

Part II - Stochastic Financial Models

Lectured by Dr M. R. Tehranchi

Artur Avameri

Michaelmas 2022

Contents

0	Introduction	2
1	One-period models	2
1.1	Mean-variance analysis	3
1.2	Capital Asset pricing model	4
1.3	Expected utility	6
1.4	Risk-aversion	7
1.5	Risk-neutral measures	9
1.6	Contingent claims	11
1.6.1	Indifference price of contingent claims.	11
1.6.2	Marginal utility pricing	13
1.7	Arbitrage	14
1.7.1	Fundamental theorem of asset pricing	15
1.8	No-arbitrage pricing	17
1.9	Attainable claims	18
2	Multiperiod models	20
2.1	Introduction to multiperiod models	20
2.2	Measurability	20
2.2.1	Properties of conditional expectations	23
2.3	Adaptedness, filtrations, martingales	24
2.3.1	Motivation for martingales in finance	25
2.3.2	Martingale transforms	26
2.4	Stopping times	27
2.5	Submartingales and supermartingales	30
2.6	Controlled Markov processes	31
2.7	The Bellman equation	32
2.7.1	Stochastic optimal control	32

0 Introduction

07 Oct 2022,
Lecture 1

The goal of the course is to discuss some financial models. We will put ourselves in the position of an investor and ask how we should invest. To do this, we first need to make some simplifying assumptions:

- No dividends will ever be paid.
- We can buy a continuous spectrum of shares, not only integer amounts.
- There is no bid-ask spread, i.e. each stock only has one price value.
- Our actions have no price impact.
- There are no transaction costs.
- There are no short-selling constraints.

For the next few lectures, we will only consider $t \in \{0, 1\}$, i.e. one-period models. After that, we will consider $t \in \{0, 1, \dots\}$, i.e. discrete-time models, and at the end of the course, $t \in \mathbb{R}_{\geq 0}$, i.e. continuous-time models.

1 One-period models

Setup: There are d risky assets, and the value of the i^{th} asset at time t is denoted by S_t^i , hence we have a vector of prices $S_t = (S_t^1, \dots, S_t^d)^T \in \mathbb{R}^d$. There is also a risk-free asset, which pays a constant interest rate r .

Introduce an investor who picks a portfolio at time 0. We say θ^i is the number of shares bought at time 0 (if $\theta^i < 0$, the investor shorts $|\theta^i|$ shares) and θ^0 is the number of shares the investor has of the risk-free asset.

Let X_t be the investor's wealth at time t . We hence have

$$\begin{aligned} X_0 &= \theta^0 + \sum_{i=1}^d \theta^i S_0^i = \theta^0 + \theta^T S_0 \\ X_1 &= \theta^0(1+r) + \theta^T S_1. \end{aligned}$$

Eliminating θ^0 , we get

$$X_1 = (1+r)X_0 + \theta^T(S_1 - (1+r)S_0).$$

Risk. r and S_0 are known constants. S_1 is unknown. We will also assume that we are certain of the distribution of S_1 .

1.1 Mean-variance analysis

Mean-variance portfolio problem. (Markovitz, 1952). Given x and m , find $\theta \in \mathbb{R}^d$ to minimize $\text{Var}(X_1)$ subject to $X_0 = x$ and $\mathbb{E}[X_1] \geq m$.

Draw a graph of variance and mean on the x - y plane. We get a subset on the right-hand side of the plane consisting of points of the form $(\text{Var}(X_1), \mathbb{E}[X_1])$.

Definition 1.1 (Mean-variance efficient frontier). The mean-variance efficient frontier is just the left boundary of the above set, i.e. for every value of m , the strategy that minimizes variance while achieving $\mathbb{E}[X_1] = m$.

Definition 1.2. A portfolio is **mean-variance efficient** if it is the optimal solution to a mean-variance portfolio problem for some x, m .

Notation/assumptions. Assume that S_1 is square integrable, let:

- $\mathbb{E}[S_1] = \mu \in \mathbb{R}^d$
- $\text{Cov}(S_1) = V$, a $d \times d$ matrix, which we assume is positive definite (it is always nonnegative definite, but we want it to be invertible). We have $V = \mathbb{E}[(S_1 - \mu)(S_1 - \mu)^T] = \mathbb{E}[S_1 S_1^T] - \mu \mu^T$.
- Assume also that $\mu \neq (1 + r)S_0$.

Theorem 1.1. The optimal solution to the mean-variance portfolio problem is

$$\theta^* = \lambda V^{-1}(\mu - (1 + r)S_0), \text{ where } \lambda = \frac{(m - (1 + r)x)^+}{(\mu - (1 + r)S_0)^T V^{-1}(\mu - (1 + r)S_0)}$$

Notation. For $a \in \mathbb{R}$, we have $a^+ = \max(0, a)$.

Before we prove this, let's dwell on it for a bit.

10 Oct 2022,
Lecture 2

Corollary 1.2 (Mutual fund theorem). A portfolio is mean-variance efficient if and only if it is a nonnegative scalar multiple of $\theta_{\text{mar}} = V^{-1}(\mu - (1 + r)S_0)$.

Definition 1.3. $\theta_{\text{mar}} = V^{-1}(\mu - (1 + r)S_0)$ is called the **market portfolio**.

Remark. What is the origin of the name "market portfolio"? Suppose there are n_i shares of asset i , $1 \leq i \leq d$. Suppose there are K investors and investor k holds portfolio θ_k . Market clearing (i.e. total supply equals total demand) implies $\sum_{k=1}^K \theta_k = n = (n_1 \ \dots \ n_d)^T$.

Make an additional assumption: assume all investors agree on μ and V and hold a mean-variance efficient portfolio. Then by the mutual fund theorem, $\theta_k = \lambda_k \theta_{\text{mar}}$ for some scalar $\lambda_k \geq 0$, hence $n = \Lambda \theta_{\text{mar}}$, where $\Lambda = \sum \lambda_k$. So θ_{mar} , the market portfolio, is really just some scaling of the entire market by some Λ .

Notation for the proof.

- $\mathbb{E}[X_1] = (1+r)x + \theta^T(\mu - (1+r)S_0)$
- $\text{Var}(X_1) = \theta^T V \theta$

So we want to minimize $\frac{1}{2}\theta^T V \theta$ subject to $\theta^T(\mu - (1+r)S_0) \geq m - (1+r)x$.

Note that if we wrote $a = \mu - (1+r)S_0$ and $b = m - (1+r)x$, the condition above would just be $a^T \theta \geq b$. Note also that

- $(\theta^*)^T(\mu - (1+r)S_0) = (m - (1+r)x)^+ \geq m - (1+r)x$.
In IB Optimization terms, this is primal feasibility. Using the above notation, this is also $a^T \theta^* \geq b$.
- $\lambda \geq 0$ (dual feasibility).
- $\lambda(m - (1+r)x - (\theta^*)^T(\mu - (1+r)S_0)) = 0$. This is $\lambda(b - a^T \theta^*) = 0$, i.e. just complementary slackness.

Proof. Now pick a feasible portfolio θ . Note $\frac{1}{2}\theta^T V \theta \geq \frac{1}{2}\theta^T V \theta + \lambda(b - a^T \theta)$ since θ is feasible. Write this as

$$\begin{aligned} \frac{1}{2}\theta^T V \theta &\geq \frac{1}{2}\theta^T V \theta + \lambda(b - a^T \theta) = \\ &= \frac{1}{2}(\theta - \theta^*)^T V (\theta - \theta^*) + \lambda b - \frac{1}{2}(\theta^*)^T V \theta^* \geq \lambda b - \frac{1}{2}(\theta^*)^T V \theta^* \end{aligned}$$

where $\theta^* = \lambda V^{-1}a$, the last inequality follows from V being positive definite, and with equality if and only if $\theta = \theta^*$. (This is just the proof of the Lagrangian sufficiency theorem). \square

1.2 Capital Asset pricing model

Recall a fact from probability: let X, Y be square-integrable random variables (i.e. they have finite variances) with $\text{Var}(X) > 0$. Then there exist unique numbers a, b such that $Y = a + bX + Z$ where $\mathbb{E}[Z] = 0$ and $\text{Cov}(X, Z) = 0$.

Proof. Let $Z = Y - a - bX$ for a, b to be determined. Then

$$\begin{aligned} \mathbb{E}[Z] = 0 &\implies \mathbb{E}[Y] = a + b\mathbb{E}[X] \\ \text{Cov}(X, Z) = 0 &\implies \text{Cov}(Y, X) = b\text{Var}(X), \end{aligned}$$

which has a unique solution $b = \frac{\text{Cov}(Y, X)}{\text{Var}(X)}$ and $a = \mathbb{E}[Y] - b\mathbb{E}[X]$. \square

Remark. Alternatively, the given values of a and b are the unique minimizers of $\mathbb{E}[(Y - a - bX)^2]$.

Setup: Our investor has initial wealth X_0 and portfolio θ . We still have $X_1 = (1+r)X_0 + \theta^T(S_1 - (1+r)S_0)$. Market capitalization will be proportional to $\theta_{\text{mar}}^T S_0$ at time 0 and $\theta_{\text{mar}}^T S_1$ at time 1.

Theorem 1.3.

$$X_1 - (1+r)X_0 = b(X_1^{\text{mar}} - (1+r)X_0^{\text{mar}}) + Z$$

where $\mathbb{E}[Z] = 0$ and $\text{Cov}(X_{\text{mar}}, Z) = 0$.

Theorem 1.4 (Reformulation with returns instead of wealths). Let $R = \frac{X_1}{X_0} - 1$, $R_{\text{mar}} = \frac{X_1^{\text{mar}}}{X_0^{\text{mar}}} - 1$. Then alpha equals zero in a mean-variance efficient market. In other words, if $R - r = \alpha + \beta(R_{\text{mar}} - r) + \epsilon$ where $\mathbb{E}[\epsilon] = 0$, $\text{Cov}(R_{\text{mar}}, \epsilon) = 0$, then $\alpha = 0$.

Remark. This says that if every market participant were mean variance efficient, then we can't "beat" the market index (everyone has the same portfolio up to scaling). This is the CAPM assumption.

Proof. We have

$$\mathbb{E}[X_1 - (1+r)X_0] = \theta^T(\mu - (1+r)S_0)$$

and

$$\begin{aligned} \text{Cov}(X_1 - (1+r)X_0, X_1^{\text{mar}} - (1+r)X_0^{\text{mar}}) &= \theta^T \text{Cov}(S_1) \theta_{\text{mar}} = \\ &= \theta^T V V^{-1} (\mu - (1+r)S_0) = \mathbb{E}[X_1 - (1+r)X_0]. \end{aligned}$$

Choosing $\theta = \theta_{\text{market}}$ in the calculation, we get

$$\mathbb{E}[X_1^{\text{mar}} - (1+r)X_0^{\text{mar}}] = \text{Var}(X_1^{\text{mar}} - (1+r)X_0^{\text{mar}}).$$

Therefore

$$a = \mathbb{E}[X_1 - (1+r)X_0] - b\mathbb{E}[X_1^{\text{mar}} - (1+r)X_0^{\text{mar}}],$$

but

$$b = \frac{\text{Cov}(X_1 - (1+r)X_0, X_1^{\text{mar}} - (1+r)X_0^{\text{mar}})}{\text{Var}(X_1^{\text{mar}} - (1+r)X_0^{\text{mar}})} = \frac{\mathbb{E}[X_1 - (1+r)X_0]}{\mathbb{E}[X_1^{\text{mar}} - (1+r)X_0^{\text{mar}}]},$$

so $a = 0$. □

Remark. Markowitz introduced mean-variance analysis, and Sharpe introduced CAPM. They won the Nobel prize in 1990.

This tells us that

12 Oct 2022,
Lecture 3

$$\mathbb{E}[X_1] = (1+r)X_0 + (\mathbb{E}[X_1^{\text{mar}}] - (1+r)X_0^{\text{mar}}) \frac{\text{Cov}(X_1^{\text{mar}}, X_1)}{\text{Var}(X_1^{\text{mar}})},$$

i.e. $\alpha = 0$, i.e. $\theta^{\text{mar}} = \text{const} \cdot (n_1 \dots n_d)^T$.

By similar reasoning to previous analysis, the optimal solution to "minimize $\text{Var}(X_1)$ wrt $X_0 = x$ and $\mathbb{E}[X_1] = m$ (notice the equals sign here instead of the inequality) is $\theta^* = \lambda \theta^{\text{mar}}$, where as before, $\theta^{\text{mar}} = V^{-1}(\mu - (1+r)S_0)$, but this time

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

(note the lack of positive part in the numerator). Plugging this in for variance, we get that the minimized variance is

$$\min \text{Var}(X_1) = \frac{(m - (1+r)x)^2}{(\mu - (1+r)S_0)^T V^{-1}(\mu - (1+r)S_0)},$$

i.e. the efficient frontier is a parabola.

1.3 Expected utility

So far, we have implicitly assumed that an agent prefers X to Y if either

- $\mathbb{E}[X] > \mathbb{E}[Y]$ and $\text{Var}(X) = \text{Var}(Y)$, or
- $\mathbb{E}[X] = \mathbb{E}[Y]$ and $\text{Var}(X) < \text{Var}(Y)$.

Aside: This is rather crude. For example, the St Petersburg paradox - toss a coin until you get heads, if it took n tries, I pay you 2^n pounds. How much should you pay to play? The expected value is infinite, yet you obviously wouldn't pay me very much. End of aside.

The solution to this is to instead consider the **expected utility** of the payout.¹

Definition 1.4. The **expected utility hypothesis** says that every agent has a **utility function** U such that the agent prefers random payout X to Y if and only if $\mathbb{E}[U(X)] > \mathbb{E}[U(Y)]$. If $\mathbb{E}[U(X)] = \mathbb{E}[U(Y)]$, we say the agent is **indifferent** between X and Y .

Aside: We can show (under some assumptions) that every agent must have a utility function. For random payouts X, Y, Z and A an event independent from X, Y, Z , we assume the von Neumann – Morgenstern axioms:

¹In the real world, another solution is to consider counterparty risk: surely most people will not be actually be in a situation to give you 2^{40} pounds, so you should never pay more than 40 pounds to play.

- Either $X \succ Y$, $X \prec Y$ or $X \sim Y$ (Completeness).
- $X \succ Y$ and $Y \succ Z \implies X \succ Z$ (Transitivity).
- $X \succeq Y$ if and only if $\mathbb{1}_A X + \mathbb{1}_{A^c} Z \succeq \mathbb{1}_A Y + \mathbb{1}_{A^c} Z$ (Independence).
- If $X \succeq Y \succeq Z$, then $\exists p \in [0, 1]$ s.t. if $\mathbb{P}(A) = p$, then $Y = X\mathbb{1}_A + Z\mathbb{1}_{A^c}$ (Continuity).

vN-M proved in 1947 that an agent has expected utility preferences if and only if the vN-M axioms (plus another technical axiom) hold. End of aside.

1.4 Risk-aversion

We have two further usual assumptions about our utility function U .

- Increasing: $x > y \implies U(x) > U(y)$. Even for random variables, if $\mathbb{P}(X > Y) = 1$, then $\mathbb{E}[U(X)] > \mathbb{E}[U(Y)]$.
- Concavity: $U(px + qy) \geq pU(x) + qU(y)$ for any $x, y, 0 \leq p = 1 - q \leq 1$.

Recall Jensen's inequality: $\mathbb{E}[U(X)] \leq U(\mathbb{E}[X])$ for a concave function U . Hence if U is concave, then $\mathbb{E}[X] \succeq X$ for any payout X .

Also note that U concave implies U' is decreasing (where we can think of U' as the **marginal utility**). Furthermore, $U' > 0$ (as the function is increasing) and $U'' < 0$.

Definition 1.5. The Arrow-Pratt **coefficient of absolute risk aversion** is

$$\frac{-U''(x)}{U'(x)}.$$

Definition 1.6. The Arrow-Pratt **coefficient of relative risk aversion** is

$$\frac{xU''(x)}{U'(x)}.$$

The most important examples of utility functions are:

- Exponential (CARA): $U(x) = -e^{-\gamma x}$ for $\gamma = \frac{-U''(x)}{U'(x)}$.
- Power (CRRA): $U(x) = \frac{1}{1-R}x^{1-R}$, $x > 0, R \geq 0, R \neq 1$ where $R = \frac{-xU''(x)}{U'(x)}$.
- Logarithmic: $U(x) = \log x$. This is CRRA with $R = 1$.

14 Oct 2022,
Lecture 4

We have the **utility maximization problem**: Given a concave, increasing utility function U and initial wealth $X_0 = x$, find θ that maximizes

$$\mathbb{E}[U(X_1)] = \mathbb{E}[U(x(1+r) - \theta^T(S_1 - (1+r)S_0))].$$

Theorem 1.5. Suppose U is concave, increasing, and differentiable (i.e. "suitably nice"), then

$$\mathbb{E}[U'(X_1^*)(S_1 - (1+r)S_0)] = 0,$$

where $X_1^* = x(1+r) + (\theta^*)^T(S_1 - (1+r)S_0)$ is the optimal time 1 wealth.

Remark. There's a hidden assumption here as well that everything is integrable, else the expected value could be infinite.

Proof. Differentiate the objective function wrt θ and pass the derivative inside the expectation. \square

Remark. Notice that $S_0 = \frac{\mathbb{E}[U'(X_1^*)S_1]}{(1+r)\mathbb{E}[U'(X_1^*)]}$. So if

$$Z = \frac{U'(X_1^*)}{(1+r)\mathbb{E}[U'(X_1^*)]},$$

then $\mathbb{E}[ZS_1] = S_0$ and $\mathbb{E}[Z] = \frac{1}{1+r}$.

Definition 1.7. Given a market model, a **state price density** or **pricing kernel** is a positive random variable Z such that $\mathbb{E}[ZS_1] = S_0$ and $\mathbb{E}[Z] = \frac{1}{1+r}$.

Our above theorem now says that the marginal utility of maximized wealth is proportional to a state-price density for the model, i.e. our Z above is a state-price model.

We can recover today's prices in terms of the expectation of tomorrow's prices times the kernel (i.e. averaging tomorrow's prices with respect to this kernel).

Aside: the origin of the name. Consider a market model with d outcomes (states) $\Omega = \{\omega_1, \dots, \omega_d\}$. Suppose asset i pays 1 if state i happens and 0 otherwise, so $S_1^i(\omega) = \begin{cases} 1 & \text{if } \omega = \omega_i \\ 0 & \text{otherwise} \end{cases}$. Let Z be a state-price density, then

$$\mathbb{E}[ZS_1] = \sum_{\omega \in \Omega} Z(\omega)S_1^i(\omega)\mathbb{P}(\{\omega\}) = Z(\omega_i)\mathbb{P}(\{\omega_i\}) = S_0^i,$$

so $Z(\omega_i) = \frac{S_0^i}{\mathbb{P}(\{\omega_i\})}$. End of aside.

1.5 Risk-neutral measures

Setup: Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with a set of outcomes Ω , a set of events \mathcal{F} (i.e. subsets of Ω), and a probability measure $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$.

Let Y be a positive random variable such that $\mathbb{E}^{\mathbb{P}}[Y] = 1$. Define a new probability measure \mathbb{Q} by the formula

$$\mathbb{Q}(A) = \mathbb{E}^{\mathbb{P}}[Y \mathbb{1}_A]$$

for any event $A \in \mathcal{F}$. It is a useful fact from measure theory that

$$\mathbb{E}^{\mathbb{Q}}[X] = \mathbb{E}^{\mathbb{P}}[YX]$$

for any \mathbb{Q} -integrable random variable X .

Notation. We write $Y = \frac{d\mathbb{Q}}{d\mathbb{P}}$, called the **density** of \mathbb{Q} with respect to \mathbb{P} , or the likelihood ratio, or the Radon-Nikodym derivative.

Definition 1.8. Two probability measures \mathbb{P} and \mathbb{Q} are **equivalent** if and only if $\mathbb{Q}(A) = 0 \iff \mathbb{P}(A) = 0$. Alternatively, $\mathbb{Q}(A) = 1 \iff \mathbb{P}(A) = 1$. Or $\mathbb{Q}(A) > 0 \iff \mathbb{P}(A) > 0$. Or even $\mathbb{Q}(A) < 1 \iff \mathbb{P}(A) < 1$.

A theorem we will not need, but explains the names above:

Theorem 1.6 (Radon-Nikodym theorem). Probability measures \mathbb{P} and \mathbb{Q} that are defined on the same measurable space (Ω, \mathcal{F}) are equivalent if and only if there exists a positive random variable Y such that $\mathbb{Q}(A) = \mathbb{E}^{\mathbb{P}}[Y \mathbb{1}_A]$ for all events A .

Example 1.1. Suppose $\Omega = \{\omega_1, \omega_2, \dots\}$. Suppose $\mathbb{P}(\omega_i) > 0, \mathbb{Q}(\omega_i) > 0 \forall i$. Then \mathbb{P} and \mathbb{Q} are equivalent, and $\frac{d\mathbb{Q}}{d\mathbb{P}}(\omega) = \frac{\mathbb{Q}(\omega)}{\mathbb{P}(\omega)}$.

Example 1.2. Suppose $X \sim \text{Exp}(\lambda)$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$. Set $Y = \frac{\mu}{\lambda} e^{(\lambda-\mu)X}$. Note $Y > 0$ and

$$\mathbb{E}[f(X)] = \mathbb{E}^{\mathbb{P}}[Y f(X)] = \int_0^\infty \frac{\mu}{\lambda} e^{(\lambda-\mu)x} f(x) \lambda e^{-\lambda x} dx = \int_0^\infty \mu e^{-\mu x} f(x) dx,$$

so X has a new distribution $\text{Exp}(\mu)$ under \mathbb{Q} .

Let us now return to finance. Setup: prices $\{S_t\}_{t \in \{0,1\}}$ are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 1.9. A **risk-neutral measure** is any probability measure \mathbb{Q} equivalent to \mathbb{P} such that

$$\mathbb{E}^{\mathbb{Q}}[S_1] = (1+r)S_0.$$

Remark. Let \mathbb{Q} be equivalent to \mathbb{P} and set $Z = \frac{1}{1+r} \frac{d\mathbb{P}}{d\mathbb{Q}}$. Then Z is a state-price density if and only if \mathbb{Q} is risk-neutral.

So finally, the main theorem about utility maximization reads: if θ^* is optimal, then $U'(X_1^*)$ is proportional to the density of a risk-neutral measure.

17 Oct 2022,
Lecture 5

Example : Utility maximization in one-period a binomial model. We search for $\max \mathbb{E}[U(X_1)]$ given $X_1 = x$ and U a concave, increasing, continuously differentiable function. The market model has interest rate r , $d = 1$ (there is only one risky asset), S_0 is given, and

$$\mathbb{P}(S_1 = (1+b)S_0) = p = 1 - \mathbb{P}(S_1 = (1+a)S_0),$$

where $-1 < a < b$ are given, and $p \in (0, 1)$.

What are the risk-neutral measures? We need to find $\mathbb{Q} \sim \mathbb{P}$ such that $\mathbb{E}^{\mathbb{Q}}[S_1] = (1+r)S_0$. Let $q = \mathbb{Q}(S_1 = (1+b)S_0)$. We need $q \in (0, 1)$ and $q(1+b)S_0 + (1-q)(1+a)S_0 = (1+r)S_0$, so $q = \frac{r-a}{b-a}$, $1-q = \frac{b-r}{b-a}$.

Note. A risk-neutral measure exists only if $a < r < b$ (i.e. this is necessary for a solution to exist to the utility maximization problem). Also, if we have a solution, it is unique.

From last time, we know that $U'(X_1^*) = \lambda \frac{d\mathbb{Q}}{d\mathbb{P}}$ for some $\lambda > 0$, so $X_1^* = (1+r)x + \theta^*(S_1 - (1+r)S_0)$, where θ^* is the optimal portfolio.

So we know that

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

Let $I = (U')^{-1}$, i.e. the inverse of the marginal utility, where I is decreasing and continuous. Hence we can look at our formula above and write it as $X_1^* = I(\lambda \frac{d\mathbb{P}}{d\mathbb{Q}})$. Now taking expected values, we find

$$\mathbb{E}^{\mathbb{Q}}[X_1^*] = (1+r)x + \theta^* \mathbb{E}[S_1 - (1+r)S_0] = (1+r)x = \mathbb{E}^{\mathbb{Q}} \left[I \left(\lambda \frac{d\mathbb{P}}{d\mathbb{Q}} \right) \right].$$

So the recipe to find the optimal solution is to let λ solve

$$\mathbb{E}^{\mathbb{Q}} \left[I \left(\lambda \frac{d\mathbb{P}}{d\mathbb{Q}} \right) \right] = (1+r)x,$$

i.e.

$$qI \left(\lambda \frac{q}{p} \right) + (1-q)I \left(\lambda \frac{1-q}{1-p} \right),$$

and then solve either $U'((1+r)x + \theta^*(b-r)S_0) = \lambda \frac{q}{p}$ or $U'((1+r)x + \theta^*(a-r)S_0) = \lambda \frac{1-q}{1-p}$ to find θ^* .

Special case: CRRA, $U(x) = \frac{1}{1-R}x^{1-R}$, $x > 0, R > 0, R \neq 1$. Then $U'(x) = x^{-R}$, so $I(y) = y^{-\frac{1}{R}}$, so in this case λ solves

$$\lambda^{-\frac{1}{R}} \mathbb{E}^{\mathbb{Q}} \left[\left(\frac{d\mathbb{P}}{d\mathbb{Q}} \right)^{-\frac{1}{R}} \right] = (1+r)x,$$

so

$$\lambda = ((1+r)x)^{-R} \left(\mathbb{E}^{\mathbb{Q}} \left[\frac{d\mathbb{P}}{d\mathbb{Q}} \right]^{\frac{1}{R}} \right)^R,$$

and then plug it in to find θ^* .

1.6 Contingent claims

Contingent claims are just another name for an asset (they are a derivative whose value depends on some other underlying asset). Idea: some assets are "fundamental", and some are contingent on the value of the fundamental asset.

Example 1.3. A **call option** in the one period-case. Given $d = 1$, a call option is the right, but not the obligation, to buy the stock at time 1 for a fixed price (strike price) K .

The call option is a contingent claim and has a price C_0 at time 0. A question we'd like to answer (and later will) is: what is C_0 ? Note that if $S_1 > K$, it is rational to exercise the option, while if $S_1 \leq K$, it is rational not to exercise the option, i.e. the buyer's profit is $(S_1 - K)^+$.

1.6.1 Indifference price of contingent claims.

Consider a utility maximizing agent. We would prefer to buy one share of the claim with time 1 payout Y at price π , if $\exists \theta$ such that

$$\mathbb{E}[U((1+r)(x-\pi) + \theta^T(S_1 - (1+r)S_0) + Y)] > \mathbb{E}[U((1+r)x + \phi^T(S_1 - (1+r)S_0))] \quad \forall \phi.$$

Notation. Let $\mathcal{X} = \{\theta^T(S_1 - (1+r)S_0) \mid \theta \in \mathbb{R}^d\}$ be the time 1 wealths attainable in the market from 0 initial wealth.

Assumption. The model parameters are such that there exists an optimal solution to any utility maximization problem we consider.

Definition 1.10. The **indifference price** $\pi(Y)$ of a contingent claim with time 1 payout Y is the unique solution to

$$\max_{X \in \mathcal{X}} \mathbb{E}[U((1+r)(x-\pi) + X + Y)] = \max_{X \in \mathcal{X}} \mathbb{E}[U((1+r)x + X)],$$

where $x = x_0$ is the time 0 initial wealth.

19 Oct 2022,
Lecture 6

We change our notation a little from last time and let

$$\mathcal{X}(x) = \{(1+r)x + \theta^\top (S_1 - (1+r)S_0) \mid \theta \in \mathbb{R}^d\}$$

be the set of time-1 wealths attainable from starting wealth x . This lets us rewrite the above indifference price equation as

$$\max_{X \in \mathcal{X}(x)} \mathbb{E}[U(X + Y - (1+r)\pi)] = \max_{X \in \mathcal{X}(x)} \mathbb{E}[U(X)]$$

Remark. $\pi \mapsto \max \mathbb{E}[U(X + Y - (1+r)\pi)]$ is decreasing. To see this, look at

$$\begin{aligned} \max \mathbb{E}[U(X + Y - (1+r)(\pi + \epsilon))] &= \mathbb{E}[U(X_\epsilon + Y - (1+r)(\pi + \epsilon))] < \\ &\mathbb{E}[U(X_\epsilon + Y - (1+r)\pi)] \leq \max_{X \in \mathcal{X}(x)} \mathbb{E}[U(X + Y - (1+r)\pi)]. \end{aligned}$$

since U is increasing (for $\epsilon > 0$, X_ϵ the value achieving the maximum). Hence the solution is unique.

Remark. $\pi \mapsto \max \mathbb{E}[U(X + Y - (1+r)\pi)]$ is also continuous, but the proof is omitted.

Warmup: suppose $\mathbb{P}(Y \geq a) = 1$. Then $\pi(Y) \geq \frac{a}{1+r}$.

Proof. For $\pi(Y)$, the indifference price of Y , we have

$$\begin{aligned} \max \mathbb{E}[U(X)] &= \max \mathbb{E}[U(X + Y - (1+r)\pi)] \geq \max \mathbb{E}[U(X + a - (1+r)\pi)] = \\ &\mathbb{E}[U(X - (1+r)(\pi - \frac{a}{1+r}))]. \end{aligned}$$

Hence $(\pi - \frac{a}{1+r}) \geq 0$ from the remark above (the function is decreasing). \square

Theorem 1.7. The indifference pricing function π is concave.

Proof. Fix Y_0, Y_1 , and let $Y_p = pY_1 + (1-p)Y_0$. Let $\pi_i = \pi(Y_i)$ for $i \in \{0, p, 1\}$, $0 \leq p \leq 1$. Then $\max \mathbb{E}[U(X + Y_p - (1+r)\pi_p)] = \max \mathbb{E}[U(X)]$ by the definition of the indifference price, but

$$\begin{aligned} \max \mathbb{E}[U(X)] &= \\ (1-p) \max \mathbb{E}[U(X + Y_0 - (1+r)\pi_0)] &+ p \max \mathbb{E}[U(X + Y_1 - (1+r)\pi_1)] = \\ (1-p) \mathbb{E}[U(X_0 + Y_0 - (1+r)\pi_0)] &+ p \mathbb{E}[U(X_1 + Y_1 - (1+r)\pi_1)] \leq \\ \mathbb{E}[U((1-p)X_0 + pX_1 + Y_p - (1+r)(p\pi_1 + (1-p)\pi_0))] \end{aligned}$$

by concavity (where X_0, X_1 are maximizers of their respective expressions).

Note $(1-p)X_0 + pX_1 \in \mathcal{X}(x)$, so the above is

$$\leq \max \mathbb{E}[U(X + Y_p - (1+r)(p\pi_1 + (1-p)\pi_0))].$$

Hence $p\pi_1 + (1-p)\pi_0 \leq \pi_p$ and we're done. \square

Remark. $\pi(Y)$ is the buyer's indifference price (bid price). Note that the seller's indifference price is $-\pi(-Y)$ (ask price).

We predict that $\pi(Y) \leq -\pi(-Y)$ (i.e. the bid price is less than the ask price). And this is true:

$$\frac{1}{2}\pi(Y) + \frac{1}{2}\pi(-Y) \leq \pi\left(\frac{Y}{2} - \frac{Y}{2}\right) = \pi(0) = 0$$

by concavity.

1.6.2 Marginal utility pricing

Fix a claim with payout Y . Let $\pi_t = \frac{\pi(tY)}{t}$. On the example sheet, we will show that $t \mapsto \pi_t$ is decreasing (using concavity). We can ask what happens at the limit, i.e. $\lim_{t \rightarrow 0} \pi_t = \sup_{t > 0} \pi_t < \pi_{-1}$ (note the limit is bounded, so exists).

Theorem 1.8.

$$\pi_0 = \lim_{t \rightarrow 0} \pi_t = \frac{1}{1+r} \mathbb{E}^{\mathbb{Q}}[Y]$$

where $\frac{d\mathbb{Q}}{d\mathbb{P}} \propto U'(X^*)$, i.e. it is proportional to the marginal utility of the time 1 optimal wealth X^* , i.e. the maximizer $\mathbb{E}[U(X)]$ over $X \in \mathcal{X}(x)$.

Remark. This provides a **linear** pricing rule for a very small number of shares of the claim.

Remark. \mathbb{Q} is a risk-neutral measure, so $\mathbb{E}^{\mathbb{Q}}[S_1] = (1+r)S_0$ (first order condition for a maximum). This means that for any $X \in \mathcal{X}(x)$,

$$\mathbb{E}^{\mathbb{Q}}[X] = (1+r)x + \theta^\top \mathbb{E}^{\mathbb{Q}}[S_1 - (1+r)S_0] = (1+r)x.$$

Proof. Let X_t be the maximizer of

$$\mathbb{E}[U(X + t(Y - (1+r)\pi_t))]$$

over $X \in \mathcal{X}(x)$ and $t > 0$. Then

$$\begin{aligned} 0 &= \frac{1}{t} \mathbb{E}[U(X_t + t(Y - (1+r)\pi_t)) - U(X_0)] \geq \\ &\quad \frac{1}{t} \mathbb{E}[U(X_0 + t(Y - (1+r)\pi_t)) - U(X_0)] \end{aligned}$$

As $t \mapsto \pi_t$ is decreasing, so the above is

$$\geq \mathbb{E} \left[\frac{U(X_0 + t(Y - (1+r)\pi_0)) - U(X_0)}{t} \right] \rightarrow \mathbb{E}[U'(X_0)](Y - (1+r)\pi_0).$$

Hence

$$\pi_0 \geq \frac{1}{1+r} \mathbb{E}^{\mathbb{Q}}[Y] = \frac{\mathbb{E}[U'(X_0)Y]}{(1+r)\mathbb{E}[U'(X_0)]}.$$

For the reverse direction, we use the following lemma:

21 Oct 2022,
Lecture 7

Lemma 1.9. For any x, y we have

$$U(y) \leq U(x) + U'(x)(y - x).$$

Proof. For $x < x + \epsilon < y$ we have $\frac{U(y) - U(x)}{y - x} \leq \frac{U(x + \epsilon) - U(x)}{\epsilon} \xrightarrow{\epsilon \rightarrow 0} U'(x)$ by concavity $U((1-p)x + py) \geq (1-p)U(x) + py$ with $p = \frac{\epsilon}{y-x}$. The case $y < x$ is similar. \square

Hence

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

where we use the fact that $\mathbb{E}[U'(X_0)X] = (1+r)X$ for any $X \in \mathcal{X}(x)$, so in particular $\mathbb{E}[U'(X_0)(X_t - X_0)] = 0$. Hence

$$\pi_t \leq \frac{\mathbb{E}[U'(X_0)Y]}{(1+r)\mathbb{E}[U'(X_0)]} = \frac{1}{1+r} \mathbb{E}^{\mathbb{Q}}[Y].$$

Taking the limit as $t \rightarrow 0$ finishes (the case $t < 0$ is similar). \square

1.7 Arbitrage

Recall our setup: we have one risk-free asset with interest rate r , and d risky assets with time t price S_t .

Definition 1.11. An **arbitrage** is a portfolio $\phi \in \mathbb{R}^d$ such that

$$\begin{aligned} \phi^\top (S_1 - (1+r)S_0) &\geq 0 \text{ almost surely, and} \\ \mathbb{P}(\phi^\top (S_1 - (1+r)S_0) > 0) &> 0. \end{aligned}$$

Fix some initial wealth $X_0 = x$ and an increasing utility function U . Consider

the problem of maximizing $\mathbb{E}[U(X_1)]$ over $X_1 \in \mathcal{X}(x)$, where

$$\mathcal{X}(x) = \{(1+r)x + \theta^\top (S_1 - (1+r)S_0) \mid \theta \in \mathbb{R}^d\}.$$

Suppose this has arbitrage ϕ . Then, given any $X \in \mathcal{X}(x)$, consider

$$X^* = X + \phi^\top (S_1 - (1+r)S_0).$$

Note that $X^* \in \mathcal{X}(x)$, but also $U(X^*) \geq U(X)$ almost surely and $\mathbb{P}(U(X^*) > U(X)) > 0$, hence $\mathbb{E}[U(X^*)] > \mathbb{E}[U(X)]$. But X was arbitrary, so there can be no maximizer!

Arbitrages are bad for our theory: if ϕ is an arbitrage, then by above, the portfolio $\theta + (n+1)\phi$ is better than $\theta + n\phi$ for all n . As n gets large, the assumption that an agent can trade with no price impact becomes more and more unrealistic.

Remarks. The definition of arbitrage does not depend on the agent's initial wealth x or utility function U . However, it does depend on the agent's beliefs through the probability measure \mathbb{P} . Nevertheless, if one agent believes ϕ is an arbitrage, then another agent with equivalent beliefs (i.e. agreeing on almost sure events) would also believe that ϕ is an arbitrage.

1.7.1 Fundamental theorem of asset pricing

So far, we know that:

- If there exists an optimal solution to a utility maximization problem, then there exists a risk-neutral measure.
- If there exists an optimal solution to a utility maximization problem, then there exists no arbitrage.

Theorem 1.10 (Fundamental theorem of asset pricing). A market model has no arbitrage if and only if there exists a risk-neutral measure.

Proof. \Leftarrow : This is the easy direction. Let ϕ be a portfolio such that

$$\mathbb{P}(\phi^\top (S_1 - (1+r)S_0) \geq 0) = 1.$$

Suppose there exists a risk-neutral measure \mathbb{Q} . Then, by equivalence, we have $\mathbb{Q}(\phi^\top (S_1 - (1+r)S_0) \geq 0) = 1$. However, by the definition of risk-neutrality, we get $\mathbb{E}^{\mathbb{Q}}[\phi^\top (S_1 - (1+r)S_0)] = \phi^\top \mathbb{E}^{\mathbb{Q}}[S_1 - (1+r)S_0] = 0$.

Hence, by the pigeonhole principle, $\mathbb{Q}(\phi^\top (S_1 - (1+r)S_0) > 0) = 0$. By equivalence, $\mathbb{P}(\phi^\top (S_1 - (1+r)S_0) > 0) = 0$, so ϕ is not an arbitrage and we're done.

\implies : Let $\xi = S_1 - (1+r)S_0$. We can assume WLOG that $\mathbb{E}[e^{-\theta^\top \xi}] < \infty$ for all $\theta \in \mathbb{R}^d$ by replacing \mathbb{P} with an equivalent measure $\tilde{\mathbb{P}}$ with density

$$\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} \propto e^{-|\xi|^2}$$

and noting that we have no \mathbb{P} -arbitrages iff we have no $\tilde{\mathbb{P}}$ -arbitrages.

Motivating aside: the existence of an optimal solution to a utility maximization problem $\implies \exists$ a risk-neutral measure with density $\frac{d\mathbb{Q}}{d\mathbb{P}} \propto U'(X_1^*)$. End of aside.

Let us try to maximize $\mathbb{E}[U(\theta^\top \xi)]$, where $U(x) = -e^{-x}$.

Let $(\theta_n)_n$ be such that $\mathbb{E}[U(\theta_n^\top \xi)] \rightarrow \sup_{\theta \in \mathbb{R}^d} \mathbb{E}[U(\theta^\top \xi)]$. We consider two cases:

- $(\theta_n)_n$ is bounded. By Bolzano-Weierstrass, there exists a convergent subsequence, so we may assume that $(\theta_n) \rightarrow \theta_0$ for some portfolio θ_0 . But

$$\mathbb{E}[U(\theta_0^\top \xi)] = \lim \mathbb{E}[U(\theta_n^\top \xi)] = \sup_{\theta \in \mathbb{R}^d} \mathbb{E}[U(\theta^\top \xi)]$$

(by continuity of $\theta \mapsto \mathbb{E}[e^{-\theta^\top \xi}]$). So θ_0 is an optimal portfolio. Hence we're done, as the marginal utility at the optimal wealth, $U'(\theta_0^\top \xi)$, is proportional to the density of a risk-neutral measure.

- Every maximizing sequence $(\theta_n)_n$ is unbounded. Assume for contradiction that there is no arbitrage. We may assume WLOG that the components of ξ are linearly independent, i.e. $(\mathbb{P}(\theta^\top \xi = 0) = 1) \implies \theta = 0$. We may assume this since the original market has no arbitrage, so any submarket also has no arbitrage, so just pass to a submarket if need be.

We may assume that $|\theta_n| \rightarrow \infty$. Let $\varphi_n = \frac{\theta_n}{|\theta_n|}$. Note $(\varphi_n)_n$ is bounded, so by Bolzano-Weierstrass we may assume $\varphi_n \rightarrow \varphi_0$ for some φ_0 .

To finish the proof, we will now show that $\varphi_0^\top \xi \geq 0$ almost surely. Since we have no arbitrage, $\varphi_0^\top \xi = 0$ almost surely. By linear independence, this means $\varphi_0 = 0$, contradicting $|\varphi_0| = 1$.

To show $\varphi_0^\top \xi \geq 0$ almost surely, it is enough to show that

$$\mathbb{P}(\varphi_0^\top \xi < -\epsilon, |\epsilon| < r) = 0$$

for all $\epsilon, r > 0$ (by continuity of probability measures).

Note on $\{\varphi_0^\top \xi < -\epsilon, |\epsilon| < r\}$, we have

$$\varphi_n^\top \xi \leq (\varphi_n - \varphi_0)^\top \xi + \varphi_0^\top \xi \leq |\varphi_n - \varphi_0|r - \epsilon.$$

Pick N such that $|\varphi_n - \varphi_0| \leq \frac{\epsilon}{2r}$ for $n \geq N$, so the above is $\leq -\frac{\epsilon}{2}$.

Now, since $\theta = 0$ is not optimal, we get

$$\begin{aligned} 1 = -\mathbb{E}[U(0)] &\geq -\mathbb{E}[U(\theta_n^\top \xi)] = \mathbb{E}[(e^{-\varphi_n \xi})^{|\theta_n|}] \geq \\ \mathbb{E}[(e^{-\varphi_n \xi})^{|\theta_n|} \mathbb{1}_{\{\varphi_0^\top \xi < -\epsilon, |\xi| < r\}}] &\geq e^{\frac{\epsilon}{2}|\theta_n|} \mathbb{P}(\varphi_0^\top \xi < -\epsilon, |\xi| < r), \end{aligned}$$

so

$$\mathbb{P}(\varphi_0^\top \xi < -\epsilon, |\xi| < r) \leq e^{-\frac{\epsilon}{2}|\theta_n|} \rightarrow 0.$$

□

Remark. The details of the above proof are all examinable. The entire proof is a bit longer than the average bookwork question, so the emphasis should be on understanding each part of the proof separately.

1.8 No-arbitrage pricing

Given a market model with r, S_0, S_1 , and the payout of a contingent claim, we want a time 0 price π for the claim. So far we can:

- Given $X_0 = x, U$, calculate the indifference price.
- Given $X_0 = x, U$, calculate the marginal utility price.

Now assume that our original market has no arbitrage, we want to find π such that the new market also has no arbitrage.

Theorem 1.11. If the original market has no arbitrage, then the augmented market also has no arbitrage if and only if $\pi = \frac{1}{1+r} \mathbb{E}^{\mathbb{Q}}[Y]$ for some risk-neutral measure in the original market.

Proof. By FTAP applied to the original market, there exist risk-neutral measures \mathbb{Q} such that $\mathbb{E}^{\mathbb{Q}}[S_1] = (1+r)S_0$. By FTAP applied to the augmented market, there is no arbitrage if and only if there exists a risk-neutral measure for the new market, i.e. $\mathbb{E}^{\mathbb{Q}} \left[\begin{pmatrix} S_1 \\ Y \end{pmatrix} \right] = (1+r) \begin{pmatrix} S_0 \\ \pi \end{pmatrix}$. □

Corollary 1.12. The set of no-arbitrage prices is an interval.

Proof. Let I be the set of no-arbitrage prices for the claim. Suppose $\pi_0, \pi_1 \in I$. By FTAP, $\pi_0 = \frac{1}{1+r} \mathbb{E}^{\mathbb{Q}_0}[Y]$, $\pi_1 = \frac{1}{1+r} \mathbb{E}^{\mathbb{Q}_1}[Y]$. Let $p \in (0, 1)$ and $\mathbb{Q}_p = p\mathbb{Q}_1 + (1-p)\mathbb{Q}_0$. Note that

$$\mathbb{E}^{\mathbb{Q}_p}[S_1] = p\mathbb{E}^{\mathbb{Q}_1}[S_1] + (1-p)\mathbb{E}^{\mathbb{Q}_0}[S_1] = p(1+r)S_0 + (1-p)(1+r)S_0 = (1+r)S_0,$$

so \mathbb{Q}_p is risk-neutral. Hence $\pi_p = \frac{1}{1+r} \mathbb{E}^{\mathbb{Q}_p}[Y] \in I$ by FTAP, but $\pi_p = p\pi_1 + (1-p)\pi_0$, so I is convex. □

1.9 Attainable claims

26 Oct 2022,

Lecture 9

Consider putting θ^0 in the bank and holding $\theta \in \mathbb{R}^d$ in risky assets. The time 1 value is $\theta^0(1+r) + \theta^\top S_1$.

Definition 1.12. A claim is **attainable** if and only if its time 1 payout Y is $Y = \theta^0(1+r) + \theta^\top S_1$ for some $\theta^0 \in \mathbb{R}, \theta \in \mathbb{R}^d$, i.e. it can be attained in our market.

Remark. We could also write $Y = (1+r)y + \theta^\top (S_1 - (1+r)S_0) \in \mathcal{X}(y)$ where $y - \theta^\top S_0 = \theta^0$.

On the example sheet we will show that there is a unique indifference price $\pi = \theta^0 + \theta^\top S_0 = y$ and it is independent of $X_0 = x$ or the utility function U . This means that there is a unique marginal utility price.

Now we will show that there is a unique no-arbitrage price.

Theorem 1.13. Suppose the original market has no arbitrage. Then an attainable claim has a unique no-arbitrage price, i.e. if it is added to the market, then there is only one initial price such that the augmented market has no arbitrage.

Proof 1, primal proof. Suppose $Y = (1+r)y + \theta^\top (S_1 - (1+r)S_0)$. Let π be the initial price.

Claim: If $\pi = y$, then there is no arbitrage, and if $\pi \neq y$, then there exists arbitrage.

Proof: Consider a portfolio $(\phi^\top, a)^\top \in \mathbb{R}^{d+1}$, a portfolio in the augmented market. Suppose $\phi^\top (S_1 - (1+r)S_0) + a(Y - (1+r)y) \geq 0$ almost surely (in the case where $\pi = y$). But the above is

$$(\phi^\top + a\theta)^\top (S_1 - (1+r)S_0) \geq 0 \text{ a.s.}$$

Since there is no arbitrage in the original market,

$$(\phi + a\theta)^\top (S_1 - (1+r)S_0) = 0 \text{ a.s.}$$

This implies that there is no arbitrage in the augmented market.

Now let us consider the case $\pi > y$. Our intuition is that we can buy it for price y and sell it for price π , which is an arbitrage.

Consider $(\theta^\top, -1)^\top \in \mathbb{R}^d$. We have

$$\theta^\top (S_1 - (1+r)S_0) - (Y - (1+r)\pi) = (\pi - y)(1+r) > 0,$$

so this is an arbitrage in the augmented market.

The case $\pi < y$ is analogous, so we're done. \square

Proof 2, "dual" proof. We know there is no arbitrage in the augmented market if and only if there exists a risk-neutral measure \mathbb{Q} for the original market such that

$$\pi = \frac{\mathbb{E}^{\mathbb{Q}}[Y]}{1+r} = \frac{\mathbb{E}^{\mathbb{Q}}[(1+r)y + \theta^{\top}(S_1 - (1+r)S_0)]}{1+r} = y + \frac{\theta^{\top}}{1+r} \mathbb{E}^{\mathbb{Q}}[S_1 - (1+r)S_0] \stackrel{*}{=} y,$$

where \star follows from the definition of a risk-neutral measure. \square

We also have a converse result:

Theorem 1.14. If a claim has a unique no-arbitrage price, then it is attainable.

Proof. Exercise on example sheet 2 (it is a consequence of FTAP). \square

Example 1.4. A forward contract is the right and the obligation to buy a given asset at time 1 for a fixed amount K (the strike price). Let $d = 1$, so our payout is $Y = S_1 - K$. This is attainable by holding one share of the asset and borrowing $\frac{K}{1+r}$ from the bank. The initial no-arbitrage price is $\pi = S_0 - \frac{K}{1+r}$.

Aside: this is the concept of the forward price of an asset. Usually, we choose K such that $\pi = 0$. This special $K = (1+r)S_0$ is called the forward price of an asset. End of aside.

Example 1.5. Consider the one-period binomial model (that we talked about in lecture 5), where S_0 becomes either $S_0(1+b)$ or $S_0(1+a)$ at time 1 with probabilities p and $1-p$. Let $Y = g(S_1)$. Turns out this is attainable for any g , since we can solve

$$\begin{aligned} \theta^0(1+r) + \theta S_0(1+b) &= g(S_0(1+b)) \\ \theta^0(1+r) + \theta S_0(1+a) &= g(S_0(1+a)) \\ \implies \theta &= \frac{g(S_0(1+b)) - g(S_0(1+a))}{S_0(b-a)}, \theta^0 = \frac{(1+b)g(S_0(1+a)) - (1+a)g(S_0(1+b))}{(1+r)(b-a)}. \end{aligned}$$

The first of these equations will appear a lot in the upcoming lectures.

Exercise: Check that

$$\theta^0 + \theta S_0 = \frac{\mathbb{E}^{\mathbb{Q}}[g(S_1)]}{1+r},$$

where

$$\mathbb{Q}(S_1 = (1+b)S_0) = \frac{r-a}{b-a}, \quad \mathbb{Q}(S_1 = (1+a)S_0) = \frac{b-r}{b-a}.$$

2 Multiperiod models

2.1 Introduction to multiperiod models

In one period, S_0 is constant and S_1 is a random vector. In two periods, both S_1 and S_2 are random. At time 1, we will know S_1 , but not S_2 . So S_2 is "more random" in a sense.

We want to have three "sets of information" $\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2$. \mathcal{F}_0 has no information, while \mathcal{F}_2 has full information (so we know what all the prices are). \mathcal{F}_1 is an intermediate situation, where we know S_1 , but not \mathcal{F}_2 .

We also have $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2$. These are filtrations, and we will talk about them in detail later.

What is information? Let g be a "set of information". Intuitively, we have a notion of conditional probability of A given g , $\mathbb{P}(A \mid g)$. So our intuitive definition is that an event A is \mathcal{G} -measurable if and only if $\mathbb{P}(A \mid \mathcal{G}) \in \{0, 1\}$ always.

Example 2.1. Let $\Omega = \{HH, HT, TH, TT\}$, and \mathcal{F}_1 the information we get from observing the first coin. Then $A = \{HH, HT\}$ is \mathcal{F}_1 -measurable, because

$$\mathbb{P}(A \mid \mathcal{G}) = \begin{cases} 1 & \text{if first flip is heads.} \\ 0 & \text{if first flip is tails.} \end{cases}$$

$$\text{But if } B = \{HT\}, \text{ then } \mathbb{P}(B \mid \mathcal{G}) = \begin{cases} \frac{1}{2} & \text{if first flip is heads.} \\ 0 & \text{if first flip is tails.} \end{cases}$$

So B is not \mathcal{F}_1 -measurable.

Idea: the set of information can be identified with the collection of measurable events.

2.2 Measurability

Definition 2.1. Given a sample space Ω (a set of outcomes), a **sigma-algebra** on Ω is a collection \mathcal{G} of events (subsets of Ω) such that

- it is nonempty;
- if $A \in \mathcal{G}$, then $A^c \in \mathcal{G}$;
- if $A_1, A_2, \dots \in \mathcal{G}$, then $\bigcup_n A_n \in \mathcal{G}$.²

Example 2.2. If $\Omega = \{HH, HT, TH, TT\}$, then $\mathcal{G} = \{\emptyset, \Omega, \{HH, HT\}, \{TH, TT\}\}$ models the information from the first coin toss.

²If it is nonempty, then some set A and its complement A^c are in the set, so their union, i.e. the empty set, and its complement Ω are in our set as well.

Definition 2.2. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ (the sample space, the collection of all events, and a probability measure) and a sub σ -algebra $\mathcal{G} \subset \mathcal{F}$, a random variable X is \mathcal{G} -**measurable** if and only if

$$\{X \leq x\} \in \mathcal{G} \quad \forall x \in \mathbb{R}.$$

So intuitively, knowing the information in \mathcal{G} allows you to measure the value of X .

Remark. If x is \mathcal{G} -measurable, then $\{x \in B\} \in \mathcal{G}$ for any "nice"³ $B \subset \mathbb{R}$.

Remark. If X is discrete, taking values in $\{x_1, x_2, \dots\}$, then X is \mathcal{G} -measurable if and only if $\{X = x_i\} \in \mathcal{G} \quad \forall i$.

Exercise. Let $\mathcal{G} = \{\emptyset, \Omega\}$. Show that if X is \mathcal{G} -measurable, then X is a constant.

Definition 2.3. The sigma-algebra generated by a random variable X is the collection of events $\{X \in B\}$ for all "nice" $B \subset \mathbb{R}$. We denote this as $\sigma(X)$.

Theorem 2.1. A random variable Y is $\sigma(X)$ -measurable if and only if we have $Y = f(X)$ for some "nice" function $f : \mathbb{R} \rightarrow \mathbb{R}$.

Recall the conditional expectation given an event: Let X be integrable and $G \in \mathcal{F}$ such that $\mathbb{P}(G) > 0$, then $\mathbb{E}[X \mid G]$ is defined to be

$$\mathbb{E}[X \mid G] = \frac{\mathbb{E}[X \mathbb{1}_G]}{\mathbb{P}(G)}.$$

Recall also: Let Y take values in $\{y_1, y_2, \dots\}$ with $\mathbb{P}(Y = y_i) > 0 \quad \forall i$. Then $\mathbb{E}[X \mid Y]$ is defined as $f(Y)$, where

$$f(y_i) = \mathbb{E}[X \mid Y = y_i].$$

Note that $\mathbb{E}[X \mid Y]$ is $\sigma(Y)$ -measurable (as it is a function of Y).

We also have a projection property:

$$\mathbb{E}[\mathbb{E}[X \mid Y]Z] = \mathbb{E}[XZ]$$

for any bounded $\sigma(Y)$ -measurable Z .

³Here, "nice" means a Borel set.

Proof. We know that $\exists g$ such that $Z = g(Y)$ (by measurability). Then

$$\begin{aligned} \text{LHS} &= \mathbb{E}[f(Y)Z] = \mathbb{E}[f(Y)g(Y)] = \sum_i \mathbb{P}(Y = y_i) f(y_i) g(y_i) = \\ &= \sum_i P(Y = y_i) \mathbb{E}[X | Y = y_i] g(y_i) = \sum_i \mathbb{E}[X \mathbb{1}_{\{Y=y_i\}} g(y_i)] = \\ &= \mathbb{E} \left[X \left(\sum \mathbb{1}_{\{Y=y_i\}} g(y_i) \right) \right] = \mathbb{E}[X g(Y)] = \mathbb{E}[X Z] = \text{RHS}. \end{aligned}$$

□

Definition 2.4. Let X be defined on $(\Omega, \mathcal{F}, \mathbb{P})$, suppose X is integrable (i.e. $\mathbb{E}[|X|] < \infty$) and let $\mathcal{G} \subset \mathcal{F}$ be a sub σ -algebra. **The conditional expectation of X given \mathcal{G}** (written as $\mathbb{E}[X | \mathcal{G}]$) is any \mathcal{G} -measurable integrable random variable Y such that

$$\mathbb{E}[X \mathbb{1}_G] = \mathbb{E}[Y \mathbb{1}_G]$$

for all $G \in \mathcal{G}$.

Theorem 2.2. Conditional expectations are almost surely unique (i.e. if Y_1 and Y_2 are conditional expectations of the same thing, then $\mathbb{P}(Y_1 = Y_2) = 1$).

Proof. We have $\mathbb{E}[X \mathbb{1}_G] = \mathbb{E}[Y_1 \mathbb{1}_G] = \mathbb{E}[Y_2 \mathbb{1}_G]$ for any $G \in \mathcal{G}$. It follows that $\mathbb{E}[(Y_1 - Y_2) \mathbb{1}_G] = 0$. Let $G = \{Y_1 > Y_2\}$. This is \mathcal{G} -measurable, since both Y_1 and Y_2 are \mathcal{G} -measurable. By pigeonhole, $\mathbb{P}(Y_1 > Y_2) = 0$, and by symmetry, $\mathbb{P}(Y_1 < Y_2) = 0$. □

Remark. $\mathbb{E}[XZ] = \mathbb{E}[\mathbb{E}[X | \mathcal{G}]Z]$ for any bounded \mathcal{G} -measurable Z . This is proved using measure theory.

Theorem 2.3. Suppose X is square-integrable (i.e. $\mathbb{E}[X^2] < \infty$). Then

$$\mathbb{E}[(X - Y)^2] \geq \mathbb{E}[(X - \mathbb{E}[X | \mathcal{G}])^2]$$

for any \mathcal{G} -measurable Y .

Sketch of proof. Given an arbitrary \mathcal{G} -measurable random variable Y , let $Z = Y - \mathbb{E}[X | \mathcal{G}]$. Then

$$\begin{aligned} \mathbb{E}[(X - Y)^2] &= \mathbb{E}[(X - \mathbb{E}[X | \mathcal{G}] + Z)^2] = \\ &= \mathbb{E}[(X - \mathbb{E}[X | \mathcal{G}])^2 + 2Z(X - \mathbb{E}[X | \mathcal{G}]) + Z^2] = \\ &= \mathbb{E}[(X - \mathbb{E}[X | \mathcal{G}])^2] + \mathbb{E}[Z^2] \geq \mathbb{E}[(X - \mathbb{E}[X | \mathcal{G}])^2]. \end{aligned}$$

□

2.2.1 Properties of conditional expectations

31 Oct 2022,
Lecture 11

Example 2.3. Suppose $\Omega = \bigcup_n G_n$ for $(G_n)_n$ disjoint and $\mathbb{P}(G_n) > 0 \forall n$. Let $\mathcal{G} = \{\bigcup_{n \in I} G_n \mid I \subset \mathbb{N}\}$ and let X be integrable. Then

$$\mathbb{E}[X \mid \mathcal{G}](\omega) = \mathbb{E}[X \mid G_n], \omega \in G_n.$$

Remark. Let $Y(\omega) = n$ if $\omega \in G_n$. Note that $\mathcal{G} = \sigma(Y)$ and $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X \mid Y]$.

Notation. We have by definition $\mathbb{E}[X \mid Y] = \mathbb{E}[X \mid \sigma(Y)]$.

Theorem 2.4. Suppose all conditional expectations are defined. We then have:

- Additivity: $\mathbb{E}[X + Y \mid \mathcal{G}] = \mathbb{E}[X \mid \mathcal{G}] + \mathbb{E}[Y \mid \mathcal{G}]$.
- Pulling out a known factor: Suppose Y is \mathcal{G} -measurable, then

$$\mathbb{E}[XY \mid \mathcal{G}] = Y\mathbb{E}[X \mid \mathcal{G}].$$

- Tower property: if $\mathcal{H} \subseteq \mathcal{G}$, then

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{H}] \mid \mathcal{G}] = \mathbb{E}[X \mid \mathcal{H}].$$

A corollary: $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[X]$ (by letting $\mathcal{H} = \{\emptyset, \Omega\}$).

- If X is independent of \mathcal{G} (so $\{X \in B\}$ is independent of G for all nice B and $G \in \mathcal{G}$, or even $\mathbb{P}(\{X \in B\} \cap G) = \mathbb{P}(X \in B)\mathbb{P}(G)$), then

$$\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X].$$

- Positivity: If $X \geq 0$ almost surely, then $\mathbb{E}[X \mid \mathcal{G}] \geq 0$ almost surely.
- Jensen's inequality: If f is convex, then $\mathbb{E}[f(X) \mid \mathcal{G}] \geq f(\mathbb{E}[X \mid \mathcal{G}])$ almost surely.

Sketch of proof. The first four: use the definition/uniqueness of conditional expectation. Positivity: use the definition again with $G = \{\mathbb{E}[X \mid \mathcal{G}] \leq 0\}$.

Jensen: Suppose f is differentiable. Then we have the supporting hyperplane theorem $f(y) \geq f(x) + f'(x)(y - x)$. Set $y = X, x = \mathbb{E}[X \mid \mathcal{G}]$ to get

$$\begin{aligned} f(X) &\geq f(\mathbb{E}[X \mid \mathcal{G}]) + f'(\mathbb{E}[X \mid \mathcal{G}])(X - \mathbb{E}[X \mid \mathcal{G}]) \\ \implies \mathbb{E}[f(X) \mid \mathcal{G}] &\geq \mathbb{E}[f(\mathbb{E}[X \mid \mathcal{G}]) \mid \mathcal{G}] + \mathbb{E}[f'(\mathbb{E}[X \mid \mathcal{G}])(X - \mathbb{E}[X \mid \mathcal{G}]) \mid \mathcal{G}] = \\ &= f(\mathbb{E}[X \mid \mathcal{G}]) + f'(\mathbb{E}[X \mid \mathcal{G}])\mathbb{E}[X - \mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{G}] = f(\mathbb{E}[X \mid \mathcal{G}]). \end{aligned}$$

□

Definition 2.5. $\mathbb{P}(A \mid \mathcal{G}) = \mathbb{E}[\mathbb{1}_A \mid \mathcal{G}]$.

Remark. Suppose $\mathbb{P}(A \mid \mathcal{G}) \in \{0, 1\}$ almost surely. Then it is an indicator function, so $\mathbb{P}(A \mid \mathcal{G}) = \mathbb{1}_B$ for some B that is \mathcal{G} -measurable.

Thus $\mathbb{P}(A) = \mathbb{E}[\mathbb{P}(A \mid \mathcal{G})] = \mathbb{E}[\mathbb{1}_B] = \mathbb{P}(B)$ and $\mathbb{P}(A \cap B) = \mathbb{E}[\mathbb{E}[\mathbb{1}_A \mathbb{1}_B \mid \mathcal{G}]] = \mathbb{E}[\mathbb{1}_B \mathbb{P}(A \mid \mathcal{G})] = \mathbb{E}[\mathbb{1}_B] = \mathbb{P}(B)$. So now

$$\mathbb{E}[(\mathbb{1}_A - \mathbb{1}_B)^2] = \mathbb{E}[\mathbb{1}_A] - 2\mathbb{E}[\mathbb{1}_A \mathbb{1}_B] + \mathbb{E}[\mathbb{1}_B] = \mathbb{P}(A) - 2\mathbb{P}(A \cap B) + \mathbb{P}(B) = 0,$$

so $\mathbb{1}_A = \mathbb{1}_B$ almost surely, so A is almost B -measurable.

2.3 Adaptedness, filtrations, martingales

Definition 2.6. A **filtration** $(\mathcal{F}_n)_{n \geq 0}$ is a sequence of σ -algebras such that $\mathcal{F}_n \subseteq \mathcal{F}_{n+1} \forall n$.

Convention. For us, \mathcal{F}_0 will always be the trivial σ -algebra $\{\emptyset, \Omega\}$. That means that \mathcal{F}_0 -measurable variables are constants.

Definition 2.7. A **(stochastic) process** $(X_n)_{n \geq 0}$ is a sequence of random variables.

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

Remark. By convention, X_0 is constant (not random).

Definition 2.9. A **martingale** is an integrable process $(X_n)_{n \geq 0}$ such that

$$\mathbb{E}[X_n \mid \mathcal{F}_{n-1}] = X_{n-1}$$

almost surely for all $n \geq 1$.

Remark. Alternatively, a martingale is an integrable process that's adapted and $\mathbb{E}[X_n - X_{n-1} \mid \mathcal{F}_n] = 0 \forall n \geq 1$.

Example 2.4. Flipping a coin twice. $\Omega = \{HH, HT, TH, TT\}$. Let

$$\mathcal{F}_0 = \{\emptyset, \Omega\}, \mathcal{F}_1 = \{\emptyset, \Omega, \{HH, HT\}, \{TH, TT\}\}, \mathcal{F}_2 = 2^\Omega.$$

Being a martingale is a lot more restrictive than being adapted: $(X_n)_n$ is adapted (where X_n is the result of the n^{th} toss), and hence can take 7 values. On the other hand, a martingale $(m_n)_n$ has a lot of structure: $m_1 = \mathbb{E}[m_2 \mid \mathcal{F}_1]$, so assuming $\mathbb{P}(\omega) = \frac{1}{4} \forall \omega \in \Omega$, if the four outcomes have values a, b, c, d , then the two outcomes in m_1 are $\frac{a+b}{2}$ and $\frac{c+d}{2}$, and m_0 is $\frac{a+b+c+d}{4}$. So if we have a finite time horizon and we know the final values, then we have completely characterized the martingale, as we can just project backwards.

02 Nov 2022,
Lecture 12

Aside on the name "risk-neutral measure". Recall U was increasing and concave, which means models prefer more to less, but are risk-averse. If $U'' = 0$, then the agent is said to be risk-neutral. Of course, in this case $U(x) = a + bx$ for $b > 0$.

Let Y be the payout of a claim and suppose there is no market. How much should a risk-neutral agent pay for Y ? The indifference price solves

$$\mathbb{E}[U((1+r)x + Y - (1+r)\pi)] = U((1+r)x),$$

which gives $\pi = \frac{1}{1+r}\mathbb{E}[Y]$. A risk-neutral measure is a fictional measure equivalent to the real measure (i.e. we reweight the probabilities of events) such that the given asset prices agree with the prices of a risk-neutral investor, i.e. exactly $S_0 = \frac{1}{1+r}\mathbb{E}^{\mathbb{Q}}[S_1]$. End of aside.

2.3.1 Motivation for martingales in finance

Definition 2.10. A probability measure \mathbb{Q} equivalent to \mathbb{P} such that

$$\mathbb{E}^{\mathbb{Q}}[S_n | \mathcal{F}_{n-1}] = (1+r)S_{n-1}$$

for all $n \geq 1$ is called **risk-neutral** where S_n is the price vector at time n .

Note. \mathbb{Q} is risk-neutral if and only if $\left(\frac{S_n}{(1+r)^n}\right)$ (the discounted prices) is a \mathbb{Q} -martingale. This is since

$$\mathbb{E}\left[\frac{S_n}{(1+r)^n} | \mathcal{F}_{n-1}\right] = \frac{S_{n-1}}{(1+r)^{n-1}} \iff \mathbb{E}[S_n | \mathcal{F}_{n-1}] = (1+r)S_{n-1}.$$

Remark. Some people call a risk-neutral measure an equivalent martingale measure (EMM).

Definition 2.11. The filtration generated by a process $(X_n)_{n \geq 0}$ is

$$\mathcal{F}_n = \sigma(X_0, \dots, X_n).$$

This is the smallest σ -algebra such that $(X_n)_n$ is adapted.

Examples of martingales.

Example 2.5. Let X_1, X_2, \dots be independent with $\mathbb{E}[X_n] = 0 \forall n$ and X_n integrable for all n . Let $S_0 = 0$ and $S_n = X_1 + \dots + X_n$. Let (\mathcal{F}_n) be generated by $(X_n)_{n \geq 1}$ where $\mathcal{F}_0 = \{\emptyset, \Omega\}$. The claim is that $(S_n)_{n \geq 0}$ is a martingale.

Proof. We can check that S_n is integrable, as $\mathbb{E}[|S_n|] \leq \mathbb{E}[|X_1| + \dots + |X_n|] < \infty$ since each term is finite. S_n is \mathcal{F}_n -measurable (it is a function of X_1, \dots, X_n),

i.e. $(S_n)_{n \geq 0}$ is adapted. As X_n is independent of \mathcal{F}_{n-1} ,

$$\mathbb{E}[S_n - S_{n-1} \mid \mathcal{F}_{n-1}] = \mathbb{E}[X_n \mid \mathcal{F}_{n-1}] = \mathbb{E}[X_n] = 0.$$

□

Example 2.6. Given an integrable random variable X and a filtration $(\mathcal{F}_n)_{n \geq 0}$, let $Z_n = \mathbb{E}[X \mid \mathcal{F}_n]$. The claim is that Z_n is a martingale.

Proof. Z_n is integrable as conditional expectation is integrable by definition.

$$\mathbb{E}[Z_n \mid \mathcal{F}_{n-1}] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{F}_n] \mid \mathcal{F}_{n-1}] \stackrel{(\star)}{=} \mathbb{E}[X \mid \mathcal{F}_{n-1}] = Z_{n-1}$$

where (\star) follows from the tower property. □

Useful fact: To check that a given process $(M_n)_{0 \leq n \leq N}$ on a finite horizon is a martingale, it is enough to check that

$$\mathbb{E}[M_n \mid \mathcal{F}_n] = M_n \quad \forall n \leq N.$$

This is sometimes easier than the equivalent

$$\mathbb{E}[M_n \mid \mathcal{F}_{n-1}] = M_{n-1} \quad \forall 1 \leq n \leq N.$$

2.3.2 Martingale transforms

Definition 2.12. A process $(H_n)_{n \geq 1}$ is **previsible** (or **predictable**) with respect to $(\mathcal{F}_n)_{n \geq 0}$ if and only if H_n is \mathcal{F}_{n-1} -measurable $\forall n \geq 1$.

Let $K_n = H_{n+1}$. Then $(K_n)_{n \geq 0}$ is adapted $\iff (H_n)_{n \geq 1}$ is previsible.

Definition 2.13. Given a previsible $(H_n)_{n \geq 1}$ and adapted $(X_n)_{n \geq 0}$, the process defined as

$$M_n = \sum_{k=1}^n H_k (X_k - X_{k-1})$$

is the **martingale transform** of $(H_n)_n$ with respect to $(X_n)_n$.

Theorem 2.5. If $(H_n)_n$ is bounded and previsible and $(X_n)_n$ is a martingale, the martingale transform $(M_n)_{n \geq 0}$ is a martingale.

Proof. M_n is integrable, as

$$\mathbb{E}[|M_n|] \leq \mathbb{E}\left[\sum_{k=1}^n \underbrace{|H_k|}_{\text{bounded}} \underbrace{|X_k - X_{k-1}|}_{\text{integrable}}\right]$$

and we have a finite number of finite terms.

M_n is also \mathcal{F}_n -measurable as it is a function of H_1, \dots, H_n and X_0, \dots, X_n which are all \mathcal{F}_n -measurable. Finally,

$$\mathbb{E}[M_n - M_{n-1} \mid \mathcal{F}_{n-1}] = \mathbb{E}[H_n(X_n - X_{n-1}) \mid \mathcal{F}_{n-1}] \stackrel{(\star)}{=} H_n \underbrace{\mathbb{E}[X_n - X_{n-1} \mid \mathcal{F}_{n-1}]}_{=0 \text{ by def. of a martingale}}$$

where (\star) follows as H_n is \mathcal{F}_{n-1} -measurable. \square

Martingale transforms in finance.

Let $(S_n)_n$ be the asset price vectors and $(\theta_n)_n$ the portfolios, with θ_n being \mathcal{F}_{n-1} -measurable (since it is the portfolio we hold between times $n-1$ and n) and let X_n be our wealth at time n . Then

$$\begin{aligned} X_n &= (1+r)X_{n-1} + \theta_n^\top (S_n - (1+r)S_{n-1}) \\ &\iff \\ \frac{X_n}{(1+r)^n} &= X_0 + \sum_{k=1}^n \theta_k^\top \left(\frac{S_k}{(1+r)^k} - \frac{S_{k-1}}{(1+r)^{k-1}} \right). \end{aligned}$$

So an investor's discounted wealth is the martingale transform of his portfolio with respect to the discounted risk asset prices.

04 Nov 2022,
Lecture 13

2.4 Stopping times

Definition 2.14. Given a filtration $(\mathcal{F}_n)_{n \geq 0}$, a stopping time is a random variable T taking values in $\{0, 1, 2, \dots\} \cup \infty$ such that $\{T \leq n\} \in \mathcal{F}_n \forall n \geq 0$.

Example 2.7. Let $(X_n)_n$ be an adapted process. Let $T = \inf\{n \geq 0 \mid X_n > 0\}$ (with the convention that $\inf \emptyset = \infty$). Then the claim is that T is a stopping time. Indeed, $\{T \leq n\} = \bigcup_{k=0}^n \{X_k > 0\}$ and $\{X_k > 0\} \in \mathcal{F}_k \subset \mathcal{F}_n$ for all $k \leq n$ by adaptedness. Since \mathcal{F}_n is closed under finite unions, we're done.

Counterexample: Let $(X_n)_{n \geq 0}$ be adapted and $T = \sup\{n \geq 0 \mid X_n > 0\}$. Then T is not a stopping time in general.

Proposition 2.6. T is a stopping time if and only if $\{T = n\} \in \mathcal{F}_n \forall n \geq 0$.

Proof. If $\{T = n\} \in \mathcal{F}_n \forall n$, then $\{T \leq n\} = \bigcup_{k=0}^n \{T = k\} \in \mathcal{F}_n$. Conversely, if T is a stopping time, then $\{T = n\} = \{T \leq n\} \cap \{T \leq n-1\}^c \in \mathcal{F}_n$. \square

Definition 2.15. Given an adapted process $(X_n)_{n \geq 0}$ and a stopping time T , the process stopped at T is the process

$$X_n^T = X_{n \wedge T},$$

where $a \wedge b = \min(a, b)$.

Proposition 2.7. $(X_n^T)_{n \geq 0}$ is the martingale transform of $(\mathbb{1}_{\{n \leq T\}})_{n \geq 1}$ with respect to $(X_n)_{n \geq 0}$. We can check that $(\mathbb{1}_{\{n \leq T\}})_{n \geq 1}$ is previsible:

$\mathbb{1}_{\{n \leq T\}}$ is \mathcal{F}_{n-1} -measurable $\iff \{n \leq T\} \in \mathcal{F}_{n-1} \iff \{T \leq n-1\}^c \in \mathcal{F}_{n-1}$ which is true, so we're done as σ -algebras are closed under complements.

Now check

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

Theorem 2.8. A stopped martingale is a martingale.

Proof. The martingale transform of a bounded previsible process with respect to a martingale is again a martingale, and $(\mathbb{1}_{\{n \leq T\}})_{n \geq 1}$ is bounded. \square

Theorem 2.9 (Optional sampling theorem). Let $(X_n)_{n \geq 0}$ be a martingale and let T be a bounded stopping time (i.e. $\exists c$ a constant such that $T \leq c$ almost surely). Then

$$\mathbb{E}[X_T] = X_0.$$

Remark. Recall that under our conventions X_0 is a constant.

Proof. Suppose that $T \leq N$ almost surely. If N is not random, then $X_T = X_{T \wedge N}$. But if $M_n = X_{n \wedge T}$, then $(M_n)_{n \geq 0}$ is a martingale. Now

$$\mathbb{E}[X_T] = \mathbb{E}[M_n] = \mathbb{E}[\mathbb{E}[M_n \mid \mathcal{F}_{n-1}]] = \mathbb{E}[M_{n-1}],$$

so by induction $\mathbb{E}[M_n] = M_0$ and we're done. \square

Aside on limit theorems in probability:

Theorem 2.10 (Monotone convergence theorem). Suppose $0 \leq X_1 \leq X_2 \leq X_3 \leq \dots$ and let $Z = \sup_n X_n = \lim_{n \rightarrow \infty} X_n$. Then $\mathbb{E}[Z] = \lim_{n \rightarrow \infty} \mathbb{E}[X_n]$. In other words,

$$\mathbb{E}[\lim_{n \rightarrow \infty} X_n] = \lim_{n \rightarrow \infty} \mathbb{E}[X_n].$$

Theorem 2.11 (Dominated convergence theorem). Suppose $X_n \rightarrow Z$ almost surely and there exists an integrable Y such that $|X_n| \leq Y$. Then $\mathbb{E}[Z] = \lim_{n \rightarrow \infty} \mathbb{E}[X_n]$.

End of aside.

Theorem 2.12 (Optional stopping theorem). Let $(X_n)_{n \geq 0}$ be a martingale. Let T be a stopping time and suppose that $|X_{n \wedge T}| \leq Y \forall n$ and $\mathbb{E}[Y] < \infty$. Then $\mathbb{E}[X_T] = X_0$.

Proof. By optional sampling theorem, $\mathbb{E}[X_{n \wedge T}] = X_0$. By dominated convergence theorem, we're done since $X_{n \wedge T} \rightarrow X_T$ almost surely. \square

Example 2.8. Let $(X_n)_n$ be a simple symmetric random walk, so $X_n = \xi_1 + \dots + \xi_n$ for $(\xi_i)_i$ i.i.d. with $\mathbb{P}(\xi_i = \pm 1) = \frac{1}{2}$. Then $(X_n)_n$ is a martingale.

Fix integers $a, b > 0$. Let $T = \inf\{n \geq 0 \mid X_n \in \{-a, b\}\}$. We have $|X_{n \wedge T}| \leq \max(a, b)$, so $\mathbb{E}[X_T] = 0$ by optional stopping theorem. Hence we arrive at $-a\mathbb{P}(X_T = -a) + b\mathbb{P}(X_T = b) = 0$, so $\mathbb{P}(X_T = -a) = \frac{b}{a+b}$ and $\mathbb{P}(X_T = b) = \frac{a}{a+b}$.

Non-example. Let $(X_n)_n$ be a simple symmetric random walk and $T = \inf\{n \geq 0 \mid X_n = -a\}$. Note that $X_T = -a$ almost surely, so $\mathbb{E}[X_T] = -a \neq 0$, seemingly a contradiction to optional stopping theorem. But $X_{n \wedge T}$ is not bounded from above, so we can't apply the optional stopping theorem.

07 Nov 2022,
Lecture 14

Example 2.9. Let $X_n = \xi_1 + \dots + \xi_n$ be a simple symmetric random walk with $\mathbb{P}(\xi = \pm 1) = \frac{1}{2}$. Fix $a > 0$ an integer and let $T = \inf\{n \geq 0 \mid X_n = -a\}$. Our goal is to compute $\mathbb{E}[z^T]$ for $z \in (0, 1)$, i.e. the probability generating function.

Claim. There exists w such that $M_n = w^{X_n} z^n$ is a martingale (in the filtration generated by the process itself, i.e. (X_n) , or equivalently (ξ_n)).

Proof. We have

$$\mathbb{E}[w^{X_n} z^n \mid \mathcal{F}_{n-1}] = z^n \mathbb{E}[w^{X_{n-1} + \xi_n} \mid \mathcal{F}_{n-1}] = z^n w^{X_{n-1}} \mathbb{E}[w^{\xi_n}]$$

which equals $z^{n-1} w^{X_{n-1}}$ if and only if $z(\frac{1}{2}w + \frac{1}{2}w^{-1}) = 1$, which we can solve to get $w = \frac{1 \pm \sqrt{1-z^2}}{z}$.

Now apply optional stopping theorem: $\mathbb{E}[M_T] = M_0 = 1$ and $\mathbb{E}[w^{X_T} z^T] = w^{-a} \mathbb{E}[z^T] = 1$, so $\mathbb{E}[z^T] = w^a$. But which value of w do we choose?

Note that $X_{T \wedge n} \geq -a$, so if $w < 1$, then $w^{X_{T \wedge n}} \leq w^{-a}$. So $M_n = w^{X_{T \wedge n}} z^{T \wedge n} \leq w^{-a}$, so it is bounded uniformly in n . So OST is applicable if we choose $w = \frac{1 - \sqrt{1-z^2}}{z}$.

Example 2.10. Let $(X_n)_n$ be a SSRW again, and let $T = \inf\{n \geq 0 \mid X_n \in \{b, -a\}\}$. Fix $z \in (0, 1)$. Our goal is to find $\mathbb{E}[z^T]$. We know that $M_n = w^{X_n} z^n$ is a martingale, where $w + w^{-1} = \frac{2}{z}$ (by the last example).

The trick is to define $M_n = w^{X_n - (\frac{b-a}{2})} z^n + w^{\frac{b-a}{2} - X_n} z^n$, which is also a martingale. Note $X_n - \frac{a+b}{2} = \frac{b-a}{2}$ or $-\frac{(b+a)}{2}$. Hence $M_T = (w^{\frac{a+b}{2}} + w^{-\frac{(a+b)}{2}}) z^T$ (and $M_0 = 2$). Thus

$$\mathbb{E}[z^T] = \frac{2}{w^{\frac{a+b}{2}} + w^{-\frac{(a+b)}{2}}},$$

where $z = \frac{2}{w+w^{-1}}$. Note that $M_{n \wedge T}$ is bounded for either choice of w , and the choice of w doesn't matter since the final formula is symmetric in w and w^{-1} .

2.5 Submartingales and supermartingales

Definition 2.16. A **submartingale** is an integrable adapted process $(X_n)_{n \geq 0}$ such that

$$\mathbb{E}[X_n | \mathcal{F}_{n-1}] \geq X_{n-1} \text{ a.s. } \forall n \geq 1.$$

Definition 2.17. A **supermartingale** is an integrable adapted process $(X_n)_{n \geq 0}$ such that

$$\mathbb{E}[X_n | \mathcal{F}_{n-1}] \leq X_{n-1} \text{ a.s. } \forall n \geq 1.$$

Proposition 2.13. If $n < N$ and $(X_n)_{n \geq 0}$ is a submartingale, then

$$\mathbb{E}[X_N | \mathcal{F}_n] \geq X_n.$$

Proof. This is true when $N = n + 1$. Assume it's true for $N = n + k - 1$, then

$$\mathbb{E}[X_{n+k} | \mathcal{F}_n] = \mathbb{E}[\mathbb{E}[X_{n+k} | \mathcal{F}_{n+k-1}] | \mathcal{F}_n] \geq \mathbb{E}[X_{n+k-1} | \mathcal{F}_n]$$

by the submartingale property and positivity of conditional expectation. \square

Proposition 2.14. Let $(X_n)_{n \geq 0}$ be a submartingale and $(H_n)_{n \geq 1}$ nonnegative, bounded and previsible. Let $M_n = \sum_{k=1}^n H_k(X_k - X_{k-1})$. Then $(M_n)_{n \geq 0}$ is a submartingale.

Proof. Integrability follows from integrability of (X_n) and boundedness of (H_n) . Adaptedness follows from the previsibility of $(H_n)_n$ and the adaptedness of $(X_n)_n$. Finally,

$$\mathbb{E}[M_n - M_{n-1} | \mathcal{F}_{n-1}] = \mathbb{E}[H_n(X_n - X_{n-1}) | \mathcal{F}_{n-1}] = H_n \mathbb{E}[X_n - X_{n-1} | \mathcal{F}_{n-1}] \geq 0.$$

\square

Theorem 2.15. Let S and T be stopping times such that $S \leq T$ almost surely. Let $(X_n)_n$ be a submartingale and let $M_n = X_{n \wedge T} - X_{n \wedge S}$. Then $(M_n)_{n \geq 0}$ is a submartingale.

Proof. Note $M_n = \sum_{k=1}^n \mathbb{1}_{S < k \leq T}(X_k - X_{k-1})$, and the indicator is bounded and previsible. \square

Theorem 2.16 (Optional sampling theorem). Suppose $S \leq T$ are bounded stopping times and $(X_n)_n$ is a submartingale. Then $\mathbb{E}[X_T] \geq \mathbb{E}[X_S]$.

Proof. Let $M_n = X_{n \wedge T} - X_{n \wedge S}$. Let N be a constant such that $T \leq N$ almost surely. Then $\mathbb{E}[M_n] \geq M_0 = 0$ because (M_n) is a submartingale. Hence we're done, since $M_N = X_{T \wedge N} - X_{S \wedge N} = X_T - X_S$. \square

2.6 Controlled Markov processes

09 Nov 2022,
Lecture 15

Review of IB Markov Chains:

Definition 2.18. We say a process $(X_n)_{n \geq 0}$ is a **Markov process** if

$$\mathbb{P}(X_n \in A \mid \mathcal{F}_{n-1}) = \mathbb{P}(X_n \in A \mid X_{n-1})$$

for any "nice" set A and where $\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n)$.

There are two ways to think about a Markov process.

- (i) Our ingredients are $X_0 = x$ (the initial condition) and the transition law $P(n, x, A) = \mathbb{P}(X_n \in A \mid X_{n-1} = x)$. $P(n, x, \cdot)$ is a probability measure $\forall n, x$.
- (ii) It is a random dynamical system. $X_0 = x$ is given and $X_n = G(n, X_{n-1}, \xi_n)$, where $G(\cdot, \cdot, \cdot)$ is given and ξ_1, ξ_2, \dots are independent random variables.

These notions are independent when $\mathbb{P}(G(n, x, \xi_n) \in A) = P(n, x, A)$.

Example 2.11. If $(X_n)_n$ is a SSRW, $X_n = \xi_1 + \dots + \xi_n$ with $(\xi_n)_n$ i.i.d. with $\mathbb{P}(\xi = \pm 1) = \frac{1}{2}$. Then $X_n = G(n, X_{n-1}, \xi_n)$ where $G(n, x, \xi) = x + \xi$.

Remark. If $(X_n)_n$ is a time-homogeneous Markov process, then $G(n, \cdot, \cdot)$ does not depend on n , and (ξ_n) are i.i.d. $\iff P(n, \cdot, \cdot)$ does not depend on n , in which case we can suppress n from the notation.

Remark. In fact, a time-homogeneous case is general, since if $(X_n)_n$ is a Markov process, then $Y_n = (n, X_n)$ is time-homogeneous. This is because $H(n, y, \xi_n) = (n, G(n, x, \xi_n)) = (y^{(1)}, G(y^{(1)}, y^{(2)}, \xi_n))$.

Definition 2.19. A process $(X_n)_n$ is a **controlled Markov process** with **control** $(U_n)_{n \geq 1}$ if $(U_n)_n$ is previsible and $X_n^U = G(n, X_{n-1}^U, U_n, \xi_n)$ for a given function G and an independent sequence $(\xi_n)_n$.

Example 2.12. An example from finance. Let r be the interest rate, $d = 1$, $S_n = S_{n-1}\xi_n$, where ξ_n are i.i.d. Then $(S_n)_n$ is Markov.

We have a wealth process $(X_n)_n$, where

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

where θ_n is \mathcal{F}_{n-1} -measurable (it is the holding of stock between times $n-1$ and n).

This is a controlled Markov process with $U_n = S_{n-1}\theta_n$ and $G(n, x, u, \xi) = (1+r)x + u(\xi - (1+r))$, which is in fact time-homogeneous.

2.7 The Bellman equation

2.7.1 Stochastic optimal control

Goal: Given (fixed, not random) N , functions f and g , and a controlled process $(X_n^U)_n$, we want to maximize

$$\mathbb{E} \left[\sum_{k=1}^N f(k, U_k) + g(x_N^U) \mid X_0 = x \right]$$

over controls $(U_n)_{1 \leq n \leq N}$.

We will make a blanket assumption that everything we need is integrable.

Definition 2.20. The **value function** for this stochastic control problem is the function

$$V(n, x) = \max \mathbb{E} \left[\sum_{k=n+1}^N f(k, U_k) + g(X_N^U) \mid X_n = x \right].$$

Note that $V(N, x) = g(x)$.

Definition 2.21. Say our controlled Markov process evolves according to the rule $X_n = G(n, X_{n-1}, U_n, \xi_n)$. Then the **Bellman equation** for this stochastic control problem is

$$\begin{aligned} V(N, x) &= g(x) \\ V(n-1, x) &= \max_u [f(n, u) + \mathbb{E}[V(n, G(n, x, u, \xi_n))]] \quad \forall 1 \leq n \leq N. \end{aligned}$$