Stochastic Financial Models *

Zhiyuan Bai

Compiled on December 29, 2021

This document serves as a set of revision materials for the Cambridge Mathematical Tripos Part II course *Stochastic Financial Models* in Michaelmas 2021. However, despite its primary focus, readers should note that it is NOT a verbatim recall of the lectures, since the author might have made further amendments in the content. Therefore, there should always be provisions for errors and typos while this material is being used.

Contents

*Based on the lectures under the same name taught by Dr. M. Tehranchi in Michaelmas 2021.

0 Introduction

To build financial models and get paid, we clearly need some unrealistic assumptions first.
First of all, we will assume that no dividend will be paid, ever (to be fair is that even a question at this stage of society, of course we are gonna assume that). We also assume that we can buy a continuous spectrum of shares because we want a easy life. Another criminal simplification we'll use is that there is no bid-ask spread, i.e. margin between ask price and bid price of a stock. We are also going to say that our action does not have any price impact, which means that the scale of our buying and selling would not affect the unit price of a share ("linear pricing"). We are also gonna take away transaction costs and short-selling constraints 'coz, y'know, capitalism.
Our standard framework is the following: There are d kinds of risky assets. The price of asset i at time t will be denoted S_t^i. For the next few lectures, we only care about $t \in \{0,1\}$ ("one-period models"). Afterwards, we might take $t \in \{0,1,\dots\}$ ("discrete-time models") or $t \in \mathbb{R}_{\geq 0}$ ("continuous models").
We will in addition assume that an agent can borrow or lend at a constant interest rate r ("risk-free assets") so that you can make sense of opportunity cost and alike.

1 One-Period Models

1.1 The Setup

Let X_t be the wealth of an agent at time t with X_0 given. Let θ^i be the number of shares of asset i hold by the agent from time 0 to time 1, so $\theta^i > 0$ implies that the agent bought θ_i shares ("long position") and $\theta^i < 0$ implies that the agent is shortselling $|\theta^i|$ shares ("short position"). Asset 0 will denote the risk-free asset (so $\theta^0 > 0$ means depositing and $\theta^0 < 0$ means borrowing). We have the budget constraint

$$X_0 = \theta^0 + \sum_{i=1}^{d} \theta^i S_0^i$$

If we write $S_t = (S_t^1, \dots, S_t^d)^\top, \theta = (\theta^1, \dots, \theta^d)^\top$, then the constraint is just $X_0 = \theta^0 + \theta^\top S_0$. At time 1, we have $X_1 = \theta^0(1+r) + \theta^\top S_1 = (1+r)X_0 + \theta^\top(S_1 - (1+r)S_0)$ by definition.
How would we model X_1? X_0, r, S_0, θ are all assumed to be known at time 0. We do not know anything about the true value of S_1, but we clearly will need

some information about it in order to get paid, so we are going to (boldly) model this as a random vector with a known distribution. There is a big industry of working out what the distribution should be via statistics, which is however not the focus of this course.

1.2 The Mean-Variance Portfolio Problem

The goal now is to maximise the earning X_1 but somehow minimise risk. The first way to formulate this is known as the mean-variance portfolio problem (Markowitz 1952): Given $X_0 = x$ and target mean m, we want to minimise $\operatorname{Var} X_1$ subject to $\mathbb{E}X_1 = m$. Markowitz's work on this problem won him the Nobel prize, so it's reasonable to deduce that he was not a mathematician.
Assume S_1 is square integrable. Let $\mu = \mathbb{E}S_1, V = \operatorname{cov} S_1 = \mathbb{E}((S_1 - \mu)(S_1 - \mu)^\top) = \mathbb{E}(S_1 S_1^\top) - \mu\mu^\top$. Different choices of θ will of course yield different values of $(\operatorname{Var} X_1, \mathbb{E}X_1)$. This is a subset of the right half-plane of $\mathbb{R}^2$. Its left side boundary is then the object of interest for the mean-variance problem. It's known as the mean-variance efficient frontier.

Definition 1.1. The mean-variance efficient frontier is the set consisting of points on the plane having the form $(\min\{\operatorname{Var}(X_1), \mathbb{E}X_1 = m\}, m), m \in \mathbb{R}$. A mean-variance efficient portfolio is a choice of θ that results in a value on the mean-variance efficient frontier, i.e. a solution to the mean-variance portfolio problem.

We shall deal with the case where $\mu \neq (1+r)S_0$ and V is positive definite (note that it is always nonnegative definite; we use this additional assumption in order for V^{-1} to exist). So we want to minimise $\theta^\top V\theta$ subject to $\theta^\top(\mu - (1+r)S_0) = m - (1+r)x$.

Theorem 1.1. *The unique optimal solution to the problem is $\theta^* = \lambda V^{-1}(\mu - (1+r)S_0)$ where*

$$\lambda = \frac{m - (1+r)x}{(\mu - (1+r)S_0)^\top V^{-1}(\mu - (1+r)S_0)}$$

Proof. One can verify directly that θ^* satisfy the desired condition ("feasible"). Suppose $\theta \neq \theta^*$ is another feasible portfolio, then let $\epsilon = \theta - \theta^* \neq 0$. We have $\theta^\top V\theta = (\theta^*)^\top V\theta^* + \epsilon^\top V\epsilon + 2\epsilon^\top V\theta^*$. But if θ is feasible then $2\epsilon^\top V\theta^* = 0$, so $\theta^\top V\theta = (\theta^*)^\top V\theta^* + \epsilon^\top V\epsilon > (\theta^*)^\top V\theta^*$ as V is positive definite. □

Definition 1.2. $\theta^{\text{mar}} = V^{-1}(\mu - (1+r)S_0)$ is called the market portfolio.

Corollary 1.2 (The Mutual Fund Theorem)**.** *A portfolio is mean-variance efficient iff it is a scalar multiple of the market portfolio.*

Corollary 1.3 (Shape of Mean-Variance Efficient Frontier)**.** *The mean-variance efficient frontier is a parabola with equation $v = (m - (1+r)x)^2/((\mu - (1+r)S_0)^\top V^{-1}(\mu - (1+r)S_0))$ (where (v, m) is the coordinate on the mean-variance graph).*

Corollary 1.4. *The accompanying problem of minimising* $\operatorname{Var} X_1$ *subject to $\mathbb{E}X_1 \geq m$ has the same solution as the mean-variance portfolio problem when $m \geq (1+r)x$ (otherwise of course the optimal strategy would be $\theta = 0$).*

1.3 Capital Asset Pricing Model

Proposition 1.5. *For square integrable random variables X, Y with $\operatorname{Var} X > 0$, there exists unique α, β such that $Y = \alpha + \beta X + Z$ with $\mathbb{E}Z = 0, \operatorname{cov}(Z, X) = 0$.*

Proof. We can simply solve the system

$$\begin{cases} 0 = \mathbb{E}Z = \mathbb{E}Y - \alpha - \beta - \mathbb{E}X \\ 0 = \operatorname{cov}(Z, X) = \operatorname{cov}(Y, X) - \beta \operatorname{cov}(X, X) \end{cases}$$

which yields $\beta = \operatorname{cov}(Y, X)/\operatorname{var}(X), \alpha = \mathbb{E}Y - \operatorname{cov}(X, Y)\mathbb{E}X/\operatorname{Var}(X)$. □

Write $S_t^{\mathrm{mar}} = (\theta^{\mathrm{mar}})^\top S_t$ (the "price of the market portfolio").

Theorem 1.6. *Let $B = \operatorname{cov}(S_1, S_1^{\mathrm{mar}})/\operatorname{Var}(S_1)$ and $Z = S_1 - (1+r)S_0 - B(S_1^{\mathrm{mar}} - (1+r)S_0^{\mathrm{mar}})$, then $\mathbb{E}Z = 0, \operatorname{cov}(Z, S_1^{\mathrm{mar}}) = 0$.*

So we can write $S_1 = (1+r)S_0 + B(S_1^{\mathrm{mar}} + (1+r)S_0^{\mathrm{mar}}) + Z$.

Proof. We have

$$\begin{aligned} \mathbb{E}(S_1^{\mathrm{mar}} - (1+r)S_0^{\mathrm{mar}}) &= (\theta^{\mathrm{mar}})^\top \mathbb{E}(S_1 - (1+r)S_0) \\ &= (\mu - (1+r)S_0)^\top V(\mu - (1+r)S_0) \\ &= (\theta^{\mathrm{mar}})^\top V \theta^{\mathrm{mar}} = \operatorname{var}(S_1^{\mathrm{mar}}) \end{aligned}$$

Also,

$$\operatorname{cov}(S_1, S_1^{\mathrm{mar}}) = \operatorname{cov}(S_1, S_1)\theta^{\mathrm{mar}} = \mu - (1+r)S_0 = \mathbb{E}(S_1 - (1+r)S_0)$$

Putting these together reveals the results. □

The capital asset pricing model (Sharpe 1964) is as follows: Let $R^i = (S_1^i/S_0^i) - 1$ be the return of asset i. Assume that there is a total of n_i shares of asset i and let $n = (n_1, \ldots, n_d)^\top$. Let

$$R^{\mathrm{index}} = \frac{n^\top S_1}{n^\top S_0} - 1$$

Suppose there are K investors and let θ^k be investor k's portfolio such that total demand equals total supply, i.e. $\sum_k \theta^k = n$ ("market clearing").
Assuming that every investor agrees on $\mu = \mathbb{E}S_1$ and $V = \operatorname{Var} S_1$ and everyone has a mean-variance efficient portfolio (with possibly different target means).

Theorem 1.7. *Write $R^i - r = \alpha^i + \beta^i(R^{\mathrm{index}} - r) + \epsilon^i$ such that $\mathbb{E}\epsilon^i = 0, \operatorname{cov}(R^{\mathrm{index}}, \epsilon^i) = 0$ (Proposition 1.5), then $\alpha^i = 0$ for all i.*

Proof. Choose λ^k such that $\theta^k = \lambda^k \theta^{\mathrm{mar}}$, so $n = (\sum_k \lambda^k)\theta^{\mathrm{mar}}$, consequently $R^{\mathrm{index}} = (S_1^{\mathrm{mar}}/S_0^{\mathrm{mar}}) - 1$. Then with the preceding theorem,

$$\begin{aligned} R^i - r &= \frac{1}{S_0^i}(S_i - (1+r)S_0) = \frac{1}{S_0^i}(B^i(S_1^{\mathrm{mar}} - (1+r)S_0^{\mathrm{mar}}) + Z^i) \\ &= \frac{B^i S_0^{\mathrm{mar}}}{S_0^i}(R^{\mathrm{index}} - r) + \frac{Z^i}{S_0^i} \end{aligned}$$

which is what we wanted. □

Remark. 1. The model gives us some kind of statistical ideas to work with, but one should note that doing statistics with economics is a little bit hard since we can't exactly keep doing the same experiment. There is a whole research area devoted to obtaining more information about the distribution of the future price given rather limited amount of information.
2. The standing assumptions under which the model evolves are not completely what happens in real life. Most of the general simplifications are reasonable, but the uniform beliefs assumption (i.e. everyone agreeing on $\mathbb{E}S_1$ and $\operatorname{Var} S_1$) and uniform preferences assumption (i.e. everyone's aim is to be mean-variance efficient) are up for debates.

1.4 Expected Utility Hypothesis

In the mean-variance portfolio problem, we have been dealing with the situation where we seek the minimisation of variance given a fixed expected return. This is a pretty restrictive idea, in a way. Maybe we'll want to use a different measurement than the variance, and maybe we'll want to deal with the case where an agent decide between strategies that have different expected returns (e.g. they might want to sacrifice return for smaller variance, or vice versa).
Let's try and find a more general framework to contain these situations, which we'll call the expected utility hypothesis: Every agent has a utility function $U = U(X)$ as a function of the payout, and they would prefer a strategy if it has a smaller $\mathbb{E}(U)$. If two strategies yield the same value of $\mathbb{E}(U)$, we say that the agent is indifferent between them.
Clearly, if we have another utility function $\tilde{U} = aU + b$ for some $a > 0$, then $\tilde{U}$ and U would give the same preference. Note also that we only need to know the marginal distributions of payouts X, Y to decide which one the agent would prefer.
The hypothesis was actually first proposed by (Daniel) Bernoulli to solve the St. Petersburg Paradox: Consider the game of tossin a coin until it comes up heads. If n tosses are required, Person A will pay Person B 2^n pounds. What's the expected earning of Person A? Infinity. But does it mean that it's a good decision for Person A to try and get into the game no matter how much Person B is going to charge them for a game entry? Doesn't seem so. Bernoulli's solution is then to introduce the utility weight on the possible earnings, since earning 2 zillion pounds compared to earning 1 zillion pounds doesn't give as much of happiness as earning 1 zillion pounds compared to earning nothing at all. [1]
Another motivation is what's known as the von Neumann-Morgenstern utility theorem.

Definition 1.3. The von Neumann-Morgenstern axioms of decision-making are the followings statements on preference of distribution functions:
1. (Completeness) Either $F \succ G$, $F \prec G$ or $F \sim G$.
2. (Transitivity) If $F \succ G$ and $G \succ H$ then $F \succ H$.
3. (Independence) If $F \succ G$ and $0 \le p \le 1$, then $pF + (1-p)H \succ pG + (1-p)H$.
4. (Continuity) If $F \succ G \succ H$, then there exists some $0 \le p \le 1$ such that $G \sim pF + (1-p)H$.

[1] In my opinion, this however is just a non-answer to a non-problem.

Theorem 1.8. *Suppose the von Neumann-Morgenstern axioms are satisfied, then there exists a utility function U such that*

$$F \succ G \iff \int_\Omega U \, \mathrm{d}F > \int_\Omega U \, \mathrm{d}G$$

Having seen enough motivations, let's now study the consequences of this hypothesis. To make our lives easier, we will sometimes restrict to the case where U is increasing, in the sense that if $X > Y$ a.s. then $\mathbb{E}U(X) > \mathbb{E}U(Y)$ (or $X \succ Y$). We sometimes also want U to be concave, i.e. $U(px + (1-p)y) \geq pU(x) + (1-p)U(y)$ for $p \in [0,1]$ ("a fix amount of money gives one less happiness when one is rich compared to when one is poor") which means that the marginal utility U' (when exists) should be decreasing. By Jensen's inequality, if this is indeed the case then $\mathbb{E}U(X) \leq U(\mathbb{E}X) \implies \mathbb{E}X \succ X$.
For an increasing, concave, twice differentiable utility U, the marginal utility $U' > 0$ would measure how much the utility increases at the given point, and $U'' < 0$ would measure the concavity of the utility at the point. They are used to measure how an agent would prefer certainty over uncertainty.

Definition 1.4. The Arrow-Pratt coefficient of absolute risk aversion at x is $-U''(x)/U'(x)$; The Arrow-Pratt coefficient of relative risk aversion at x is $-xU''(x)/U'(x)$.

Example 1.1. A CARA (Constant Absolute Risk Aversion) utility is one that has constant coefficient of absolute risk aversion, i.e. $U(x) = -Ce^{-\gamma x}$ (for some constant $C > 0$ which we will almost always taken as 1 when we talk about "the" CARA utility) where γ is the said coefficient.
A CRRA (Constant Relative Risk Aversion) utility is one that has constant coefficient of relative risk aversion, i.e. $U(x) = C(1-R)^{-1}x^{1-R}, x > 0$ (for some constant $C > 0$ which, again, we will almost always taken as 1 when we talk about "the" CRRA utility) where $R \in \mathbb{R}_{>0} \setminus \{1\}$ is the said coefficient.
When the utility has coefficient 1 for constant relative risk aversion, it has to be a multiple of log (defined on the positive reals). $U(x) = \log x$ is called the logarithmic utility.

Under this setup, we can now formulate the utility maximisation problem: Given a utility U, how would we maximise $\mathbb{E}U(X_1)$ subject to $X_0 = x$? Having nothing better to do, let's expand the expression to get $\mathbb{E}U(X_1) = \mathbb{E}U(x(1+r) + \theta^\top(S_1 - (1+r)S_0))$.

Theorem 1.9. *If U is suitably nice and θ^* is optimal, then* $\mathbb{E}(U'(X_1^*)(S_1 - (1+r)S_0)) = 0$.

Proof. Differentiate $\mathbb{E}U(X_1)$ with respect to θ and use dominated convergence theorem to pass derivative inside $\mathbb{E}$. □

Remark. One sufficient niceness condition is U being concave, differentiable, and $U(x(1+r) + \theta^\top(S_1 - (1+r)S_0))$ is integrable for all θ in a neighbourhood of θ^*.

It seems that there's nothing more we can do. Fearing losing our jobs, we introduce new definitions so as to appear to be doing something.

Definition 1.5. The state price density (or pricing kernel) is a positive random variable Z such that $\mathbb{E}Z = 1/(1+r)$ and $\mathbb{E}(ZS_1) = S_0$

The terminology "pricing kernel" is easy to understand: Adjoining it (and taking the expectation) brings S_1 back to S_0. But why is it also called the state price density? Consider a market model on the asset space/sample space $\Omega = \{\omega_1, \ldots, \omega_d\}$ with $S_1^i = 1_{\{\omega_i\}}$. $(S_1^1, \ldots, S_1^d)$ are called Arrow-Debreu securities. So if Z is a state price density, then easily $Z(\omega_i) = S_0^i/\mathbb{P}\{\omega_i\}$ which justified the name.
Now we can rephrase our previous theorem to get

Theorem 1.10. *$Z = U'(X_1^*)/((1+r)\mathbb{E}U'(X_1^*))$ is a state price density.*

1.5 Risk-Neutral Measures

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and $Y : \mathcal{F} \to \mathbb{R}_{>0}$ be a positive (as in a.s. positive) random variable with expected value $\mathbb{E}Y = \mathbb{E}^{\mathbb{P}}Y = 1$. Let $\mathbb{Q} : \mathcal{F} \to [0,1]$ be such that $\mathbb{Q}(A) = \mathbb{E}^{\mathbb{P}}(Y 1_A)$, then the monotone convergence theorem shows that $\mathbb{Q}$ is also a probability measure on $(\Omega, \mathcal{F})$. Also, $\mathbb{P}(A) = 0$ iff $\mathbb{Q}(A) = 0$ (hence $\mathbb{P}(A) = 1$ iff $\mathbb{Q}(A) = 1$, $\mathbb{P}(A) > 0$ iff $\mathbb{Q}(A) > 0$, $\mathbb{P}(A) < 1$ iff $\mathbb{Q}(A) < 1$).

Definition 1.6. Probability measures on the same measurable space are equivalent if they have the same null sets (i.e. sets with probability 0).

Can there be equivalent probability measures that would not arise in the way we described previously?

Theorem 1.11 (Radon-Nikodym)**.** *$\mathbb{P}, \mathbb{Q}$ are equivalent iff there exists a positive random variable Y ($\mathbb{P}, \mathbb{Q}$-a.s. unique) such that $\mathbb{Q}(A) = \mathbb{E}^{\mathbb{P}}(Y 1_A)$.*

So we are all good. For equivalent $\mathbb{P}, \mathbb{Q}$, we write $\mathbb{P} \sim \mathbb{Q}$ and $Y = \mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}$ ("the density of $\mathbb{Q}$ with respect to $\mathbb{P}$"). Y is also known as the Radon-Nikodym derivative or likelihood ratio.

Example 1.2. Suppose $\Omega = \{\omega_1, \ldots, \omega_n\}$ and $\mathbb{P} \sim \mathbb{Q}$ with $\mathbb{P}\{\omega_i\} > 0$ for all i, then $Y\{\omega_i\} = \mathbb{Q}\{\omega_i\}/\mathbb{P}\{\omega_i\}$.

If Z is $\mathbb{Q}$-integrable (i.e. $\mathbb{E}^{\mathbb{Q}}(|Z|) < \infty$), then $\mathbb{E}^{\mathbb{Q}}(Z) = \mathbb{E}^{\mathbb{P}}(YZ)$ (using the dominated convergence theorem). Also, if $\mathbb{E}f(X) = \mathbb{E}f(Y)$ for all bounded f, then X, Y have the same law.

Example 1.3. Let $X \sim \mathcal{N}(\mu, \sigma^2)$ under $\mathbb{P}$ and $Y = e^{X-\mu-\sigma^2/2}$, then $Y > 0$ and $\mathbb{E}^{\mathbb{P}}Y = 1$. Suppose $\mathbb{Q}(A) = \mathbb{E}^{\mathbb{P}}(Y 1_A)$, then for any bounded f,

$$\mathbb{E}^{\mathbb{Q}} f(X) = \mathbb{E}^{\mathbb{P}}(e^{X-\mu-\sigma^2/2} f(X)) = \mathbb{E}^{\mathbb{P}}(f(X + \sigma^2))$$

So $X \sim \mathcal{N}(\mu + \sigma^2, \sigma^2)$ under $\mathbb{Q}$, i.e. the law of X under $\mathbb{Q}$ is exactly the law of $X + \sigma^2$ under $\mathbb{P}$.

Definition 1.7. Given a market model, a risk-neutral measure is a probability measure $\mathbb{Q}$ equivalent to the original probability measure $\mathbb{P}$ such that $\mathbb{E}^{\mathbb{Q}}(S_1/(1+r)) = S_0$.

So $\mathbb{Q}$ is risk-neutral iff $(\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P})/(1+r)$ is a state price density.

Example 1.4 (Utility Maximisation in the Binomial Model)**.** Consider $d = 1$ (i.e. there is only one kind of risk asset) and $\mathbb{P}(S_1 = (1+b)S_0) = p = 1-\mathbb{P}(S_1 = (1+a)S_0)$. A risk neutral measure $\mathbb{Q}$ (say with $\mathbb{Q}(S_1 = (1+b)S_0) = q = 1-\mathbb{Q}(S_1 = (1+a)S_0))$ would satisfy $\mathbb{E}^{\mathbb{Q}}S_1 = (1+r)S_0$, or $qS_0(1+b)+(1-q)S_0(1+a) = (1+r)S_0 \implies q = (r-a)/(b-a)$. So a risk-neutral measure exists iff $a < r < b$, in which case it is unique and is given by the above discussion. Consider a increasing, concave, differentiable utility U. We want to maximise $\mathbb{E}U(X_1)$ subject to $X_0 = x$ where $X_1 = (1+r)x+\theta(S_1-(1+r)S_0)$. The optimal portfolio, as we've seen, has $U'(X_1^*)$ is proportional to the state price density, which is proportional to the risk-free measure, if exists. So

$$\begin{cases} U'(x(1+r)+\theta^* S_0(b-r)) = (\lambda/p)((r-a)/(b-a)) \\ U'(x(1+r)+\theta^* S_0(a-r)) = (\lambda/(1-p))((b-r)/(b-a)) \end{cases}$$

for some proportionality constant $\lambda > 0$. This simplifies to

$$\frac{U'(x(1+r)+\theta^* S_0(b-r))}{U'(x(1+r)+\theta^* S_0(a-r))} = \frac{(r-a)(1-p)}{(b-r)p}$$

1.6 Contingent Claim Pricing

We are back to our usual model with d assets of prices S_0, S_1 and interest rate r (the "fundamental market"), except this time we add a contingent claim, i.e. an extra asset with payout Y at time 1. We want to know how to correctly price this new asset at time 0. Usually Y is taken to be $g(S_1)$ for some function g. One important example of this is the call option, which is the right (but not the obligation) to buy an asset at a predetermined price K at time 1. Of course, if $S_1 > K$ then it's rational to exercise the right, and if $S_1 \leq K$ then it's rational not to, so the payout has the form

$$Y = (S_1 - K)^+ = \begin{cases} S_1 - K & \text{if } S_1 > K \\ 0 & \text{otherwise} \end{cases}$$

We will write $\mathcal{X}(x) = \{(1+r)x + \theta^\top(S_1 - (1+r)S_0) : \theta \in \mathbb{R}^d\}$ to denote the set of possible time 1 wealth in the fundamental market (i.e. without the contingent claim) fixing the times 0 wealth x. An agent with utility U and $X_0 = x$ would be willing to buy the claim at time 0 for price π if there exists $X^* \in \mathcal{X}(x-\pi)$ with $\mathbb{E}U(X^*+Y) \geq \mathbb{E}U(X)$ for all $X \in \mathcal{X}(x)$.

Definition 1.8. The utility indifference price (or the reservation bid price) $\pi = \pi(Y)$ of payout Y is the largest π such that

$$\sup\{\mathbb{E}U(X+Y) : X \in \mathcal{X}(x-\pi)\} \geq \sup\{\mathbb{E}U(X) : X \in \mathcal{X}(x)\}$$

Remark. 1. Note that $X \in \mathcal{X}(x-\pi)$ iff $X+(1+r)\pi \in \mathcal{X}(x)$, so $\sup\{\mathbb{E}U(X^*+Y) : X^* \in \mathcal{X}(x-\pi)\} = \sup\{\mathbb{E}U(X-(1+r)\pi+Y) : X \in \mathcal{X}(x)\}$.
2. If $\pi(Y)$ is the indifference bid price, then $-\pi(-Y)$ is known as the indifference ask price, i.e. the fair price one should sell the contingent claim for.

Fix the initial wealth x and utility function U that is assumed to be increasing, concave and differentiable. We will only be interested in the case where the payout of the contingency claim is in the (convex) set of random variables Y such that for every π, there exists an optimal solution of maximising $\mathbb{E}U(X+Y)$ subject to $X \in \mathcal{X}(x-\pi)$.

Theorem 1.12. *$Y \mapsto \pi(Y)$ is concave.*

Remark. In other words, the theorem is saying that $\pi(tY_1+(1-t)Y_0) \geq t\pi(Y_1)+(1-t)\pi(Y_0)$, i.e. an investor will value diversification.

Proof. $F_Y(\pi) = \sup\{\mathbb{E}U(X+Y) : X \in \mathcal{X}(x-\pi)\}$ is decreasing and continuous. We also have $F_Y(\pm\infty) = U(\mp\infty)$ and $\pi = \pi(Y)$ is the unique solution to $F_Y(\pi) = F_0(0)$.
Fix payouts Y_0, Y_1 and $Y_t = tY_1 + (1-t)Y_0$. Let $X_t = X$ be the maximiser of $\mathbb{E}U(X + Y_t)$ subject to $X \in \mathcal{X}(x - \pi(Y_t))$.

$$\begin{aligned}
F_{Y_t}(\pi(Y_t)) = F_0(0) &= tF_{Y_1}(\pi(Y_1)) + (1-t)F_{Y_0}(\pi(Y_0)) \\
&= t\mathbb{E}U(X_1 + Y_1) + (1-t)\mathbb{E}U(X_0 + Y_0) \\
&\leq \mathbb{E}U(tX_1 + (1-t)X_0 + Y_t) \leq F_{Y_t}(t\pi(Y_1) + (1-t)\pi(Y_0))
\end{aligned}$$

This means that $\pi(Y_t) \geq t\pi(Y_1) + (1-t)\pi(Y_0)$ as F_{Y_t} is decreasing. □

Observe that $t \mapsto \pi(tY)/t$ is decreasing by concavity. Let $\pi_t = \pi(tY)/t$ and we define π_0 as the limit of π_t as $t \downarrow 0$.

Theorem 1.13.
$$\pi_0 = \frac{\mathbb{E}(U'(X_0)Y)}{(1+r)\mathbb{E}U'(X_0)}$$
where $X_0 = X$ is the maximiser of $\mathbb{E}U(X)$ subject to $X \in \mathcal{X}(x)$.

Remark. This means that, when t is small, then the indifference price is almost linear, so the state price density is proportional to the marginal utility of the optimal time 1 wealth in the fundamental market.

Lemma 1.14. *If $X^* = X$ maximises $\mathbb{E}U(X+Y)$ subject to $X \in \mathcal{X}(x)$, then $\mathbb{E}(U'(X^*+Y)(S_1 - (1+r)S_0)) = 0$.*

Proof. Analogous to Theorem 1.9. □

Hence $\mathbb{E}(U'(X^*+Y)(X^* - X)) = 0$ for all $X \in \mathcal{X}(x)$.

Lemma 1.15 (Supporting Hyperplane)**.** *Suppose that U is concave and differentiable, then $U(y) - U(x) \leq U'(x)(y-x)$.*

Proof. Suppose $x < x + \epsilon < y$, then
$$\frac{U(y)-U(x)}{y-x} \leq \frac{U(x+\epsilon)-U(x)}{\epsilon}$$
by concavity. Sending $\epsilon \to 0$ shows the result. □

Proof of Theorem 1.13. Suppose $X_t = X$ maximises $\mathbb{E}U(X + tY)$ subject to $X \in \mathcal{X}(x - t\pi_t) = \mathcal{X}(x - \pi(tY))$, then

$$\begin{aligned}
0 &= \frac{\mathbb{E}(U(X_t + tY) - U(X_0))}{t} \geq \frac{\mathbb{E}(U(X_0 - t\pi_t(1+r) + tY) - U(X_0))}{t} \\
&\geq \frac{\mathbb{E}(U(X_0 - t\pi_0(1+r) + tY) - U(X_0))}{t} \to \mathbb{E}(U'(X_0)(Y - (1+r)\pi_0))
\end{aligned}$$

by dominated convergence theorem. So $\mathbb{E}(U'(X_0)(Y-(1+r)\pi_0)) \leq 0$. Conversely, by the preceding lemma,

$$\begin{aligned}
0 &= \frac{\mathbb{E}(U(X_t+tY)-U(X_0))}{t} \leq \frac{\mathbb{E}(U'(X_0)(X_t+tY-X_0))}{t} \\
&= \mathbb{E}(U'(X_0)(Y-(1+r)\pi_t)) + \frac{1}{t}\mathbb{E}(U'(X_0)(X_t+t\pi_t(1+r)-X_0)) \\
&= \mathbb{E}(U'(X_0)(Y-(1+r)\pi_t))
\end{aligned}$$

by Lemma 1.14. This means that $\mathbb{E}(U'(X_0)(Y-(1+r)\pi_0)) \geq 0$, so $\mathbb{E}(U'(X_0)(Y-(1+r)\pi_0)) = 0$. The theorem follows. ☐

1.7 Arbitrage; Fundamental Theorem of Asset Pricing

We now move away from the theory of contingent claims and go back to the fundamental market. Daydreaming, we want to earn money in the market with positive probability but no risk.

Definition 1.9. A portfolio ϕ is an arbitrage if and only if $\phi^\top(S_1-(1+r)S_0) \geq 0$ a.s. and $\mathbb{P}(\phi^\top(S_1-(1+r)S_0)>0)>0$.

Remark. The definition of arbitrage does not depend on utility functions nor initial wealth, but it does depend on the probability measure $\mathbb{P}$ (the "belief"). If you squint hard enough, you'll realise that it only depends on $\mathbb{P}$ through its null sets. So equivalent probability measures have the same set of arbitrages.

Consider the utility maximisation problem with a increasing utility function U and an arbitrage ϕ, then for any portfolio θ, $(1+r)x+(\theta+\phi)^\top(S_1-(1+r)S_0) \geq (1+r)x+\theta^\top(S_1-(1+r)S_0)$ a.s. and strict inequality happens with positive probability. If we then take the (U-weighted) expected value, then we get

$$\mathbb{E}U((1+r)x+(\theta+\phi)^\top(S_1-(1+r)S_0)) > \mathbb{E}U((1+r)x+\theta^\top(S_1-(1+r)S_0))$$

which contradicts the optimality of θ, not that anyone is surprised. So the existence of an arbitrage means that the utility maximisation problem does not have a solution. In other words, if the utility maximisation problem has a solution, then there cannot be an arbitrage.
So when an arbitrage does exist, then a investor will just buy any amount of asset as a multiple of the arbitrage, which is sadly incompatible with the standing assumption of linear pricing – not that we can't do maths with it.
Recall that if there exists a solution to the utility maximisation problem, then there exists a risk-neutral measure, namely the one induced by

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \frac{U'(X_1^*)}{\mathbb{E}U'(X_1^*)}$$

Theorem 1.16 ((First) Fundamental Theorem of Asset Pricing)**.** *The nonexistence of arbitrage is equivalent to the existence of a risk-neutral measure.*

Proof. Suppose there exists a risk-neutral measure $\mathbb{Q}$ and an arbitrage ϕ, then $\phi^\top(S_1-(1+r)S_0) \geq 0$ $\mathbb{P}$-a.s. (hence $\mathbb{Q}$-a.s.). We will show that $\phi^\top(S_1-(1+r)S_0) = 0$ $\mathbb{P}$-a.s. which will give a contradiction. $\mathbb{E}^{\mathbb{Q}}(\phi^\top(S_1-(1+r)S_0)) = \phi^\top\mathbb{E}^{\mathbb{Q}}(S_1-(1+r)S_0) = 0$ by risk-neutrality of $\mathbb{Q}$. So $\phi^\top(S_1-(1+r)S_0) = 0$

$\mathbb{Q}$-a.s. (pigeonhole principle or whatever it's called), hence $\mathbb{P}$-a.s. as desired.
Conversely, suppose there is no arbitrage, we shall construct a risk-neutral measure. Let $\xi = S_1 - (1+r)S_0$. Assume WLOG that $(\xi^i)_i$ are linearly independent (if not, take a maximal linearly independent subset and everything works out nicely) and that $\mathbb{E}e^{-\theta^\top \xi}$ is finite for all $\theta \in \mathbb{R}^d$ (if not, replace $\mathbb{P}$ by $\tilde{\mathbb{P}}$ with $\mathrm{d}\tilde{\mathbb{P}}/\mathrm{d}\mathbb{P} \propto e^{-\|\xi\|^2}$).
Let $(\theta_n)_n$ be a sequence such that $\mathbb{E}e^{-\theta_n^\top \xi} \to \inf_\theta \mathbb{E}e^{-\theta^\top \xi}$. If $(\theta_n)_n$ is bounded, then by Bolzano-Weierstrass, we can assume WLOG (by passing to a subsequence) that $(\theta_n)_n$ converges, say to θ_0. Then $\mathbb{E}e^{-\theta_n^\top \xi} \to \mathbb{E}e^{-\theta_0^\top \xi}$ by the continuity of moment generating functions. That is, $\theta_0 = \theta$ minimises $\mathbb{E}e^{-\theta^\top \xi}$, hence $\mathbb{E}(e^{-\theta_0{}^\top \xi}\xi) = 0$, so there is a risk-neutral measure $\mathbb{Q}$ derived from $\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P} \propto e^{-\theta_0^\top \xi}$.
What if $(\theta_n)_n$ is unbounded? WLOG $\|\theta_n\| \to \infty$. $\phi_n = \theta_n/\|\theta_n\|$ is bounded and hence WLOG converges to some ϕ_0 with $\|\phi_0\| = 1$. We shall show that $\phi_0^\top \xi \geq 0$ a.s. which will mean that $\phi_0^\top \xi = 0$ a.s. by the nonexistence of arbitrage. This will, by the linear independence of $(\xi^i)_i$, mean that $\phi_0 = 0$ which is a contradiction.
To show the inequality, observe that for large enough n we have $1 = \mathbb{E}e^{-0} \geq \mathbb{E}e^{-\theta_n^\top \xi} = \mathbb{E}e^{-\|\theta_n\|\phi_n^\top \xi}$. For $r, \xi > 0$, there is some $N \in \mathbb{N}$ such that $\|\phi_n - \phi_0\| \leq \xi/(2r)$ for all $n \geq N$. When $\phi_0^\top \xi < -\epsilon, \|\xi\| < r$, we have $\phi_n^\top \xi = (\phi_n - \phi_0)^\top \xi + \phi_0^\top \xi \leq \|\phi_n - \phi_0\|\|\xi\| + \phi_0^\top \xi \leq -\epsilon/2$, so

$$1 \geq \mathbb{E}e^{-\|\theta_n\|\phi_n^\top \xi} \geq \mathbb{E}e^{-\|\theta_n\|\phi_n^\top \xi} 1_{\phi_0^\top \xi < -\epsilon, \|\xi\| < r} \geq e^{\epsilon\|\theta_n\|/2}\mathbb{P}(\phi_0^\top \xi < -\epsilon, \|\xi\| < r)$$

Consequently $\mathbb{P}(\phi_0^\top \xi < -\epsilon, \|\xi\| < r) \to 0$ as $n \to \infty$. So $\mathbb{P}(\phi_0^\top \xi < 0) = 0$ which is what we wanted. $\square$

1.8 No-Arbitrage Pricing

In the contingency claim model, we must know about the utility U and initial wealth x to obtain the indifference pricing. This is fine for theory, but not very computationally efficient. We now introduce a new pricing method of contingency claims that is comparatively easier to calculate.

Theorem 1.17. *Given a fundamental market with no arbitrage. Suppose we augment the market by adding a contingency claim with payout Y and initial price π. Then there is no arbitrage in the augmented market iff $\pi = (1+r)^{-1}\mathbb{E}^{\mathbb{Q}}(Y)$ where $\mathbb{Q}$ is a risk-neutral measure in the fundamental market.*

Proof. By Theorem 1.16, the claim is equivalent to the existence of an equivalent measure $\mathbb{Q}$ with $\mathbb{E}^{\mathbb{Q}}((S_1, Y)^\top)/(1+r) = (S_0, \pi)^\top$. The first row is saying that $\mathbb{Q}$ has to be risk-neutral for the fundamental market, and the second row is exactly the identity claimed in the theorem. $\square$

Theorem 1.18. *The set of all possible no-arbitrage prices π is an interval.*

Proof. Suppose $\mathbb{Q}_0$ and $\mathbb{Q}_1$ are risk-neutral for the fundamental market giving no-arbitrage prices π_0 and π_1 respectively. Then $\mathbb{Q}_t = t\mathbb{Q}_1 + (1-t)\mathbb{Q}_0$ is also risk-neutral for all $0 \leq t \leq 1$. So $\pi_t = (1+r)^{-1}\mathbb{E}^{\mathbb{Q}_t}(Y) = t\pi_1 + (1-t)\pi_0$ is a no-arbitrage price. $\square$

Definition 1.10. A contingency claim Y is called attainable if there exists some $a \in \mathbb{R}$ and $b \in \mathbb{R}^d$ with $Y = a + b^\top S_1$.

That is, $Y = \mathcal{X}(y)$ where $y = (1+r)^{-1}a + b^\top S_0$.

Theorem 1.19. *The unique no-arbitrage price at an attainable claim with payout $Y \in \mathcal{X}(y)$ is $\pi = y$.*

Proof. $\pi = \mathbb{E}^{\mathbb{Q}}(Y)/(1+r)$ for some risk-neutral $\mathbb{Q}$. By risk neutrality we have $\mathbb{E}^{\mathbb{Q}}(Y) = \mathbb{E}^{\mathbb{Q}}(a + b^\top S_1) = a + (1+r)b^\top S_0$. □

Example 1.5. For $d = 1, r = 0, S_0 = S_0^1 = 5$ and $S_1 = S_1^1$ equals either 4 or 7, each with probability 1/2. Let Y be a "strike 6" call option, i.e. you get to buy the stock with price 6 at time 1, so the payout Y has is either 0 or 1, each with probability 1/2. Y is actually attainable, since $Y = (-4 + S_1)/3$. So the unique no-arbitrage price for Y is 1/3. Why? Suppose the time 0 price for Y is $\pi > 1/3$, an arbitrage can be obtained by buying 1/3 stock and selling 1 call. On the other hand, if the price is $\pi < 1/3$, then selling 1/3 stock and buying 1 call would be an arbitrage.
Suppose we have $S_1 = 7$ with risk-neutral probability p and $S_1 = 4$ with risk-neutral probability $q = 1 - p$, then since $\mathbb{E}^{\mathbb{Q}}(S_1)/(1+r) = S_0$ we have $p = 1/3, q = 2/3$. This is another justification of the no-arbitrage price being 1/3 and demonstrated the duality between arbitrages and risk-neutral measures guaranteed by Theorem 1.16.

Theorem 1.20. *If a contingent claim has a unique no-arbitrage price, then it is attainable.*

Proof. Example sheet. □

1.9 Complete Models

Definition 1.11. A market model is called complete if every contingent claim is attainable.

Theorem 1.21. *In a complete market model, S_1 can take at most $d+1$ values.*

Proof. For any disjoint events $A_1, \dots, A_n$, the indicators 1_{A_i} are clearly linearly independent. By completeness, since 1_{A_i} is a random variable, $1_{A_i} = a_i + b^\top S_1$ for some $a_i \in \mathbb{R}, b_i \in \mathbb{R}^d$. In particular, $\mathrm{Span}\{1_{A_1}, \dots, 1_{A_n}\} \subset \mathrm{Span}\{1, S_1^1, \dots, S_1^d\}$, so $n \le d+1$ and we are done. □

Theorem 1.22 (Second Fundamental Theorem of Asset Pricing)**.** *A market model with no arbitrage is complete iff there is a unique risk-neutral measure.*

Proof. Suppose the market is complete and $\mathbb{Q}_0, \mathbb{Q}_1$ are risk-neutral, then let $Y = \mathrm{d}\mathbb{Q}_1/\mathrm{d}\mathbb{P} - \mathrm{d}\mathbb{Q}_0/\mathrm{d}\mathbb{P}$. Choose a, b such that $Y = a + b^\top S_1$, then $\mathbb{E}^{\mathbb{Q}_i}(Y) = a + b^\top S_0(1+r)$ for both i, so $\mathbb{E}^{\mathbb{P}}(Y^2) = \mathbb{E}^{\mathbb{Q}_1}(Y) - \mathbb{E}^{\mathbb{Q}_0}(Y) = 0$, so necessarily $Y = 0$ a.s., i.e. the risk-neutral measure is unique.
Conversely, suppose the risk-neutral measure is unique, then every claim has a unique no-arbitrage price, so every claim is attainable by Theorem 1.20. □

2 Multi-Period Models

2.1 Motivation

How would we model a market with more than one period? Say if we want to model a 2-period market, then as usual S_0 should be given. S_2 is certainly random, but S_1 should be random at $t = 0$ but is no longer random at $t \geq 1$. This is a problem in general: We need to model the phenomenon that the price random variables lose their randomness as time goes on (i.e. as we obtain more information about the market).
The idea is the following: Suppose $\mathcal{G}$ is some collection of events (with some good set-theoretic properties), representing a certain set of information, then naturally we will want to say an event A becomes certain after $\mathcal{G}$ ("$\mathcal{G}$-measurable") if $\mathbb{P}(A|\mathcal{G}) \in \{0, 1\}$.

Example 2.1. Suppose we toss a coin twice. Let $\mathcal{G}$ be the information available after the first toss, then then event $\{HH, HT\}$ would become certain after $\mathcal{G}$ but $\{TT\}$ would not.

What do we want from this notion of $\mathcal{G}$-measurability? Naturally, $\varnothing$ should be $\mathcal{G}$-measurable. ALso, if A is $\mathcal{G}$-measurable, we also want A^c to be $\mathcal{G}$-measurable. Also, if $A_1, A_2, \ldots$ are $\mathcal{G}$-measurable, we certainly also want $\bigcup_n A_n$ to be measurable.

2.2 Some Basic Measure Theory

Despite having used these notions implicitly already, let's recall again what some of them mean.

Definition 2.1. Let Ω be a set. A σ-algebra on Ω is a nonempty collection $\mathcal{G}$ of subsets of Ω such that:
1. $\varnothing \in \mathcal{G}$.
2. $A \in \mathcal{G} \implies A^c \in \mathcal{G}$.
3. $A_1, A_2, \ldots \in \mathcal{G} \implies \bigcup_n A_n \in \mathcal{G}$.

Example 2.2. 1. $\{\varnothing, \Omega\}$ is called the trivial σ-algebra and contains no information.
2. 2^Ω is also a σ-algebra, and it can be interpreted as a set of all information.

Definition 2.2. A random variable (i.e. a function) $X : \Omega \to \mathbb{R}$ is called $\mathcal{G}$-measurable if $X^{-1}(B) \in \mathcal{G}$ whenever $B \in \mathcal{B}$, where $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}$. [2]

Remark. It is easy to show (given that you know how $\mathcal{B}$ is constructed) X is $\mathcal{G}$-measurable iff $\{X \leq x\} \in \mathcal{G}$ for all $x \in \mathbb{R}$. It is also equivalent to $\{X < x\} \in \mathcal{G}$ for all $x \in \mathbb{R}$.

Example 2.3. If X is $\{\varnothing, \Omega\}$-measurable, then X is not random (i.e. is constant).

Definition 2.3. The σ-algebra generated by a random variable $X : \Omega \to \mathbb{R}$, denoted as $\sigma(X)$, is the collection of all events of the form $X^{-1}(B), B \in \mathcal{B}$

[2] Stay calm and do a course in measure theory.

One can easily verify that this is indeed a σ-algebra.

Theorem 2.1. *If Y is $\sigma(X)$-measurable, then there exists a (Borel measurable) function g such that $Y = g(X)$.*

Proof. Lol. $\square$

Remark. The converse is also true: $g(X)$ is $\sigma(X)$-measurable whenever $g : \mathbb{R} \to \mathbb{R}$ is Borel measurable.

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ (where $\mathcal{F}$ is a σ-algbera on Ω), let $G \in \mathcal{F}$ be an event with $\mathbb{P}(G) > 0$. Recall that the conditional expectation of an event $A \in \mathcal{G}$ is defined by $\mathbb{P}(A|G) = \mathbb{P}(A \cap G)/\mathbb{P}(G)$. This inspires us to define the conditional expectation of a random variable X (implicitly assumed to be $\mathcal{F}$-measurable and $\mathbb{P}$-integrable) given $G \in \mathcal{F}$ to be $\mathbb{E}(X|G) = \mathbb{E}(X1_G)/\mathbb{P}(G)$.
Now suppose we have some other random variable Y taking values in a discrete set, we can make sense of a function constructed as $f(y) = \mathbb{E}(X|Y = y)$. This has the "projection property": For any bounded $\sigma(Y)$-measurable Z, we have $\mathbb{E}(XZ) = \mathbb{E}(f(Y)Z)$ by the preceding theorem.
Motivated by this,

Definition 2.4. Given a random variable X ($\mathcal{F}$-measurable and $\mathbb{P}$-integrable) and a σ-algebra $\mathcal{G} \subset \mathcal{F}$. A random variable $\tilde{X}$ is called a conditional expectation of X given $\mathcal{G}$ iff it is $\tilde{G}$-measurable and $\mathbb{E}(XZ) = \mathbb{E}(\tilde{X}Z)$ for any bounded, $\mathcal{G}$-measurable Z.

Proposition 2.2. *The conditional expectation of X given $\mathcal{G}$ exists and is a.s. unique.*

Proof. Existence is, as expected (pun intended), omitted.
Suppose $\tilde{X}_0, \tilde{X}_1$ are conditional expectations of X given $\mathcal{G}$, then $\mathbb{E}(\tilde{X}_0 Z) = \mathbb{E}(XZ) = \mathbb{E}(\tilde{X}_1 Z)$ for any bounded $\mathcal{G}$-measurable Z. Let $Z = 1_{\tilde{X}_0 < \tilde{X}_1}$, then the identity means that $\mathbb{E}((\tilde{X}_1 - \tilde{X}_0)1_{\tilde{X}_0 < \tilde{X}_1}) = 0$, so $\mathbb{P}(\tilde{X}_0 < \tilde{X}_1) = 0$. Symmetrically $\mathbb{P}(\tilde{X}_0 > \tilde{X}_1) = 0$, so $\tilde{X}_0 = \tilde{X}_1$ a.s.. $\square$

So we can confidently write $\mathbb{E}(X|\mathcal{G})$ to denote the conditional expectation of X given $\mathcal{G}$.

Definition 2.5. For random variables X, Y, the conditional expectation of X given Y is defined as $\mathbb{E}(X|Y) = \mathbb{E}(X|\sigma(Y))$.

Proposition 2.3. *Suppose X is square-integrable and let $\tilde{X} = \mathbb{E}(X|\mathcal{G})$ for some σ-algebra $\mathcal{G}$, then $\mathbb{E}((X - \tilde{X})^2) \le \mathbb{E}((X - Z)^2)$ for all square-integrable $\mathcal{G}$-measurable Z.*

Proof. $\mathbb{E}((X - Z)^2) = \mathbb{E}((X - \tilde{X})^2) + 2\mathbb{E}((X - \tilde{X})(X - Z)) + \mathbb{E}((\tilde{X} - Z)^2)$.
$\mathbb{E}((X - \tilde{X})(\tilde{X} - Z)) = 0$ as $\tilde{X} - Z$ is $\mathcal{G}$-measurable by the projection property (which technnically only works for bounded $\tilde{X} - Z$ but we can extend using measure theoretic arguments) which gives the result. $\square$

So in a sense $\tilde{X}$ is the "best prediction of X given $\mathcal{G}$".

Proposition 2.4. *Suppose all conditional expectations of interest here are well-defined.*
(i) $\mathbb{E}(X+Y|\mathcal{G}) = \mathbb{E}(X|\mathcal{G}) + \mathbb{E}(Y|\mathcal{G})$.
(ii) If Y *is already* $\mathcal{G}$*-measurable, then* $\mathbb{E}(XY|\mathcal{G}) = Y\mathbb{E}(X|\mathcal{G})$.
(iii) If X *is a.s. nonnegative, then* $\mathbb{E}(X|\mathcal{G})$ *is also a.s. nonnegative.*
(iv) Suppose X *is independent of* $\mathcal{G}$ *(in the sense that any* $A \in \sigma(X)$ *and* $G \in \mathcal{G}$ *are independent). Then* $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}X$.
(v) (Jensen's Inequality) If f *is convex, then* $\mathbb{E}(f(X)|\mathcal{G}) \geq f(\mathbb{E}(X|\mathcal{G}))$ *a.s..*
(vi) (Tower Property) Suppose $\mathcal{G} \subset \mathcal{H}$ *are* σ*-algebras and* X *is integrable, then* $\mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H}) = \mathbb{E}(\mathbb{E}(X|\mathcal{H})|\mathcal{G}) = \mathbb{E}(X|\mathcal{G})$.

Proof. For (i), let $\tilde{X} = \mathbb{E}(X|\mathcal{G})$ and $\tilde{Y} = \mathbb{E}(Y|\mathcal{G})$, then for all bounded $\mathcal{G}$-measurable W we have $\mathbb{E}((\tilde{X}+\tilde{Y})W) = \mathbb{E}((X+Y)W)$, so $\tilde{X}+\tilde{Y} = \mathbb{E}(X+Y|\mathcal{G})$ by uniqueness. (ii) is similar. (iii) follows from interpreting $\mathbb{E}(X1_{\mathbb{E}(X|\mathcal{G})<0}) = \mathbb{E}(\mathbb{E}(X|\mathcal{G})1_{\mathbb{E}(X|\mathcal{G})<0})$. (iv) is immediate from definiton. To see (v), we use the fact that any convex f admits some $\lambda = \lambda(x)$ such that $f(y) \geq f(x)+\lambda(x)(y-x)$ (a more general form of Lemma 1.15). (vi) is again just checking definitions. □

Example 2.4. Suppose $(A_n)_n$ is a countable collection of events partitioning Ω with $\mathbb{P}(A_i) > 0$. Then $\mathcal{G} = \{\bigcup_{i\in I} A_n : I \subset \mathbb{N}\}$ is a σ-algebra and a random variable Y on Ω is $\mathcal{G}$-measurable iff it is constant on each A_i. Also, if Y takes distinct values on distinct A_i's, then $\mathcal{G} = \sigma(Y)$. We have $\mathbb{E}(X|\mathcal{G}) = \sum_i \mathbb{E}(X|A_i)1_{A_i}$ which equals $\mathbb{E}(X|Y) = \sum_i \mathbb{E}(X|Y=y_i)1_{Y=y_i}$ when $\sigma(Y) = \mathcal{G}$.

2.3 Martingales

Definition 2.6. A filtration on Ω is an increasing sequence $(\mathcal{F}_n)_n$ of σ-algebras (i.e. $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots$) on Ω.

By convention, $\mathcal{F}_0 = \{\varnothing, \Omega\}$ is the trivial σ-algebra.
We want to use the idea of filtration to capture the intuition of a random process whose information is gradually revealed over time.

Definition 2.7. A (discrete-time) stochastic process $(X_n)_n$ is just a sequence of random variables.

Definition 2.8. A stochastic process $(X_n)_n$ is adapted to a filtration $(\mathcal{F}_n)_n$ if X_n is $\mathcal{F}_n$-measurable for all $n \geq 0$.

Our convention that $\mathcal{F}_0$ is trivial means that X_0 is not random, if $(X_n)_n$ is adapted to $(\mathcal{F}_n)_n$. Note also that $\mathbb{E}(Z|\mathcal{F}_0) = \mathbb{E}Z$ for any random variable Z.

Definition 2.9. The filtration $(\mathcal{F}_n)_n$ generated by a given stochastic process $(X_n)_n$ is defined by $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$ (the σ-algebra generated by $X_0, \ldots, X_n$), i.e. the smallest filtration to which the process $(X_n)_n$ is adapted.

Definition 2.10. A stochastic process $(X_n)_n$ is a martingale with respect to a filtration $(\mathcal{F}_n)_n$ if X_n is integrable for every n and $\mathbb{E}(X_n|\mathcal{F}_{n-1}) = X_{n-1}$.

Proposition 2.5. *The followings are equivalent:*
(i) $(X_n)_n$ *is an* $(\mathcal{F}_n)_n$*-martingale.*
(ii) $(X_n)_n$ *is adapted to* $(\mathcal{F}_n)_n$ *and* $\mathbb{E}(X_n - X_{n-1}|\mathcal{F}_{n-1}) = 0$ *for all* $n \geq 1$.
(iii) $\mathbb{E}(X_n|\mathcal{F}_m) = X_m$ *for all* $0 \leq m \leq n$.

The third equivalent condition will allow us to generalise this theory to continuous-time.

Proof. The only nontrivial implication is that (i) $\implies$ (iii), which follows from the tower property $\mathbb{E}(X_{m+k}|\mathcal{F}_m) = \mathbb{E}(\mathbb{E}(X_{m+k}|\mathcal{F}_{m+k-1})|\mathcal{F}_m)$. □

Example 2.5. 1. Suppose $\Omega = \{HH, HT, TH, TT\}$ is the sample space of flipping two (fair) coins. Let $\mathcal{F}_0 = \{\varnothing, \Omega\}, \mathcal{F}_1 = \{\varnothing, \Omega, H = \{HH, HT\}, T = \{TH, TT\}\}$ and $\mathcal{F}_n = 2^\Omega$ for $n \geq 2$. Suppose $(X_n)_n$ is a stochastic process adapted to $(\mathcal{F}_n)_n$, then X_0 is constant, X_1 can take 2 values, and X_2 can take 4 values. The branching happens exactly with the information of the two coin tosses being given. If $(X_n)_n$ is a martingale and $X_2(HH) = a, X_2(HT) = b, X_2(TH) = c, X_2(TT) = d$, then necessarily $X_1(H) = (a + b)/2, X_1(T) = (c + d)/2$ and $X_0 = (a + b + c + d)/4 = \mathbb{E}X_2$.
One can generalise the moral of this example to the infinite coin-tossing space by considering filtration generated by $(Y_n)_n$ where $Y_0 = 0$ and Y_n is the indicator of the success of the n^{th} tossing.
2. Let Z be integrable and $(\mathcal{F}_n)_n$ be a filtration. Let $X_n = \mathbb{E}(Z|\mathcal{F}_n)$, then $(X_n)_n$ is an $(\mathcal{F}_n)_n$-martingale (in a way, this is similar to the previous example in spirit). The integrability of X_n is due to Jensen's inequality: $\mathbb{E}|X_n| \leq \mathbb{E}(\mathbb{E}(|Z||\mathcal{F}_n)) = \mathbb{E}|Z| < \infty$. For the martingale property, suppose $m < n$, then by the tower property we have $\mathbb{E}(X_n|\mathcal{F}_m) = \mathbb{E}(\mathbb{E}(Z|\mathcal{F}_n)|\mathcal{F}_m) = \mathbb{E}(Z|\mathcal{F}_m) = X_m$.

The above examples are backward-in-time constructions of martingales. We can also construct them in a forward-in-time fashion.

Example 2.6. 1. Let $(X_n)_n$ be independent and integrable with $\mathbb{E}X_n = 0$ for all n. Let $S_0 = 0$ and $S_n = X_1 + \cdots + X_n$, then S_n is a martingale with respect to the filtration $(\mathcal{F}_n)_n$ generated by $(X_n)_n$. Indeed, S_n must be $\mathcal{F}_n$-measurable, so $(S_n)_n$ is adapted to $(\mathcal{F}_n)_n$. Clearly each S_n must be integrable by triangle inequality, so we can safely calculate $\mathbb{E}(S_n - S_{n-1}|\mathcal{F}_{n-1}) = \mathbb{E}(X_n|\mathcal{F}_{n-1}) = \mathbb{E}(X_n) = 0$.
2. Let $(X_n)_n$ be independent and integrable with $\mathbb{E}X = 1$ for all n. Let $(\mathcal{F}_n)_n$ be the filtration generated by $(X_n)_n$, then the stochastic process $(M_n)_n$ given by $M_0 = 1, M_n = X_1 \cdots X_n$ is an $(\mathcal{F}_n)_n$-martingale.

Why are we interested in martingales in finance?

Definition 2.11. For a multi-period model with prices $(S_n)_{n\geq 0}$ and interest rate r, a probability measure $\mathbb{Q}$ is risk-neutral if it is equivalent to the original measure and $S_n/(1 + r)^n$ is a martingale under $\mathbb{Q}$ (with a filtration fixed by the model).

We will show an analogue of Theorem 1.16: The existence of a risk-free measure is equivalent to the non-existence of multi-period arbitrages (which we'll define in a moment).

2.4 Stopping Time

Definition 2.12. A stopping time for a filtration $(\mathbb{F}_n)_n$ is a random variable T taking values in $\{0, 1, 2, \ldots\} \cup \{\infty\}$ such that the $\{T \leq n\} \in \mathcal{F}_n$ for all n.

Example 2.7. 1. Suppose $(X_n)_n$ is adapted to $(\mathcal{F}_n)_n$, then $T = \inf\{n \geq 0 : X_n > 0\}$ (with the convention that $\inf \varnothing = \infty$) is a stopping time since $\{T \leq n\} = \bigcup_{k=0}^{n} \{X_k > 0\} \in \mathcal{F}_n$.
2. (non-example) Not everything is a stopping time. Let $(X_n)_n$ be as above and $T = \sup\{n \geq 0 : X_n > 0\}$, then $\{T \leq n\} = \bigcap_{k=n+1}^{\infty} \{X_k \leq 0\}$ which isn't in general $\mathcal{F}_n$-measurable.

Definition 2.13. Given a process $X = (X_n)_n$ and a random variable T, the process X^T ("X but it stopped at T") is defined by

$$X_n^T = X_{n \wedge T} = X_{\min\{n,T\}} = X_0 + \sum_{k=1}^{n} 1_{k \leq T}(X_k - X_{k-1})$$

As intuition suggests, we are usually interested in the case where T is a stopping time with respect to some filtration that X adapts to.

Proposition 2.6. *(i) When X is integrable (in the sense that each X_n is integrable), so is X^T.*
(ii) Suppose X is adapted to a filtration $(\mathcal{F}_n)_n$ to which T is a stopping time, then X^T is also adapted to $(\mathcal{F}_n)_n$.

Proof. (i) Triangle inequality.
(ii) $1_{k \leq T}$ is $\mathcal{F}_n$-measurable for any $1 \leq k \leq n$ since $\{k \leq T\} = \{T \leq k-1\}^c \in \mathcal{F}_{k-1} \subset \mathcal{F}_n$. $\square$

Theorem 2.7 (Optional Sampling Theorem)**.** *Fix a filtration $(\mathcal{F}_n)_n$. Suppose X is a martingale and T is a stopping time, then X^T is also a martingale. In particular, if T is a.s. bounded, then $\mathbb{E}X_T = X_0$ (recall that X_0 is a.s. constant by the cconvention that $\mathcal{F}_0 = \{\varnothing, \Omega\}$).*

Proof. We already know that X^T is integrable and adapted. It remains to check that $\mathbb{E}(X_n^T - X_{n-1}^T | \mathcal{F}_{n-1}) = 0$. Indeed,

$$\begin{aligned}\mathbb{E}(X_n^T - X_{n-1}^T | \mathcal{F}_{n-1}) &= \mathbb{E}(1_{n \leq T}(X_n - X_{n-1}) | \mathcal{F}_{n-1}) \\ &= 1_{n \leq T} \mathbb{E}(X_n - X_{n-1} | \mathcal{F}_{n-1}) = 0\end{aligned}$$

since $\{n \leq T\} = \{T \leq n-1\}^c \in \mathcal{F}_{n-1}$ and X is a martingale.
When T is a.s. bounded, there exists some N such that $T \leq N$ a.s.. So $\mathbb{E}X_T = \mathbb{E}X_{T \wedge N} = \mathbb{E}X_N^T = X_0$ as X^T is a martingale. $\square$

Proposition 2.8. *Let X be a simple symmetric random walk on $\mathbb{Z}$ starting at $X_0 = 0$, which is a martingale with respect to the filtration generated by it. Let $T = \inf\{n \geq 1 : X_n \in \{-a, b\}\}$ for integers $-a < 0 < b$, then*

$$\mathbb{P}(X_T = -a) = \frac{b}{a+b} = 1 - \mathbb{P}(X_T = b)$$

Proof. It is very tempting to write $(-a)\mathbb{P}(X_T = -a) + b\mathbb{P}(X_T = b) = \mathbb{E}(X_T) = X_0 = 0$ using the preceding theorem, but T is not bounded. The proposition is true regardless, which can be proved either using elementary methods or with dominated convergence theorem, which we've quoted a few times already but nonetheless stated below. $\square$

Theorem 2.9 (Dominated Convergence Theorem)**.** *Suppose $(Z_n)_n$ is a sequence of random variables such that $Z_n \to Z$ a.s. and $|Z_n| \leq Y$ for some fixed integrable random variable Y, then $\mathbb{E}Z_n \to \mathbb{E}Z$.*

Example 2.8 (Non-example)**.** The requirement that T is bounded is necessary, although it didn't cause problem in the preceding proposition since it was in a form where we can use dominated convergence theorem. When T behaves badly, however, problems can occur: If one take instead that $T = \inf\{n \geq 0 : X_n = b\}$ (i.e. taking $a = \infty$), then by techniques from Markov chains we know that $X_T = b$ a.s., but then $\mathbb{E}(X_T) = b \neq 0$.

Theorem 2.10 (Optional Stopping Theorem)**.** *Fix a filtration $(\mathcal{F}_n)_n$ (again with the convention that $\mathcal{F}_0 = \{\varnothing, \Omega\}$). Let X be a martingale and T a stopping time. Suppose there is an integrable random varible Y with $|X_n^T| \leq Y$ for all n, then $X_n^T \to X_T$ as $n \to \infty$ and $\mathbb{E}X_T = X_0$.*

Proof. The convergence when T is a.s. finite is clear. The general case uses the martingale convergence theorem, which exceeds the scope of this course. $\mathbb{E}(X_T) = X_0$ is a consequence of Theorem 2.7 and the dominated convergence theorem. □

Let's state some more measure theoretical result which we will not prove.

Theorem 2.11 (Monotone Convergence Theorem)**.** *If $0 \leq X_n \leq X_{n+1}$ a.s. for all n, then $\lim_{n\to\infty} \mathbb{E}X_n = \mathbb{E}(\lim_{n\to\infty} X_n)$.*

Example 2.9. Let X be the simple symmetric random walk on $\mathbb{Z}$ and $T = \inf\{n \geq 0 : X_n \in \{-a, b\}\}$ (with respect to the filtration generated by X), then $\mathbb{E}T = \mathbb{E}X_T^2 = a^2(b/(a+b)) + b^2(a/(a+b)) = ab$. Indeed, $Q_n = X_n^2 - n$ is a martingale (example sheet). By Theorem 2.7, $\mathbb{E}Q_n^T = 0$, so $\mathbb{E}(X_{T\wedge n}^2) = \mathbb{E}(n \wedge T) \to \mathbb{E}T$ as $n \to \infty$ by monotone convergence theorem. On the other hand, $\mathbb{E}(X_{T\wedge n}^2) \to \mathbb{E}_T^2$ by dominated convergence theorem, hence our claim.
When $a = \infty$, however, we would have $X_T = b$ a.s., so $\mathbb{E}X_T = b \neq 0 = X_0$ which would mean $\mathbb{E}(\min_{n\leq T} X_n) = -\infty$ (otherwise we'll be able to draw a contradiction from the dominated convergence theorem).

2.5 Supermartingales

Definition 2.14. With respect to a filtration $(\mathcal{F}_n)_n$, a stochastic process $(X_n)_n$ is a supermartingale (resp. submartingale) if it is an adapted integrable process with $\mathbb{E}(X_n|\mathcal{F}_{n-1}) \leq X_{n-1}$ (resp. $\mathbb{E}(X_n|\mathcal{F}_{n-1}) \geq X_{n-1}$) a.s. for all $n \geq 1$.

Theorem 2.12 (Optional Sampling Theorem for Supermartingales)**.** *Fix a filtration $(\mathcal{F}_n)_n$. Suppose X is a supermartingale and S, T are two (a.s.) bounded stopping times with $S \leq T$ a.s., then $\mathbb{E}X_S \geq \mathbb{E}X_T$.*

Proof. We have

$$X_{n\wedge T} - X_{n\wedge S} = \sum_{k=1}^{n} 1_{S<k\leq T}(X_k - X_{k-1})$$

Note that $\{S < k\}, \{k \leq T\} \in \mathcal{F}_{k-1}$, so $\{S < k \leq T\} \in \mathcal{F}_{k-1}$, so

$$\begin{aligned}
\mathbb{E}(X_{n\wedge T} - X_{n\wedge S}) &= \sum_{k=1}^{n} \mathbb{E}(1_{S<k\leq T}(X_k - X_{k-1})) \\
&= \sum_{k=1}^{n} \mathbb{E}(\mathbb{E}(1_{S<k\leq T}(X_k - X_{k-1})|\mathcal{F}_{k-1})) \\
&= \sum_{k=1}^{n} \mathbb{E}(1_{S<k\leq T}\mathbb{E}(X_k - X_{k-1}|\mathcal{F}_{k-1})) \leq 0
\end{aligned}$$

As S, T are bounded, this shows the theorem. □

Definition 2.15. Let H, X be two processes. The martingale transform of H with respect to X is the process

$$(H \cdot X)_n = \sum_{k=1}^{n} H_k(X_k - X_{k-1})$$

Definition 2.16. A process $(H_n)_{n\geq 1}$ is called previsible (or predictable) with respect to a filtration $(\mathcal{F}_n)_n$ if H_n is $\mathcal{F}_{n-1}$-measurable for all $n \geq 1$.

Theorem 2.13. *Fix a filtration $(\mathcal{F}_n)_n$. Let X be a martingale and H a (termwise) bounded previsible process, then $H \cdot X$ is a martingale.*

Proof. $H \cdot X$ is clearly integrable by the boundedness of H, and is adapted since H is previsible and X is adapted. It remains to compute

$$\begin{aligned}
\mathbb{E}((H \cdot X)_n - (H \cdot X)_{n-1}|\mathcal{F}_{n-1}) &= \mathbb{E}(H_n(X_n - X_{n-1})|\mathcal{F}_{n-1}) \\
&= H_n\mathbb{E}(X_n - X_{n-1}|\mathcal{F}_{n-1}) = 0
\end{aligned}$$

as X is a martingale. □

Theorem 2.14. *Fix a filtration $(\mathcal{F}_n)_n$. Let X be a supermartingale and H a nonnegative, (termwise) bounded, previsible process, then $H \cdot X$ is a supermartingale.*

Proof. Similar. □

How do these constructions relate to finance?
Let $(S_n)_n$ be the price process adapted to a filtration $(\mathcal{F}_n)_n$ and r a riskfree interest rate. Let θ_n be the portfolio held between time $n-1$ and n and X_n be the wealth of the agent at time n. The budget constraint is then $X_{n-1} = \theta_n^0 + \theta_n^\top S_{n-1}$ and we have $X_n = \theta_n^0(1+r) + \theta_n^\top S_n$. A simple calculation shows that

$$\frac{X_n}{(1+r)^n} = X_0 + \sum_{k=1}^{n} \theta_k^\top \left(\frac{S_k}{(1+r)^k} - \frac{S_{k-1}}{(1+r)^{k-1}} \right)$$

So the discounted wealth is exactly the martinagle transform of portfolio with respect to discounted prices.

2.6 Controlled Markov Processes

Recall that a process $(X_n)_n$ is Markov iff $\mathbb{E}(f(X_n)|\mathcal{F}_{n-1}) = \mathbb{E}(f(X_n)|X_{n-1})$ for any bounded f, where $(\mathbb{F}_n)_n$ is the filtration generated by X_n. The Markov property means that we only need to know $P(x, A) = \mathbb{P}(X \in A|X = x)$ in order to build a Markov process.
From a different point of view, one can view this process as a random dynamical system: Let $(\xi_n)_n$ be an i.i.d. sequence taking values in some space $\mathcal{V}$ and a function $G : \mathcal{X} \times \mathcal{V} \to \mathcal{X}$. Then $X_n = G(X_{n-1}, \xi_n)$ gives a Markov process. It's clear that these two viewpoints give the same class of objects, and they are linked together by $\mathbb{P}(G(x, \xi_1) \in A) = P(x, A)$.

Example 2.10. Let $(\xi_n)_n$ be a sequence of i.i.d. Unif$[0, 1]$ random variables and

$$G(x, v) = \begin{cases} x+1 & \text{if } v \in [0, 1/2) \\ x-1 & \text{if } v \in [1/2, 1] \end{cases}$$

Tnen $X_0 = 0, X_n = G(X_{n-1}, \xi_n)$ is the symmetric random walk on $\mathbb{Z}$.

The idea of a controlled Markov process evolves from the addition of an accompanying previsible process (with respect to a given filtration). Given a sequence of i.i.d. random variables, a previsible process $(U_n)_n$ (taking values in $\mathcal{U}$) and function $G : \mathcal{X}\times\mathcal{U}\times\mathcal{V} \to \mathcal{X}$, the process $X_n = G(X_{n-1}, U_n, \xi_n)$ is called a controlled Markov process. Alternatively, these processes can be characterised by the controlled transition probabilities $P(x, u, A) = \mathbb{P}(G(x, u, \xi_1) \in A) = \mathbb{P}(X_1 \in A : X_0 = x, U_1 = u)$.

Example 2.11. Consider a multi-period market model with one risk asset. Suppose $S_n = S_{n-1}\xi_n$ where ξ_n are i.i.d., then

$$X_n = (1+r)X_{n-1} + \theta_n^\top(S_n - (1+r)S_{n-1}) = (1+r)X_{n-1} + \theta_n^\top S_{n-1}(\xi_n - (1+r))$$

So (X_n) is a controlled Markov process with $U_n = \theta_n^\top S_{n-1}$ and $G(x, u, v) = (1 + r)x + u(v - (1 + r))$.

Motivated by the example, we are interested in the optimisation of

$$\mathbb{E}\left(\sum_{k=1}^{N} f(k, U_k) + g(X_N)\middle| X_0 = x\right)$$

over controls $(U_k)_{1\le k\le N}$.

Definition 2.17. The value function of the problem is defined as

$$V(n, x) = \sup_{(U_k)} \mathbb{E}\left(\sum_{k=n+1}^{N} f(k, U_k) + g(X_N)\middle| X_n = x\right)$$

We know that $V(N, x) = g(x)$ and we want the value of $V(0, x)$. This is screaming dynamical programming, so let's do some of that. A hunch of this flavour is Bellman's equation

$$V(n-1, x) = \sup_u (f(n, u) + \mathbb{E}V(n, G(x, u, \xi)))$$

where $X_n = G(X_{n-1}, U_n, \xi_n)$.

Theorem 2.15. *Suppose f, g, G are given and V solves Bellman's equation*

$$V(n-1, x) = \sup_u (f(n, u) + \mathbb{E}V(n, G(x, u, \xi)))$$

subject to $V(N, x) = g(x)$. Suppose $u^*(n, x)$ is the optimal solution to the maximisation of $f(n, u) + \mathbb{E}V(n, G(x, u, \xi))$. Write $U_n^* = u^*(n, X_{n-1}^*), X_n^* = G(X_{n-1}^*, U_n^*, \xi_n)$ with $X_0^* = x$ given. Assuming everything is integrable, then $(U_n^*)_{1 \le n \le N}$ is the optimal control for the optimisation of

$$\mathbb{E}\left(\sum_{k=1}^N f(k, U_k) + g(X_N) \middle| X_0 = x\right)$$

In particular, the optimal control also has the Markov property.

Proof. Fix $X_0 = x$. Let $(U_n)_{1 \le n \le N}$ be a previsible process and set $X_n = G(X_{n-1}, U_n, \xi_n)$. Let $M_n = \sum_{k=1}^n f(k, U_k) + V(n, X_n)$. We shall show that $(M_n)_n$ is a supermartingale. It is clearly adapted and we have assumed integrability. For the supermartingale property, $\mathbb{E}(M_n - M_{n-1} | \mathcal{F}_{n-1}) = \mathbb{E}(f(n, u_n) + V(n, X_n) - V(n-1, X_{n-1}) | \mathcal{F}_{n-1}) = f(n, U_n) + \mathbb{E}(V(n, X_n) | \mathcal{F}_{n-1}) - V(n-1, X_{n-1}) \le 0$ since V satisfies Bellman's equation. So

$$\mathbb{E}\left(\sum_{k=1}^N f(k, U_k) + g(X_N) \middle| \mathcal{F}_n\right) = \mathbb{E}(M_n | \mathcal{F}_n) \le M_n = \sum_{k=1}^n f(k, U_k) + V(n, X_n)$$

That is,

$$V(n, X_n) \ge \mathbb{E}\left(\sum_{k=n+1} f(k, U_k) + g(X_N) \middle| X_n\right)$$

with equality iff $U = U^*$, so V is the value function and U^* is optimal. □

Consider $X_n = X_{n-1}(1+r) + \theta_n^\top (S_n - (1+r)S_{n-1})$ with $d = 1$ and $S_n = S_{n-1}\xi_n$ where $(\xi)_n$ are i.i.d. copies of some random variable ξ. Then $X_n = X_{n-1}(1+r) + \eta_n(\xi_n - (1+r))$ where $\eta_n = \theta_n S_{n-1}$ is a controlled Markov process.

Example 2.12. We want to maximise $\mathbb{E}(U(X_N) | X_0 = x)$. Bellman's equation becomes

$$V(n-1, x) = \sup_\eta \mathbb{E}V(n, (1+r)x + \eta(\xi - (1+r)))$$

subject to $V(N, x) = U(x)$. This is pretty much all we can do without further assumptions, so let's make some. Let $U(x) = -e^{-\gamma x}$ be a CARA utility. Then an educated guess would be $V(n, x) = U(x(1+r)^{N-n})A_n$ for some constant A_n. This clearly holds for $n = N$ with $A_N = 1$. Suppose the formula holds for some n, then

$$\begin{aligned} V(n-1, x) &= \sup_\eta \mathbb{E}(A_n U(((1+r)x + \eta(\xi - (1+r)))(1+r)^{N-n})) \\ &= U((1+r)^{N-n}x) A_n \inf_t \mathbb{E}e^{t(\xi - (1+r))} \end{aligned}$$

So the induction holds if $A_n = \alpha^{N-n}$ where $\alpha = \inf_t \mathbb{E}e^{t(\xi-(1+r))}$. Consequently, $\theta_n^* = t^*/(\gamma(1+r)^{N-n}S_{n-1})$ where $t = t^*$ is the minimiser of $\mathbb{E}e^{t(\xi-(1+r))}$.

What if we allow the investor to consume C_n (which gives a certain utility) during the interval $(n-1, n]$? The model then becomes

$$\begin{aligned} X_n &= (1+r)(X_{n-1} - C_n) + \theta_n^\top (S_n - (1+r)S_{n-1}) \\ &= (1-r)(X_{n-1} - C_n) + \eta_n(\xi_n - (1+r)) \end{aligned}$$

where as usual we care about the case $d = 1, S_n = \xi_n S_{n-1}$.

Example 2.13. We want to maximise

$$\mathbb{E}\left(\sum_{k=1}^{N} U(C_k) + U(X_N) \middle| X_0 = x\right)$$

with Bellman equation

$$V(n-1, x) = \sup_{C,\eta}(U(C) + \mathbb{E}V(n, (1+r)(x-C) + \eta(\xi - (1+r))))$$

subject to $V(N, x) = U(x)$. Suppose $U(x) = (1-R)^{-1}x^{1-R}$ is a CRRA utility with $R > 0, R \neq 1$. We guess the solution has a separated form $V(n, x) = U(x)A_n, A_N = 1$. After a disgusting amount of pointless algebra, we get

$$V(n-1, x) = x^{1-R} \sup_s (U(s) + A_n U(1-s)\alpha)$$

where $\alpha = (1-R)\sup_t \mathbb{E}U((1+r) + t(\xi - (1+r)))$. The maximiser of s is, of course, $s^* = 1/(1 + (A_n x)^{1/R})$ and we eventually get

$$A_n = \left(\frac{1 - \alpha^{(N-n+1)/R}}{1 - \alpha^{1/R}}\right)^R$$

So

$$C_n^* = \frac{X_{n-1}^*}{1 + (A_n x)^{1/R}}, \theta_n^* = \frac{\eta_n^*}{S_{n-1}} = \frac{t^*(X_{n-1}^* - X_n^*)}{S_{n-1}}$$

where $t = t^*$ is the maximiser of $\mathbb{E}U((1+r) + t(\xi - (1+r)))$.

2.7 Infinite Horizon Problems

Consider a controlled Markov process $X_n = G(X_{n-1}, U_n, \xi_n)$ where $(\xi_n)_n$ are i.i.d. copies of ξ. The goal is to maximise

$$\mathbb{E}\left(\sum_{k=1}^{\infty} \beta^{k-1} f(U_k) \middle| X_0 = x\right)$$

We can't really do dynamical programming anymore (phew) since we no longer have a boundary condition. But Bellman's equation (noooooooooo) can still be written down

$$V(x) = \max_u (f(u) + \beta \mathbb{E}V(G(x, u, \xi)))$$

Theorem 2.16. *Suppose $f(x) \geq 0$ for all n and $V(x) \geq 0$ solves the Bellman equation. Let $u^*(x)$ be the maximiser of $f(u) + \beta\mathbb{E}V(G(x, u, \xi))$. Consider the process $U_n^* = u^*(X_{n-1}^*), X_n^* = G(X_{n-1}^*, u_n^*, \xi_n)$ with $X_0^* = x$. Suppose $\beta^n \mathbb{E}V(X_n^*) \to 0$ as $n \to \infty$, then $(U_n^*)_{n\geq 1}$ is the optimal control and V the value function of the optimisation problem.*

Proof. GIven any control $(U_n)_{n\geq 1}$, let $(X_n)_n$ be its controlled process. Set

$$M_n = \sum_{k=1}^{n} \beta^{k-1} f(U_k) + \beta^n V(X_n)$$

which, as one can verify, is a supermartingale and is a martingale when $U = U^*$. So

$$\mathbb{E}\left(\sum_{k=1}^{n} \beta^{k-1} f(u_k)\right) = \mathbb{E}(M_n - \beta^n V(X_n)) \leq V(x) - \beta^n \mathbb{E} V(X_n)$$

with equality if $u = u^*$. By monotone convergence theorem,

$$\mathbb{E} \sum_{k=1}^{\infty} \beta^{k-1} f(u_k) = \lim_{n\to\infty} \mathbb{E} \sum_{k=1}^{n} \beta^{k-1} f(u_k)$$

As $V(x) \geq 0$, we have

$$\mathbb{E} \sum_{k=1}^{\infty} \beta^{k-1} f(u_k) \leq V(x)$$

with equality iff $u = u^*$ since $\beta^n \mathbb{E}(V(X_n^*)) \to 0$ as $n \to \infty$. $\square$

2.8 Optimal Stopping Problems

Let (X_n) be a Markov process (a priori not controlled). Fix constant "horizon" $N \geq 0$. We want to maximise $\mathbb{E} g(X_T)$ over stopping times $T \leq N$.
Suppose $X_n = G(X_{n-1}, \xi_n)$ where $(\xi_n)_n$ are i.i.d. copies of some ξ. Let $Z_n = X_{n\wedge T}$ where T is a stopping time $T \leq N$, so $Z_n = Z_{n-1} 1_{T\leq n-1} + G(Z_{n-1}, \xi_n) 1_{T\geq n}$. So this turns into an optimal control problem with control $U_n = 1_{T\leq n-1}$. Our goal is then to maximise $\mathbb{E} g(Z_N)$.
Bellman's equation becomes

$$V(n-1, x) = \max\{g(x), \mathbb{E} V(n, G(n, \xi))\}$$

subject to $V(N, x) = g(x)$.

Theorem 2.17. *Suppose V solves the Bellman's equation as stated above. Let $T^* = \min\{n \geq 0 : V(n, X_n) = g(X_n)\}$, then T^* is optimal.*

Proof. Pretty much identical as before. The process $M_n = V(n, X_n)$ is always a supermartingale. $V(n, x) \geq g(x)$ for all n, x, so for any stopping time $T \leq N$ we have $\mathbb{E} g(X_T) \leq \mathbb{E}(M_T) \leq M_0 = V(0, X_0)$ by Theorem 2.12. Now T^* is clearly a stopping time and $(M_{n\wedge T^*})_n$ is a martingale. So $\mathbb{E} g(X_{T^*}) = \mathbb{E} M_{T^*} = M_0 = V(0, X_0)$ by Theorem 2.7 and done. $\square$

2.9 Arbitrage in Multi-Period Models

Recall our usual setting: We have d assets with an adapted process $(S_n)_{n\geq 0}$, risk-free interest rate r_0 and a self-financing constraint $X_n = X_{n-1}(1+r) + \theta_r^\top (S_n - (1+r) S_{n-1})$ and $(\theta_n)_n$ is a previsible process valued in $\mathbb{R}$.
Fix a finite horizon N.

Definition 2.18. $(\theta_n)_{1\le n\le N}$ is an arbitrage if

$$\mathbb{P}\left(\sum_{k=1}^{N}\theta_k^\top\left(\frac{S_k}{(1+r)^k}-\frac{S_{k-1}}{(1+r)^{k-1}}\right)\ge 0\right)=1$$

and

$$\mathbb{P}\left(\sum_{k=1}^{N}\theta_k^\top\left(\frac{S_k}{(1+r)^k}-\frac{S_{k-1}}{(1+r)^{k-1}}\right)> 0\right)>0$$

Definition 2.19. A risk-neutral measure (or an equivalent martingale measure) is an equivalent measure $\mathbb{Q}$ under which the discounted prices $(S_n/(1+r)^n)_n$ are martingales.

That is, $\mathbb{E}^{\mathbb{Q}}(S_n|\mathcal{F}_{n-1})=(1+r)S_{n-1}$ for all $n\ge 1$.

Theorem 2.18 ((Multi-Period) Fundamental Theorem of Asset Pricing)**.** *In a finite-horizon market model, the existence of a risk-neutral measure is equivalent to the nonexistence of an arbitrage.*

Proof. Suppose $\theta=(\theta_n)_n$ is a previsible process with $X_N\ge 0$ $\mathbb{P}$-a.s.. If θ is bounded and there exists a risk-neutral measure $\mathbb{Q}$, then $X_N\ge 0$ $\mathbb{Q}$-a.s. since $\mathbb{P},\mathbb{Q}$ are equivalent. Also, $(X_n/(1+r)^n)_{1\le n\le N}$ is a $\mathbb{Q}$-martingale since it is the martingale transform of θ with respect to the martingale $(S_n/(1+r)^n)_n$. Hence $\mathbb{E}^{\mathbb{Q}}(X_N/(1+r)^N)=X_0=0$, so $X_N=0$ $\mathbb{Q}$-a.s., which then means that $X_N=0$ $\mathbb{P}$-a.s., so θ cannot be an arbitrage.
If θ is not necessarily bounded, we can use a stopping time trick to get the same result, which is sadly omitted here.
The other direction is similar as in the proof of Theorem 1.16. □

2.10 European Contingency Claims in Binomial Model

Take $d=1$ and $S_n=S_{n-1}\xi_n$ where $(\xi_n)_n$ are i.i.d. copies of some ξ where $\mathbb{P}(\xi=1+b)=p=1-\mathbb{P}(\xi=1+a)$ where $b>a$ and $p\in(0,1)$. This is called the Cox-Ross-Rubinstein binomial model.

Proposition 2.19. *There is no arbitrage iff $b>r>a$.*

Proof. For the "only if" direction, suppose $r\le a$, then $\theta_1=1$ is a one-period arbitrage. Similarly, if $r\ge b$, then $\theta_1=-1$ would be an arbitrage.
Conversely, we will construct a risk-neutral measure when $b>r>a$. Indeed, this is given uniquely by

$$\mathbb{Q}(\xi=1+b)=q=\frac{r-a}{b-a}=1-\mathbb{Q}(\xi=1-a)$$

It's an easy exercise to check this works and is indeed unique. □

Definition 2.20. A European contingency claim with expiry date N is an $\mathcal{F}_N$-measurable payout Y payable at time N.

Definition 2.21. We call a European contingency claim Y vanilla if $Y=g(S_N)$ for some function g.

Suppose we have a payout $Y=g(S_N)$ in the binomial model.

Theorem 2.20. *The unique time-n no-arbitrage price of Y is*

$$\pi_n = \frac{1}{(1+r)^{N-n}}\mathbb{E}^{\mathbb{Q}}(g(S_N)|\mathcal{F}_n)$$

Proof. $\pi_n/(1+r)^n$ is a $\mathbb{Q}$-martingale and the risk-neutral measure $\mathbb{Q}$ is unique. □

Theorem 2.21. *Let*

$$V(a,s) = \frac{1}{(1+r)^{N-n}}\mathbb{E}^{\mathbb{Q}}(g(S_N)|S_n = s)$$

$$\theta_n = \frac{V(n, S_{n-1}(1+b)) - V(n, S_{n-1}(1+a))}{S_{n-1}(b-a)}$$

Let $X_0 = V(0, S_0)$ and $X_n = X_{n-1}(1+r) + \theta_n(S_n - (1+r)S_{n-1})$, then $X_N = g(S_N)$.

$(\theta_n)_n$ are called the replicating strategies.

Remark. Note that $V(n, S_n) = \pi_n$ by the Markov property of $(S_n)_n$ under $\mathbb{Q}$. From the general theory of Markov chains, V satisfy the recurrence $(1+r)V(n-1, s) = qV(n, s(1+b)) + (1-q)V(n, s(1+a))$ subject to $V(N, s) = g(s)$.

Proof. Suppose $X_{n-1} = V(n-1, S_{n-1})$, then

$$\begin{aligned}
X_n &= (1+r)V(n-1, S_{n-1}) + \theta_n(S_n - (1+r)S_{n-1}) \\
&= qV(n, \xi_{n-1}(1+b)) + (1-q)V(n, \xi_{n-1}(1+a)) \\
&\quad + \frac{V(n, \xi_{n-1}(1+b)) - V(n, \xi_{n-1}(1+a))}{b-a}(\xi_n - (1+r)) \\
&= V(n, \xi_{n-1}(1+a))\frac{(1+b) - \xi_n}{b-a} + V(n, \xi_{n-1}(1+b))\frac{\xi_n - (1+a)}{b-a} \\
&= V(n, S_n)
\end{aligned}$$

So by induction $X_n = V(n, S_n)$ for all n, in particular for $n = N$. □

Let's take the concrete example where Y is a call option.

Definition 2.22. A (European) call option is the right, but not the obligation, to buy a given asset at a fixed price (strike) K at a fixed time N (expiry/maturity date).

More precisely, we are just looking for $Y = g(S_N)$ where $g(s) = (s-K)_+$.

Theorem 2.22. *Let $g(s) = (s-K)_+$ and θ_n be its replicating strategy. Then $\theta_n \in [0,1]$ and*

$$\frac{\theta_{n+1} - \theta_n}{S_n - (1+r)S_{n-1}} \geq 0$$

Proof. Let

$$V(n,s) = \frac{1}{(1+r)^{N-n}}((S_n - K)_+|S_n = s)$$

Note that $S_N = S_n Z$ where $Z = \xi_{n+1} \cdots \xi_N$. So

$$V(n,s) = \frac{1}{(1+r)^{N-n}} \mathbb{E}^{\mathbb{Q}}((sZ - K)_+)$$

Since $s \mapsto (sZ - K)_+$ is increasing and convex, so is $s \mapsto V(n,s)$, so $\theta_n \geq 0$. On the other hand

$$\begin{aligned} V(n,s) &= \frac{1}{(1+r)^{N-n}} \mathbb{E}^{\mathbb{Q}}(S_N - K + (K - S_N)_+ | S_n = s) \\ &= s - \frac{K}{(1+r)^{N-n}} + \frac{1}{(1+r)^{N-n}} \mathbb{E}^{\mathbb{Q}}((K - Zs)_+) \end{aligned}$$

As $S_n/(1+r)^n$ is a $\mathbb{Q}$-martingale, $s \mapsto V(n,s) - s$ is decreasing, so $\theta_n \leq 1$. The last inequality follows from (some calculations) and the convexity of $s \mapsto V(n,s)$. $\square$

How about put options?

Definition 2.23. A (European) put option is the right, but not the obligation, to sell a given asset at a fixed price (strike) at a fixed time N (expiry/maturity date).

That is, $Y = g(S_N)$ where $g(s) = (K - s)_+$.

Proposition 2.23. *A European put option of strike K can be replicated by holding 1 European call option of the same strike and expiry date (or the replication of such a call option), shorting one stock, and putting $K/(1+r)^{N-n+1}$ in the bank during time $(n-1, n]$. Consequently,the no-arbitrage price P_n of the put option equals $K/(1+r)^{N-n} - S_n + C_n$ (the "put-call parity formula") where C_n is the no-arbitrage price of the corresponding call option.*

Proof. $(K - s)_+ = K - s + (s - K)_+$. $\square$

2.11 American Contingency Claims

Fix a horizon (expiry date) N.

Definition 2.24. An American contingency claim is an adapted process $(Y_n)_n$ where Y_n is the payout of the claim if it is exercised at time n.

Definition 2.25. An American call option with strike K is given by $Y_n = (S_n - K)_+$ and an American put option with strike K is given by $Y_n = (K - S_n)_+$.

The time-n price π_n of an American claim $(Y_n)_n$ is given by

$$\pi_n = \max_{T \text{ stopping time in } [n,N]} \mathbb{E}^{\mathbb{Q}} \left(\frac{1}{(1+r)^T} Y_T \middle| \mathcal{F}_n \right)$$

in a complete market with risk-neutral measure $\mathbb{Q}$. Of course,

$$\pi_{n-1} = \max \left\{ Y_{n-1}, \frac{1}{1+r} \mathbb{E}^{\mathbb{Q}}(\pi_n | \mathcal{F}_{n-1}) \right\}$$

and an optimal stopping time is given by $T^* = \min\{n \in [0, N] : \pi_n = Y_n\}$.

Proposition 2.24. *Suppose $(Y_n/(1+r)^n)_n$ is a submartingale under $\mathbb{Q}$, then it is optimal to let the contingency claim expire.*

Proof. We claim that $\pi_n = \mathbb{E}(Y_N/(1+r)^{N-n}|\mathcal{F}_n)$. Indeed, this is true for $n = N$ as $\pi_N = Y_N$. Suppose it is true for n, then

$$\begin{aligned}\pi_{n-1} &= \max\left\{Y_{n-1}, \frac{1}{1+r}\mathbb{E}^{\mathbb{Q}}(\pi_n|\mathcal{F}_{n-1})\right\}\\ &= \max\left\{Y_{n-1}, \frac{1}{(1+r)^{N-n+1}}\mathbb{E}^{\mathbb{Q}}(Y_N|\mathcal{F}_{n-1})\right\}\\ &= \mathbb{E}^{\mathbb{Q}}\left(\frac{Y_N}{(1+r)^{N-n+1}}\middle|\mathcal{F}_{n-1}\right)\end{aligned}$$

by the submartingale property. $\square$

Example 2.14. $(S_n - K)_+/(1+r)^n$ is a submartingale, so NEVER exercise an American call early!

3 Continuous-Time Models

There really is no need to motivate why we want continuous-time model – we all are longing for it for a while now.

3.1 Brownian Motion

Consider the binomial model $S_n = S_0\xi_1\cdots\xi_n$ with $(\xi_i)_i$ i.i.d.. For $\delta > 0$ small, we set $t = n\delta$ and $\log\hat{S}_t = \log S_0 + \mu t + \sigma W_t$ where $\mu = \delta^{-1}\mathbb{E}(\log\xi)$, $\sigma^2 = \delta^{-1}\operatorname{Var}(\log\xi)$, and of course

$$W_t = \frac{\log(\hat{S}_t/S_0) - \mu t}{\sigma}$$

$W_t - W_s$ (for $t > s$) should be independent of $(W_u)_{0\le u\le s}$, so the central limit theorem should give some results of the form $(W_t - W_s)/\sqrt{t-s} \approx \mathcal{N}(0,1)$.

Definition 3.1. A (standard) Brownian motion is a process $(W_t)_{t\ge 0}$ (i.e. a collection of random variables indexed by $\mathbb{R}_{\ge 0}$) such that $W_0 = 0$, $W_t - W_s$ is independent of $(W_u)_{0\le u\le s}$, $W_t - W_s \sim \mathcal{N}(0, t-s)$, and $(W_t)_{t\ge 0}$ is continuous (in the sense that $t \mapsto W_t$ is a.s. continuous).

Theorem 3.1 (Wiener 1923)**.** *Brownian motion exists.*

Proof. Omitted. $\square$

Definition 3.2. A process $(X_t)_{t\ge 0}$ is Gaussian if $(X_{t_1}, \ldots, X_{t_n})$ is Gaussian for every $0 \le t_1 < \ldots < t_n$.

Theorem 3.2. *A continuous process $(W_t)_{t\ge 0}$ is a Brownian motion iff it is a Gaussian process with $\mathbb{E}W_t = 0, \mathbb{E}(W_sW_t) = \min\{s,t\}$ for any $s,t \ge 0$.*

Proof. Suppose (W_t) is a Brownian motion, then it is clearly Gaussian since $W_{t_2} - W_{t_1}, \ldots, W_{t_n} - W_{t_{n-1}}$ are independent normals for $0 \leq t_1 < \ldots < t_n$. $\mathbb{E}W_t = 0$ as $W_t \sim \mathcal{N}(0,t)$ and (suppose $s < t$) $\mathbb{E}(W_s W_t) = \mathbb{E}(W_s^2) + \mathbb{E}(W_s)\mathbb{E}(W_t - W_s) = s$.
Conversely, suppose (W_t) is Gaussian and $\mathbb{E}W_t = 0, \mathbb{E}(W_s W_t) = \min\{s,t\}$ for any $s, t \geq 0$. Then whenever $t > s$, $W_t - W_s$ is normal with mean $\mathbb{E}(W_t - W_s) = 0$ and variance $\mathrm{Var}((W_t - W_s)^2) = t - 2s + s = t - s$. So indeed $W_t - W_s \sim \mathcal{N}(0, t-s)$. Now any $u < s < t$ would have $\mathbb{E}((W_t - W_s)W_u) = u - u = 0$, so $W_t - W_s$ is uncorrelated with $(W_u)_{0 \leq u < s}$, which means that they are independent as they are Gaussians. □

Theorem 3.3. *If $(W_t)_{t \geq 0}$ is a Brownian motion, so are:*
(i) $\tilde{W}_t = cW_{t/c^2}$ for $c \in \mathbb{R}$.
(ii) $\tilde{W}_t = W_{t+T} - W_T$ for $T \geq 0$.
(iii) $\tilde{W}_t = tW_{1/t}, \tilde{W}_0 = 0$.

Proof. Check all of them are Gaussian with the correct mean and covariance. The only technicality is to check that $W_t/t \to 0$ a.s. as $t \to \infty$ in (iii), which requires a form of strong law of large numbers. □

Theorem 3.4. *Any Brownian motion is a (continuous-time) Markov process.*

Proof. Let $(W_t)_{t \geq 0}$ be a Brownian motion and $\mathcal{F}_s = \sigma(W_u : u \leq s)$.

$$\begin{aligned}\mathbb{E}(g(W_t)|\mathcal{F}_s) = \mathbb{E}(g(W_t - W_s + W_s)|\mathcal{F}_s) &= \int_{\mathbb{R}} g(z + W_s) \frac{e^{-z^2/2(t-s)}}{\sqrt{2\pi(t-s)}} \, \mathrm{d}z \\ &= \mathbb{E}(g(W_t)|W_s)\end{aligned}$$

which is the Markov property. □

Definition 3.3. A collection of σ-algebras $(\mathcal{F}_t)_t$ indexed by $\mathbb{R}_{\geq 0}$ is a filtration if $\mathcal{F}_s \subset \mathcal{F}_t$ whenever $s < t$. The filtration generated by a process $(X_t)_{t \geq 0}$ is $\mathcal{F}_t = \sigma(X_s : s \leq t)$.

Definition 3.4. A continuous process $(X_t)_{t \geq 0}$ is a martingale with respect to the filtration $(\mathcal{F}_t)_{t \geq 0}$ if $\mathbb{E}(X_t|\mathcal{F}_s) = X_s$ for all $0 \leq s \leq t$.

Theorem 3.5. *Any Brownian motion is a martingale with respect to the filtration generated by itself.*

Proof. $\mathbb{E}(W_t|\mathcal{F}_s) = \mathbb{E}(W_t - W_s + W_s|\mathcal{F}_s) = \mathbb{E}(W_t - W_s) + W_s = W_s$ whenever $0 \leq s \leq t$. □

Theorem 3.6. *A continuous process $(W_t)_{t \geq 0}$ with $W_0 = 0$ is a Brownian motion iff $(e^{\theta W_t - \theta^2 t/2})_{t \geq 0}$ is a martingale for all $\theta \in \mathbb{R}$.*

Proof. The "only if" part is clear. Conversely, suppose $(e^{\theta W_t - \theta^2 t/2})_{t \geq 0}$ is a martingale, then $\mathbb{E}(e^{\theta(W_t - W_s)}|\mathcal{F}_s) = e^{\theta^2(t-s)/2}$, so $W_t - W_s \sim \mathcal{N}(0, t-s)$ and is independent of $\mathcal{F}_s$. □

Theorem 3.7 (Lévy)**.** *A continuous process $(W_t)_{t \geq 0}$ with $W_0 = 0$ is a Brownian motion iff $(W_t)_{t \geq 0}$ and $(W_t^2 - t)_{t \geq 0}$ are martingales.*

Proof. Omitted. $\square$

Theorem 3.8. *Let $T_a = \inf\{t \geq 0 : W_t = a\}$, then $T_a < \infty$ a.s. for any $a \in \mathbb{R} \setminus \{0\}$.*

Proof. We'll deal with the situation where $a > 0$. The other case is similar. We want to show that $\sup_{t\geq 0} W_t > a$ a.s. for all $a > 0$. Indeed, for any $c > 0$,

$$\mathbb{P}\left(\sup_{t\geq 0} W_t > a\right) = \mathbb{P}\left(\sup_{t\geq 0} \tilde{W}_{tc^2} > ca\right) = \mathbb{P}\left(\sup_{t\geq 0} W_t > ca\right)$$

where $\tilde{W}_t = cW_{t/c^2}$. Sending $c \downarrow 0$ shows that $Z = \sup_{t\geq 0} W_t \in \{0, \infty\}$. $\hat{Z} = \sup_{t\geq 1}(W_t - W_1)$ has the same law as Z. We have $p = \mathbb{P}(Z = 0) = \mathbb{P}(Z = 0, \hat{Z} = 0) \leq \mathbb{P}(W_1 \leq 0, \hat{Z} = 0) = p/2$, so $p = 0$. $\square$

3.2 Reflection Principle

Theorem 3.9. *Let T be a finite stopping time and W a Brownian motion. Then the process $\tilde{W}_t = W_{t+T} - W_T$ is also a Brownian motion.*

Proof. Omitted but based on strong Markov property. $\square$

Theorem 3.10 (Reflection Principle)**.** *Let W be a Brownian motion and set $T_a = \inf\{t \geq 0 : W_t = a\}$. Then $\tilde{W}_t = W_t 1_{t\leq T_a} + (2a - W_t)1_{t\geq T_a}$ is a Brownian motion.*

Proof. $\tilde{W}_{t+T_a} = W_{T_a} + (W_{T_a} - W_{t+T_a})$. But $W_{T_a} - W_{t+T_a}$ has the same law as $W_{t+T_a} - W_{T_a}$, so we are essentially done. $\square$

Lemma 3.11. *Let $M_t = \max_{0\leq s\leq t} W_s$, then $\mathbb{E}(g(W_t)1_{W_t\leq a, M_t\geq a}) = \mathbb{E}(g(2a - W_t)1_{W_t\geq a})$.*

Proof. Let $\tilde{W}_t = W_t 1_{t\leq T_a} + (2a - W_t)1_{t\geq T_a}$ and $\tilde{M}_t = \max_{0\leq s\leq t} \tilde{W}_s$, then $\{M_t \geq a\} = \{\tilde{M}_t \geq a\}$. Since $W_t = 2a - \tilde{W})t$ when $t \geq T_a$, we have

$$\begin{aligned}\mathbb{E}(g(W_t)1_{W_t\leq a, M_t\geq a}) &= \mathbb{E}(g(W_t)1_{W_t\leq a, \tilde{M}_t\geq a}) = \mathbb{E}(g(2a - \tilde{W}_t)1_{\tilde{W}_t\geq a, \tilde{M}_t\geq a}) \\ &= \mathbb{E}(g(2a - \tilde{W}_t)1_{\tilde{W}_t\geq a}) = \mathbb{E}(g(2a - W_t)1_{W_t\geq a})\end{aligned}$$

by the preceding theorem. $\square$

Corollary 3.12. *Whenever $a \geq b$ and $a \geq 0$, $\mathbb{P}(W_t \leq b, M_t \geq a) = \mathbb{P}(W_t \geq 2a - b)$.*

Proof. Take $g(w) = 1_{w\leq b}$. $\square$

Theorem 3.13. *For $a \geq b, a \geq 0$,*

$$f_{M_t, W_t}(a, b) = -f'_{W_t}(2a - b) = \frac{2(2a - b)}{t^{3/2}\sqrt{2\pi}} e^{-(2a-b)^2/(2t)}$$

Proof. We have

$$\begin{aligned}\mathbb{P}(M_t \geq a) &= \mathbb{P}(M_t \geq a, W_t \geq a) + \mathbb{P}(M_t \geq a, W_t \leq a) \\ &= \mathbb{P}(W_t \geq a) + \mathbb{P}(W_t \geq a) = 2\Phi(-a/\sqrt{t})\end{aligned}$$

which means that M_t has the same law as $|W_t|$ for all fixed t. We also have $\mathbb{P}(M_t \geq a) = \mathbb{P}(T_a \leq t)$, so this gives $f_{T_a}(t) = \frac{a}{t^{3/2}\sqrt{2\pi}} e^{-a^2/(2t)}$. $\square$

3.3 Cameron-Martin Theorem

Recall from example sheet that if $Z \sim \mathcal{N}(0,1)$ and $a \in \mathbb{R}$, then

$$\mathbb{E}(e^{aZ-a^2/2} f(Z)) = \mathbb{E} f(Z+a)$$

This of course has a multidimensional analogue given by

$$\mathbb{E}(e^{a^\top Z - \|a\|^2/2} f(Z)) = \mathbb{E} f(Z+a)$$

whenever $Z \sim \mathcal{N}_n(0, I), a \in \mathbb{R}^n$.

Theorem 3.14 (Cameron-Martin)**.** *Let $g : C([0,t]) \to \mathbb{R}$ be a sufficiently nice functional (measurable, integrable, etc.). Suppose W is a Brownian motion and $c \in \mathbb{R}$, then*

$$\mathbb{E} g((W_s + cs)_{0\le s\le t}) = \mathbb{E}(e^{cW_t - c^2 t/2} g((W_s)_{0\le s\le t}))$$

Proof. It suffices to consider the case where $g(w) = G(w(t_1), \ldots, w(t_n))$ by a monotone class argument. We can further modify this to the form

$$g(w) = H\left(\frac{W(t_1)}{\sqrt{t_1}}, \frac{W(t_2) - W(t_1)}{\sqrt{t_2 - t_1}}, \ldots, \frac{W(t_n) - W(t_{n-1})}{\sqrt{t_n - t_{n-1}}}\right)$$

Then if we write $Z_k = (W(t_k) - W(t_{k-1}))/\sqrt{t_k - t_{k-1}}$ (which are by definition i.i.d. $\mathcal{N}(0,1)$ random variables),

$$\begin{aligned}
\mathbb{E} g((W_s + cs)_{0\le s\le t}) &= \mathbb{E} H(Z_1 + c\sqrt{t_1}, \ldots, Z_n + c\sqrt{t_n - t_{n-1}}) \\
&= \mathbb{E}(e^{\sum_i c\sqrt{t_i - t_{i-1}} Z_i - \sum_i c^2 (t_i - t_{i-1})/2} H(Z_1, \ldots, Z_n)) \\
&= \mathbb{E}(e^{cW_t - c^2 t/2} g((W_s)_{0\le s\le t}))
\end{aligned}$$

as desired. □

Corollary 3.15.

$$\mathbb{P}\left(\max_{0\le s\le t}(W_s + cs) \le a\right) = \mathbb{E}\left(e^{cW_t - c^2 t/2} 1_{\max_{0\le s\le t} W_s \le a}\right)$$

Corollary 3.16. *Let $(W_t)_{t\ge 0}$ be a Brownian motion under a probability measure $\mathbb{P}$. Fix $c > 0$ and horizon $T > 0$. Set*

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = e^{cW_T - c^2 T/2}$$

Then $\tilde{W}_t = W_t - ct$ is a $\mathbb{Q}$-Brownian motion.

Proof.

$$\mathbb{E}^{\mathbb{Q}}(g((\tilde{W}_t)_{0\le t\le T})) = \mathbb{E}^{\mathbb{P}}(e^{cW_T - c^2 T/2} g((W_t - ct)_{0\le t\le T})) = \mathbb{E}^{\mathbb{P}}(g(W_t)_{0\le t\le T})$$

Consequently the law of $\tilde{W}$ under $\mathbb{Q}$ is exactly the same as the law of W under P, which is a Brownian motion. □

3.4 Black-Scholes Model

In the binomial model, we had $S_n = S_0\xi_1 \cdots \xi_n$ for i.i.d. $\xi_k \sim \xi$. We motivated continuous-time models by considering $\log \hat{S}_t = \log S_0 + \mu t + \sigma W^t$ where $t = n\delta, \mu = \delta^{-1}\mathbb{E}\log\xi, \sigma^2 = \delta^{-1}\operatorname{Var}(\log\xi)$. We (heuristically) argued that $(W_t)_{t\geq 0} \approx \mathcal{N}(0, t-s)$ as $\delta \to 0$.
This inspires us to take a Brownian motion $(W_t)_{t\geq 0}$ and study the continous process $S_t = S_0 e^{\mu t + \sigma W_t}$. This is the premise of the Black-Scholes model. μ is known as the drift and σ the volatility.
We haven't really talked about how interest rates should be modelled in continuous time. As a motivation, one might want to do this by writing down $(1+r)^n = (1+\hat{r}\delta)^{t/\delta} \approx e^{\hat{r}t}$ where $\hat{r} = r/\delta$. To put into practice, we let the bank account at time t be worth e^{rt} where r is a continuous-time compound interest rate.

Proposition 3.17. *There is a risk-neutral measure in the Black-Scholes model.*

That is, there is a probability measure under which $(e^{-rt}S_t)_{0\leq t\leq T}$ is a $\mathbb{Q}$-martingale.

Proof. Let $c = (r-\mu)/\sigma - \sigma/2$ and $\tilde{W}_t = W_t - ct$. Then $\mathbb{Q}$ as defined by

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = e^{cW_T - c^2T/2}$$

is risk-neutral since $S_t e^{-rt} = S_0 e^{-\sigma^2 t/2 + \sigma\tilde{W}_t}$ is a martingale as $\tilde{W}_t$ is a $\mathbb{Q}$-Brownian motion. □

Definition 3.5. Let Y be the payout of a European contingent claim with maturity horizon T. The Black-Scholes price π_t at time t is defined by

$$\pi_t = e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}(Y|\mathcal{F}_t)$$

where $\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P} = e^{cW_T - c^2T/2}, c = (r-\mu)/\sigma - \sigma/2$.

It follows that $(e^{-rt}\pi_t)_{0\leq t\leq T}$ is a $\mathbb{Q}$-martingale.
Suppose $Y = g(S_T)$ is a vanilla claim, then

$$\pi_t = e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}(g(S_t e^{(r-\sigma^2/2)(T-t)+\sigma(\tilde{W}_T - \tilde{W}_t)})|\mathcal{F}_t)$$

Write this expression as $V(t, S_t)$ which has an explicit formula

$$V(t,s) = \int_{-\infty}^{\infty} g(se^{(r-\sigma^2/2)(T-t)+\sigma\sqrt{T-t}z})\frac{e^{-z^2/2}}{\sqrt{2\pi}}\,\mathrm{d}z$$

When Y is a European call, i.e. $g(s) = (s-K)_+$, plugging this into the above formula gives $V(t,s) = s\Phi(d_1) - e^{-r(T-t)}K\Phi(d_2)$ (the Black-Scholes formula) where

$$d_1 = \frac{-\log(K/s)}{\sigma\sqrt{T-t}} + \left(\frac{r}{\sigma} + \frac{\sigma}{2}\right)\sqrt{T-t}, d_2 = \frac{-\log(K/s)}{\sigma\sqrt{T-t}} + \left(\frac{r}{\sigma} - \frac{\sigma}{2}\right)\sqrt{T-t}$$

How did Black and Scholes get their price formula for π_t in the first place? Like above we set $V(t,s)$ be the Black-Scholes price of a vanilla claim when $S_t = s$. Then V solves

$$\frac{\partial V}{\partial t} + rs\frac{\partial V}{\partial s} + \frac{1}{2}\sigma^2 s^2\frac{\partial^2 V}{\partial s^2} = rV$$

subject to $V(T,s) = g(s)$. This is known as the Black-Scholes PDE. We will justify the formula for the Black-Scholes price by deriving this PDE from the binomial model.

3.5 The Binomial Heuristics for Black-Scholes PDE

As usual, we set $\mathbb{E}\log\xi = \mu\delta, \mathrm{Var}(\log\xi)$. So the binomial parameters are $a = -\sigma\sqrt{\delta} + \mu\delta + O(\delta^2), b = \sigma\sqrt{\delta} + \mu\delta + O(\delta^2), r^{\text{bin}} = \delta r$ and therefore $q = (r\delta - a)/(b-a)$. We then have $(1+r\delta)V(t-\delta, s) = qV(t, s(1+b)) + (1-q)V(t, s(1+a))$ which gives

$$\begin{aligned}
V + (rV - \frac{\partial V}{\partial t})\delta = q &\left(V + \frac{\partial V}{\partial s} sb + \frac{1}{2}\frac{\partial^2 V}{\partial s^2}(sb)^2\right) \\
&+ (1-q)\left(V + \frac{\partial V}{\partial s} sa + \frac{1}{2}\frac{\partial^2 V}{\partial s^2}(sa)^2\right) \\
&= V + \frac{\partial V}{\partial s} s(qb + (1-q)a) + \frac{1}{2}\frac{\partial^2 V}{\partial s^2} s^2(qb^2 + (1-q)a^2) \\
&= V + \frac{\partial V}{\partial s} sr\delta + \frac{1}{2}\frac{\partial^2 V}{\partial s^2} s^2\sigma^2\delta
\end{aligned}$$

up to $o(\delta)$. Taking $\delta \to 0$ gives the Black-Scholes PDE.
$\partial V/\partial S$ is called the Black-Scholes Delta (of a vanilla European claim), which is approximately the binomial replicating portfolio.

Definition 3.6. Delta-hedging is the strategy of holding the Black-Scholes Delta amount of the underlying stock.

Example 3.1. The Black-Scholes Delta of a European call is $\Phi(d_1)$ where, as usual,

$$d_1 = \frac{-\log(K/s)}{\sigma\sqrt{T-t}} + \left(\frac{r}{\sigma} + \frac{\sigma}{2}\right)\sqrt{T-t}$$

$\partial^2 V/\partial s^2$ is called the Black-Scholes Gamma, which indicates the amount of share one should buy if the underlying price moves.

Definition 3.7. If g is increasing, then the Black-Scholes Delta is nonnegative; If g is convex, then the Black-Scholes Gamma is nonnegative.

Proof. Recall that

$$V(t,s) = e^{-r(T-t)}\mathbb{E}g(se^{(r-\sigma^2/2)(T-t)+\sigma\sqrt{T-t}Z})$$

for $Z \sim \mathcal{N}(0,1)$. □

Delta and Gamma are examples of Black-Scholes Greeks, which are the partial derivatives of V. They are usually, but not always, denoted by Greek letters. When they are not, they are usually named after television celebrities.

3.6 The Heat Equation Heuristics

Proposition 3.18. *Fix suitable f and let $u(\tau, x) = \mathbb{E}f(x + \sqrt{\tau}Z)$, then $u_\tau = u_{xx}/2$.*

Proof. Move the derivative inside. □

Suppose V is the value function given by the Black-Scholes formula. Let

$$u(t,x) = e^{r\tau/\sigma^2} V\left(T - \frac{\tau}{\sigma^2}, e^{x-(r/\sigma^2-1/2)\tau}\right) = \mathbb{E}g(e^{x+\sqrt{\tau}Z})$$

Then u solves the heat equation. Reverse engineering this process gives the very same Black-Scholes PDE.

3.7 Exotic Claims

Example 3.2 (Forward Strike Call Option)**.** We now set the strike price to be S_{T_0} for maturity $T_1 > T_0$, i.e. the payout is $(S_{T_1} - S_{T_0})_+$. Our previous discussion solves to the Black-Scholes price for $t \geq T_0$, when it essentially reduces to a call with known strike $K = S_{T_0}$. Before T_0, the Black-Scholes price should be

$$\begin{aligned}
\pi_t &= e^{-r(T_1-t)}\mathbb{E}^{\mathbb{Q}}((S_{T_1} - S_{T_0})_+ | \mathbb{F}_t) \\
&= e^{-r(T_1-t)}\mathbb{E}^{\mathbb{Q}}\left(S_{T_0}\left(\frac{S_{T_1}}{S_{T_0}} - 1\right)_+ \middle| \mathcal{F}_t\right) \\
&= e^{-r(T_1-t)}\mathbb{E}^{\mathbb{Q}}(S_{T_0}|\mathcal{F}_t)\mathbb{E}^{\mathbb{Q}}\left(\frac{S_{T_1}}{S_{T_0}} - 1\right)_+ \\
&= e^{-r(T_1-t)}e^{r(T_0-t)}S_t\mathbb{E}^{\mathbb{Q}}\left(\frac{S_{T_1}}{S_{T_0}} - 1\right)_+
\end{aligned}$$

Consider a European claim with payout Y. Then "up-and-in" version of the claim is a claim whose payout has the form $Y1_{\max_{0\leq t\leq T} S_t \geq B}$ where T is the maturity date and B is a fixed barrier. The "up-and-out" version is $Y1_{\max_{0\leq t\leq T} S_t < B}$, "down-and-in" is $Y1_{\min_{0\leq t\leq T} S_t < B}$, and "down-and-out" is $Y1_{\min_{0\leq t\leq T} S_t \geq B}$.
So for example an "up-and-in" European call is the right, but not the obligation, to buy the stock for price K at time T, assuming that the stock price exceeds B at some earlier time.

Proposition 3.19. *In the Black-Scholes model, the initial price of an up-and-out vanilla claim has payout $g(S_T)1_{\max_{0\leq t\leq T} S_t < B}$ is the same as the initial price of the vanilla claim with payout*

$$g(S_T)1_{S_T<B} - \left(\frac{B}{S_0}\right)^{2r/\sigma^2-1} g\left(\frac{B^2 S_T}{S_0^2}\right) 1_{S_T<S_0^2/B}$$

Proof. The Black-Scholes price of a European claim Y is $e^{-rt}\mathbb{E}^{\mathbb{Q}}(Y)$, so it suffices to show that the two random variable has the same $\mathbb{Q}$-expectation. Note that $\{\max_t S_t < B\} \subset \{S_T < B\}$, so $g(S_T)1_{\max_t S_t<B} = g(S_T)1_{S_T<B} - g(S_T)1_{S_T<B, \max_t S_t \geq B}$. Writing $\mathbb{E} = \mathbb{E}^{\mathbb{Q}}$, $b = \log(B/S_0)/\sigma$ and $S_t = S_0 e^{\sigma(W_t+ct)}$

where $c = r/\sigma - \sigma/2$, W is a $\mathbb{Q}$-Brownian motion, then

$$\begin{aligned}
\mathbb{E}(g(S_T)1_{S_T<B,\max_t S_t\geq B}) &= \mathbb{E}(g(S_0e^{\sigma(W_T+cT)})1_{W_T+cT<B,\max_t(W_t+ct)\geq B}) \\
&= \mathbb{E}(e^{cW_T-c^2T/2}g(S_0e^{\sigma W_T})1_{W_T<b,\max_t W_t\geq b}) \\
&= \mathbb{E}(e^{c(2b-W_T)-c^2T/2}g(S_0e^{\sigma(2b-W_T)})1_{W_T>b}) \\
&= e^{2bc}\mathbb{E}(e^{cW_T-c^2T/2}g(e^{2b\sigma}S_0e^{\sigma W_T})1_{W_T<-b}) \\
&= e^{2bc}\mathbb{E}(g(e^{2b\sigma}S_0e^{\sigma(W_T+cT)})1_{W_T+cT\leq -b})
\end{aligned}$$

by Theorem 3.14. Translating this to the original notations gives the result. $\square$

3.8 Numerical Schemes

We are interested in the Black-Scholes PDE which can be modified into the heat equation after changing variables. It certainly would be useful if we can find a good numerical scheme to solve it.
To be precise, the problem we're trying to solve is $u_t = u_{xx}/2$ where $u(0,x) = h(x)$ for $x \in [0, 2\pi]$, $u(t,0) = a(t), u(t, 2\pi) = b(t)$.
We will use the grid method. Fix step sizes Δt and $\Delta x = 2\pi/N$ where N is an integer. We want to approximate $u(n\Delta t, k\Delta x)$ by $U_{n,k}$ where $U_{0,k} = h(k\Delta x), U_{n,0} = a(n\Delta t), U_{n,N} = b(n\Delta t)$. To obtain a sensible recursive scheme for U, we certainly want a way to discretise derivatives.
There are three most popular ways of doing it, namely

$$\frac{U_{n+1,k} - U_{n,k}}{\Delta t} = \frac{U_{n,k+1} - 2U_{n,k} + U_{n,k-1}}{2(\Delta x)^2}$$

$$\frac{U_{n+1,k} - U_{n,k}}{\Delta t} = \frac{U_{n+1,k+1} - 2U_{n+1,k} + U_{n+1,k-1}}{2(\Delta x)^2}$$

$$\frac{U_{n+1,k} - U_{n,k}}{\Delta t} = \frac{U_{n,k+1} - 2U_{n,k} + U_{n,k-1}}{4(\Delta x)^2} + \frac{U_{n+1,k+1} - 2U_{n+1,k} + U_{n+1,k-1}}{4(\Delta x)^2}$$

Where the first is known as the forward-in-time method, the second is the backwards-in-time method, and the third, which is the average of the two, is known as the Crank-Nicolson method. The forward-in-time method is very natural, but it has numerical disadvantage.

Definition 3.8. A numerical scheme for the heat equation is stable (in the von Neumann sense) if $\sup_\beta |r(\beta)| \leq 1$ where $r(\beta)$ is the unique real number such that $U_{n,k} = r(\beta)^n e^{i\beta k}$ is a solution to the recursion defining the scheme.

For the forward-in-time method $r(\beta) = 1 - \nu(1 - \cos(\beta\Delta x))$ where $\nu = \Delta t/(\Delta x)^2$, which means that it is not stable unless $\nu \leq 1$. The backward-in-time method however is unconditionally stable as it has $r(\beta) = (1 + \nu(1 - \cos(\beta\Delta x)))^{-1}$. The Crank-Nicolson method is also always stable with $r(\beta) = (2 - 2(1 - \cos(\beta\Delta t)))/(2 + 2(1 - \cos(\beta\Delta x)))$.