

$$A = \bigcap_{n=1}^{\infty} \bigcup_{\substack{0 \leq q \leq t \\ q\text{rational}}} \left\{ m - \frac{1}{n} < X() < m + \frac{1}{n} \right\}$$

ad we claim $(\tau_m \leq t) \subset A$, countable additivity gives the first side. On the other hand, A

is non-empty; let $\omega \in A$; then $\forall\, n \in \mathbb{N}, \exists q_n \in \mathbb{Q}, q_n \leq t$ such that

$$m - \frac{1}{n} < X(q_n, \omega) < m + \frac{1}{n}$$

Take a subsequence q_{n^*} with an accumulation point so that $q_{n^*} \to q$, which can't be anything but τ_m; a bounded sequence of real number, given the inequalities above, this must be τ_m. This says that $A \subset (\tau_m \leq t)$.

We now have the following important theorem: let $X(t)$ be a stochastic process and denote the **stopped process** $X(\tau \wedge t)$:

Theorem 8.1 (Optional sampling)
A stopped martingale is a martingale (and accordingly for supermartingale or submartingale).

Perpetual American Put

In this case, the underlying security has a price described in the usual way by a geometric Brownian motion,

$$\mathrm{d}S = rS(t)\,\mathrm{d}t + \sigma S(t)\,\mathrm{d}\widetilde{W}(t)$$

where r, σ are positive constants with payoff $(K - S(t))_+$. It make sense to maximize this quantity over all stopping time $\tau \in \mathcal{T}$,

$$v_*(x) = \max_{\tau \in \mathcal{T}} \widetilde{\mathsf{E}}\left(e^{-r\tau}(K - S(\tau))_+\right)$$

with $S(0) = x$. We want to determine a particular value of $v_*(x)$. As the holder of the option, we need an optimal rule, which is not time dependent. This has to be a rule that is a function of x only. This will be characterized by some set; since it is a put, and it is a decreasing function of the asset price; we should exercise below a threshold. We restrict to stopping times which satisfy the threshold condition L_*, so the maximum over τ_{L_*}. Subsequently, we will then show that the more general problem has to be less than or equal to the one obtained from the restricted problem L_*.

Lemma 8.2 (Laplace transform for first passage time of drifted Brownian motion)
Let W be a Brownian motion, and let $\mu \in \mathbb{R}, m \geq 0$ defining $X(t) = \mu t + W(t)$ and corresponding stopping time τ_m for first passage time at m. We look for

$$m = \mu\tau_m + W(\tau_m).$$

or $W(\tau_m) = m - \mu\tau_m$ which can be interpreted as the first passage to a line with intercept

m and slope μ in the space. If μ is negative, there will be a positive probability of not reaching the slope.

We use the stopped exponential martingale for this purpose: let

$$M(t) = \exp\left[\sigma W(t \wedge \tau_m) - \frac{1}{2}\sigma^2(t \wedge \tau_m)\right]$$

we want σ to be such that taking $t \to \infty$, we get the result we want. We are after $\mathsf{E}\left(e^{-\lambda\tau_m}\right)$ and the solution will be $\sigma = -\mu + \sqrt{\mu^2 + 2\lambda} > 0$ for $\lambda > 0$ or equivalently solving for λ, one has $\mu^2 + 2\lambda = (\mu + \sigma)^2$ thus $\lambda = \mu\sigma + \frac{1}{2}\sigma^2$. We want to look at the MGF and take $t \to \infty$ to recover the value of τ_m:

$$\begin{aligned}\exp\left(\sigma X(t) - \lambda t\right) &= \exp\left(\sigma X(t) - \mu\sigma t - \frac{1}{2}\sigma^2 t\right)\\ &= \exp\left(\sigma W(t) + \mu\sigma t - \mu\sigma t - \frac{1}{2}\sigma^2 t\right)\end{aligned}$$

which goes back to the definition of $M(t)$, except the latter is stopped, yet both are martingales. For $t < \tau_n$, we have $W(t) < m - \mu t$ The expectation of M is $1 = M(0) = \mathsf{E}\left(M(n)\right)$, which is all that is needed. This is

$$\begin{aligned}\mathsf{E}\left(M(n)\right) &= \mathsf{E}\left(\exp\left(\sigma(n \wedge \tau_m) - \frac{1}{2}\sigma^2(n \wedge \tau_m)\right)\right)\\ &= \mathsf{E}\left(\mathrm{I}\left(\tau_m \leq n\right)e^{\sigma\left(m - \mu\tau_m - \frac{1}{2}\sigma^2\tau_m\right)} + \mathrm{I}\left(\tau_m > n\right)e^{\sigma W(n) - \frac{1}{2}\sigma^2 n}\right)\end{aligned}$$

and now we let $n \to \infty$. We have $\mathrm{I}\left(\tau_m > n\right)e^{\sigma W(n)} < \mathrm{I}\left(\tau_m > n\right)e^{\sigma(m - \mu n)}$ Thus

$$\mathsf{E}\left(M(n)\right) = \mathsf{E}\left(\mathrm{I}\left(\tau_m \leq n\right)e^{\sigma(m - \lambda\tau_m)} + \mathrm{I}\left(\tau_m > n\right)e^{\sigma X(n) - \lambda n}\right)$$

where

$$\sigma W(n) - \frac{1}{2}\sigma^2 n = \sigma X(n) - \sigma\mu n - \frac{1}{2}\sigma^2 n$$

and we will bound the second term, let $\lambda \to 0$ to get the probability $\mathsf{P}\left(\tau_m \leq n\right)$.

We have a Brownian motion with drift, of the form $X(t) = \mu t + W(t)$; we are after the first passage time of $m > 0$, labeled τ_m. To know when to exercise the option, we look after the MGF; for $\lambda > 0$, one thus look for $\mathsf{E}\left(e^{-\lambda\tau_m}\right)$. We are looking after a martingale of the form $\exp(\sigma X(t) - \lambda t)$; we require this to involve undrifted Brownian motion:

$$\exp(\sigma X(t) - \lambda t) = \exp(\sigma W(t) - (\lambda - \sigma\mu)t) \tag{8.8}$$

that is if $\lambda - \sigma\mu = \frac{1}{2}\sigma^2$ (an exponential martingale). We need to solve the corresponding equation

$$\sigma^2 + 2\mu\sigma - 2\lambda = 0, \qquad \sigma = -\mu \pm \sqrt{\mu^2 + 2\lambda}$$

and we choose the root $\sigma = -\mu + \sqrt{\mu^2 + 2\lambda}$; with that, we can then assert that (8.8) is a martingale. We consider then

$$\exp\left(\sigma W(t \wedge \tau_m) - (\lambda - \sigma\mu)(t \wedge \tau_m)\right)$$

our stopped martingale. We have

$$\mathsf{E}\left(\exp\left(\sigma W(t \wedge \tau_m) - (\lambda - \sigma\mu)(t \wedge \tau_m)\right)\right) = 1$$

and want to split this when $\min(t, \tau_m) = \tau_m$. Let wlog $t \in \mathbb{Z}^+, t \equiv n$, and consider

$$\mathsf{E}\left(\mathrm{I}\left(\tau_m \leq n\right)\exp\left(\sigma W(\tau_m) - (\lambda - \sigma\mu)\tau_m\right)\right) + \mathsf{E}\left(\mathrm{I}\left(\tau_m > n\right)\exp\left(\sigma W(n) - (\lambda - \sigma\mu)n\right)\right)$$

and since $X(\tau_m) = m$ by definition, then for $t < \tau_m, X(t) < m$ and $W(\tau_m) = m - \mu\tau_m$ thus the term in the exponential is

$$\sigma W(\tau_m) - (\lambda - \sigma\mu)\tau_m = \sigma m - \lambda\tau_m.$$

We thus have

$$e^{\sigma m}\mathsf{E}\left(\mathrm{I}\left(\tau_m \leq n\right)e^{-\lambda\tau_m}\right) + \dagger;$$

where $\dagger$ stands as placeholder for the second term which can be bounded; now $t < \tau_m, X(t) < m$ and thus $W(t) \leq m - \mu t$; this gives

$$\sigma W(n) - (\lambda - \sigma\mu)n \leq \sigma m - \lambda n$$

Thus

$$\dagger \leq \mathsf{E}\left(\mathrm{I}\left(\tau_m > n\right)\right)e^{\sigma m - \lambda n}$$

Let now n (or t) $\to \infty$. Using Dominated convergence or monotone convergence, $e^{-\lambda n} \to 0$ and $\mathrm{I}\left(\tau_m > n\right) \searrow 0$, so we conclude $\mathsf{E}\left(\lim_{n\to\infty}(\cdot)\right) = 0$. We are left with the first term

$$e^{\sigma m}\mathsf{E}\left(\mathrm{I}\left(\tau_m < \infty\right)e^{-\lambda\tau_m}\right).$$

If $\lambda \to 0$, then

$$e^{-\sigma m} = \mathsf{E}\left(\mathrm{I}\left(\tau_m < \infty\right)\right) = \mathsf{P}\left(\tau_m < \infty\right) = \lim_{\lambda \to 0} \exp\left(-(-\mu + \sqrt{\mu^2 + 2\lambda})m\right) = \exp\left(-|\mu| + \mu\right)$$

and we get that if $\mu > 0$, then $\mathsf{P}\left(\tau_m < \infty\right) = 1$, while if $\mu < 0$, then $\mathsf{P}\left(\tau_m < \infty\right) = e^{-2|\mu|} < 1$. In other words, if the barrier is positive but we drift away from it, there is a positive probability of not getting to it. Recall we have

$$\mathsf{E}\left(e^{-\lambda \tau_m}\right) = \mathsf{E}\left(\mathrm{I}\left(\tau_m < \infty\right) e^{-\lambda \tau_m}\right) + \mathsf{E}\left(\mathrm{I}\left(\tau_m = \infty\right) e^{-\lambda \tau_m}\right)$$

using complementary events; the second term is zero if $\tau_m = \infty$. Replacing in λ, we get the MGF

$$\mathsf{E}\left(e^{-\lambda \tau_m}\right) = e^{-\sigma m} = e^{m(\mu - \sqrt{\mu^2 + 2\lambda}}$$

as opposed to what we had before (which was $e^{-m\sqrt{2\lambda}}$).

Let us go back now to the perpetual American put option. If the strike price is greater than the market price, you make profit by exercising (selling) the option (stock). We want to do this optimally: we are after

$$\max_{\tau} \mathsf{E}\left(e^{-r\tau}(K - S(\tau))_+\right) = V(x)$$

with $S(0) = x$. If S is our usual geometric Brownian motion, than this depend on the initial value. Our optimality criterion will be to wait until $S(t) = L < K$ so as to get $e^{-r\tau_L}(K - L)$ in return. In the following, the expectation will always be under the risk-neutral measure and $W(t), \mathsf{E}$ are understood as $\tilde{W}(t), \tilde{\mathsf{E}}$. For the price process

$$\mathrm{d}S = rS \, \mathrm{d}t + \sigma S \, \mathrm{d}W$$

the solution is found to be

$$S(t) = S(0) \exp\left(\sigma W(t) + \left(r - \frac{1}{2}\sigma^2\right) t\right)$$

the discounted value of the exponential martingale. If $S(t) = L$, we get

$$W(t) + \left(r - \frac{1}{2}\sigma^2\right) \frac{t}{\sigma} = -\frac{1}{\sigma} \log\left(\frac{x}{L}\right)$$

and

$$v_L(x) = (K - L)\mathsf{E}\left(e^{-r\tau_L}\right)$$

If we make the change of variable

$$m = \frac{1}{\sigma} \log \left(\frac{x}{L} \right)$$

and so

$$\mu = -\frac{1}{\sigma} \left(r - \frac{1}{2}\sigma^2 \right) t$$

and $\lambda = r$; we get

$$\frac{1}{\sigma}\left(r - \frac{1}{2}\sigma^2\right) + \sqrt{\frac{1}{\sigma^2}\left(r^2 + r\sigma^2 + \frac{1}{4}\sigma^4\right)} = \frac{1}{\sigma}\left(r - \frac{1}{2}\sigma^2\right) + \sqrt{\frac{1}{\sigma^2}\left(r + \frac{1}{2}\sigma^2\right)^2} = \frac{2r}{\sigma}.$$

We then have

$$\mathsf{E}\left(e^{-r\tau_L}\right) = \exp\left(-\frac{2r}{\sigma^2} \log\left(\frac{x}{L}\right)\right) = \left(\frac{x}{L}\right)^{-\frac{2r}{\sigma^2}}$$

and at last, we get

$$v_L(x) = (K - L)\left(\frac{x}{L}\right)^{-\frac{2r}{\sigma^2}}.$$

For more on this, see Shreve; what is done is sensible and illuminating.

We pursue our quest of optimal stopping rule; if $x < L$, then we should exercise immediately and get $K - x$. For $x \geq L$, we have in Figure 8.3.1 a picture of the lines. The condition $K - x$ gives us the first line segment. If the curve v_L is below, then we are wrong (lower value in waiting). As in most cases in economics, we want L^* to be at the point of tangency; any value of L that is bigger or lower is suboptimal, and one can convince oneself that exercise should have been taking place before. The tangency requirement gives us two conditions:

$$v_L(L) = K - L \qquad \text{and} \qquad v_L'(L) = -1$$

as the graph and our line are at the same ordinate at L with the same slope; thus $v_L(L0$ holds automatically. For the second,

$$v_L'(x) = -\frac{2r}{\sigma^2}(K - L)x^{-\frac{2r}{\sigma^2}-1}L^{\frac{2r}{\sigma^2}}$$

and so

$$v_L'(L) = -\frac{2r}{\sigma^2}\frac{(K - L)}{L}$$

from which we get that $\frac{K}{L} = \frac{\sigma^2}{2r} + 1 = \frac{\sigma^2+2r}{2r}$ and

$$L = \frac{2rK}{\sigma^2 + 2r}$$

is our optimal condition.

Because of the time translation invariance, we are allowed to look back at the discounted value $e^{-rt}v_L(S(t))$; we examine its differential using the Itō-Doeblin formula

$$\mathrm{d}(e^{-rt}v_L(S(t))) = e^{-rt}\left[-rv_L(S(t))\,\mathrm{d}t + v_L'(S(t))\,\mathrm{d}S(t) + \frac{1}{2}v_L''(S(t))\,\mathrm{d}S(t)\,\mathrm{d}S(t)\right]$$

where

$$\mathrm{d}S = rS\,\mathrm{d}t + \sigma S\,\mathrm{d}W$$

under the risk neutral measure. Thus $\mathrm{d}(e^{-rt}v_L(S(t)))$ will involve a $\mathrm{d}t$ term. The above becomes

$$= e^{-rt}\left[\left(-rv_L\,\mathrm{d}t + rSv_L' + \frac{v_L''}{2}\sigma^2S^2\right)\,\mathrm{d}t + v_L'\sigma S\,\mathrm{d}W\right]$$

Now

$$\begin{aligned} v_L(x) &= (K-L)\left(\frac{x}{L}\right)^{-\frac{2r}{\sigma^2}} \\ v_L''(x) &= \frac{2r}{\sigma^2}\left(\frac{2r}{\sigma^2}+1\right)(K-L)L^{\frac{2r}{\sigma^2}}x^{-\frac{2r}{\sigma^2}-2}\mathrm{I}\left(x > L\right) \end{aligned}$$

while for $x < L$

$$\begin{aligned} v_L(x) &= K - L \\ v_L'(x) &= -1 \\ v_L''(x) &= 0 \end{aligned}$$

and so

$$-rv_L + rxv_L' = -r(K-x) - rx = -rK$$

which is not zero.

$$
\begin{aligned}
&\text{If } x > L: \qquad rv_L - rxv_L' - \frac{1}{2}\sigma^2 x^2 v_L'' = 0, \quad v > K - x \\
&\text{If } x < L: \qquad rv_L - rxv_L' - \frac{1}{2}\sigma^2 x^2 v_L'' = rK, \quad v = K - x
\end{aligned}
$$

which we refer to as the linear complementary conditions. The term with $\mathrm{d}t$ is a supermartingale, since the term is negative, we have a downward drift. If $e^{-rt}v(S(t))$, which in general is a supermartingale, is stopped, giving $e^{-r(t\wedge\tau_m)}v(S(t\wedge\tau_m))$ – and this is a martingale.

To show that one cannot do any better (no other optimal stopping rule), we use the fact that $e^{-rt}v(S(t))$ is a supermartingale under the risk-neutral measure. For arbitrary τ,

$$
\mathsf{E}\left(e^{-r(t\wedge\tau_m)}v(S(t\wedge\tau_m)) \mid \mathcal{F}_0\right) \leq v_L(S(0))
$$

but we have of course that

$$
\mathsf{E}\left(e^{-r(t\wedge\tau_L)}v(S(t\wedge\tau_L)) \mid \mathcal{F}_0\right) = v_L(S(0))
$$

and so this τ_{L^*} is optimal.

Shreve rightfully so points out that if you are short on these options, you want to hedge using the delta-hedging rule while in the martingale phase – when in phase is where the holder would have exercised and have $v = K - x$, you go short one unit of the asset and have to be able to buy one unit at K; if things have gone too far, sit with the one short option, you can start consuming the interest on the money market after the optimal time - you can consume K at rate r while still being able to meet your obligations. To consider the process more generally, since the only information available is the stock price, consider a stopping set $\mathcal{S}$ and a continuation set $\mathcal{C}$ forming the whole of $\mathbb{R}^+$ with a threshold. Using the time-translation invariance, on the finite horizon model, we have t limited to times $0 \leq t < T$; we get the graph in Figure 8.4.1. We have x, τ and stopped as $x \leq L(\tau)$ – this will be our new stopping rule (which intuitively makes sense- the further from T, the more you are ready to wait). If $x < K$ with the complementary-slackness conditions, we have

$$
rv - v_t - rxv_x - \frac{1}{2}\sigma^2 x^2 v_{xx} = 0
$$

on top for $\mathcal{C}$ with $v(t,x) > K - x$ while underneath the curve $x = L(\tau)$, $v(t,x) = K - x = rK$. We have smoothness in the $\mathcal{C}^1$ sense; the solutions are equal at v_L and the first derivative coincide; this is not the case for higher derivatives. The rest of the details are in the book.

For the American call $(S(t) - K)_+$, it is not optimal to exercise before maturity of the option. Consider a function $h(x)$ which is convex that is

$$h(x_1\lambda + (1-\lambda)x_2) \leq \lambda h(x_1) + (1-\lambda)h(x_2)$$

Since $(x - K)_+$ is convex, if $h(0) = 0$, this tells us that $h(\lambda x) \leq \lambda h(x)$. What we have to consider is the expectation under the risk-neutral measure of the discounted value of exercise.

$$\mathsf{E}\left(e^{-rt}h(S(t)) \mid \mathcal{F}(s)\right) \geq h\left(\mathsf{E}\left(e^{-rt}S(t) \mid \mathcal{F}(s)\right)\right) = h\left(e^{-rs}S(s)\right)$$

for $s < t$ as $e^{-rt}S(t)$ is a martingale. In short,

$$\mathsf{E}\left(e^{-r(t-s)}h(S(t)) \mid \mathcal{F}(s)\right) \geq h(S(s))$$

this process is now a submartingale and will tend to go up, so we have no advantage to exercise early. The only exception to this rule will be when dividend (whether in discrete or continuous time) are paid, since the value of the underlying asset go down with the payment (it is the person hedging (not the holder) who would be getting the dividend). Then, it would be optimal to exercise just before the dividend payment.

Section 9
Introduction to Jump Processes

We will be interested mostly in Poisson processes, which will be our new source of randomness. A jump process is one in which jumps happen at a random time. The waiting time is governed by the exponential distribution

$$F(t) = 1 - e^{-t}$$

or introducing a scale parameter $F(t) = 1 - e^{-t/s}$. Indeed, if $X \sim \mathcal{E}(1)$, then sX will have expectation $\mathsf{E}\,(sX) = s$. We could also characterize the distribution with a rate parameter λ, in which case $F(t) = 1 - e^{-\lambda t}$. For the $\mathcal{E}(1)$ distribution, we have $f(t) = e^{-t}$ for $t \geq 0$. One can easily verify that $\int_0^\infty e^{-t}\,\mathrm{d}t = 1$ with corresponding expectation

$$\int_0^\infty te^{-t}\,\mathrm{d}t = -\int_0^\infty t\,\mathrm{d}(e^{-t}) = -te^{-t}\Big|_0^\infty + \int_0^\infty e^{-t} = 1$$

integrating by parts and the second moment

$$\int_0^\infty t^2e^{-t}\,\mathrm{d}t = -\int_0^\infty t^2\,\mathrm{d}(e^{-t}) = 2\int_0^\infty e^{-t}t\,\mathrm{d}t = 2$$

or correspondingly $\frac{2}{\lambda^2}$ for rate λ, $2s^2$ if we have a scale parameter s. The exponential distribution, as mentioned earlier, is used for waiting time; jump between exponential time amount. The sum S_n is the time before n jumps,

$$S_n = \sum_{i=1}^n \tau_i, \quad \tau_i \sim \mathcal{E}(1)$$

where $\tau_i \perp\!\!\!\perp \tau_j$ are independent. The sum of exponential distribution follows a Gamma distribution. $Y \sim \gamma(a)$ has a PDF of the form

$$f(y) = \frac{1}{\Gamma(a)} y^{a-1} e^{-y}$$

where

$$\Gamma(z) = \int_0^\infty e^{-y} y^{z-1}\,\mathrm{d}y$$

is the gamma function; it has a set of poles at negative integers and $\Gamma(-n) = \infty$; $\Gamma(1) = \int_0^\infty e^{-y}\,\mathrm{d}y = 1$ and for positive integers $\Gamma(n) = (n-1)!$. The Gamma function is nonelementary. We can also derive an interesting recurrence relation, for $\Gamma(n+1)$ where $n \in \mathbb{N}$

$$\begin{aligned}\Gamma(n+1) &= \int_0^\infty e^{-y} y^n \,\mathrm{d}y \\ &= e^{-y} \int_0^\infty e^{-y} \,\mathrm{d}(y^n) \\ &= n \int_0^\infty e^{-y} y^{n-1} \,\mathrm{d}y = n\Gamma(n).\end{aligned}$$

We now derive the moment generating function of τ, which is then $\mathsf{E}\left(e^{-u\tau}\right) = int_0^\infty e^{-u\tau} e^{-\tau} \,\mathrm{d}\tau = (1+u)^{-1}$. The moments are of the form

$$\frac{1}{\Gamma(n)} \int_0^\infty e^{-y} y^{n+k-1} \,\mathrm{d}y = \frac{\Gamma(n+k)}{\Gamma(n)}.$$

For the sum, we have the MGF is $\mathsf{E}\left(e^{-uS_n}\right) = (1+u)^{-n}$; the Gamma moment generating function is on the other hand

$$\frac{1}{\Gamma(n)} \int_0^\infty e^{-uy} y^{n-1} \,\mathrm{d}y = \frac{1}{\Gamma(n)} \int_0^\infty e^{-z} \frac{z^{n-1}}{(1+u)^n} \,\mathrm{d}y = \left(\frac{1}{1+u}\right)^n$$

making the change of variable $z = (1+u)y, \ \mathrm{d}y = \frac{\mathrm{d}z}{1+u}$.

We can easily generate random numbers for integer value using the probability integral transform for $F(t) = 1 - e^{-t}$, we have $F^{-1}(p) = -\log(1-p)$ and we can take sum of independent draws for the γ distribution.

Poisson process

We now define the **Poisson process** corresponding to the number of jumps before or at time t as

$$N(t) = \begin{cases} 0 & \text{if } t < s_1 \\ 1 & \text{if } s_1 \leq t < s_2 \end{cases}$$

and the correspondence $\{N(t) \geq n\} \Leftrightarrow \{S_n \leq t\}$. We thus have

$$\begin{aligned}\mathsf{P}\left(N(\tau) \geq n\right) &= \mathsf{P}\left(S_n \leq t\right) \\ &= \frac{1}{\Gamma(n)} \int_0^t e^{-s_n} (s_n)^{n-1} \,\mathrm{d}s_n\end{aligned}$$

which is the incomplete gamma function. Since $N(t)$ is discrete, we want a PMF taking values $0, 1, \ldots, n$. We have

$$\mathsf{P}\left(N(t) \geq n\right) - \mathsf{P}\left(N(t) \geq n+1\right) = \frac{1}{\Gamma(n)} e^{-s} s^{n-1} \,\mathrm{d}s - \frac{1}{\Gamma(n+1)} \int_0^t e^{-s} s^n \,\mathrm{d}s$$

and once again, integration by parts saves the day. We take the second term with

$$-\frac{1}{\Gamma(n+1)} \int_0^t s^n \,\mathrm{d}(e^{-s}) = -\frac{t^n e^{-t}}{n!} + \frac{1}{\Gamma(n+1)} \int_0^t e^{-s} \,\mathrm{d}(s^n)$$

which gives

$$\frac{1}{\Gamma(n)} e^{-s} s^{n-1} \,\mathrm{d}s - \frac{1}{\Gamma(n)} e^{-s} s^{n-1} \,\mathrm{d}s + \frac{t^n e^{-t}}{n!}$$

so

$$\mathsf{P}\left(N(t) = n\right) = \frac{t^n e^{-t}}{n!}$$

and

$$\sum_{n=0}^{\infty} \mathsf{P}\left(N(t) = n\right) = e^{-t} \sum_{n=0}^{\infty} \frac{t^n}{n!} = 1$$

–we will use this as the base distribution to characterize our process. We will be interested in the survival function

$$\mathsf{P}\left(\tau > t \mid \tau > s\right) = \frac{\mathsf{P}\left((\tau > t) \wedge (\tau > s)\right)}{\mathsf{P}\left(\tau > s\right)} = e^{-(t-s)}$$

which is the same distribution as if the process started at time s; this is the memoryless property of the exponential distribution. We are now set to extend stochastic calculus to jump processes.

If we looked at $N(t) \sim \mathcal{P}(t)$ corresponding to the number of jumps before time t, so

$$\mathsf{P}\left(N(t) = n\right) = \frac{e^{-t} t^n}{n!}$$

If we look at this as a function of t, then we will have to deal with covariance and correlation. However, the memoryless property gives us independent increments, that is for $s \leq t$, $N(t) - N(s) \perp\!\!\!\perp N(s)$ and $N(t) - N(s) \stackrel{d}{=} N(t-s)$. This allows us to work out little bits and pieces. We can think of a filtration $\mathcal{F}(t)$, such that $N(t)$ is $\mathcal{F}(t)$-measurable and $N(t+s) - N(t) \perp\!\!\!\perp \mathcal{F}(t)$, analogous to the properties we have for Brownian motion.

We will want to have evolution of price governed by both the path of a Brownian motion and the jump process, both adapted to the filtration $\mathcal{F}(t)$; in that case, the Brownian motion and jump processes have to be independent. Another thing that we find is that we can extend the Itō-Doeblin formula and deal with stochastic differential equation akin to what we had before.

Consider as a generalization an extension to $\mathcal{P}(\lambda t)$ random variables $N(t)$. This will give a Poisson process of intensity λ. If λ is large, the expectation is very small and vice-versa. The density becomes

$$\mathsf{P}\left(N(t) = n\right) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}$$

which could just be a redefinition of time (used in financial markets and sometimes called θ time, so to have constant transaction occurrence). It makes sense to have a martingale. Consider $N(t) - \lambda t = M(t)$. It is obvious that $\mathsf{E}\left(M(t)\right) = 0$ and the martingale property follows from the independent increment property

$$\mathsf{E}\left(M(t+s) \mid \mathcal{F}(t)\right) = M(t).$$

We next define compound process, since we do not expect to have jump of the same size in the same direction continuously happening. Instead, if the jumps have a particular distribution, which are IID generalization of a distribution; let Y_i be the i^{th} jump– when $N(t) = i$. We define the **compound Poisson process**

$$Q(t) = \sum_{i=1}^{N(t)} Y_i$$

and $N \perp\!\!\!\perp Y_i$, where N stands for the whole process. More generally, we would be interested in $\mathsf{E}\left(Q(t)\right) = \beta N(t)$, if $\mathsf{E}\left(Y_i\right) = \beta$. Also, it is clear that $Q(t) - \lambda\beta t$ is a martingale. The latter is called **compensated Poisson process**. The moment generating function is

$$\begin{aligned}
\mathsf{E}\left(e^{uQ(t)}\right) &= \mathsf{E}\left(\exp\left(u\sum_{i=1}^{N(t)} Y_i\right)\right) \\
&= \mathsf{E}\left(\prod_{i=1}^{N(t)} e^{uY_i}\right) \\
&= \mathsf{E}\left((M_Y(u))^{N(t)}\right) \\
&= \sum_{n=0}^{\infty} M_Y^n(u) e^{-\lambda t}\frac{(\lambda t)^n}{n!} \\
&= e^{-\lambda t} e^{\lambda t M_Y(u)} \\
&= e^{\lambda t (M_Y(u)-1)}
\end{aligned}$$

in the compound case if we let $M_Y(u) := \mathsf{E}\left(e^{uY_i}\right)$ and for simple $\mathcal{P}(\lambda t)$ random variable, the MGF is

$$M(u) = e^{\lambda t(e^u - 1)}.$$

If the possible jumps are in the set $y_1, \ldots, y_M$ with $p(y_m)$ for $m = 1, \ldots, M$ corresponding probability such that $\sum_{i=1}^{M} p(y_m) = 1$, then we can say that the total number of jumps is

$$\begin{aligned}
N(t) &= \sum_{m=1}^{M} N_m(t) \\
Q(t) &= \sum_{i=1}^{N(t)} Y_i = \sum_{m=1}^{M} y_m N_m(t)
\end{aligned}$$

First, if $Q(t)$ is as above, a combination of simple Poisson process, the MGF will be

$$M_Y(u) = \mathsf{E}\left(e^{uY}\right) = \sum_{m=1}^{M} p(y_m) e^{uy_m}$$

and we get

$$\begin{aligned}
\exp(-\lambda t)\exp\left(\sum_{m=1}^{M} p(y_m)e^{uy_m}\right) &= e^{-\lambda t}\prod_{m=1}^{M}\exp\left(\lambda t p(y_m) e^{uy_m}\right) \\
&= \prod_{m=1}^{M}\exp\left((\lambda p(y_m) t(e^{uy_m} - 1)\right)
\end{aligned}$$

since the exponential of sums is the product of exponentials; for later use take $\lambda_m \equiv \lambda p(y_m)$. Suppose we start with M different processes with intensity λ_m. We have then $\overline{N}_m(t) \sim \mathcal{P}(\lambda_m)$ mutually independent. We define

$$\overline{Q}(t) = \sum_{m=1}^{M} y_m \overline{N}_m(t).$$

Once again, we have

$$\begin{aligned}\mathsf{E}\left(e^{u\overline{Q}(t)}\right) &= \prod_{m=1}^{M} \mathsf{E}\left(e^{uy_m}\overline{N}_m(t)\right)\\ &= \prod_{m=1}^{M} \exp\left(\lambda_m t(e^{uy_m} - 1)\right)\end{aligned}$$

Earlier, we worked out integrals of adapted functions and Brownian motion W. We introduce J, arbitrary jump processes until we have specific models using compound or simple Poisson processes. Before that, we have few definitions and properties which follow from those.

Consider a stochastic process

$$X(t) = X(0) + \int_0^t \mathit{\Theta}(s)\,\mathrm{d}s + \int_0^t \mathit{\Gamma}(s)\,\mathrm{d}W(s) + J(t),$$

$\mathit{\Theta}, \mathit{\Gamma}$ adapted to $\sigma(W(t))$; we will want regularity conditions such that $J(t)$ is also adapted. We require that in any finite time interval, the number of jumps is finite – we rule out the case of singular distribution, since having explosion is not plausible from a financial point of view. Let

$$X(t) = X^c(t) + J(t)$$

a pure jump part and a piecewise constant part, which moves only at jump times. We now consider integrals:

$$\int_0^t \mathit{\Phi}(s)\,\mathrm{d}X(s) = \int_0^t \mathit{\Phi}(s)\,\mathrm{d}X^c(s) + \sum_{s<t} \mathit{\Phi}(s)\Delta J(s)$$

where

$$\Delta J(s) = J(s) - J(s-)$$

so that J is right-continuous and cadlag. For the quadratic variation, first look at the continuous part, viz

$$[X^c, X^c](t) = \int_0^t \Gamma^2(s)\,\mathrm{d}s$$

so $\mathrm{d}X^c\,\mathrm{d}X^c = \Gamma^2\,\mathrm{d}t$ and we would interpret as a Stieltjes integral

$$\Phi(t)\,\mathrm{d}X(t) = \Phi(t)\,\mathrm{d}X^c(t) + \Phi(t)\,\mathrm{d}J(t)$$

Unless we take particular precaution, even for adapted integrand, the general $X(t)$ process involving a jump part will not be a martingale, as the following example illustrates.

Example 9.1

Let $M(t) = M(t) - \lambda t$ again the compensated Poisson process and $\Phi(s) = \Delta N(s)$ and consider $\int_0^t \Phi(s)\,\mathrm{d}M(s)$.

$$\int_0^t \Phi(s)\,\mathrm{d}M(s) = -\lambda \int_0^t \Phi(s)\,\mathrm{d}s + (\Delta N(s))^2 = N(t)$$

as $\Delta N(s) \leq 1$, so we need that $\Delta N(0) = 0$, so there is no jump at time 0. We do not have a martingale; by the first Fundamental theorem of asset pricing, if we find the risk-neutral measure for $M(t)$, we could rule out arbitrage which is happening in this example. To get out with adaptivity, we will require that the integrand be left-continuous and the integrator is right-continuous. We have $\Phi(s-)\,\mathrm{d}M(s)$. In such case, if we integrate with respect to a martingale, then we get back a martingale. We could use predictable (that is require unpredictability of the integrand, which is a weaker condition than the left-continuity.)

Let $\Phi(s) = \mathrm{I}\,(s \leq S_1)$, for S_1 the first jump time. Then, we get a martingale integrating against $M(s)$.

Now that we established cross quadratic-variation between jump processes and Brownian motion, which will turn out to be zero. If $X_k(t) = X_k^c(t) + J_k(t)$; we want

$$\begin{aligned}
&\lim_{\|\Pi\|\to 0} \left| \sum_{i=1}^n (X_1(t_i) - X_1(t_{i-1}))(X_2(t_i) - X_2(t_{i-1})) \right| \\
&\quad = \lim_{\|\Pi\|\to 0} \left| \sum_{i=1}^n (X_1^c(t_i) - X_1^c(t_{i-1}) + J_1(t_i) - J_1(t_{i-1}))(X_2^c(t_i) - X_2^c(t_{i-1}) + J_2(t_i) - J_2(t_{i-1})) \right| \\
&\quad = \lim_{\|\Pi\|\to 0} \left| \sum_{i=1}^n (X_1^c(t_i) - X_1^c(t_{i-1}))(X_2^c(t_i) - X_2^c(t_{i-1})) + [\cdots] \right|
\end{aligned}$$

where the first term corresponds to $[X_1^c, X_2^c](t)$. Using the triangle inequality and looking now at $[\cdots]$, involving the cross-term, we have

$$\begin{aligned}
&\sum_{i=1}^{n} \left|\left(X_1^c(t_i) - X_1^c(t_{i-1})\right)\left(J_2(t_i) - J_2(t_{i-1})\right)\right| \\
&\qquad \leq \max_i \left|\left(X_1^c(t_i) - X_1^c(t_{i-1})\right)\right| \sum_{i=1}^{n} \left|J_2(t_i) - J_2(t_{i-1})\right| \\
&\qquad =\to 0
\end{aligned}$$

so we conclude that $[X_1^c, J_2](t) = 0$ and $\mathrm{d}X_1^c \,\mathrm{d}J_2 = 0$. Similarly for the terms involving X_2^c, J_1. The last terms in the cross-product will be

$$\begin{aligned}
&\sum_{i=1}^{n} \left(J_1(t_i) - J_1(t_{i-1})\right)\left(J_2(t_i) - J_2(t_{i-1})\right) \\
&\qquad = \sum_{s<t} \Delta J_1(s) \Delta J_2(s)
\end{aligned}$$

and this is non-zero provided that the jumps are not simultaneously occurring. If $N_m \perp\!\!\!\perp N_n$, then the cross quadratic variation will be $[J_1, J_2](t) = 0$, otherwise the product of of the jumps where they are simultaneous. If we have a Brownian motion, the variation $[W, J](t) = 0$ by an argument analogous to above.

Let us introduce the notation $\widetilde{X}_k(t) = \widetilde{X}_k(0) + \int_0^t \varPhi_k(s)\,\mathrm{d}X(s)$ for $k = 1, 2$, for the integral of such process. It can also be expressed with a pure jump process added to a continuous process, so as to get

$$\widetilde{X}_k = \widetilde{X}_k^c + \widetilde{J}_k$$

and

$$\begin{aligned}
[\widetilde{X}_1, \widetilde{X}_2](t) &= [\widetilde{X}_1^c, \widetilde{X}_2^c](t) + [\widetilde{J}_1, \widetilde{J}_2](t) \\
&= \int_0^t \varPhi_1(s)\varPhi_2(s)\,\mathrm{d}X_1^c\,\mathrm{d}X_2^c + \sum_{s \leq t} \varPhi_1(s)\varPhi_2(s)\Delta\widetilde{J}_1(s)\Delta\widetilde{J}_2(s) \\
&= \int_0^t \varPhi_1(s)\varPhi_2(s)\,\mathrm{d}[X_1, X_2](s) + \sum_{s \leq t} \varPhi_1(s)\varPhi_2(s)\Delta\widetilde{J}_1(s)\Delta\widetilde{J}_2(s)
\end{aligned}$$

Proposition 9.1 (Jump Itō-Doeblin formula)
If $f \in \mathcal{C}^2$, then

$$f(X(t)) = f(X(0)) + \int_0^t f'(X(s))\,\mathrm{d}X^c(s) + \frac{1}{2}\int_0^t f''(X(s))\,\mathrm{d}X^c(s)\,\mathrm{d}X^c(s) + \sum_{0\le s\le t}[f(X(s)) - f(X(s-))]$$

Geometric Poisson process

Since the compensated Poisson process can be negative, we consider the **geometric Poisson process**, which is of the form for $\sigma > -1$

$$S(t) = S(0)\exp\left(N(t)\log(\sigma+1) - \lambda\sigma t\right) = S(0)e^{-\lambda\sigma t}(\sigma+1)^{N(t)}$$

and we have to show that this is a martingale

Proposition 9.2 (Martingale property of the geometric Poisson process)
Using the Itō-Doeblin formula, consider

$$X(t) = N(t)\log(\sigma+1) - \lambda\sigma t$$

where $\lambda\sigma t$ is the continuous part and $N(t)\log(\sigma+1)$ the jump part. It is easy to see that

$$S(s) = S(s-)(\sigma+1), \qquad \Rightarrow \qquad \Delta S(s) - \sigma S(s-) = \sigma S(s-)\Delta N(s)$$

We can see that with $f(x) = e^x$, we get

$$\begin{aligned}
&1 - \lambda\sigma\int_0^t S(s-)\,\mathrm{d}s + \sum_{0\le s\le t} S(s) - S(s-)\\
&= 1 - \lambda\sigma\int_0^t S(s)\,\mathrm{d}s + \sigma\int_0^t S(s-)\,\mathrm{d}N(s)\\
&= 1 + \sigma\int_0^t S(s-)\left(\mathrm{d}(-\lambda s + N(s))\right)
\end{aligned}$$

for $S(0) = 1$; this gives a martingale, since we have the compensated process on the right and a left-continuous integrand $S(s-)$.

Corollary 9.3
Let $W(t)$ be a Brownian motion and $N(t)$ be a Poisson process adapted to one and the same filtration $\mathcal{F}(t)$. Then $N(t)$ and $W(t)$ are independent.

Proof Look in the book, consider the crucial

$$Y(t) = \exp\left(u_1 W(t) + u_2 N(t) - \frac{1}{2}u_1^2 t - \lambda(e^{u_2} - 1)t\right)$$

the joint characteristic function; because of the separability of the characteristic function factors for any time t and since the above is a martingale with expectation 1, we can get that they are independent for each t. ■

Proposition 9.4 (Itō-Doeblin formula for two jump processes)
We have functions of two arguments $f(t, x_1, x_2)$ with $f(t, X_1(t), X_2(t))$; the formula is then

$$X_1(t)X_2(t) \;=\; X_1(0)X_2(0) \;+\; \int_0^t X_2(s-)\,\mathrm{d}X_1(s) \;+\; \int_0^t X_1(s-)\,\mathrm{d}X_2(s) \;+\; [X_1, X_2](t)$$

We look now at change of measures to the risk-neutral measure for Poisson processes. For stochastic processes without jump, we can invoke Girsanov theorem to change the measure using the Radon-Nikodym theorem to the risk-neutral measure. Let

$$\begin{aligned} Z(t) &= \exp\left(-\int_0^t \Gamma(s)\,\mathrm{d}W(s) - \frac{1}{2}\int_0^t \Gamma^2(s)\,\mathrm{d}s\right) \\ &= \exp\left(X^c(t) - \frac{1}{2}[X^c, X^c](t)\right) \end{aligned}$$

and

$$\begin{aligned} \mathrm{d}Z(t) &= -Z(t)\Gamma(t)\,\mathrm{d}W(t) \\ &= Z(t)\,\mathrm{d}X^c(t) \end{aligned}$$

where

$$X^c(t) = -\int_0^t \Gamma(s)\,\mathrm{d}W(s).$$

Now, for the corresponding process with jump, we have

$$\begin{aligned} \mathrm{d}Z^X(t) &= Z^X(t-)\,\mathrm{d}X(t) \\ \Delta Z^X(s) &= Z^X(s-)\Delta X(s) \\ Z^X(s) &= Z^X(s-) + \Delta Z^x(s) = Z^X(s-)(1 + \Delta X(s)) \end{aligned}$$

To correspond with what we had for $Z(t)$ up here, we define the following process

$$Z^X(t) = \exp\left(X^c(t) - \frac{1}{2}[X^c, X^c](t)\right) \prod_{0\leq s\leq t} (1 + \Delta X(s)) \tag{9.9}$$

which we claim to be the solution of $\mathrm{d}Z^X(t)$. Indeed,

$$Z^X(t) = 1 + \int_0^t Z^X(s-)\,\mathrm{d}X(s)$$

where we have the back of the book answer. Again, due to lack of time, we refer the reader to the book. Writing $Z^X(t) = Y(t)K(t)$ where $K(t)$ is the product $\prod_{0\leq s\leq t}(1+\Delta X(s))$, we can use the product rule, and it goes through in a straightforward fashion using Proposition 9.4. Note in passing that (9.9) is named (after Catherine Doléan) the Doléan-Dale exponential of $X(t)$.

Change of measure

Let $N(t)$ be a Poisson process in the usual setting. Let $\widetilde{\lambda}$ be positive and define

$$Z(t) = e^{(\lambda-\widetilde{\lambda})t}\left(\frac{\widetilde{\lambda}}{\lambda}\right)^{N(t)}$$

and $M(t) = N(t) - \lambda t$ the compensated counterpart, so that

$$\mathrm{d}Z(t) = \frac{\widetilde{\lambda}-\lambda}{\lambda} Z(t-)\,\mathrm{d}M(t)$$

where

$$\begin{aligned} X^c(t) &= (\lambda - \widetilde{\lambda})t \\ J(t) &= \frac{\widetilde{\lambda}-\lambda}{\lambda} N(t) \end{aligned}$$

and the jump $1+\Delta X(t) = \frac{\widetilde{\lambda}}{\lambda}$ which allows us to $Z(t)$ itself. To show that this is martingale, we show that it is the Doléan-Dale exponential of $X(t)$ defined as

$$X(t) = \frac{\widetilde{\lambda}-\lambda}{\lambda} M(t).$$

Now

Proposition 9.5

Under the risk-neutral measure $\widetilde{\mathsf{P}}$, we have $N(t) \sim \mathcal{P}(\widetilde{\lambda}t)$. Indeed,

$$\widetilde{\mathsf{P}}(A) = \int_A Z(T)\,\mathrm{d}P$$

for T positive. We want to look at

$$\widetilde{\mathsf{E}}\left(e^{uN(t)}\right) = \mathsf{E}\left(Z(t)e^{uN(t)}\right) = \cdots = \exp\left(\widetilde{\lambda}t(e^u - 1)\right)$$

Example 9.2

Consider a change of measure to risk-neutral for a geometric Poisson process model, of the form

$$\begin{aligned} S(t) &= S(0)e^{(\alpha-\lambda\sigma)t}(\sigma+1)^{N(t)} \\ \mathrm{d}S &= \alpha S\,\mathrm{d}t + \sigma(t-)S\,\mathrm{d}M(t) \end{aligned}$$

as to get

$$\mathrm{d}S = rS\,\mathrm{d}t + \sigma(t-)S\,\mathrm{d}\widetilde{M}(t),$$

we need the relationship $S(\alpha - \lambda\sigma) = S(r - \widetilde{\lambda}\sigma)$ so that

$$\widetilde{\lambda} = \lambda - \frac{\alpha - r}{\sigma}$$

which requires to have $\lambda > \frac{\alpha - r}{\sigma}$. If not, then we have the possibility of an arbitrage.

If we had now different intensity, we could have

$$Z_m(t) = e^{(\lambda_m - \widetilde{\lambda}_m)t}\left(\frac{\widetilde{\lambda}_m}{\lambda_m}\right)^{N_m(t)}$$

for the m^{th} process. We use the fact for $Z(t) = \prod_{m=1}^{M} Z_m(t)$ using Itō product rule and extending to more by induction. This holds for the discrete case, with some more work and time, we can show that

$$Z_2(t) = e^{(\lambda-\widetilde{\lambda})t}\prod_{i=1}^{N(t)}\frac{\widetilde{\lambda}\widetilde{f}(Y_i)}{\lambda f(Y_i)}$$

which is also a martingale and $\mathsf{E}\left(Z_2(t)\right) = 1$. If now

$$Z_1(t) = \exp\left(-\int_0^t \Theta(s)\,\mathrm{d}W(s) - \frac{1}{2}\int_0^t \Theta^2(s)\,\mathrm{d}s\right)$$

then $Z_1(t)Z_2(t) = Z(t)$ is also a martingale with expectation 1.

We can get more than one martingale, dealing with one asset with two martingales, for example one Poisson and one Brownian motion, then we do not have a unique risk-neutral measure (it does exist, but we cannot edge except in the average). For three asset, two Poisson processes and one Brownian motion, we may fail to have risk-neutral price. The market price risk equation may then be solved.

We finish with an European call model driven solely by Poisson jump processes. We have

$$\begin{aligned} S(t) &= S(0)e^{(\alpha-\lambda\sigma)t}(\sigma+1)^{N(t)} \\ \mathrm{d}S(t) &= \alpha S(t)\,\mathrm{d}t + \sigma S(t-)\,\mathrm{d}M(t) \end{aligned}$$

and we make the change of measure to get

$$\mathrm{d}S(t) = -rS(t)\,\mathrm{d}t + \sigma S(t-)\,\mathrm{d}\widetilde{M}(t)$$

and now the value of the portfolio under the risk neutral measure is

$$V(t) = \widetilde{\mathsf{E}}\left(e^{-r(T-t)}(S(T)-K)_+ \;\middle|\; \mathcal{F}(t)\right)$$

and

$$S(T) = S(t)e^{r-\widetilde{\lambda})(T-t)}(\sigma+1)^{N(T)-N(t)}$$

where now $N(T) - N(t) \equiv N(T-t)$ and using the independence lemma, we get

$$V(t) = c(t, S(t))$$

where

$$c(t,x) = \sum_{j=0}^{\infty}\left(xe^{-\widetilde{\lambda}\sigma\tau}(\sigma+1)^j - Ke^{-r\tau}\right)_+ \frac{\widetilde{\lambda}^j\tau^j}{j!}e^{-\widetilde{\lambda}\tau}$$

and we can get a corresponding PDE as in Feynman-Kac using Itō-Doeblin rule with $\mathrm{d}\widetilde{M}(t)$ with $e^{-rt}c(t,S(t))$ to get (answer)

$$-rc(t,x) + c_t(t,x) + (r-\widetilde{\lambda}\sigma)xc_x(t,x) + \widetilde{\lambda}(c(t,(\sigma+1)x) - c(t,x)) = 0$$

a difference equation.

Index